

Вэнь Тонг
Пейин Чжу

Сети



Путь от 5G к 6G глазами разработчиков

Сети 6G.

Путь от 5G к 6G

глазами разработчиков

6G: The Next Horizon

From Connected People and Things to Connected Intelligence

Edited by

WEN TONG

Huawei Technologies Co., Ltd., Canada

PEIYING ZHU

Huawei Technologies Co., Ltd., Canada



CAMBRIDGE
UNIVERSITY PRESS

Сети 6G. Путь от 5G к 6G глазами разработчиков

От подключенных людей и вещей
к подключенному интеллекту

Редакторы

ВЭНЬ ТОНГ

Huawei Technologies Co., Ltd., Canada

ПЕЙИН ЧЖУ

Huawei Technologies Co., Ltd., Canada



Москва, 2022

УДК 654
ББК 32.88
В29

В29 Вэнь Тонг, Пейин Чжу (редакторы)

Сети 6G. Путь от 5G к 6G глазами разработчиков. От подключенных людей и вещей к подключенному интеллекту. / пер. с англ. В. С. Яценкова. – М.: ДМК Пресс, 2022. – 624 с.: ил.

ISBN 978-5-97060-995-8

Это первая книга о беспроводной связи 6G, которая подготавливает читателя к грядущему технологическому прорыву – в этой новой картине мира центральное место займут интеллектуальные устройства, входящие в единую сеть.

В книге рассматриваются новые радиоинтерфейсы и сетевые технологии, интеграция услуг радарного сканирования и связи, а также сетей наземного и аэрокосмического базирования. Обсуждаются новые сетевые архитектуры, обеспечивающие функционирование ИИ, и нативная защита данных.

Издание адресовано исследователям, инженерам, разработчикам, менеджерам, работающим в области беспроводной связи, и всем, кто интересуется технологиями будущего.

УДК 654
ББК 32.88

Copyright © Cambridge University Press, 2021. This translation of title is published by arrangement with Cambridge University Press.

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

Материал, изложенный в данной книге, многократно проверен. Но, поскольку вероятность технических ошибок все равно существует, издательство не может гарантировать абсолютную точность и правильность приводимых сведений. В связи с этим издательство не несет ответственности за возможные ошибки, связанные с использованием книги.

ISBN 978-1-10898981-7 (англ.)
ISBN 978-5-97060-995-8 (рус.)

© Cambridge University Press, 2021
© Оформление, перевод на русский язык, издание,
ДМК Пресс, 2022

Оглавление

Вступление от издательства.....	14
Авторы	15
Предисловие	17
Словарь сокращений	19
Часть I. Вступление	29
Глава 1. Мобильная связь до 2030 года и далее.....	30
1.1. Эволюция систем мобильной связи	30
1.2. Ключевые факторы потребности в 6G.....	33
1.3. Общая концепция	41
1.3.1. Ключевые технологические тенденции	48
1.3.2. Типичные варианты использования	57
1.3.3. Ключевые показатели эффективности	61
1.4. Структура книги.....	65
1.5. Источники	67
Часть II. Сценарии использования и целевые KPI.....	69
Введение в часть II	70
Глава 2. Максимально достоверный эффект погружения	71
2.1. Облачная VR высокой степени достоверности	72
2.1.1. Требования к задержке передачи.....	73
2.1.2. Требования к пропускной способности	76
2.1.3. Обзор основных требований для полной VR.....	78
2.2. Тактильное и мультисенсорное общение	79
2.2.1. Дистанционная работа в высокодинамичных средах.....	80
2.2.2. Основные требования к телеуправлению в высокодинамичных средах	82
2.3. Дисплеи 3D без стекла и голографические дисплеи	83
2.3.1. Глубина восприятия и 3D-дисплеи без стекла.....	83
2.3.2. Методы реконструкции трехмерных изображений без использования стекла.....	84
2.3.3. Требования к разрешению и задержке	85
2.3.4. Основные требования к 3D-дисплеям без стекла	86
2.4. Источники	86
Глава 3. Сканирование, локализация и визуализация.....	88
3.1. Высокоточная локализация.....	89
3.1.1. Абсолютная локализация.....	90
3.1.2. Относительная локализация.....	91
3.1.3. Семантическая локализация	92
3.2. Визуализация, картирование и локализация в реальном времени	93

3.2.1. Одновременная локализация и картирование.....	94
3.2.2. Визуализация и картирование внутри помещений.....	95
3.2.3. Визуализация и картирование на открытом воздухе	96
3.3. Расширенное человеческое восприятие	97
3.3.1. За пределами возможностей глаза – сверхвысокое разрешение	98
3.3.2. За пределами возможностей глаза – сделать невидимое видимым.....	99
3.3.3. За пределами возможностей глаза – спектральное распознавание	100
3.4. Распознавание жестов и действий	101
3.4.1. Бесконтактное управление на макроуровне	102
3.4.2. Бесконтактное управление на микроуровне	104
3.5. Источники	105
Глава 4. Полнофункциональная индустрия 4.0 и выше	106
4.1. Фабрика будущего.....	108
4.2. Управление движением.....	110
4.3. Совместная деятельность роботов в группе	111
4.4. От интеллектуальных коботов к киборгам	112
4.5. Источники	113
Глава 5. Умный город и умная жизнь.....	114
5.1. Умный транспорт	114
5.2. Умное здание.....	116
5.3. Умное здравоохранение	118
5.4. Интеллектуальные сервисы на базе БПЛА	119
5.5. Источники	121
Глава 6. Глобальное покрытие сети мобильного доступа	122
6.1. Широкополосный беспроводной доступ для всех	123
6.1.1. Мобильная широкополосная связь в любом месте	123
6.1.2. Широкополосное соединение в движении	124
6.1.3. Связь служб быстрого реагирования и помощь при стихийных бедствиях	125
6.2. Широкий спектр услуг интернета вещей в труднодоступных местах.....	126
6.3. Высокоточное позиционирование и навигация.....	127
6.4. Мониторинг земной поверхности в реальном времени.....	128
6.5. Источники	129
Глава 7. Подключенное машинное обучение и сетевой ИИ	130
7.1. Услуги и операции 6G, дополненные искусственным интеллектом.....	131
7.1.1. Качество услуг сети 6G, дополненной искусственным интеллектом	132
7.1.2. Управление сетью, дополненной ИИ.....	133
7.2. Сервисы искусственного интеллекта с поддержкой 6G.....	134
7.2.1. Совместный анализ и управление в реальном времени.....	134
7.2.2. 6G для крупномасштабного интеллекта.....	136
7.3. Источники.....	137
Резюме части II.....	138

Часть III. Теоретические основы	139
Введение в часть III	140
Глава 8. Теоретические основы искусственного интеллекта и машинного обучения	142
8.1. Фундаментальная теория искусственного интеллекта	142
8.1.1. Определения	142
8.1.2. Таксономия машинного обучения	145
8.1.3. Теоретико-информационный принцип DNN	148
8.1.4. Реализации DNN	150
8.2. Теория распределенного ИИ	151
8.3. Теория динамической байесовской сети	156
8.4. Источники	162
Глава 9. Теоретические основы беспроводных сетей большой емкости ...	165
9.1. Теория электромагнитной информации	165
9.1.1. Физический канал распространения	167
9.1.2. Диаграммы направленности антенных решеток и эффект взаимного влияния	168
9.1.3. Электромагнитные физические характеристики	169
9.2. Теория крупномасштабной коммуникации	171
9.3. Источники	177
Глава 10. Теоретические основы беспроводных сетей большой емкости	183
10.1. Теория семантической коммуникации	183
10.2. Теория сверхвысокого разрешения	188
10.3. Источники	193
Глава 11. Теоретические основы энергоэффективных систем	194
11.1. Энергоэффективные коммуникации и теория вычислений	194
11.2. Теория зеленого ИИ	197
11.3. Источники	200
Резюме части III	202
Часть IV. Новые элементы	205
Введение в часть IV	206
Глава 12. Новый спектр	207
12.1. Глобальное распределение спектра 5G до 2020 года	208
12.2. Требования к спектру 6G	210
12.3. Средние диапазоны – наиболее экономичный способ широкого охвата	212
12.4. Полосы миллиметрового диапазона в эпоху 6G	215
12.5. Новые возможности для сканирования и связи в ТГц-диапазоне	218
12.6. Источники	221
Глава 13. Новые каналы	223
13.1. Новые требования к моделированию канала 6G	224

13.2. Канальные измерения в 6G	228
13.2.1. Канальные измерения в новом спектре.....	228
13.2.2. Канальные измерения в новых сценариях	231
13.3. Источники.....	233
Глава 14. Новые материалы.....	235
14.1. Развитие кремниевой платформы	235
14.2. Гетерогенная платформа на материалах типа III–V.....	237
14.3. Реконфигурируемый материал.....	237
14.4. Фотонный кристалл	239
14.5. Фотоэлектрические материалы и фотодетектор	240
14.6. Плазмонный материал	240
14.7. Источники.....	242
Глава 15. Новые антенны.....	246
15.1. Антенна с фотопроводящей линзой	247
15.2. Отражающие и передающие решетки.....	248
15.3. Метаповерхности	249
15.4. Нанопотодетекторы	251
15.5. Антенна на кристалле и антенна в корпусе	252
15.6. Орбитальный угловой момент	253
15.7. Источники.....	254
Глава 16. Современные технологии ТГц-диапазона.....	257
16.1. Компоненты ТГц-систем	258
16.1.1. Электронный подход	258
16.1.2. Гибридный и фотонный подходы.....	264
16.2. Системы ТГц-диапазона	266
16.2.1. Терагерцовые системы связи.....	266
16.2.2. ТГц-визуализация и сенсорные системы.....	268
16.3. Проблемы.....	270
16.4. Источники.....	272
Глава 17. Вычислительная техника после закона Мура	280
17.1. Постмуровская эра	280
17.2. Нейроморфные вычисления.....	282
17.3. Квантовые вычисления.....	284
17.4. Новые вычислительные архитектуры.....	286
17.5. Источники.....	288
Глава 18. Новые устройства	290
18.1. Мобильные устройства будущего	290
18.2. Интерфейс мозга и устройств будущего	297
18.3. Новые носимые устройства.....	300
18.4. Источники.....	301
Резюме части IV.....	304

Часть V. Технологии-кандидаты для применения в радиointерфейсе 6G	305
Введение в часть V	306
Глава 19. Структура интеллектуального радиointерфейса	312
19.1. Почему нужен новый радиointерфейс	312
19.2. Обзор существующих технологий	313
19.2.1. Использование спектра и энергоэффективность в NR	314
19.2.2. ИИ и машинное обучение на физическом уровне	315
19.2.3. Машинное обучение на уровне доступа к среде (MAC)	317
19.3. Ожидания от новых разработок и потенциальные направления исследований	318
19.3.1. Персонализированный радиointерфейс с поддержкой ИИ	319
19.3.2. Организация сквозного канала на основе ИИ и нерешенные проблемы	329
19.4. Источники	330
Глава 20. Интеграция наземной и неназемной связи	334
20.1. Почему нужна интеграция наземной и неназемной связи	334
20.2. Обзор существующих решений	336
20.3. Ожидания от новой системы и потенциальные направления исследований	341
20.3.1. Интегрированная многоуровневая сеть	341
20.3.2. Улучшенная неназемная связь	345
20.4. Источники	350
Глава 21. Интеграция сканирования и связи	351
21.1. Почему нужна интеграция сканирования и связи	351
21.2. Обзор существующих решений	353
21.3. Ожидания от новой сети и потенциальные направления исследований	357
21.3.1. Аспекты проектирования интегрированной системы сканирования/связи	358
21.3.2. Конструкция и алгоритмы оборудования радиочастотного сканирования	366
21.4. Источники	371
Глава 22. Новые формы сигналов и схемы модуляции	374
22.1. Почему нужны новые формы сигнала и схемы модуляции	374
22.2. Обзор существующих решений	376
22.2.1. Сигналы с несколькими несущими	377
22.2.2. Формы сигналов с одной несущей	384
22.2.3. Схемы модуляции	386
22.2.4. Формы сканирующих сигналов	388
22.3. Ожидания от новых разработок и потенциальные направления исследований	390
22.4. Источники	394

Глава 23. Новое канальное кодирование	399
23.1. Почему нужно новое кодирование	399
23.2. Схемы канального кодирования	401
23.2.1. История вопроса	401
23.2.2. Целевые KPI канального кодирования 6G	402
23.2.3. Принципы проектирования канального кодирования 6G.....	404
23.3. Совместное кодирование источника и канала	410
23.3.1. Предыстория исследований.....	410
23.3.2. JSCC на основе машинного обучения.....	412
23.3.3. Принципы проектирования JSCC для 6G	414
23.4. Сетевое кодирование на физическом уровне	415
23.4.1. Основы.....	415
23.4.2. Принципы проектирования сетевого кодирования 6G	418
23.5. Источники.....	420
Глава 24. Новый множественный доступ.....	427
24.1. Почему нужен новый множественный доступ	427
24.2. Обзор существующих решений	429
24.2.1. Ортогональный множественный доступ	429
24.2.2. Неортогональный множественный доступ.....	431
24.2.3. Безгрантовый множественный доступ	439
24.3. Ожидания от новых разработок и потенциальные направления исследований	443
24.3.1. Множественный доступ для услуг URLLC большой емкости	443
24.3.2. Множественный доступ для дешевых и маломощных устройств.....	444
24.3.3. Множественный доступ и сверхмассивное подключение	445
24.3.4. Множественный доступ и надежное формирование луча.....	446
24.3.5. Множественный доступ с поддержкой ИИ	446
24.4. Источники.....	447
Глава 25. Ультрамассивный MIMO.....	451
25.1. Почему нужен ультрамассивный MIMO	451
25.2. Обзор существующих решений	452
25.2.1. Технологии MIMO для FR1	452
25.2.2. Технологии MIMO для FR2	454
25.2.3. Совместный MIMO.....	455
25.3. Новые технологии MIMO	458
25.3.2. Реконфигурируемые интеллектуальные поверхности	461
25.3.3. Антенные решетки со сверхбольшой апертурой.....	463
25.3.4. MIMO с искусственным интеллектом.....	464
25.3.5. Другие потенциальные технологии MIMO	466
25.4. Ожидания от новых разработок и потенциальные направления исследований	471
25.4.1. MIMO со сканированием.....	472
25.4.2. Управляемый радиоканал и топология сети.....	474
25.4.3. MIMO на FR2 и терагерцовые частоты	475
25.4.4. Антенные решетки со сверхбольшой апертурой.....	477

25.4.5. MIMO с поддержкой ИИ	478
25.5. Источники.....	481
Глава 26. Интеграция каналов доступа и прямых каналов связи	489
26.1. Почему нужна интеграция каналов	489
26.2. Обзор существующих решений	492
26.3. Ожидания от новых разработок и потенциальные направления исследований	494
26.3.1. Ключевые технологии суперканалов.....	494
26.3.2. Интеграция суперканалов с каналами доступа как единая конструкция.....	495
26.4. Источники.....	497
Резюме части V	499
Часть VI. Новые подходы к проектированию сетевой архитектуры 6G.....	501
Введение в часть VI	502
Глава 27. Технологии для сетевой архитектуры искусственного интеллекта	508
27.1. История вопроса.....	508
27.2. Соображения и принципы проектирования	510
27.2.1. Основные требования.....	510
27.2.2. Ключевое различие между 5G и 6G.....	511
27.3. Особенности архитектуры	512
27.3.1. Общий обзор.....	512
27.3.2. Связь, ориентированная на задачи	515
27.3.3. Глубококонвергентные вычисления и граничные коммуникации	518
27.3.4. Эксплуатация и управление службами искусственного интеллекта.....	521
27.4. Источники.....	523
Глава 28. Архитектура, ориентированная на пользователя	524
28.1. Обзор технологии UCN	524
28.2. Соображения и принципы проектирования UCN.....	525
28.2.1. Уроки, извлеченные из существующих сетей.....	525
28.2.2. Основные требования	528
28.3. Особенности архитектуры.....	533
28.3.1. Децентрализованная архитектура, ориентированная на пользователя.....	533
28.3.2. Слияние физического и кибернетического миров.....	536
28.3.3. Управление цифровыми активами.....	539
28.4. Источники.....	540
Глава 29. Механизмы обеспечения благонадежности	541
29.1. Понятие благонадежности системы	541
29.1.1. От философии к обществу.....	541
29.1.2. От общества к промышленности.....	542
29.2. Благонадежность сложных систем связи	543

29.3. Правила достижения благонадежности.....	546
29.3.1. Основные принципы	546
29.3.2. Цели	547
29.4. Технологии благонадежности	549
29.4.1. Модель многостороннего доверия	550
29.4.2. Технология распределенного реестра	552
29.4.3. Постквантовая криптография.....	554
29.4.4. Автономная безопасность.....	555
29.5. Источники.....	556
Глава 30. Архитектура управления данными в сети 6G	559
30.1. Почему необходимо управление данными	559
30.2. Идеи и принципы проектирования системы управления данными... ..	560
30.3. Особенности архитектуры управления данными	562
30.3.1. Независимый уровень данных	562
30.3.2. Многопользовательские роли в управлении данными.....	564
30.3.3. Ресурс данных	564
30.3.4. Сбор данных	566
30.3.5. Аналитика данных.....	566
30.3.6. Десенсибилизация данных	568
30.4. Источники.....	569
Глава 31. Архитектура многосторонней экосистемы.....	570
31.1. Почему нужна новая архитектура экосистемы.....	570
31.2. Идеи и принципы проектирования.....	572
31.3. Детали многосторонней архитектуры.....	573
31.3.1. Технология распределенного реестра	574
31.3.2. Платформа многостороннего пользования	577
31.3.3. Управление идентификацией.....	578
31.3.4. Управление данными	579
31.3.5. Сетевое управление	581
31.3.6. Поддержка операционной деятельности и бизнеса.....	584
31.4. Источники.....	585
Глава 32. Интегрированная архитектура неназемных сетей	587
32.1. Предпосылки к интеграции сетей	587
32.2. Идеи и принципы проектирования.....	590
32.2.1. Спутниковая группировка	591
32.2.2. Низкая задержка в глобальном масштабе	593
32.2.3. Обеспечение подключения	593
32.2.4. Мультисервисное обслуживание	594
32.3. Особенности архитектуры.....	594
32.3.1. Задержка.....	594
32.3.2. Модели подключения	602
32.3.3. Маршрутизация в космосе	604
32.3.4. Эксплуатация, администрирование и обслуживание	606
32.4. Источники.....	607

Резюме части VI.....	608
Часть VII. Подведение итогов и планы на будущее	609
Глава 33. Экосистема 6G и дорожная карта развития.....	610
33.1. Инициативы и экосистема 6G.....	610
33.1.1. Инициативы ITU-R	610
33.1.2. Региональные мероприятия	611
33.1.3. Мнения промышленности и научных кругов.....	613
33.2. Дорожная карта до 2030 года и далее	616
33.3. Источники.....	618
Предметный указатель	621

Вступление от издательства

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв на нашем сайте www.dmkpress.com, зайдя на страницу книги и оставив комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу dmkpress@gmail.com; при этом укажите название книги в теме письма.

Если вы являетесь экспертом в какой-либо области и заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу http://dmkpress.com/authors/publish_book/ или напишите в издательство по адресу dmkpress@gmail.com.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы обеспечить высокое качество наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг – возможно, ошибку в основном тексте или программном коде, – мы будем очень благодарны, если вы сообщите нам о ней. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Если вы найдете какие-либо ошибки в коде, пожалуйста, сообщите о них главному редактору по адресу dmkpress@gmail.com, и мы исправим это в следующих тиражах.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательства «ДМК Пресс» и Cambridge University Press очень серьезно относятся к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконной публикацией какой-либо из наших книг, пожалуйста, пришлите нам ссылку на интернет-ресурс, чтобы мы могли применить санкции.

Ссылку на подозрительные материалы можно прислать по адресу электронной почты dmkpress@gmail.com.

Мы высоко ценим любую помощь по защите наших авторов, благодаря которой мы можем предоставлять вам качественные материалы.

Авторы

Арашмид Ахаваин

Huawei Technologies Co., Ltd., Канада

Сюэли Ан

Huawei Technologies Co., Ltd., Германия

Хади Балих

Huawei Technologies Co., Ltd., Канада

Алиреза Байесте

Huawei Technologies Co., Ltd., Канада

Жан-Клод Бельфиоре

Huawei Technologies Co., Ltd., Франция

Сяоянь Би

Huawei Technologies Co., Ltd., Китай

Ян Чен

Huawei Technologies Co., Ltd., Канада

Веншуань Данг

Huawei Technologies Co., Ltd., Китай

Меруан Дебба

Huawei Technologies Co., Ltd., Франция

Ицюнь Гэ

Huawei Technologies Co., Ltd., Канада

Хуанхуань Гу

Huawei Technologies Co., Ltd., Канада

Максим Гийо

Huawei Technologies Co., Ltd., Франция

Гаонинг Хэ

Huawei Technologies Co., Ltd., Китай

Цзя Хэ

Huawei Technologies Co., Ltd., Китай

Артур Хеккер

Huawei Technologies Co., Ltd., Германия

Хуан Хуанг

Huawei Technologies Co., Ltd., Китай

Ронг Ли

Huawei Technologies Co., Ltd., Китай

Сюй Ли

Huawei Technologies Co., Ltd., Канада

Янчунь Ли

Huawei Technologies Co., Ltd., Франция

Чжунфэн Ли

Huawei Technologies Co., Ltd., Китай

Хуэй Линь

Huawei Technologies Co., Ltd., Китай

Инбэй Линь

Huawei Technologies Co., Ltd., Германия

Фэй Лю

Huawei Technologies Co., Ltd., Сингапур

Юн Лю

Huawei Technologies Co., Ltd., Китай

Цзяньминь Лу

Huawei Technologies Co., Ltd., Китай

Хэцзя Луо

Huawei Technologies Co., Ltd., Китай

Цзяцзинь Ло

Huawei Technologies Co., Ltd., Китай

Юнся Лу

Huawei Technologies Co., Ltd., Китай

Цзянлей Ма

Huawei Technologies Co., Ltd., Канада

Мэньяо Ма

Huawei Technologies Co., Ltd., Китай

Амин Маарэф

Huawei Technologies Co., Ltd., Канада

Майкл Майер

Huawei Technologies Co., Ltd., Канада

Руи Ни

Huawei Technologies Co., Ltd., Китай

Чэнхуэй Пэн

Huawei Technologies Co., Ltd., Китай

Моррис Репета

Huawei Technologies Co., Ltd., Канада

Сюэлян Ши

Huawei Technologies Co., Ltd., Китай

Хуан Сунь

Huawei Technologies Co., Ltd., Китай

Роб Сан

Huawei Technologies Co., Ltd., Канада

Дэнни Кай Пин Тан

Huawei Technologies Co., Ltd., Китай

Хао Тан

Huawei Technologies Co., Ltd., Китай

Вэй Тан

Huawei Technologies Co., Ltd., Китай

Вен Тонг

Huawei Technologies Co., Ltd., Канада

Гуанцзянь Ван

Huawei Technologies Co., Ltd., Китай

Цзянь Ван

Huawei Technologies Co., Ltd., Китай

Цзюнь Ван

Huawei Technologies Co., Ltd., Китай

Лэй Ван

Huawei Technologies Co., Ltd., Китай

Дэвид Вессель

Huawei Technologies Co., Ltd., Канада

Цзяньцзюнь Ву

Huawei Technologies Co., Ltd., Китай

Сюнь Сяо

Huawei Technologies Co., Ltd., Германия

Сюцян Сюй

Huawei Technologies Co., Ltd., Китай

Сюэцян Янь

Huawei Technologies Co., Ltd., Китай

Ченчен Ян

Huawei Technologies Co., Ltd., Китай

Сюнь Ян

Huawei Technologies Co., Ltd., Китай

Ziming Yu

Huawei Technologies Co., Ltd., Китай

Кун Цзэн

Huawei Technologies Co., Ltd., Китай

Чунцин Чжан

Huawei Technologies Co., Ltd., Китай

Ханг Чжан

Huawei Technologies Co., Ltd., Канада

Хуази Чжан

Huawei Technologies Co., Ltd., Китай

Лицин Чжан

Huawei Technologies Co., Ltd., Канада

Миньюй Чжао

Huawei Technologies Co., Ltd., Китай

Пэйин Чжу

Huawei Technologies Co., Ltd., Канада

Предисловие

Беспроводные революции, безусловно, невозможны без разработки передовых технологий и революционных применений для них. Когда эти две силы встречаются, возникает новое поколение беспроводных технологий. Именно это случилось, когда произошла конвергенция мобильной голосовой и цифровой радиосвязи и когда мобильный интернет объединился с радиочастотной технологией повышенной пропускной способности, адаптированной к протоколу IP. Новая сеть 5G призвана сделать беспроводными все каналы связи – как высоконагруженные, так и сверхвысоконадежные, в конечном итоге соединяя все, чем мы пользуемся, и ускоряя цифровую трансформацию каждого бизнеса. Опираясь на фундамент 5G, беспроводная сеть 6G поставила своей целью повсеместную интеллектуальную революцию. Фактически 6G будет служить нейронной сетью в масштабе человечества и связующим звеном между двумя мирами, физическим и цифровым. Искусственный интеллект (ИИ), основанный на машинном обучении, станет основой 6G, и в этой сфере наше общество полностью перейдет от подключенных людей и подключенных вещей к *подключенному интеллекту* (connected intelligence). Иначе говоря, беспроводная сеть 6G нацелена на предоставление услуги ИИ каждому человеку, дому и бизнесу, что, в свою очередь, приведет к появлению *всеобщего интеллекта* (intelligence of everything). С точки зрения беспроводных технологий, нам впервые предоставляется возможность *ощущать* окружающую среду и предметы, используя радиоволны беспроводной связи. Таким образом, помимо передачи битов данных, беспроводная сеть 6G может служить в качестве *сетевых органов чувств* – сенсоров, извлекающих знания и большие данные из физического мира в реальном времени. Эта извлеченная информация не только будет иметь большое значение для улучшения передачи данных, но также будет способствовать машинному обучению для служб ИИ. Еще один новый аспект, о котором определенно стоит упомянуть, – это расширение группировок *сверхнизкоорбитальных спутников* (VLEO, very low Earth orbit), которые вращаются вокруг Земли на очень малой высоте в очень больших количествах, в конечном итоге образуя «небесные» беспроводные сети 6G. Благодаря им беспроводные услуги и приложения, доступные в любой точке нашей планеты, станут обыденным явлением. Разумеется, это видение будущего чрезвычайно амбициозно, а его реализация существенно повлияет как на наше общество, так и на экономику. Вдобавок ко всему беспроводную сеть 6G не получится создать без применения революционных

технологических инноваций в области связи, вычислений, материалов и алгоритмов. Это одна из причин, по которой наше путешествие в мир будущего растянется на десять лет.

В этой книге мы даем полное представление о беспроводной сети 6G через призму технологий. Наша цель – познакомить заинтересованных читателей с первыми результатами исследований и размышлений о беспроводной сети 6G. Мы не ограничиваемся описанием перспективных беспроводных услуг и технологий; мы также исследуем требования, возможности и практические применения этой сети, уделяя особое внимание новому радиоинтерфейсу и сетевой архитектуре. Эта книга является результатом коллективных усилий нашей исследовательской группы по выработке концепции 6G. Инновации никогда не прекращаются, поэтому нашу работу следует рассматривать только в качестве отправной точки. Траектория развития 6G в конечном итоге будет формироваться экспертами по всему миру, поскольку мы твердо убеждены в том, что открытые инновации и единый глобальный унифицированный стандарт являются фундаментом, на котором будет построен успех 6G. Как и в случае с ее предшественниками, успех беспроводной сети 6G приведет к успеху открытой и глобальной экосистемы.

Наконец, беспроводная революция, которую мы наблюдаем сегодня, продолжается уже более четырех десятилетий, но ее влияние на общество и технологии по-прежнему превосходит все ожидания. Поэтому мы уверены, что потенциал беспроводного будущего невозможно недооценить или переоценить. В связи с этим уместно вспомнить слова Гульельмо Маркони, который заявил в 1932 году, что «вводить ограничения на беспроводную связь опасно».

Словарь сокращений

3GPP	3rd Generation Partnership Project / Партнерский проект третьего поколения
5G	5th generation of mobile communication / 5-е поколение мобильной связи
5GAA	5G automotive association / Автомобильная ассоциация 5G
5G-ACIA	5G alliance for connected industries and automation / Альянс 5G для подключенных отраслей и автоматизации
ACLR	Adjacent channel leakage ratio / Коэффициент утечки по соседнему каналу
ADC	Analog-to-digital converter / Аналого-цифровой преобразователь
AE	Auto encoder / Автоматический кодировщик, автокодировщик
AGV	Automated guided vehicle / Автоматизированный управляемый автомобиль
AI	Artificial intelligence / Искусственный интеллект
AIaaS	AI as a service / ИИ как услуга
AMC	Adaptive modulation and coding / Адаптивная модуляция и кодирование
AMI	Antagonist myoneural interface / Мионевральный (нервно-мышечный) интерфейс антагониста
AMP	Approximate message passing / Приближенная передача сообщения (алгоритм)
ANC	Analog network coding / Кодирование аналоговой сети
ANN	Artificial neural network / Искусственная нейронная сеть
AoA	Angle of arrival / Угол прихода волны (луча)
AoD	Angle of departure / Угол выхода волны (луча)
AoSA	Array of sub-arrays / Решетка из антенных подрешеток
APSK	Amplitude phase shift keying Амплитудно-фазовая манипуляция
AR	Augmented reality / Дополненная реальность
A-SIC	Application-specific integrated circuits / Интегральные схемы для конкретных приложений
ATIS	Alliance for telecommunications industry solutions / Альянс по решениям для телекоммуникационной отрасли
AWGN	Additive white Gaussian noise / Аддитивный белый гауссов шум
B2B	Business-to-business / Корпоративный сегмент, «бизнес для бизнеса»
BCH	Bose-Chaudhuri-Hocquenghem / Код Боуза-Чоудхури-Хоквингема (БЧХ)

BCI	Brain-computer interface / Интерфейс «мозг–компьютер»
BER	Bit error rate / Коэффициент битовых ошибок
BICM	Bit-interleaved coded modulation / Кодированная модуляция с чередованием битов
BLAST	Bell Laboratories layered space time / Многослойное пространство-время Bell Laboratories
BLER	Block error rate / Коэффициент блочных ошибок
BN	Bayesian network / Байесовская сеть
BP	Belief propagation / Распространение доверия
BPSK	Binary phase shift keying / Двоичная фазовая манипуляция
BSS	Business supporting system / Система обеспечения бизнеса
CA	Carrier aggregation / Агрегация несущих частот
CAGR	Compound annual growth rate / Совокупные темпы годового роста
CAPEX	Capital expenditure / Капитальные затраты
CAPL	Concurrent passive and active localization / Одновременная пассивная и активная локализация
CC	Component carrier / Компонентная несущая
CCSA	China communications standards association / Китайская ассоциация стандартов связи
CDMA	Code division multiple access / Множественный доступ с кодовым разделением каналов
CEM	Computational electromagnetics / Моделирование задач электромагнетизма
CJT	Coherent joint transmission / Когерентная совместная передача
CN	Core network / Базовая сеть, опорная сеть
CNN	Convolutional neural network / Сверточная нейронная сеть
CNT	Carbon nanotube / Карбоновые нанотрубки
CoMP	Coordinated multipoint transmission / Скоординированная многоточечная передача
CP	Cyclic prefix / Циклический префикс
CRS	Cell reference signal / Опорный сигнал соты
CS	Compressed sensing / Сжатая выборка
CSI	Channel state information / Информация о состоянии канала
CT	Computed tomography / Компьютерная томография
CU	Control unit / Блок управления
CWS	Continuous wave spectroscopy / Спектроскопия с непрерывным источником
D2D	Device-to-device / Межприборное взаимодействие, связь устройство–устройство
DAC	Digital-to-analog converter / Цифроаналоговый преобразователь
DBN	Dynamic Bayesian network / Динамическая байесовская сеть

DC	Dual connectivity / Двойное подключение
DFT	Discrete Fourier transform / Дискретное преобразование Фурье
DL	Downlink / Нисходящий канал
DLT	Distributed ledger technology / Технология распределенного реестра (блокчейн)
DMRS	Demodulation reference signals / Опорные сигналы демодуляции
DNN	Deep neural network / Глубокая нейросеть
DoF	Degrees of freedom / Число степеней свободы
DoU	Data of usage / Данные об использовании
DPC	Dirty paper coding / «Кодирование на грязной бумаге» (предварительное кодирование Коста)
DRL	Deep reinforcement learning / Глубокое обучение с подкреплением
DRX	Discontinuous reception / Прерывистый прием
DSS	Dynamic spectrum sharing / Динамическое совместное использование спектра
E2E	End-to-end / Сквозной
EEG	Electroencephalography / Электроэнцефалография
EIT	Electromagnetic information theory / Теория электромагнитной информации
ELAA	Extremely large aperture array / Антенная решетка с экстремально большой апертурой
eMBB	Enhanced mobile broadband / Улучшенный мобильный широкополосный доступ
EMG	Electromyography / Электромиография
ESPRIT	Estimation of signal parameters via rotational invariance techniques / Оценка параметров сигнала с помощью методов вращательной инвариантности
EVM	Error vector magnitude / Амплитуда вектора ошибки
FBMC	Filter bank multi-carrier / Банк фильтров в системах со множественными несущими
FDD	Frequency division duplex / Дуплексная связь с частотным разделением каналов
FDMA	Frequency division multiple access / Множественный доступ с частотным разделением
FEC	Forward error correction / Прямая коррекция ошибок
FFT	Fast Fourier transform / Быстрое преобразование Фурье
FinFET	Fin field-effect transistor / Многозатворный полевой транзистор (fin-транзистор)
FMCW	Frequency modulated continuous wave / Нерывное излучение с частотной модуляцией
f-OFDM	Filtered OFDM / OFDM с фильтрованием
FOV	Field of view / Поле зрения

FPGA	Field programmable gate arrays / Программируемая вентильная матрица
FPS	Frames per second / Количество кадров в секунду
GaAs	Gallium arsenide / Арсенид галлия
GAN	Generative adversarial network / Генеративная состязательная сеть
GB	Grant-based / На основе грантов (схема доступа)
GBSM	Geometry-based stochastic model / Стохастическая модель на основе геометрии
GDP	Gross domestic product / Валовый внутренний продукт, ВВП
GDPR	General Data Protection Regulation / Общие правила защиты данных
GEM	General expectation maximization / Общая максимизация ожидания
GEO	Geostationary Earth orbit / Геостационарная околоземная орбита
GF	Grant-free / Безгрантовый (доступ)
GFDM	Generalized frequency-division multiplexing / Обобщенное мультиплексирование с частотным разделением каналов
GNSS	Global navigation satellite system / Глобальная навигационная спутниковая система
GPS	Global positioning system / Спутниковая навигационная система
GPU	Graphics processing unit / Графический процессор
GQD-PD	Graphene photodetector sensitized with semiconducting quantum dots photodetectors / Графеновый фотоприемник, сенсibilизированный полупроводниковыми фотоприемниками на квантовых точках
GSMA	Global System for Mobile Communications Association / Ассоциация глобальной системы мобильной связи
HAP	High-altitude platform / Высотная платформа
HAPs	High-altitude platform stations / Станции на высотных платформах
HARQ	Hybrid automatic repeat request / Гибридный автоматический повторный запрос
HBC	Human bond communication / Поддержка естественных человеческих связей
HBT	Heterojunction bipolar transistors / Биполярные транзисторы с гетеропереходом
HEMT	High electron mobility transistors / Транзисторы с высокой подвижностью электронов
HMI	Human-machine interface / Человеко-машинный интерфейс
HMM	Hidden Markov model / Скрытая марковская модель
Holo-MIMO	Holographic MIMO / Голографический MIMO

IaaS	Infrastructure as a service / Инфраструктура как услуга
IAB	Integrated access and backhaul / Интегрированный доступ и обратная связь
IB	Information bottleneck / Узкое место информационного потока
ICI	Inter-carrier interference / Помехи между несущими
ICT	Information and communications technology / Информационные и коммуникационные технологии
IDFT	Inverse discrete Fourier transform / Обратное дискретное преобразование Фурье
IDMA	Interleave-division multiple access / Множественный доступ с чередованием и разделением каналов
IFFT	Inverse fast Fourier transform / Обратное быстрое преобразование Фурье
IGMA	Interleave-grid multiple access / Множественный доступ с чередованием сетки
IMT	International mobile telecommunication / Международная мобильная связь
InH	Indoor hotspot / Точка доступа в помещении
InP	Indium phosphide / Фосфид индия
IoE	Internet of everything / Интернет всего
IRS	Intelligent reflecting surface / Интеллектуальная отражающая поверхность
ISAC	Integrated sensing and communication / Интегрированное сканирование и связь
JSCC	Joint source and channel coding / Совместное кодирование источников и каналов
ISLR	Integrated sidelobe ratio / Интегрированное отношение боковых лепестков
KPI	Key performance indicator / Ключевой показатель эффективности
LDPC	Low-density parity-check code / Код низкой плотности с контролем четности
LED	Light emitting diode / Светодиод
LEO	Low-earth-orbit / Низкая околоземная орбита
LIDAR	Light detection and ranging / Активный дальномер оптического диапазона, лидар
LIS	Large intelligent surface / Большая интеллектуальная поверхность
LLR	Log likelihood ratio / Логарифмическое отношение правдоподобия
LNA	Low-noise amplifier / Малошумящий усилитель
Log-MAP	Logarithmic maximum a posteriori / Апостериорный логарифмический максимум

LOS	Line-of-sight / Прямая видимость
LSTM	Long short-term memory / Долгосрочная кратковременная память
LTE	Long-term evolution / Долгосрочная эволюция
LTE-A	LTE-advanced / LTE-расширенный
LVDM	Lagrange-Vandermonde division multiplexing / Мультиплексирование с разделением по Лагранжу–Вандермонду
M2M	Machine-to-machine / Межмашинное взаимодействие, межмашинные коммуникации
MA	Multiple access / Множественный доступ
MAC	Media access control / Управление доступом к среде
MARL	Multi-agent reinforcement learning / Многоагентное (распределенное) обучение с подкреплением
MCS	Modulation and coding scheme / Схема модуляции и кодирования
MDP	Markov decision process / Марковский процесс принятия решений
MEC	Mobile edge computing / Мобильные граничные вычисления
MEO	Medium earth orbit / Средневысотная околоземная орбита
MIMO	Multiple input – multiple output / Многоканальный вход, многоканальный выход
ML	Machine learning / Машинное обучение
MMSE	Minimum mean square error / Минимальная среднеквадратическая ошибка
mMTC	Massive machine type of communication / Поточковая связь машинного типа
mm Wave	Millimeter wave / Миллиметровые волны
MNO	Mobile network operator / Оператор мобильной сети
MOSFET	Metal-oxide-semiconductor field-effect transistor / Полевой транзистор со структурой металл–оксид–полупроводник
MP	Matching pursuit / Аппроксимация с преследованием
MR	Mixed reality / Смешанная реальность
MRI	Magnetic resonance imaging / Магнитно-резонансная визуализация (томография)
MTP	Motion-to-photon / «Инициация фотона движением» (инженерный жаргон)
MUD	Multi-user detection / Многопользовательское детектирование
MUSA	Multi-user shared access / Многопользовательский совместный доступ
MUSIC	Multiple signal classification / Классификация групповых сигналов
MUST	Multi-user superposition transmission / Многопользовательская передача с наложением

NASA	National aeronautics and space administration / Национальное управление по авиации и исследованию космического пространства
NB-IoT	Narrowband Internet of things / Широкополосный интернет вещей
NCJT	Non-coherent joint transmission / Некогерентная совместная передача
NDF	Narrowband digital forward / Узкополосное цифровое распространение (информации)
NEP	Noise-equivalent power / Эквивалентная мощность шума
NFI	Near-field imaging / Визуализация в ближнем поле
NFV	Network functions virtualization / Виртуализация сетевых функций
NIST	National Institute of Standards and Technology / Национальный институт по стандартизации и технологии
NLP	Natural language processing / Обработка (понимание) естественного языка
NMR	Nuclear magnetic resonance / Ядерный магнитный резонанс
NOMA	Non-orthogonal multiple access / Неортогональный множественный доступ
NR	New radio / Новое радио
NTN	Non-terrestrial network / Неназемная сеть
OA&M	Operations, administration, and maintenance / Эксплуатация, управление и техническое обслуживание
OAM	Orbital angular momentum / Орбитальный угловой момент
OAMP	Orthogonal approximate message passing / Ортогональная приближенная передача сообщения
OFDM	Orthogonal frequency-division multiplexing / Ортогональное мультиплексирование с частотным разделением
OFDMA	Orthogonal frequency-division multiple access / Множественный доступ с ортогональным частотным разделением
OMA	Orthogonal multiple access / Ортогональный множественный доступ
OMP	Orthogonal matching pursuit / Ортогональная аппроксимация с преследованием
OOBE	Out-of-band emission / Внеполосное излучение
OPEX	Operational expenditure / Операционные расходы
OQAM	Offset QAM / Смещение QAM
OSS	Operation support system / Система обеспечения эксплуатации
OT	Operational technology / Эксплуатационная технология
OTFS	Orthogonal time-frequency space / Ортогональное пространство время–частота
OTP	One-time pad / Одноразовый блокнот

OTT	Over-the-top / Сверхвысокий уровень, оказание услуги поверх сети мобильной связи
OWC	Optical wireless communication / Оптическая беспроводная связь
P2P	Point-to-point / Прямая коммуникация, сквозная передача
PA	Power amplifier / Усилитель мощности
PaaS	Platform as a service / Платформа как услуга
PAE	Power-added efficiency / КПД преобразования энергии постоянного тока в энергию излучения
PAPR	Peak-to-average power ratio / Отношение пикового уровня мощности сигнала к среднему
PAR	Packet arrival rate / Частота поступления пакетов
PC	Photonic crystal / Фотонный кристалл
PCA	Photoconductive antenna / Фотопроводящая антенна
PCE	Photo conversion efficiency / Эффективность фотопреобразования
PDMA	Pattern division multiple access / Множественный доступ с разделением по шаблону
PDR	Packet drop rate / Коэффициент отбрасывания пакетов
PDSCH	Physical downlink shared channel / Физический уплотненный канал нисходящей линии связи
PHY	Physical layer / Физический уровень
PLE	Path loss exponent / Степень потерь в тракте
PNC	Physical-layer network coding / Кодирование на уровне физической сети
PPD	Pixel per angular degree / Количество пикселей на угловой градус
PQC	Post quantum cryptography / Постквантовая криптография
PSLR	Peak to sidelobe ratio / Отношение пиковых и боковых лепестков
QAM	Quadrature amplitude modulation / Квадратурная амплитудная модуляция
QD	Quantum dot / Квантовая точка
QHA	Quadrifilar helix antenna / Четырехзаходная спиральная антенна
QoS	Quality of service / Качество обслуживания
RACH	Random access channel / Канал произвольного доступа
RAN	Radio access network / Сеть радиодоступа
RCN	Radio computing node / Вычислительный узел радиосети
RE	Resource element / Ресурсный элемент
ReLU	Rectified linear unit / Линейный выпрямленный блок (функция)
RF	Radio frequency / Радиочастота
RIS	Reconfigurable intelligent surface / Реконфигурируемая интеллектуальная поверхность

RL	Reinforcement learning / Обучение с подкреплением
RM	Reed-Muller / Рида–Маллера (код)
RNN	Recurrent neural network / Рекуррентная нейросеть
RRC	Radio resource control / Протокол управления радиоресурсами
RRH	Remote radio head / Вынесенная радиоголовка (радиоблок)
RS	Reference signal / Опорный сигнал
RSMA	Resource spread multiple access / Множественный доступ с расширением ресурсов
RSRP	Reference signal received power / Принимаемая мощность эталонного сигнала
RTT	Round-trip time / Время приема-передачи
SAE	Society of Automotive Engineers / Ассоциация инженеров автомобилестроения
SCL	Successive cancellation list / Список последовательной отмены
SCMA	Sparse code multiple access / Множественный доступ с разреженным кодом
SDG	Sustainable development goal / Цель устойчивого развития
SDMA	Space division multiple access / Множественный доступ с пространственным разделением
SDN	Software-defined network / Программно-конфигурируемая сеть
SEEG	Stereoelectroencephalography / Стереозлектроэнцефалография
SeGW	Security gateway / Шлюз безопасности
SIC	Successive interference cancellation / Последовательное подавление помех
SiN	Silicon neuron / Кремниевый нейрон
SLAM	Simultaneous localization and mapping / Одновременные локализация и картографирование
SLM	Spatial light modulator / Пространственный модулятор света
SNR	Signal-to-noise ratio / Отношение сигнал–шум
SP-OFDM	Spectrally precoded OFDM / OFDM с предварительным спектральным кодированием
SPP	Surface plasmon polariton / Поверхностный плазмон-поляритон
SRR	Split-ring-resonator / Разрезной кольцевой резонатор
SRS	Sounding reference signal / Зондирующий опорный сигнал
SSB	Synchronization signal blocks / Блоки сигнала синхронизации
STBC	Space-time block codes / Пространственно-временное блочное кодирование
SUL	Supplementary uplinks / Дополнительные восходящие каналы связи
SWC	Surface wave communication / Связь на поверхностных волнах
TCA	Tightly coupled array / Сильно связанный массив
TDD	Time division duplex / Дуплексная связь с временным разделением

TDL-C	Tapped-delay line channel / Канал на основе линии задержки с отводами
TDMA	Time division multiple access / Множественный доступ с временным разделением каналов
TDS	Time domain spectroscopy / Спектроскопия во временной области
THz	Terahertz / Терагерц, ТГц
TN	Transport network / Транспортная сеть
TPC	Terminal-pipe-cloud / Архитектура «оконечное устройство – сетевой канал – облако»
TR	Time reversal / Обращение времени, перестановка во времени
TRP	Transmit/receive point / Точка (узел) приема/передачи
TTI	Transmission time interval / Интервал передачи
UAV	Unmanned aerial vehicle / Беспилотный летательный аппарат
UCN	User-centric network / Ориентированная на пользователя сеть
UCNC	User-centric no cell / Ориентация на пользователя без ячейки
UDR	User data repository / Репозиторий данных пользователя
UFMC	Universal filtered multi-carrier / Универсальная фильтрация с несколькими несущими
UN	United Nations / Организация Объединенных Наций
URLLC	Ultra-reliable low-latency communication / Сверхнадежная система связи с малой задержкой
UTC-PD	Uni-traveling-carrier photodiode / Фотонный диод с односторонним движением носителя
UWB	Ultra-wideband / Сверхширокополосный
V2I	Vehicle-to-infrastructure / Транспортное средство – инфраструктура
V2V	Vehicle-to-vehicle / Связь транспортное средство – транспортное средство
V2X	Vehicle-to-everything / Транспортное средство – внешний мир
VAE	Variational auto encoder / Вариационный автокодировщик
VLC	Visible light communication / Связь в видимой части спектра
VLEO	Very low earth orbit / Очень низкая околоземная орбита
VLSI	Very-large-scale integration / Интеграция сверхвысокого уровня
VR	Virtual reality / Виртуальная реальность
W-OFDM	Windowed OFDM / Оконное OFDM
WRC	World Radiocommunication Conference / Всемирная конференция по радиосвязи
XOR	Exclusive OR / Исключающее ИЛИ

Часть I



Вступление

Глава 1

Мобильная связь до 2030 года и далее

Сегодня мы находимся на гребне огромной цифровой волны, за которой следует непрерывный поток инноваций и новых возможностей. Перемены затронут буквально все – каждого человека, его семью, автомобиль, изменят целые отрасли промышленности во многих странах, переопределят то, как мы живем, работаем, учимся и бережем здоровье. Сегодня, когда глобальное развертывание сетей 5G набирает ход и открывает невообразимые возможности, мы наблюдаем, как 5G меняет все аспекты нашей жизни, отрасли и общества. Давайте вместе заглянем в будущее до 2030 года и далее и попробуем представить, чего мы можем ожидать от следующего поколения мобильной связи.

1.1. Эволюция систем мобильной связи

Системы мобильной связи претерпели значительные изменения с 1980-х годов, при этом новое поколение появляется приблизительно каждые 10 лет. В то же время на достижение зрелости основных услуг, предоставляемых сетями мобильной связи, и полноценного ввода в эксплуатацию новых частотных диапазонов обычно уходят два поколения – или 20 лет. Как показано на рис. 1.1, каждое новое поколение обеспечивает значительный рост возможностей по сравнению с предыдущим поколением за счет внедрения новых технологий, новых принципов проектирования и новых архитектур в сетях радиодоступа и базовых сетях.

Для сетей 2G и 3G основной движущей силой роста были мобильные подписки, в первую очередь ориентированные на услуги голосовой связи. Когда уровень проникновения мобильных телефонов и использования голосовых услуг достиг точки насыщения, бизнес-модель, основанная на подписке, начала выходить на плато.

В промежутке между поколениями 3G и 4G быстро рос объем услуг передачи данных, и доминирующей услугой для 4G стала мобильная широкополосная связь. За последние 10 лет значительные достижения в области мобильной передачи данных радикально изменили образ жизни общества. Например, смартфоны со всевозможными приложениями глубоко проникли в каждый аспект жизни многих людей. Поэтому операторы сетей 4G в основном полагаются на объем трафика, а не на подписки – рост потребления трафика на душу населения способствует росту бизнеса.

Сеть 4G оказала огромное влияние на нашу жизнь – технологические возможности, которые она принесла, привели к революционным переменам в нашей повседневной жизни. Одним из показательных примеров этого в Китае является массовый переход от наличных денег к онлайн-платежам. Сегодня и молодые, и пожилые люди предпочитают использовать такие способы оплаты, как AliPay и WeChat Pay, считая, что им удобнее оплачивать что угодно, от покупки продуктов до парковки, при помощи смартфона, не имея в кармане наличных. Другой пример – рост социальных сетей. Теперь любой может делиться фотографиями и видео со смартфона в режиме реального времени, с кем угодно и где угодно, фактически становясь ведущим персональной службы новостей, что ускоряет распространение информации.

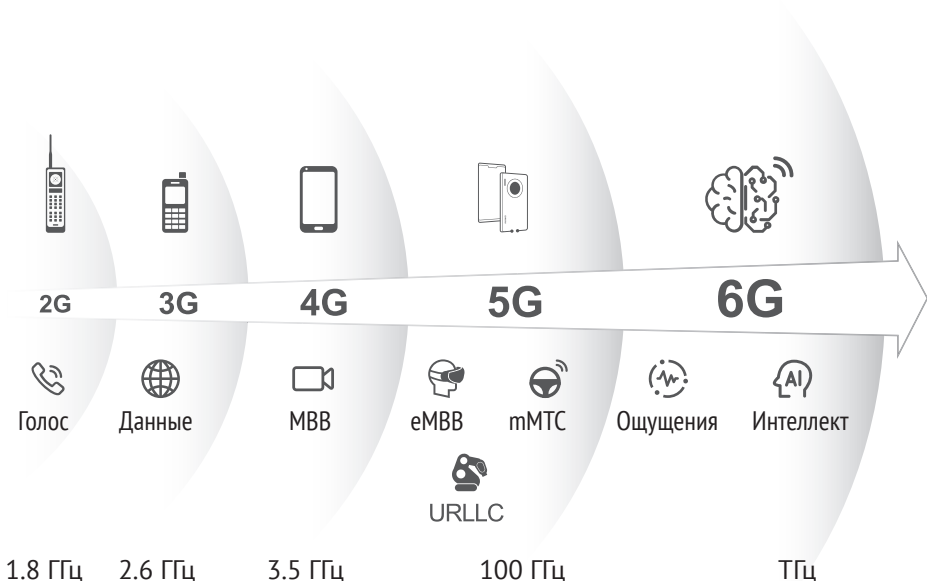


Рис. 1.1. Эволюция мобильных сетей

Эта тенденция сохраняется в 5G, поскольку появляются все больше и больше приложений, критически зависимых от высокой пропускной способности. К таким приложениям относятся видео высокой четкости и мультимедийные приложения с эффектом присутствия, такие как *дополненная реальность* (AR), *виртуальная реальность* (VR) и *смешанная реальность* (MR). В настоящее время в мире используется около 3,8 млрд смартфонов. Мы ожидаем, что к 2025 году это число достигнет 8 млрд, и к этому времени будет более 6,5 млрд пользователей мобильного интернета, 80 % из которых будут иметь мобильный широкополосный доступ. Вдобавок к этому будет 440 млн пользователей AR/VR, и 40 % автомобилей будут подключены к сети.

По мере стандартизации *узкополосного интернета вещей* (NB-IoT), промышленного интернета вещей и *полностью подключенных транспортных средств* (vehicle to everything, V2X) мобильные сети сместили акцент с подключения людей с помощью *расширенной мобильной широкополосной связи* (enhanced mobile broadband, eMBB) к подключению вещей с помощью *сверхнадежных систем связи с малой задержкой* (ultra-reliable low-latency communication, URLLC) и *поточковой межмашинной связи* (massive machine type of communication, mMTC). Это, в свою очередь, должно послужить фундаментом цифровизации бизнеса, без которой невозможна следующая волна экономического роста. Коммерческое развертывание 5G первоначально было сосредоточено на потребительских услугах, но более поздние версии стандарта 3GPP 5G (например, версии 16 и 17) были направлены на стимулирование развития вертикальных приложений, таких как V2X и промышленный IoT. С целью обеспечения реализации всех уровней технологии беспилотного управления автомобилями и индустрии 4.0 для широкого круга предприятий и отраслей мобильная индустрия тесно сотрудничает с такими консорциумами, как 5G-ACIA [1] и 5GAA [2]. Ожидается, что автономное вождение 4-го уровня¹ будет реализовано примерно в 2024 году, и многие ждут повышения эффективности грузоперевозок в связи с распространением технологии V2X. Оптимизированные бизнес-процессы и эффективность производства станут ключевыми стимулами будущего роста валового внутреннего продукта (ВВП).

В то время как 5G открывает двери для *всеобъемлющего интернета* (Internet of everything, IoE), мы прогнозируем, что сеть 6G – наследник 5G – станет платформой для *подключенного интеллекта*, где мобильная сеть соединяет огромное количество *интеллектуальных устройств* – и соединяет их интеллектуально. Ожидается, что следующая волна цифровизации создаст больше инноваций, направленных на

¹ Уровень автономности, при котором наличие водителя в кабине не является обязательным, однако в некоторых случаях он может присутствовать в ней и брать управление на себя. – Прим. перев.

удовлетворение всех аспектов наших потребностей, какие мы только можем представить. С помощью *искусственного интеллекта* (ИИ) и *машинного обучения* (machine learning, ML) мы сможем установить связь между физическим и цифровым мирами в реальном времени, что позволит нам собирать, извлекать и получать доступ к большим объемам информации и знаний в режиме реального времени и таким образом превратить подключенный мир вещей в подключенный интеллект. Кроме того, сенсорное сканирование и распределенные вычисления вместе с передовыми интегрированными *неназемными сетями* (non-terrestrial network, NTN) и технологиями беспроводной связи на короткие расстояния заложат основу для построения интеллектуальных сетей мобильной связи в будущем.

1.2. Ключевые факторы потребности в 6G

Мы выделяем три ключевых фактора, определяющих потребность в новом поколении подключенного интеллекта (рис. 1.2). Давайте рассмотрим эти факторы подробнее.



Рис. 1.2. Ключевые факторы потребности в 6G – поколении подключенного интеллекта

Фактор 1: новые применения и бизнесы

Сегодня доход бизнеса зависит от роста потребления трафика на одну подписку². На рис. 1.3 показан ожидаемый рост среднего глобального мобильного трафика на подписку в месяц (сплошные линии) и количества подписок (столбцы) с 2020 по 2030 год для *межмашинных коммуникаций* (machine-to-machine, M2M) и устройств, не поддерживающих M2M. Данные, использованные на рисунке, были взяты из отчета ITU-R M.2370 [3]. По рисунку видно, что рост подписок на смартфоны до-

² Под подпиской (subscription) здесь подразумевается абонентский договор с оператором мобильной связи. – Прим. перев.

стиг насыщения в 2020 году; *совокупный годовой темп роста* (compound annual growth rate, CAGR) с 2020 по 2030 год составит около 6 %. Кроме того, Ассоциация глобальной системы мобильной связи (Global System for Mobile Communications Association, GSMA) ожидает, что уровень проникновения уникальных мобильных абонентов увеличится всего на три процентных пункта, то есть с 67 % до 70 %, с 2019 по 2025 год [4]. Тем не менее, несмотря на это, ожидается, что трафик мобильных данных на подписку MBV вырастет в 50 раз за 10-летний период, с 5,3 ГБ в месяц в 2020 году до 257 ГБ в месяц в 2030 году.

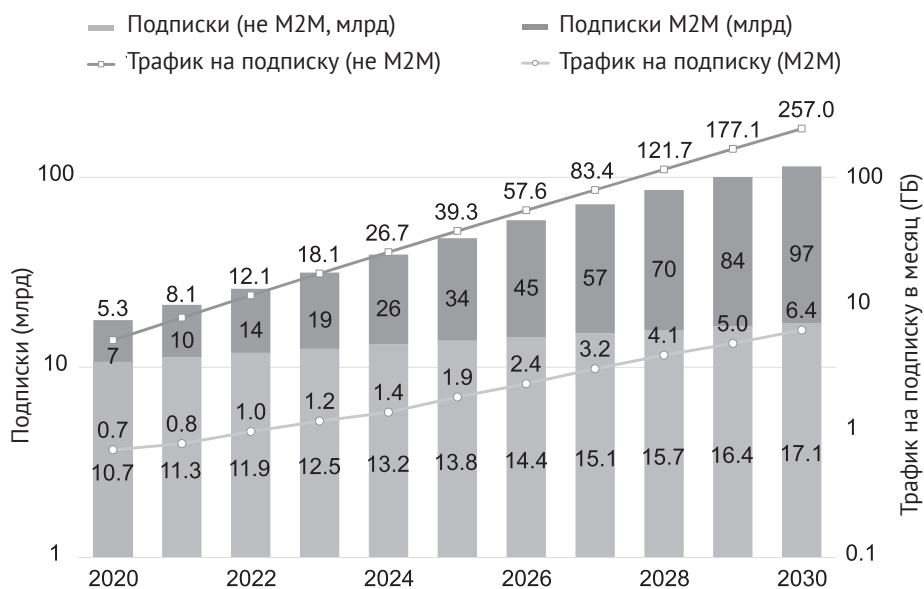


Рис. 1.3. Прогноз роста подписок и мобильного трафика с 2020 по 2030 год

Платформа 5G поддерживает широкий спектр приложений, требующих высокой пропускной способности, что увеличило объем трафика и повысило спрос на пропускную способность сети. В эпоху 6G появятся еще больше приложений, и облачные сервисы *расширенной реальности* (extended reality, XR) вместе с тактильной обратной связью и голографическим дисплеем, вероятно, станут основными приложениями для демонстрации VR-фильмов с круговым обзором, удаленных сервисов с поддержкой AR, виртуальных учебных курсов в 3D, передачи тактильных ощущений, телемедицины и удаленных телеопераций. В отчете Huawei о глобальном видении отрасли [5] прогнозируется, что к 2025 году наголовными устройствами VR/AR будут пользоваться свыше 337 млн пользователей, в то время как более 10 % предприятий будут использовать

технологии AR/VR для бизнес-операций, и эти показатели обязательно увеличатся к 2030 году. По мере увеличения количества и популярности облачных XR-приложений, а также увеличения разрешения, размера и частоты кадров требования к пропускной способности и задержке могут превзойти возможности сетей 5G даже с учетом их эволюции. Экспоненциальный рост потребности в трафике в пересчете на одно устройство вместе с высокими требованиями к задержке и надежности станет серьезной проблемой при проектировании сети 6G с точки зрения необходимой огромной ресурсоемкости. Кроме того, безлимитные тарифные планы на передачу данных, которые предлагают многие операторы, стали ключевой бизнес-моделью и тоже будут способствовать потенциальному росту потребления данных.

По рис. 1.3 также видно, что в 2030 году устройств M2M будет примерно в 13 раз больше, чем в 2020 году, и количество подключений к интернету вещей как для предприятий, так и для потребителей будет увеличиваться. В своем отчете по мобильной экономике за 2020 год [4] GSMA прогнозирует, что к 2024 году корпоративный интернет вещей обгонит потребительский. Как следствие ИИ станет движущей силой для всех видов автоматизации и будет потреблять большие объемы данных для преобразования ситуационной осведомленности в реальном времени в принимаемые решения (также в режиме реального времени). Огромное количество широкополосных датчиков будет задействовано в таких сценариях применения, как умный дом, умное здоровье, умный автомобиль, умный город, умное здание и умная фабрика, чтобы получить огромный объем данных, необходимых для обучения и работы ИИ. Большие данные являются основой успеха машинного обучения, и они становятся основной причиной увеличения пропускной способности сети 6G на порядок. Кроме того, новые технологические решения, такие как *сети как сенсоры* (networks-as-sensors) и неназемная связь, станут неотъемлемой частью мобильных систем 6G, что позволит осуществлять мониторинг окружающей среды и съемку на больших территориях в режиме реального времени с помощью еще большего количества сенсоров.

В дополнение к этому высокопроизводительные промышленные приложения интернета вещей предъявляют высокие требования к пропускной способности беспроводной сети с точки зрения детерминированной задержки и джиттера и требуют наличия гарантированного доступа и надежности. Например, высокая пропускная способность необходима для оперативного управления и контроля, а также для координации движений и совместной работы нескольких роботов. Такие варианты использования можно рассматривать как факторы развития экстремального быстрого действия, которым должны обладать сети 6G.

Фактор 2: расширение искусственного интеллекта

В ближайшие десятилетия цифровая экономика по-прежнему будет основной движущей силой экономического роста во всем мире и продолжит расти намного быстрее, чем мировая экономика в целом. В 2019 году цифровая экономика росла в 3,5 раза быстрее, чем мировая, и достигла 15,60 трлн долларов США, что составляет 19,7% мировой экономики. Ожидается, что к 2025 году этот показатель достигнет 24,3%. С точки зрения инвестиционного плеча, анализ показывает, что за последние 30 лет увеличение цифровых инвестиций на 1 доллар приводит к увеличению ВВП на 20 долларов по сравнению со средним показателем 1:3 для нецифровых инвестиций [6].

Являясь одним из наиболее динамичных секторов индустрии информационных и коммуникационных технологий (information and communications technology, ICT), мобильная связь оказала глубокое влияние на жизнь людей, помогла уменьшить цифровой разрыв и внесла значительный вклад в общую продуктивность общества и экономический рост. Ожидается, что к 2024 году мобильные технологии и услуги будут генерировать 4,9% мирового ВВП (приближаясь к 5 трлн долларов США), при этом большее количество отраслей получают выгоду от повышения производительности и эффективности, вызванного более широким внедрением мобильных услуг [4].

Мы полагаем, что эта тенденция сохранится до 2030 года и далее. В частности, по мере того как всеобъемлющий интеллект станет ключевым фактором, способствующим развитию бизнес-моделей и экономических моделей в будущем, изменения парадигмы в радиотехнологиях и сетевой архитектуре будут определяться четырьмя критическими факторами, как показано на рис. 1.4.

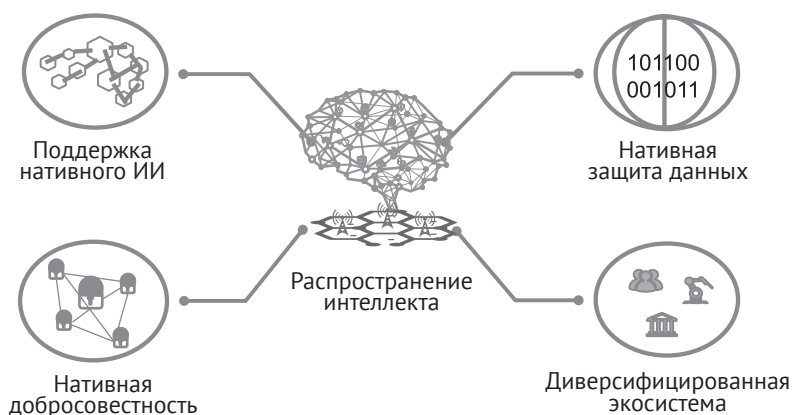


Рис. 1.4. Бизнес-факторы, порождаемые всеобъемлющим интеллектом и большими данными

- **Нативный³ искусственный интеллект:** хотя конструкция базовой сети 5G поддерживает интеллектуальные функции, вводя новый тип сетевых функций (например, аналитику сетевых данных), возможности использования ИИ в сетевых операциях и управлении ограничены. Вместо этого 5G предоставляет ИИ как услугу *сверхвысокого уровня* (OTT). И наоборот, когда речь идет о 6G, системы мобильной связи спроектированы с нативной поддержкой ИИ и машинного обучения – не только в качестве базовой функциональности, но и для обеспечения оптимальной эффективности. С точки зрения архитектуры, запуск распределенного ИИ на периферии может обеспечить максимальную производительность, одновременно решая проблемы владения данными отдельных лиц и предприятий, а также соблюдая региональные и национальные нормативные ограничения. Встроенная поддержка ИИ в 6G направлена на предоставление услуг ИИ в любом месте и в любое время и будет постоянно улучшать производительность системы и удобство работы пользователей за счет постоянной оптимизации. Следовательно, по-настоящему всеобъемлющий интеллект в сочетании с глубоко конвергентными системами ICT, обеспечивающими разнообразные возможности подключения, вычислений и ресурсов периферийного хранения данных, станет естественным свойством сети 6G. Соответствующие возможности (такие как алгоритмы, нейронные сети, базы данных и API) должны быть интегрированы в систему 6G как часть реализации сети. Сетевая архитектура 6G с нативной поддержкой ИИ принесет «сетевой ИИ», отойдя от сегодняшнего централизованного «облачного ИИ».
- **Нативная защита данных:** в дополнение к средствам безопасности, разработанным в 5G и более ранних поколениях, соблюдение конфиденциальности будет критическим требованием и принципом проектирования сетей 6G. Защита конфиденциальности во всех аспектах работы сети и использования данных 6G будет иметь исключительное значение. С одной стороны, ключевым требованием к сети является владение данными и право на доступ к данным, и возникает вопрос, как сетевая архитектура может обеспечить защиту конфиденциальности. С другой стороны, встроенный ИИ нуждается в возможности обрабатывать данные и получать доступ к ним распределенным образом. Вместо того чтобы полагаться на поставщиков сетевых и прикладных услуг для обеспечения защиты данных, мы ожидаем, что поль-

³ Нативный (native) с технической точки зрения – изначально заложенный на уровне конструкции и реализации и являющийся неотъемлемым, «врожденным» конструктивным элементом или функцией системы. – *Прим. перев.*

зователи – которые могут быть как людьми, так и машинами – будут наделены полномочиями субъектов данных с правами управления и использования. В устройстве системы следующего поколения гарантия конфиденциальности данных должна быть главным приоритетом, а не дополнительной функцией. Подобная система должна обеспечивать надлежащие права субъектов данных, гарантировать контроль и обработку данных и включать поддержку таких политик, как *Общий регламент по защите данных* (GDPR), тем самым определяя фундаментальные руководящие принципы для разработки и использования технологий в будущем.

- **Нативная добросовестность:** для соответствия разнообразным вариантам использования и рынкам важно иметь настраиваемую, поддающуюся проверке и измеряемую *добросовестность* (trustworthiness). Владение и функционирование, основанное на законах, характерное для текущего и предыдущих поколений сетей, будет перерастать в многостороннее доверие со стороны партнеров, участников и действующих лиц. Этот бизнес-фактор будет способствовать созданию новой архитектуры добросовестности, включающей в себя множество факторов. Инклюзивная модель многостороннего доверия будет более насыщенной, чем модель одиночного доверия. Помимо ориентации на будущее, архитектура многостороннего доверия должна включать защиту, конфиденциальность, отказоустойчивость, безопасность и надежность.
- **Диверсифицированная экосистема:** три основных элемента ИИ – это данные, алгоритмы и вычисления. Однако отдельные компании могут не обладать полными возможностями для достижения цифровой трансформации с помощью всеобъемлющего интеллекта и быстрых технологических инноваций. Следовательно, для достижения успеха в бизнесе важно создать открытую, устойчивую и многостороннюю экосистему для совместной работы.

Кроме того, ожидается, что по мере постепенного расширения возможностей 5G в 2020-х годах вертикальный рынок беспроводной связи будет расти. Игроки в секторах ICT и *операционных технологий* (operational technology, OT) ищут новые способы сотрудничества, чтобы генерировать новые источники дохода. По мере приближения к эре 6G было бы полезно обзавестись универсальным решением ICT, способным предложить всеобъемлющую перспективу для всех отраслей и тем самым ускорить сотрудничество и конвергенцию секторов ICT и OT.

Первая волна коммерческого использования 6G, вероятно, приведет к росту как потребительского, так и вертикального рынков.

Фактор 3: социальная ответственность и устойчивость

С точки зрения социальной ответственности и устойчивости возьмем в качестве примера пандемию COVID-19. Этот глобальный кризис затронул почти всех во всем мире, за это время индустрия ICT активизировалась и сыграла значительную роль в спасении человеческих жизней. Технологии беспроводной связи и определения местоположения использовались для отслеживания инфицированных пациентов и распространения заболевания, а необходимость минимизации воздействия на медицинский персонал привела к значительным инновациям в 5G для автоматизации здравоохранения. Чтобы ограничить скопление людей и одновременно сохранить экономическую активность, многие страны использовали различные приложения для работы по беспроводным сетям, включая такие приложения, как телемедицина, телеобразование, удаленный офис, промышленная автоматизация и электронная коммерция. Поскольку мобильная индустрия поддерживала различные секторы мировой экономики и общества во время этой пандемии, найденные в результате варианты использования внесли свой вклад в будущую технологическую эволюцию мобильной индустрии.

Мобильные сети способны изменить бизнес, образование, государственное управление, здравоохранение, сельское хозяйство, производство и окружающую среду, а также наши способы взаимодействия с другими людьми. Они стали одной из ключевых движущих сил социальной эволюции и могут изменить само наше существование. Согласно GSMA [4], мобильная связь занимает центральное место в достижении целей устойчивого развития (sustainable development goals, SDG), установленных Организацией Объединенных Наций в 2015 году для преобразования нашего мира, и является мощным инструментом для достижения этих целей. Индустрия мобильной связи оказала значительное влияние на все 17 SDG – влияние, которое продолжает расти, обеспечивая прочную основу для цифровой экономики и выступая в качестве катализатора для разнообразного и инновационного спектра услуг.

Huawei и ООН совместно работали над контрольным показателем SDG в области ICT [7], чтобы количественно измерить степень, в которой развитие ICT способствует прогрессу в достижении целей устойчивого развития. Оценка 2019 года показала сильную корреляцию между зрелостью ICT и прогрессом в достижении SDG, достигнув $R^2 = 0,86$. Среди них, как показано на рис. 1.5, SDG 3 (хорошее здоровье и благополучие) и SDG 4 (качественное образование) показали самую сильную

корреляцию с ICT, что указывает на области, в которых цифровые технологии обладают наибольшим потенциалом для ускорения развития экономики страны.

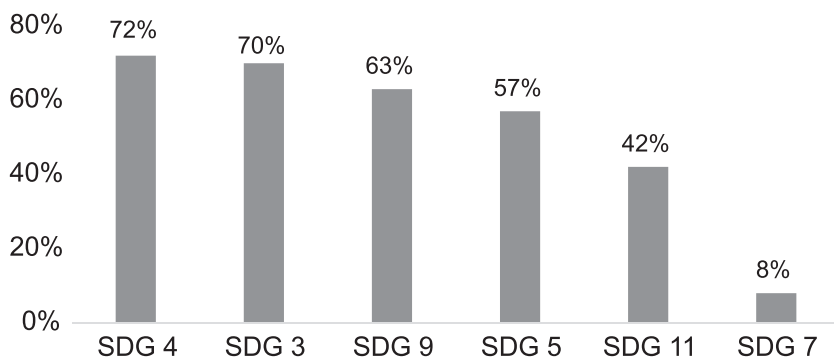


Рис. 1.5. Основные корреляции между отдельными целями устойчивого развития и уровнем ICT. Из совместного отчета Huawei и ООН [7].

SDG 4: качественное образование;

SDG 3: хорошее здоровье и благополучие;

SDG 9: промышленность, инновации и инфраструктура;

SDG 5: гендерное равенство; SDG 11: устойчивые города и сообщества;

SDG 7: доступная и чистая энергия

Что касается экологической устойчивости, развитие ICT имеет решающее значение для достижения целей SDG 7 (доступная и чистая энергия) и SDG 11 (устойчивые города и сообщества). Мир становится все более урбанизированным, и прогнозы показывают, что к 2030 году 5 млрд человек будут жить в городах, занимая всего 3 % пригодной для проживания земли, при этом на их долю будет приходиться 60–80 % мирового потребления энергии [8]. По мере достижения все более высокой энергоэффективности на бит ICT станут приближаться к нулевому балансу выбросов углерода, а решения на базе ICT (например, интеллектуальные сети, интеллектуальная логистика и интеллектуальная промышленность) помогут преобразовать мир в направлении более устойчивого и энергоэффективного будущего [9].

По состоянию на 2020 год около 90 % людей во всем мире имеют доступ к сетям 3G или 4G. Что касается оставшихся 10 %, то и 5G, и 6G будут стремиться подключить их с помощью таких технологий, как спутниковая связь. Например, 5G уже пыталась интегрировать технологии неназемного доступа в 5G New Radio (NR), в то время как амбициозные планы по созданию группировки на сверхнизкой околоземной орбите (very low earth orbit, VLEO) – созвездия из десятков тысяч спутников VLEO – могут быть реализованы в эпоху 6G. Услуги, приложения и контент, предоставляемые мобильными сетями, помогают расширить финансовую

и социальную интеграцию, в то время как новые технологии, такие как интернет вещей, большие данные, искусственный интеллект и машинное обучение, интегрируются в сетевую инфраструктуру, демонстрируя свой потенциал революционного преобразования общества и среды. Сильная корреляция между ICT и SDG означает, что мы должны учитывать все аспекты SDG при проектировании систем и сетей связи 6G.

1.3. Общая концепция

Мобильная связь за 40 лет полностью изменила мир. Сегодня мы сильно зависим от беспроводной связи как в работе, так и в жизни; она стала ключевым фактором цифровой трансформации каждого бизнеса. По мере распространения 5G – современного поколения беспроводной связи – к сети будет подключено все, включая отдельных людей; с помощью этой гиперкоммуникации мы сможем автоматизировать все аспекты жизни общества. Более того, стремление к инновациям в беспроводной связи набирает обороты. Как было сказано в [10], «Наука не имеет границ, беспроводная связь по-настоящему безгранична».

Инновации беспроводных технологий, несомненно, продолжатся в течение ближайших 10 лет, и двумя основными катализаторами новых технологических прорывов станут рост ИИ, основанного на машинном обучении, и создание *цифровых двойников* (то есть проекций объектов физического мира на кибернетический). Возникшая в результате этого прогресса сеть 6G изменит правила игры как с точки зрения экономики, так и с точки зрения общества – она заложит прочную основу для будущего всеобъемлющего интеллекта.

Сеть 6G станет следующим поколением беспроводной связи, перейдя от эпохи подключенных людей и подключенных вещей к эпохе подключенного интеллекта. По мере того как общество движется ко всеобъемлющему интеллекту, 6G станет ключевой технологией распространения ИИ, соединяя с распределенным интеллектом каждого человека, дом, автомобиль и бизнес.

Функционируя как распределенная нейронная сеть с каналами связи, 6G объединит физический и кибернетический миры. Это будет не просто канал для перекачки битов, а сеть, вобравшая в себя все сенсоры, все подключения и весь ИИ. Иными словами, 6G будет представлять из себя машинообучаемую сеть сенсоров, где центры обработки данных станут нейронными центрами с машинным обучением, распространяющимся по всей сети. Примерно так будет выглядеть кибермир для будущего всеобъемлющего интеллекта.

Сеть 6G станет ключевым фактором в достижении полномасштабной цифровой трансформации всех вертикальных предприятий. Обладая исключительными характеристиками, такими как скорость передачи

данных в несколько Тбит/с, задержка менее миллисекунды и надежность в семь девяток (99,99999 %), 6G обеспечит значительные улучшения с точки зрения ключевых показателей производительности (key performance indicators, KPI) – в некоторых случаях рост более чем на порядок – по сравнению с 5G. Она обеспечит универсальную высокопроизводительную связь, сопоставимую с оптоволоконными кабелями по скорости и надежности, за исключением того, что это будет беспроводная связь. Свободная от функциональных ограничений и ограничений производительности, 6G будет универсальной платформой, которая поддерживает создание любой службы и любого приложения.

Основное отличие 6G от предыдущих поколений заключается в сочетании прорывных технологий и инноваций. Далее мы рассмотрим лишь некоторые из ключевых особенностей 6G, которые окажут огромное влияние на технологии и общество в течение ближайших десятилетий:

- сеть 6G изначально подразумевает наличие собственного распределенного искусственного интеллекта, а сетевая архитектура будет включать возможности машинного обучения, в частности распределенное машинное обучение. Проще говоря, 6G – это сеть, предназначенная для поддержки ИИ, или сеть, включающая ИИ, в которой многие элементы сети будут выполнять функции ИИ и машинного обучения;
- *беспроводное сканирование*⁴ (wireless sensing), основанное на естественных механизмах распространения радиоволн, станет ключевой прорывной технологией в 6G, использующей радиоволны и эхо для сканирования (или *ощущения*) физического мира. Предыдущие поколения беспроводных систем использовали радиоволны в основном для передачи информации. Однако для поддержки ИИ и машинного обучения нам необходимо собирать чрезвычайно большие объемы данных из физического мира; базовые станции 6G могут работать для этой цели как сенсоры. В частности, за счет использования более высоких частот, таких как миллиметровые волны и терагерцовый (ТГц) спектр, 6G обеспечит сканирование с высоким разрешением;
- важной отличительной чертой 6G станет интеграция спутниковых группировок VLEO с наземными сетями. Плотнo размещен-

⁴ В русском языке нет устоявшегося и полностью адекватного перевода термина *sensing*, который часто употребляется в зарубежных публикациях про беспилотные автомобили и сети 6G. В том и другом случаях речь идет о «прошупывании» окружающей среды электронными устройствами, поэтому мы решили, что наиболее подходящим переводом в этом контексте будет сканирование. Применяемый в робототехнике перевод «ощущение» имеет несколько иное и более общее значение, а перевод «зондирование» означает измерение параметров канала связи 5G и в нашем случае будет вносить путаницу. – Прим. перев.

ные небольшие спутники позволят создать «беспроводную сеть в небе» для покрытия всей земной поверхности. Это стало экономически выгодным благодаря достижениям Space-X в передовых технологиях запуска спутников, что значительно снизило стоимость создания массивных спутниковых группировок. Эта новая неназемная беспроводная инфраструктура будет полностью интегрирована с существующей наземной системой мобильной связи, что станет еще одним ключевым фактором 6G;

- сетевая архитектура 6G будет значительно отличаться от предыдущих поколений. В основе 6G лежат данные, а также полученные на их основе сведения и знания. Сетевая архитектура будет спроектирована таким образом, чтобы обеспечить надежность с учетом достижений в области технологий безопасности, сохранения конфиденциальности и управления данными. Фактически архитектуру сети 6G нужно разработать заново, чтобы в ней нашлось место для всеобщего интеллекта. Кроме того, 6G будет использовать новые права собственности на данные, модели доверия и конструкции безопасности, устойчивые к атакам на основе квантовых вычислений;
- устойчивость является центральной темой для 6G, особенно с точки зрения энергопотребления всей сети и связанной с ней инфраструктуры и устройств ICT. Устройство сети 6G должно соответствовать строгим требованиям в этом отношении. В частности, общее энергопотребление инфраструктуры 6G должно быть намного ниже, чем у предыдущих поколений, и в приоритете должна быть реализация архитектуры E2E, которая является одновременно устойчивой и энергоэффективной. Как глобальная инфраструктура ICT, сеть 6G будет ориентирована на достижение социальной, экологической и экономической устойчивости. Будущее интеллекта должно соответствовать нашей общей цели – сделать нашу планету лучше для жизни.

Таким образом, *главный смысл создания 6G заключается в переходе от подключенных людей и подключенных вещей к подключенному интеллекту*. Этот руководящий принцип заложен в основу сценариев использования, схем и технологий сетей 6G. Искусственный интеллект, слияние физического и кибернетического миров, а также приоритет связности (connectivity supremacy) – это новые столпы, на которых мы стремимся построить общество с распределенным всеобъемлющим интеллектом.

Мы предполагаем, что технологии 6G предложат следующие шесть новых возможностей для решения потенциальных проблем, с которыми столкнется эра 6G:

6G достигнет предельного уровня приоритета связности; будут задействованы все радиочастоты до терагерцового диапазона или даже видимого света.

Традиционно инфраструктура сотовой сети отдавала предпочтение более низкочастотному спектру из доступных ради достижения максимального покрытия. С другой стороны, рост трафика обычно вызывает потребность в выделении новых участков спектра. По мере развития нескольких поколений мобильной связи для модернизации сетей выделялось все большее количество участков спектра. В дополнение к миллиметровому диапазону 6G впервые будет использовать терагерцовый (ТГц) диапазон или даже видимый свет, а это означает, что для достижения максимальной возможности подключения будут задействованы все потенциально доступные участки спектра. Инновационные решения для обеспечения покрытия, такие как рой спутников VLEO, образующих новую инфраструктуру покрытия, и работающие по запросу *станции на высотных платформах* (high-altitude platform stations, HAPS), которые обеспечивают временную инфраструктуру покрытия, требуют инновационных подходов для агрегирования доступных спектров, чтобы реализовать необходимую пропускную способность 6G и сетевых решений. Сюда входят определение моделей владения и использования спектра, таких как совместное использование, гибкое распределение спектра и дуплексирование. Благодаря этому сеть 6G будет обладать почти бесконечной пропускной способностью с беспрецедентной скоростью беспроводного соединения. Новый радиоинтерфейс 6G может объединить технологии физического уровня eMBB, URLLC и mMTC, поскольку пропускная способность и задержка больше не будут узкими местами, препятствующими разработке подлинно *настраиваемого беспроводного соединения* для каждого пользователя, службы, приложения и сценария использования;

6G будет нативно поддерживать ИИ, соединяя интеллектуальные объекты и интеллектуально соединяя объекты.

Одна из основных задач 6G – поддержка повсеместного ИИ, где 6G будет транспортной системой E2E, предназначенной для поддержки сервисов и приложений на основе ИИ. Что принципиально важно, эта функциональность не является ни дополнением, ни надстройкой над сетью; вместо этого сама система 6G должна быть самой эффективной платформой для ИИ. Однако это создает новые проблемы с точки зрения реализации минимальных затрат как на коммуникацию, так и на вычисления, каждая из которых является отдельным ключевым показателем эффективнос-

ти и нуждается в дальнейшем исследовании. Для минимальных затрат на связь необходимо разработать систему 6G, способную передавать огромные объемы больших данных для обучения ИИ с минимальными энергозатратами. С точки зрения минимизации затрат на вычисления необходимо оптимально распределять вычислительные ресурсы по сетям в стратегических местах, где мы можем наилучшим образом использовать мобильные *границные вычисления* (edge computing). Для поддержки машинного обучения и создания кибернетического мира 6G необходимо будет обеспечить сбор массивных данных из физического мира. Это значительное увеличение объемов данных, что создает серьезную проблему для 6G. Новой и важной областью исследований будет достижение эффективного сжатия обучающих данных на основе теорий информации и обучения. Еще одна проблема – снижение вычислительной нагрузки, связанной с обучением ИИ, за счет совместного обучения. На сетевом уровне разделение данных и разделение моделей будет включено в архитектуру 6G, где распределенное и федеративное обучение будет использоваться не только для оптимизации вычислительных ресурсов, локального и глобального обучения, но и для соответствия новым требованиям к локальному хранению данных. Что касается сетевой архитектуры, основные сетевые функции будут переведены в *глубокую периферийную сеть* (deep-edge network), в то время как облачное программное обеспечение сместится в сторону машинного обучения. Кроме того, 6G RAN перейдет от радиодоступа, ориентированного на нисходящий канал, к радиодоступу, ориентированному на восходящий канал, потому что массивные обучающие данные, используемые в машинном обучении, требуют значительно более высокой пропускной способности в восходящем канале. Более того, радиоинтерфейс 6G может быть разработан с использованием новых возможностей машинного обучения для реализации интеллектуальной связи;

6G фактически станет сетевым ИИ, в корне меняющим определение сетей и вычислений.

Искусственный интеллект будет изначально встроен в структуру 6G и приведет к появлению новых сетевых и вычислительных архитектур. Например, используемые сегодня облачные центры обработки данных превратятся в сетевые нейронные центры искусственного интеллекта. Это приведет к переходу от вычислений на базе обычных процессоров (central processing unit, CPU) к вычислениям на базе графических процессоров (graphical processing unit, GPU). В большинстве случаев вычислительное оборудование,

предназначенное для ИИ, должно быть спроектировано и оптимизировано под использование алгоритмами ИИ. Однако развитие ИИ создает серьезные проблемы с точки зрения вычислений. В среднем человеческий мозг достигает скорости передачи данных 20 000 Тбит/с и может хранить 200 ТБ информации, потребляя при этом всего 20 Вт. Сегодня вычислительная мощность ИИ удваивается каждые два месяца, что намного превышает закон Мура. В момент времени, близкий к концу закона Мура, чтобы достичь тех же возможностей, что и человеческий мозг, нейронный центр будет потреблять в 1000 раз больше энергии. Чтобы нейронные центры могли заменить центры обработки данных и полностью реализовали потенциал ИИ, необходимо использовать передовые технологии машинного обучения, которые способствуют устойчивому развитию 6G на основе ИИ [11]. Стандартизованный подход к реализации вычислительной архитектуры нейронного центра и согласования программного обеспечения является ключевым условием, от которого зависит открытость платформы и экосистемы 6G;

6G будет функционировать как сетевой датчик, позволяющий объединить кибернетический, физический и биологический миры.

Сканирование – новая и важнейшая функция 6G – это новый канал, с помощью которого мы можем связать физический и биологический миры с кибермиром. Чтобы создать параллельную кибервселенную, которая является точной копией физической, в частности создать цифрового двойника, нам необходимо сканирование в реальном времени. Радиоволна 6G будет играть роль органа чувств на всех узлах и устройствах радиодоступа, включая базовые станции и мобильные устройства. Данные сканирования, собранные сетью и устройствами, могут использоваться для двух целей: для улучшения связи, особенно для миллиметрового и ТГц-диапазонов, и для нужд машинного обучения и искусственного интеллекта. В обоих случаях данные сканирования содержат информацию в реальном времени и знания о физическом и биологическом мирах. Таким образом, мы можем рассматривать сеть 6G как распределенный сетевой сенсор, кибернетический орган чувств, который радикально отличается от беспроводных систем предыдущих поколений, просто передающих информацию. Сканирование на основе сети и устройств может обеспечивать глобальное и локальное сканирование соответственно. Благодаря такой функциональности 6G выведет ИИ и машинное обучение в реальном времени на новый уровень;

6G с интегрированными наземными и неназемными сетями обеспечит полное покрытие всей земной поверхности, устраняя цифровой разрыв.

Интеграция неназемных сетей, особенно спутниковых мегасозвездий (роев) VLEO, в 6G является очень привлекательной перспективой. Спутниковая система VLEO, помимо обеспечения покрытия всей земной поверхности, предлагает ряд новых возможностей и преимуществ. Например, она устраняет проблему с задержкой связи, присущую традиционным спутниковым системам на геостационарной околоземной орбите (geostationary Earth orbit, GEO) и средней околоземной орбите (middle-earth orbit, MEO). Она также может обеспечивать покрытие зон, не охваченных наземными сетями, предлагая дополнительный радиодоступ. Одно из уникальных преимуществ спутниковой системы VLEO заключается в том, что она обеспечивает глобальный канал связи с малой задержкой, что очень важно для критически важных приложений, таких как алгоритмический частотный трейдинг. Спутниковые системы VLEO также могут обеспечивать более точное позиционирование, что имеет решающее значение для автономного вождения и важно для сканирования Земли и получения изображений земной поверхности. Помимо спутниковой связи, радиоузлы нового типа, такие как малогабаритные дроны, беспилотные летательные аппараты (БПЛА) и платформы HAPS, станут неотъемлемой частью 6G, функционируя либо как мобильные оконечные устройства, либо как временные узлы инфраструктуры;

6G будет поддерживать сетевую архитектуру, ориентированную на потребителей, а не на операторов, охватывая инклюзивную открытую экосистему.

Сеть 6G приведет к смене парадигмы участия, поскольку она стимулирует экономические и социальные изменения на основе достижений в области виртуализации и искусственного интеллекта. Наличие встроенного в сеть 6G интеллекта позволит реализовать подход к созданию сетей и предоставлению услуг на основе долевого вклада. Произойдет переопределение интеллектуальной инфраструктуры подключения как динамического пула, состоящего из ресурсов всех участвующих пользователей. Это радикальный сдвиг парадигмы от традиционной схемы, ориентированной на оператора, к всеобъемлющей схеме, ориентированной на *просьюмеров*⁵. Благодаря совместной модели, объединяющей

⁵ Просьюмер (prosumer) – это тот, кто одновременно и производит (produce), и потребляет (consume) некоторый продукт. – Прим. перев.

множество сетей, ключевые аспекты, такие как многостороннее владение, владение данными и конфиденциальность, а также модели доверия вовлеченных игроков, могут (и должны) быть разработаны как нативные, а не дополнительные функции. Кроме того, для обеспечения локального управления данными и сетевого суверенитета 6G будет применять новую модель доверия и технологии безопасности.

В инклюзивной модели, ориентированной на потребителей, каждый участник системы может как поставлять, так и потреблять ресурсы и услуги. Благодаря использованию технологии ИИ и машинного обучения сети 6G будут полностью автономными, не требующими ручного вмешательства. В этом смысле сети 6G будут скорее индивидуально подстраиваемыми, чем проприетарными, что приведет к появлению концепции «моя сеть».

1.3.1. Ключевые технологические тенденции

Основываясь на концепциях 6G, рассмотренных в предыдущем разделе, мы можем сделать следующие выводы:

- 6G расширит возможности человеческого общения, обеспечивая максимальное погружение с достоверными ощущениями в произвольном месте;
- 6G станет новой формой межмашинного взаимодействия, предоставляя интеллектуальную связь для эффективного машинно ориентированного доступа. 6G будет полностью включать в себя как машинное обучение, так и ИИ;
- 6G выйдет за рамки простого средства коммуникации. Сеть будет включать в себя новые функции, такие как сканирование и вычисления, предоставлять новые услуги и использовать расширенные знания о среде для машинного обучения;
- 6G откроет новое и уникальное поколение беспроводных сетей с поддержкой всеобщего интеллекта. Наличие искусственного интеллекта, надежность и энергоэффективность будут нативными свойствами 6G.

Ниже мы обсудим шесть основных технологических тенденций.

Тенденция 1: новый спектр в ТГц-диапазоне и оптическая беспроводная связь для чрезвычайно высоких скоростей передачи данных.

Для поддержки новых приложений, таких как AR/VR/MR и голографическая связь, необходимы сверхвысокие скорости передачи

данных до десятков Тбит/с. Субтерагерцовый и миллиметровый диапазоны будут базовыми спектрами в сотовых сетях 6G, в то время как полоса нижнего ТГц-диапазона (0,3–1,0 ТГц) будет основным кандидатом для передачи данных на короткие расстояния, например для использования внутри помещений или в каналах устройство–устройство (D2D). ТГц-диапазон обеспечивает сверхширокую полосу пропускания, превышающую десятки ГГц. ТГц-диапазон позволит использовать широкий спектр приложений, требовательных к данным и чувствительных к задержкам; в дополнение к этому ТГц-диапазон можно использовать для беспроводного сканирования.

Терагерцовая связь – это новая беспроводная технология, внедрение которой сопряжено со множеством проблем. В настоящее время проводятся исследования по разработке усилителей, новых материалов для антенн, силовых радиочастотных транзисторов, архитектуры ТГц-трансивера на кристалле, моделирования каналов и обработки массивов сигналов. Будет ли терагерцовая технология успешно применена в 6G, зависит от технических достижений в конструировании и производстве таких компонентов, связанных с ТГц-диапазоном, как электронные, фотонные и гибридные приемопередатчики и антенные решетки на кристалле.

Связь в диапазоне видимого света (visible light communication, VLC) – это технология передачи за пределами радиочастотного диапазона, которая позволяет минимизировать воздействие электромагнитного поля на окружающие объекты. Однако для достижения скоростей передачи данных, приближающихся к десяткам Тбит/с для связи на короткие расстояния с низким энергопотреблением, малым форм-фактором и недорогим оборудованием потребуется разработать технологию изготовления и использования *микросветодиодных массивов* (micro-LED array). Кроме того, связь в видимом свете позволяет использовать нелицензируемую часть спектра, но возможность успешного использования VLC в 6G зависит от нескольких проблем, связанных с передачей по восходящей линии связи, управлением переподключением и высокоэффективными приемопередатчиками.

Тенденция 2: интегрированное сканирование и связь (ISAC) для новых услуг и улучшенной беспроводной связи.

Традиционно сканирование окружающей среды является отдельной функцией, требующей наличия специализированных устройств и оборудования, таких как радар, лидар, компьютер-

ный томограф (КТ) и магнитно-резонансный томограф (МРТ). Позиционирование мобильного телефона в мобильных системах с помощью сигналов радиointерфейса и измерений на основе устройств отчасти напоминает технологию сканирования. Однако за счет использования миллиметрового и ТГц-диапазонов, которые предлагают более широкую полосу пропускания и меньшую длину волны, 6G позволит интегрировать функцию сканирования в систему связи. В полной системе ISAC функции сканирования и связи будут дополнять друг друга, предлагая следующие ключевые преимущества:

- **сотовая связь как датчик:** сигналы связи будут использоваться для новых применений сканирования, таких как высокоточная локализация, распознавание жестов и движений, обнаружение и отслеживание объектов, съемка и реконструкция объектов окружающей среды;
- **улучшение связи с помощью сканирования:** сканирование помогает улучшить качество обслуживания (QoS) и эффективность связи, в том числе за счет выбора пути, прогнозирования канала и выравнивание луча.

Интегрированное сканирование и связь позволяют услугам сканирования в 6G выйти далеко за рамки простого позиционирования. Ему на смену придут новые сервисы, включая дополнительные функции сканирования с повышенной точностью (которые описывают разницу между измеренными и реальными значениями диапазона, угла, скорости и т. д.) и разрешением сканирования (которое описывает возможность различать несколько объектов по диапазону, углу, скорости и т. д.). Этот вопрос более детально рассмотрен в главе 3.

По сравнению с традиционной технологией радиолокации, сканирование 6G, использующее широкополосный спектр и большие антенные решетки, обеспечит технологические инновации, такие как крупномасштабное взаимодействие между базовыми станциями и пользовательскими устройствами, общую конструкцию устройств связи и сканирования, передовые методы подавления помех и использование данных сканирования для содействия ИИ. Сканирование потенциально станет одной из самых революционных инноваций 6G, на базе которой могут быть созданы многочисленные приложения для машинного обучения и искусственного интеллекта в реальном времени.

Благодаря сверхширокой полосе пропускания при измерении в ТГц-диапазоне можно достичь гораздо более высокой точности

и разрешения. Благодаря эффекту терагерциндуцированной вибрации молекул при помощи терагерцового сканирования можно снимать вторичную спектрограмму для определения компонентов продуктов питания, лекарств и загрязнителей воздуха. Благодаря компактному форм-фактору и отсутствию ионизирующего излучения устройства для ТГц-сканирования могут быть интегрированы в мобильные устройства и даже элементы одежды для определения количества калорий в пище и обнаружения скрытых объектов. Чувствительные устройства 6G откроют перспективы для реализации многочисленных инновационных приложений искусственного интеллекта.

Тенденция 3: ИИ как услуга и функция в системе связи 6G для интеллектуального подключения «умных» устройств.

Ключевой задачей проектирования в 6G является конвергенция беспроводных технологий и технологий искусственного интеллекта с самого начала, а не разработка сначала беспроводной системы, а затем добавление искусственного интеллекта. Использование искусственного интеллекта для улучшения беспроводной системы 6G создает возможности для развития постшенноновской теории связи и инноваций в беспроводных технологиях.

В 6G будут реализованы две концепции: (1) *ИИ для сети* использует приложения ИИ в качестве инструментов для оптимизации сети; (2) *сеть для ИИ* подстраивает сеть для поддержки и оптимизации приложений ИИ. Сеть также помогает предоставлять функциональные возможности ИИ или даже выполнять такие функции самостоятельно; например, сеть может обрабатывать логические выводы и машинные вычисления. Конечно, ИИ можно применять в качестве универсального инструмента для функциональной оптимизации, как в случае с сетью 5G и продуктами ее эволюции.

Ниже представлены более подробные описания этих концепций.

- **ИИ для сети:** технологии искусственного интеллекта, которые по своей сути управляются данными, могут быть интегрированы в классическую структуру систем связи на основе моделей, чтобы справиться с ситуациями, когда система на основе модели является слишком сложной или не может обеспечить достаточно высокую точность. Используя ИИ для сети, можно создать интеллектуальную линию связи, которая адаптируется к динамической среде сквозной передачи (end-to-end, E2E). Кроме того, полностью интегрируя обработку сигналов и анализ данных, мы можем упростить и унифицировать архитектуру вычислений и логического вывода, а также преоб-

разовать сеть из архитектуры, основанной на динамической обработке и отклике, в сеть, в которой возможно упреждающее прогнозирование и принятие решений. Хотя это многообещающая область, она также создает множество проблем. Например, проектирование таких систем затруднено из-за отсутствия надежных аналитических инструментов и универсальной архитектуры нейронной сети, что затрудняет поиск оптимального баланса между параметрами системы. Следовательно, достижение глубокого теоретического понимания ИИ – важное направление для дальнейших исследований. Чтобы добиться максимальной поддержки искусственного интеллекта в сетях 6G, может потребоваться более революционный подход – возможно, нам придется пересмотреть фундаментальное понимание того, как система связи передает информацию. Для более глубокого понимания постшеннонских коммуникаций потребуются обширные теоретические исследования, которые еще предстоит выполнить.

- **Сеть для ИИ:** сеть 6G будет развиваться в направлении более распределенной архитектуры со встроенными возможностями МЕС для локального сбора данных, обучения, рассуждений и логических выводов вместе с глобальным обучением и логическими выводами для надежной защиты конфиденциальности, снижения задержки и снижения потребления полосы пропускания. Одной из возможных технологий, которые позволят реализовать такие функции, является *федеративное обучение*. На начальном этапе проектирования радиоинтерфейсов 6G и сетевой архитектуры важно заложить распределенное обучение и логический вывод ИИ, чтобы итеративно реализовать эффективный крупномасштабный интеллект. Более радикальный подход – изучение теории узких мест информационных потоков в контексте ИИ – представляет собой новое направление исследований 6G. Это поможет нам сжать огромные объемы обучающих данных, отправляемых по сети, с использованием минимальных ресурсов, таких как пропускная способность и память.

Тенденция 4: надежность 6G, основанная на многосторонней модели доверия и новых криптографических технологиях.

В 6G мобильные устройства станут порталом в кибермир, который является точной копией физического мира. Поскольку наша зависимость от 6G и предлагаемых им услуг будет только расти, надежность сети и услуг будет иметь решающее значение. Входя-

щая в состав сетевой архитектуры надежная система безопасности послужит основой для установления доверительных отношений между различными объектами в сети. Функции безопасности должны разрабатываться совместно с функциями ядра сети и удовлетворять требованиям к услугам как индивидуальных пользователей, так и бизнес-клиентов. Ниже обсуждаются два аспекта надежной доверенной системы 6G:

- **многосторонняя модель доверия:** любая архитектура безопасности, независимо от того, централизованная она или децентрализованная, имеет как преимущества, так и недостатки. Централизованная архитектура использует набор очень сильных механизмов безопасности, которые применяют строгие политики безопасности, но с точки зрения диалога безопасности это связано с более высокой сложностью во время операций роуминга, передачи обслуживания или повторного входа в систему, выполняемых в сети. Следовательно, вероятность атаки выше, потому что чем больше интерфейсов, тем более уязвимой становится система. Децентрализованная архитектура с гибкими и настраиваемыми механизмами поддерживает различные требования и может ограничиваться близлежащими сервисами. В случае атаки вредоносное воздействие ограничится пределами небольшой локализованной области. В случаях, когда не все заинтересованные стороны являются доверенными, для реализации устойчивости может использоваться многосторонняя модель доверия, чтобы другие заинтересованные стороны оставались доверенными. По сравнению с централизованной архитектурой в распределенной архитектуре отсутствует эффективная синхронизация унифицированных политик безопасности. Следовательно, необходима унифицированная архитектура безопасности с несколькими атрибутами безопасности, применимая как к централизованной, так и к децентрализованной архитектуре.

Более инклюзивная многосторонняя модель доверия послужит основой будущих систем безопасности. Эта модель поможет создать устойчивую и естественную архитектуру надежности, охватывающую весь жизненный цикл 6G. Она также будет гибко реализовывать централизованные политики безопасности, распределенные механизмы на основе консенсуса, а также проверенные сторонние ссылки и верификацию;

- **новые криптографические алгоритмы:** по мере развития квантовых вычислений возникают проблемы, связанные с классической криптографией, которая основана на матема-

тических задачах, таких как факторизация большого числа простых чисел и дискретные алгоритмы. Алгоритмы генерации и обмена ключами – два незаменимых элемента криптографии. В 6G для защиты от атак, основанных на квантовых вычислениях в полнодуплексной двунаправленной связи на физическом уровне, может использоваться шифрование с одноразовым блокнотом (one-time pad, OTP). Одним из важных аспектов криптографии является то, что 6G требует, чтобы криптографические алгоритмы выполняли операции в течение микросекунд, дабы обеспечить сверхнизкую задержку. Когда квантовые вычисления станут реальностью, ожидается, что технологии квантовой связи станут более безопасными из-за квантовой запутанности. Кроме того, облегченные криптографические алгоритмы и алгоритмы соблюдения конфиденциальности – это лишь некоторые из возможностей, которые заслуживают дальнейшего исследования в 6G.

Некоторые из ключевых вопросов, связанных с реализацией предшествующих собственных механизмов надежного доверия, заключаются в следующем:

- (1) Как новые технологии будут интегрированы с традиционными механизмами безопасности?
- (2) Как децентрализованные технологии будут интегрированы с архитектурой беспроводной сети?
- (3) Как будут реализованы открытые и прозрачные стандарты безопасности и защиты данных?

Все эти вопросы потребуют дальнейшего изучения.

Тенденция 5: интеграция наземных и неназемных сетей для повсеместного доступа на всей земной поверхности.

Сегодня, даже в развитых странах, во многих сельских и отдаленных районах по-прежнему отсутствует высокоскоростное подключение к интернету. Ситуация в развивающихся странах еще хуже. Фактически более 3 миллиардов человек во всем мире все еще не имеют доступа к интернету, что создает серьезный цифровой разрыв между подключенными и неподключенными сегментами человечества [4].

В настоящее время основным препятствием для достижения бесшовного глобального покрытия являются экономические, а не технические факторы. Ожидается, что интеграция наземных и неназемных сетей станет экономически эффективным решением и позволит обеспечить беспрепятственное покрытие услугами

высокоскоростного мобильного интернета независимо от географических ограничений.

По мере снижения стоимости производства и запуска спутников реальностью станет огромный парк малых спутников на низкой околоземной орбите (low-earth-orbit, LEO) или VLEO. Кроме того, использование БПЛА и HAPS будет означать, что покрытие, обеспечиваемое будущей мобильной системой, больше не будет плоским (двухмерным). Вместо этого трехмерная гибридная сетевая архитектура, включающая несколько уровней и множество движущихся точек доступа, позволит предоставлять услуги связи и навигации в любом месте и в любое время. Это означает радикальное изменение протоколов связи в части планирования соединения, захвата и передачи абонентского соединения, а также *беспроводного обратного соединения базовой станции с центром управления сетью (wireless backhaul)*.

В настоящее время БПЛА и HAPS проектируются и эксплуатируются отдельно, но ожидается, что в будущих сетях 6G их функции и операции, а также управление ресурсами и мобильностью будут тесно интегрированы. Такая интегрированная система будет идентифицировать каждое пользовательское устройство с помощью уникального идентификатора, унифицировать процессы выставления счетов и постоянно предоставлять высококачественные услуги через оптимальные точки доступа.

Чтобы легко интегрировать новый БПЛА или спутник LEO без необходимости ручной настройки, интегрированная сеть должна обладать функцией самоорганизации. Благодаря интеллектуальному устройству радиоинтерфейса добавление и удаление точки доступа будет прозрачным для пользовательских устройств с точки зрения процедур физического уровня (таких как формирование луча, измерение и обратная связь). Учитывая, что развертывание, обслуживание и энергоснабжение спутников полностью отличаются от наземных сетей, можно ожидать появления новых операционных и бизнес-моделей.

Тенденция 6: экологичные и устойчивые сети с низкой совокупной стоимостью владения для устойчивого развития во всем мире.

Увеличение числа подключенных устройств, базовых станций и сетевых узлов не только приведет к огромному всплеску трафика данных, но и к значительному увеличению энергопотребления во всех частях сети. Долгое время основной целью проектирования была энергоэффективность передачи данных, определяемая как

бит/джоуль. При проектировании сети 6G этот критерий станет еще более важным – высокая энергоэффективность больше не будет просто полезным свойством; скорее, это будет обязательным требованием для мобильных сетей 6G.

На сегодняшний день на ICT приходится около 2 % мировых выбросов парниковых газов (из которых на мобильные сети приходится около 0,256 %) [12]. Ожидается, в последующие годы этот процент будет неуклонно увеличиваться. Для 6G, помимо энергоэффективности, важно будет снизить энергопотребление сетей. Это необходимо не только для сокращения расходов на электроэнергию, но и для сокращения выбросов парниковых газов, что является важным социальным обязательством. В то же время необходимо будет учитывать как капитальные затраты (capital expenditure, CAPEX), так и операционные расходы (operational expenditure, OPEX). Хотя проектирование рентабельных и энергоэффективных сетей направляет отрасль ICT на путь устойчивого развития, ICT в целом может сыграть важную роль в сокращении глобальных выбросов CO₂ ради более чистой и здоровой среды обитания. Ожидается, что ICT могут сократить глобальные выбросы CO₂ на 20 % к 2030 году по сравнению с уровнями 2015 года [13]. В то же время система связи 6G должна поддерживать постоянное расширение перечня приложений и сценариев использования в бизнесе, чтобы облегчить работу в других отраслях, а также обеспечить устойчивое социальное развитие.

Так называемая *зеленая радиосеть* (green radio network) – обширная исследовательская дисциплина. Потенциальные энергоэффективные технологии охватывают архитектуры, материалы, аппаратные компоненты, алгоритмы, программное обеспечение и протоколы. Развертывание плотных сетей (ведущее к более короткому расстоянию распространения), централизованная архитектура RAN (приводящая к меньшему количеству сотовых узлов и повышению эффективности использования ресурсов), энергосберегающие протоколы и взаимодействие между пользователями и базовыми станциями – вот некоторые факторы, которые необходимо тщательно учитывать, создавая энергоэффективную систему связи 6G. Еще одна важная проблема – снижение эффективности выходного усилителя мощности по мере перехода к использованию все более высоких частот. Кроме того, не следует забывать про возобновляемые источники энергии и технологии сбора энергии в радиочастотном диапазоне.

С другой стороны, по мере распространения искусственного интеллекта в центрах обработки данных, граничных узлах и даже

на мобильных устройствах потребление энергии, связанное с обучением и освоением искусственного интеллекта, становится ключевой проблемой, которую необходимо решать уже сейчас. В [14] указывалось, что обучение одной модели ИИ выделяет столько же углерода, сколько пять автомобилей производят за все время своей эксплуатации. Одни только центры обработки данных потребляли более 2 % мировой электроэнергии в 2018 году, и ожидается, что их доля будет увеличиваться по мере того, как к 2030 году появится больше граничных узлов и устройств с поддержкой ИИ. Некоторые недавние исследования показывают, что обучение универсальной сети, которая поддерживает различные конфигурации архитектуры, значительно снизит выбросы CO₂ по сравнению с поиском специализированной нейронной сети и ее обучением с нуля для каждого варианта использования [15].

1.3.2. Типичные варианты использования

По мере внедрения новых технологий в системы беспроводной связи многие аспекты нашей повседневной жизни претерпят изменения за счет сверхвысокоскоростных беспроводных соединений, искусственного интеллекта и передовых сенсорных технологий. Проще говоря, изменятся и способы, и содержание нашего взаимодействия с технологиями.

В дополнение к широкополосной передаче данных сеть 5G стала шагом вперед к высоконадежному беспроводному доступу с малой задержкой, что позволило реализовать набор вертикальных приложений и приложений интернета вещей. ITU-R⁶ определил три типа сценариев использования 5G (eMBB, URLLC и mMTC) в дорожной карте IMT-2020 [16]. По мере того как в 6G вводятся интеллектуальные и сенсорные функции, а покрытие выходит за пределы земной поверхности, сети следующего поколения будут создавать новые приложения и улучшать существующие. Некоторые из этих приложений, возможно, уже обсуждались в концепции 5G, даже если они не были включены в действующие сети 5G из-за технологических ограничений или незрелости рынка. Мы сосредоточимся на вариантах использования, которые сеть 5G не может поддерживать (например, сканирование), а также на сценариях, которые обсуждались, но не получили широкого распространения в 5G, таких как интеллект и улучшения трех применений 5G (eMBB+, URLLC+, и mMTC+). Мы выделяем шесть категорий потенциальных вариантов использования 6G, как показано на рис. 1.6.

⁶ International Telecommunication Union – Radio, Международный союз электросвязи, МСЭ.

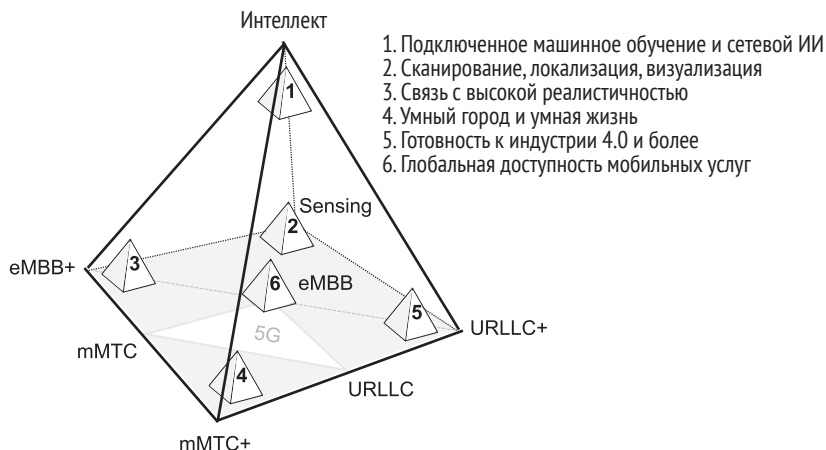


Рис. 1.6. Обзор сценариев использования 6G

- **Связь с высокой реалистичностью:** постоянное стремление к улучшению пользовательского опыта коммуникаций. Чтобы обеспечить опыт удаленного присутствия с погружением и ориентированных на человека применений, таких как AR, VR и MR, а также голографической связи, нам необходимо стремиться повысить разрешение дисплеев до предела человеческого восприятия. Подобные дисплеи нуждаются в сверхвысоких скоростях передачи данных до Тбит/с, а сегодняшняя сеть 5G не способна этого достичь. Чтобы избежать укачивания (головокружение и усталость) и получать тактильную обратную связь в реальном времени при телеуправлении, требуется чрезвычайно низкая задержка сети E2E, максимально приближенная к пределам возможностей человеческих органов чувств и восприятия.
- **Сканирование, локализация и съемка:** помимо связи, использование диапазонов более высоких частот (ТГц и миллиметровые волны) обеспечивает другие возможности, такие как сканирование, визуализация⁷ (imaging) и локализация. В результате в обиход войдут различные инновационные приложения, такие как высокоточная навигация, распознавание жестов, отображение и реконструкция изображений. По сравнению со связью, для оценки качества сканирования, локализации и визуализации используют другие критерии, такие как разрешение сканирования и точность измерения диапазона, угла или скорости. Они также

⁷ Под визуализацией мы здесь понимаем создание цифровых образов (image) объектов окружающей среды (зданий, помещений, предметов). Для визуализации может применяться ТГц-сканирование, а цифровой образ не обязательно визуально понятен человеку, т. к. чаще предназначен для ИИ. – Прим. перев.

включают новый набор показателей качества работы, таких как вероятность неправильного обнаружения и ложной тревоги.

- **Полнофункциональная индустрия 4.0 и выше:** отраслевые сценарии использования 5G достаточно хорошо изучены [1]. Хотя 5G был разработан с низкой задержкой и высокой надежностью, некоторые из этих сценариев предъявляют чрезвычайно высокие требования (например, точное управление движением), превышающие возможности 5G. В свою очередь, 6G позволит реализовать эти варианты использования за счет технологий, реализующих сверхвысокую надежность и чрезвычайно низкую задержку. Вместе с тем, по мере того как новые методы взаимодействия человека и машины на основе ИИ станут жизнеспособными, будущие автоматизированные производственные системы будут ориентированы на совместную работу роботов, коботов⁸ или даже киборгов. Взаимодействие между роботами и людьми в режиме реального времени требует еще меньших задержек и более высокой надежности по сравнению с 5G.
- **Умный город и умная жизнь:** огромное количество датчиков будет развернуто для нужд умного транспорта, строительства, здравоохранения, автомобилей, заводов и городов. Эти датчики будут собирать большие данные для алгоритмов ИИ, которые затем используются для предоставления ИИ как услуги (AI as a service, AIaaS). Ожидается, что мы будем жить в мире, где физическая реальность отображается на цифровых двойниках в кибермире. Другими словами, автоматизация и интеллект будут созданы в кибермире и доставлены в физический мир через беспроводные сети 6G. Для реализации этой концепции крайне важно разработать новые средства – сенсоры, алгоритмы сканирования – для извлечения больших данных, востребованных при обучении глубоких нейронных сетей (deep neural network, DNN). Для этого, в свою очередь, тоже нужны беспроводные каналы с очень высокой пропускной способностью, позволяющие собирать данные сканирования в реальном времени. Следовательно, для умного города и умной жизни требуется огромное количество надежных и безопасных подключений.
- **Глобальное покрытие для мобильных услуг:** чтобы предоставлять бесперебойные и безразрывные мобильные услуги на всей земной поверхности, в сетях 6G предусмотрена интеграция

⁸ Cobot (кобот), или collaborative robot (совместный робот) – это специальный робот, предназначенный для совместной работы рядом с людьми, в общем пространстве, в отличие от обычных промышленных роботов, которые изолированы от контакта с людьми. – *Прим. перев.*

наземной и неназемной связи. В такой интегрированной системе мобильный пользователь с помощью одного устройства может получить доступ к услугам мобильного широкополосного доступа как в городских, так и в сельских районах или даже на движущихся самолетах⁹ и кораблях. В этих сценариях постоянно происходит динамический выбор оптимальных каналов наземных и неназемных сетей без прерывания текущих услуг. Беспилотные транспортные средства получают возможность использовать безразрывную высокоточную навигацию в любой точке планеты. К возможным вариантам использования глобального покрытия относятся также IoT-подключения для защиты окружающей среды в реальном времени и точного ведения сельского хозяйства.

- **Подключенное машинное обучение и сетевой ИИ:** полноценное использование возможностей ИИ является фундаментальным принципом функционирования 6G. По сути, с одной стороны, возможности ИИ могут быть расширены и интегрированы в большинство функций и приложений 6G. С другой стороны, почти все приложения 6G будут основаны на ИИ, при этом ИИ также может расширять все предыдущие варианты использования беспроводных сетей, внося в них продвинутую автоматизацию. В то же время существуют проблемы и ограничения, связанные с предоставлением ИИ в качестве OTT-сервиса. Первая проблема заключается в том, что машинное обучение требует передачи огромных объемов данных в центры обработки данных, особенно для настраиваемых сервисов искусственного интеллекта. Вторая проблема напрямую связана с требованиями к территориальному управлению данными. В частности, в большинстве стран передача данных в зарубежные центры обработки данных не разрешена. Третья проблема связана со взаимодействием различных агентов ИИ (даже расположенных рядом) через сеть 6G. В качестве основного варианта рассматривается полное подключение распределенных агентов машинного обучения через сеть 6G для формирования сетевого интеллекта и лучшей защиты конфиденциальности данных. С этой точки зрения подключенное машинное обучение и сетевой ИИ, по существу, будут опираться на следующие аспекты: нацеленность 6G на максимальное увеличение возможностей машинного обучения; сетевая архитектура, поддерживающая распределенный

⁹ Наличие мобильной связи на борту самолета можно отнести к покрытию земной поверхности, потому что даже низкоорбитальные спутники находятся значительно выше самолета. – Прим. перев.

характер ИИ и граничные вычисления для служб ИИ в реальном времени; высокая пропускная способность, низкая задержка и высоконадежный логический вывод ИИ. Кроме того, ИИ изначально будет встроенной функцией 6G, облегчая разработку будущих схем передачи, интеллектуального контроля и управления ресурсами, а также «самонастройку» сети.

1.3.3. Ключевые показатели эффективности

Чтобы обеспечить наилучший опыт пользователя во всех сферах применения, сеть 6G должна значительно улучшить ключевые показатели по сравнению с предшественниками. Исходя из опыта обновления предыдущих поколений мобильной связи, можно ожидать, что ключевые показатели 6G увеличатся в 10–100 раз по сравнению с 5G. Наши предположения схематически показаны на рис. 1.7 и подробно описаны ниже.



Рис. 1.7. Краткий перечень ключевых показателей эффективности для сетей радиодоступа

Очень высокая скорость передачи данных и эффективность использования спектра

Для реализации коммуникаций с опытом погружения потребуются очень высокая пропускная способность. Для передачи 360-градусной AR/VR- и голографической информации, в зависимости от таких факторов, как разрешение, размер и частота обновления изображений, может потребоваться скорость передачи данных от нескольких Гбит/с до нескольких Тбит/с.

В IMT-2020 (5G) минимальные требования ITU-R к пиковой и удобной для пользователя скорости передачи данных составляют 10–20 Гбит/с и 100 Мбит/с соответственно [17]. В 6G пиковая и удобная для пользователя скорость передачи данных должны составлять 1 Тбит/с и 10–100 Гбит/с соответственно. Кроме того, ожидается, что 6G будет и дальше развивать использование спектра, улучшая пиковую эффективность использования спектра в 5–10 раз по сравнению с 5G.

Очень большая емкость и сверхмассивная подключенность

Пропускная способность области – это общая пропускная способность трафика для географической области. Это произведение *плотности подключения в области* (общее количество устройств на единицу площади) и средней скорости передачи данных для пользователей. В 5G минимальное требование ITU-R к плотности подключения составляет 1 000 000 устройств на км² [17]. В 6G, где мы должны поддерживать такие сценарии использования, как индустрия 4.0 с подключенным интеллектом и умный город, в ближайшие 10 лет и позже плотность подключения должна быть увеличена примерно в 10–100 раз, т. е. до 100 млн/км². Такое огромное количество подключений должно быть способно поддерживать разнообразные типы услуг с разными характеристиками (например, с разной пропускной способностью, задержкой и QoS). В связи с этим пропускная способность системы 6G должна быть в 1000 раз больше, чем 5G, чтобы предоставлять качественные услуги для большого количества подключений.

Очень низкая задержка и джиттер, а также сверхвысокая надежность

В некоторых сценариях использования интернета вещей, таких как беспилотные автомобили и промышленная автоматизация, критически важно, чтобы данные доставлялись вовремя (с низкой задержкой и джиттером). Радиоинтерфейсы 6G обеспечивают задержку 0,1 мс и джиттер $\pm 0,1$ с. Исходя из требований к удаленным службам XR, общая задержка передачи E2E туда и обратно должна составлять 1–10 мс. Помимо низкой задержки, приложениям интернета вещей также требуется надежность (то есть правильная передача информации). ITU-R требует, чтобы надежность для услуг URLLC в 5G составляла 99,999%. В 6G, где ожидается более широкое распространение различных вертикальных применений, необходимо многократное повышение надежности, достигающее 99,99999%.

Очень высокая локализация, точность сканирования и разрешение

Сканирование, локализация и съемка – это новые функции в 6G, знаменующие значительный шаг вперед для подключенного интеллекта, о чем пойдет речь в главе 9. Ожидается, что благодаря расширенному до терагерц частотному диапазону и передовым технологиям сканирования 6G обеспечит сверхвысокую точность определения местоположения – 50 см для применений на открытом воздухе и 1 см внутри помещений. Для других сервисов сканирования, как будет описано в главе 5, максимальная точность и разрешение могут достигать 1 мм и 1 см соответственно.

Очень широкое покрытие и очень высокая мобильность

Для предоставления высококачественных услуг мобильного интернета с более широким покрытием *бюджет канала* (энергетический баланс) для радиоинтерфейсов 6G должен быть увеличен как минимум на 10 дБ по сравнению с 5G. Но покрытие 6G не должно определяться только бюджетом канала. Технология 6G должна обеспечить 100%-ное покрытие земной поверхности и охват населения путем интеграции как наземных, так и неназемных сетей. Кроме того, 6G будет поддерживать покрытие для самолетов, движущихся со скоростью около 1000 км/ч, что намного выше, чем у 5G (500 км/ч, в основном для высокоскоростных поездов).

Очень высокая энергоэффективность и экономичность

Энергопотребление – один из самых сложных аспектов в системах 6G. С одной стороны, это связано с передачей в очень высокочастотных диапазонах, при очень большой полосе пропускания и очень большом количестве антенн. Более низкая эффективность усилителей мощности и увеличенное количество звеньев радиотракта – две ключевые проблемы, которые необходимо решить в ближайшее время. С другой стороны, по мере роста конвергенции связи и вычислений и поддержки собственного ИИ его обучение и логический вывод в сетях 6G будут потреблять больше энергии. Это означает, что для достижения сопоставимых уровней общего энергопотребления потребление энергии на бит в 6G должно быть как минимум в 100 раз ниже, чем в 5G. С точки зрения устройств с повышенными скоростями передачи данных, необходимо соответственно повысить энергоэффективность обработки сигналов. Кроме того, источники питания сенсоров для использования в умном городе, умном здании, умном доме и в сценариях умного здоровья должны служить до 20 лет.

Нативный ИИ

Как говорилось в разделе 1.3.1, нативная поддержка ИИ в системах мобильной связи 6G включает два аспекта: ИИ для сети и сеть для ИИ.

- **ИИ для сети** образует интеллектуальный фреймворк для проектирования радиоинтерфейсов и сетевых функций, поддерживает динамическую передачу E2E, работу сети без вмешательства человека и автоматическое создание специализированных сетевых сегментов для различных услуг и предприятий;
- **сеть для ИИ** требует более распределенной архитектуры со встроенными возможностями мобильных граничных вычислений, чтобы сочетать локальный сбор данных, обучение и рассуждения/логические выводы с глобальным обучением и логическими выводами для лучшей защиты конфиденциальности и снижения задержки или потребления полосы пропускания.

Нативная добросовестность

6G укрепит связь между физическим и цифровым мирами, став неотъемлемой частью нашей жизни. Понятие *добросовестности* (trustworthiness) в мобильной связи включает в себя защищенность, конфиденциальность, отказоустойчивость, безопасность и стабильность [18]:

- **защищенность** – это состояние, которое является результатом выполнения защитных мероприятий, позволяющих организации выполнять свою миссию или критические функции, несмотря на риски, связанные с угрозами использования ею систем, как определено в [19]. Защитные меры могут включать сочетание сдерживания, предотвращения, обнаружения, восстановления и исправления, которые должны составлять часть подхода организации к управлению рисками;
- **конфиденциальность** – это свобода от вторжения в частную жизнь или дела человека, когда это вторжение является результатом ненадлежащего или незаконного сбора и использования данных об этом человеке, как определено в [20];
- **отказоустойчивость** – это способность быстро адаптироваться и восстанавливаться после любых известных или неизвестных изменений в окружающей среде посредством целостного внедрения управления рисками, непредвиденных обстоятельств и планирования непрерывности, как определено в [21];

- **безопасность** – это свобода от обстоятельств, которые могут привести к смерти, травмам, профессиональному заболеванию, повреждению или потере оборудования или собственности, или ущербу окружающей среде, как это определено в [22];
- **стабильность** – это способность системы или компонента функционировать в заданных условиях в течение определенного периода времени, как определено в [23].

1.4. Структура книги

Эта книга состоит из семи частей, посвященных конкретным темам, связанным с 6G. *Часть I*, состоящая из главы 1, описывает эволюцию мобильной связи от 2G к 6G и общую концепцию 6G, определяя шесть технологических тенденций, которые являются центральными для трех фундаментальных движущих сил будущего подключенного интеллекта.

В *части II* мы обсуждаем возможные варианты использования 6G и анализируем ключевые требования к производительности. Варианты использования варьируются от эволюции сетей 5G, которые приобретут популярность и станут более зрелыми на этапе 6G благодаря большей пропускной способности, меньшей задержке и более высокой надежности, до совершенно новых применений, ставших возможными благодаря новым функциям и возможностям, которые предлагает 6G. В части II типичные варианты использования разделены на шесть категорий, каждая из которых описана в отдельной главе. Они охватывают ориентированное на человека общение с эффектом полного погружения (глава 2); высокоточное сканирование, локализацию и визуализацию плюс расширение восприятия человека (глава 3); полнофункциональную индустрию 4.0 с подключенным интеллектом (глава 4); умный город и умную жизнь (глава 5); глобальное 3D-покрытие для мобильных услуг с интегрированной наземной и неназемной связью (глава 6) и нативную поддержку ИИ во всех случаях использования (глава 7).

В *части III* исследуются объем и пределы конструкторских разработок 6G, обсуждаются теоретические основы радиотехнологий 6G и сетевых технологий, а затем исследуется ряд эффективных технологий, обладающих потенциалом достижения заданных KPI. Представленный материал охватывает теоретические основы нативного ИИ и машинного обучения (глава 8), теоретические основы обеспечения всеобщей подключенности (глава 9), теоретические основы будущих межмашинных коммуникаций (глава 10), а также теоретические основы построения энергоэффективных систем (глава 11).

В *части IV* мы анализируем будущий спектр Международных мобильных телекоммуникаций (international mobile telecommunications,

ИМТ) с точки зрения связи и сканирования (глава 12), а также соответствующие методологии моделирования каналов и некоторые примеры канальных измерений (глава 13). Затем, чтобы обеспечить широкое и всестороннее понимание того, как 6G будет развиваться в течение следующих 10 лет, мы описываем перспективные новые материалы для производства оборудования (глава 14), новые антенные структуры для сверхмассивных систем со множеством входов и множеством выходов (multiple-input multiple-output, ММО) (глава 15), новые радиочастотные компоненты для работы в ТГц-диапазоне (глава 16), эволюцию вычислений после выхода за рамки закона Мура (глава 17), а также новые требования к окончательным устройствам (глава 18).

Часть V фокусируется на общих принципах проектирования и перспективных технологиях для радиоинтерфейсов 6G. Во введении к этой части мы описываем сдвиги парадигмы в разработке радиоинтерфейсов по сравнению с 5G и более ранними поколениями, а затем обсуждаем ряд сопутствующих технологий. Эти перспективные технологии включают интеллектуальный радиоинтерфейс (глава 19), интеграцию наземной и неназемной связи (глава 20), интеграцию сканирования и связи (глава 21), новые типы волны и модуляции (глава 22), новое кодирование (глава 23), новую технологию множественного доступа (глава 24), сверхмассивные ММО (глава 25) и связь на сверхмалых расстояниях (глава 26). Для каждой технологии мы описываем предпосылки и обоснование для ее использования, исследуем существующие решения, проясняем ожидания от новых конструктивных решений и выделяем потенциальные исследовательские проблемы и направления для будущих исследований.

Аналогично, в *части VI* основное внимание уделяется принципам проектирования и перспективным технологиям, позволяющим проектировать архитектуру сети 6G. Она также начинается с введения, посвященного смене парадигмы при проектировании сетевой архитектуры. После этого мы углубленно рассматриваем несколько новых основных функций и технологий, используемых в сетевых архитектурах 6G. Это архитектурные решения для сетевого ИИ (глава 27), ориентированная на пользователя сеть (глава 28), нативная надежность (глава 29), управление данными (глава 30), многопользовательские экосистемы (глава 31) и интегрированные неназемные сети (глава 32).

Часть VII (глава 33) завершает книгу описанием текущего состояния экосистемы 6G в мире, включая исследовательские проекты, платформы, семинары и документы по 6G, а затем предлагает наше видение дорожной карты до 2030 года.

1.5. Источники

- [1] 5G-ACIA. <https://www.5g-acia.org/>.
- [2] 5GAA. <https://5gaa.org/>.
- [3] ITU-R, *IMT traffic estimates for the years 2020 to 2030*, Report ITU-R M.2370-0, July 2015.
- [4] *The mobile economy 2020*, Intelligence, GSMA, 2020.
- [5] *Touching an intelligent world*, Sept. 2019. https://www.huawei.com/minisite/giv/Files/whitepaper_en_2019.pdf.
- [6] *Terahertz spectroscopic system TAS7400 product specification*, https://www.advantest.com/documents/11348/146157/spec_TAS7400_EN.pdf.
- [7] *2019 ICT Sustainable development goals benchmark*, 2019. https://www.huawei.com/minisite/giv/Files/whitepaper_en_2019.pdf.
- [8] *United Nations Development Programme, Goal 11: Sustainable cities and communities*. <https://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-11-sustainable-cities-and-communities.html>.
- [9] *ICTs for a sustainable world #ICT4SDG*, ITU, 2019. <https://www.itu.int/en/sustainable-world/Pages/default.aspx>.
- [10] V. Bush, *Science, the endless frontier*. Ayer Company Publishers, 1995.
- [11] N. C. Thompson, K. Greenewald, K. Lee, G. F. Manso, *The computational limits of deep learning*, arXiv preprint arXiv:2007.05558, 2020.
- [12] *ICT sector helping to tackle climate change*, Dec. 2016. <https://unfccc.int/news/ict-sector-helping-to-tackle-climate-change>.
- [13] *SMARTer2030: ICT solutions for 21st century challenges*, The Global eSustainability Initiative (GeSI), Brussels, Brussels-Capital Region, Belgium, Technical Report, 2015.
- [14] E. Strubell, A. Ganesh, and A. McCallum, *Energy and policy considerations for deep learning in NLP*, arXiv preprint arXiv:1906.02243, 2019.
- [15] H. Cai, C. Gan, T. Wang, Z. Zhang, and S. Han, *Once-for-all: Train one network and specialize it for efficient deployment*, arXiv preprint arXiv:1908.09791, 2019.
- [16] ITU-R, *IMT Vision – framework and overall objectives of the future development of IMT for 2020 and beyond*, Recommendation ITU-R M.2083-0, Sept. 2015.
- [17] ITU-R, *Minimum requirements related to technical performance for IMT-2020 radio interface(s)*, Report ITU-R M.2410-07, Nov. 2017.
- [18] E. R. Griffor, C. Greer, D. A. Wollman, and M. J. Burns, *Framework for cyber-physical systems: Volume 1, overview*, Version 1.0, NIST Special Publication, 2017.

- [19] C. Dukes, *Committee on national security systems (CNSS) glossary*, CNSSI, Fort Meade, MD, USA, Technical Report, vol. 4009, 2015.
- [20] S. L. Garfinkel, *De-identification of personal information*, National Institute of Standards and Technology (US Department of Commerce), 2015.
- [21] M. Swanson, *Contingency planning guide for federal information systems*. DIANE Publishing, vol. 800, 2011.
- [22] *System safety*, Department of Defense Standard Practice, May 11, 2012. <https://e-hazard.com/wp-content/uploads/2020/08/departement-of-defense-standard-practice-system-safety.pdf>.
- [23] IEEE Standards Coordinating Committee, *IEEE standard glossary of software engineering terminology (IEEE Std 610.12-1990)*. Los Alamitos, CA, CA: IEEE Computer Society, vol. 169, 1990.

Часть II



Сценарии использования и целевые KPI

Введение в часть II

В части II более подробно обсуждаются типичные варианты использования 6G, включая различные аспекты их влияния на нашу жизнь и работу в будущем. Поскольку для развития сервисов требуется как минимум два поколения, некоторые службы, анонсированные в 5G, достигнут зрелости только в 6G, к 2030 году или позже. Кроме того, технические требования некоторых новых служб будут превосходить возможности 5G. В следующих главах мы представляем шесть категорий вариантов использования вместе с ключевыми требованиями к каналам радиопередачи и сетевой архитектуре. В первую категорию входят варианты, требующие чрезвычайно полного эффекта погружения человека в виртуальную реальность. XR-видео, тактильная и мультисенсорная информация и трехмерные голографические изображения будут предоставлять пользователям глубокоиммерсивный опыт, мгновенно преодолевающий расстояния в физическом мире. Вторая категория – сканирование, локализация и визуализация. Эти новые возможности принесут улучшения качества жизни во многие области и даже в новые регионы, которые никогда не были охвачены предыдущими системами мобильной связи. Следующие две категории – автоматизация промышленности и умный город / умная жизнь. Высоконадежные сети беспроводной связи со сверхмалой задержкой, а также достижения в области IoT и ICT помогут внедрить на будущих предприятиях гораздо бóльшую степень автоматизации, чем когда-либо раньше. Как в промышленности, так и в повседневной жизни будут использоваться разнообразные новые типы роботов, а города и окружающая среда станут умными. Пятая категория – это глобальный охват земной поверхности мобильными услугами. За счет интеграции наземных и неназемных сетей 6G обеспечит покрытие каждого уголка земли. Помимо услуг беспроводной широкополосной связи и связи IoT, это трехмерное полное покрытие дает новый импульс для таких вариантов использования, как навигация и наблюдение за поверхностью Земли. Последняя, шестая категория – это подключенное машинное обучение и сетевой ИИ. Эта категория наглядно свидетельствует, что машинное обучение и искусственный интеллект являются ключевыми факторами, способствующими повышению быстродействия и качества работы сети 6G. В то же время непревзойденная производительность и интегрированные функции сканирования, связи и вычислений в сети 6G становятся ключевым фактором для распространения услуги подключенного нативного искусственного интеллекта.

Глава 2

Максимально достоверный эффект погружения

На протяжении поколений от 1G до 4G, как показано на рис. 2.1, беспроводные сети в основном использовались для удовлетворения наших потребностей в достаточно быстром дистанционном обмене информацией, посредством голосовых вызовов, текстовых сообщений, потоковой передачи видео или звука. Масштабное развертывание сети 5G раскрывает потенциал приложений XR, включая VR, AR и MR.

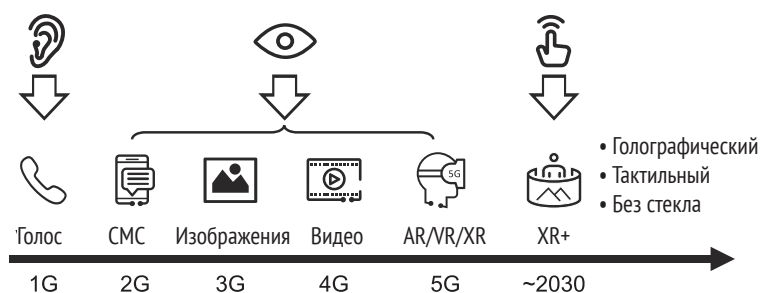


Рис. 2.1. Развитие ориентированных на человека применений в беспроводной связи

Однако для достижения удовлетворительных результатов с точки зрения достоверности погружения в удаленные виртуальные сценарии необходимо достичь аудиовизуальных ощущений высокой четкости и тактильного восприятия людей и объектов в целевой среде. Мы полагаем, что беспроводная сеть 6G будет поддерживать весьма захватывающий пользовательский интерфейс, который будет развиваться в следующих трех ключевых направлениях:

- полноэкранный чрезвычайно достоверный режим XR с полным обзором в 360° с очень высоким разрешением и частотой кадров видео, близкой к пределу человеческого восприятия;

- интерактивная тактильная и мультисенсорная связь для телеуправления с новыми человеко-машинными интерфейсами;
- 3D-дисплей без стекла и голографический дисплей, интегрированный с XR.

В следующем разделе мы подробно рассмотрим возможные варианты использования новых устройств и их требования к мобильной сети.

2.1. Облачная VR высокой степени достоверности

Пример 2.1 Кейт любит играть в футбол. Однако у нее остается не так много времени на игру, потому что ей часто приходится ездить в командировки. Но теперь благодаря устройствам VR 360° она может виртуально играть в футбол со своими друзьями в любом месте и в любое время. Благодаря облачным технологиям виртуальной реальности, которые обеспечивают оптимальное визуальное погружение и чрезвычайно низкую интерактивную задержку, Кейт чувствует себя так, как будто она на самом деле играет в футбол на стадионе, и может делать это в течение длительного времени, не испытывая головокружения и дезориентации, как показано на рис. 2.2.

Кейт не только играет в виртуальный футбол, но и любит смотреть футбольные матчи в прямом эфире с помощью устройств VR 360°. Они позволяют ей смотреть на матчи с точки зрения судьи, благодаря чему она испытывает эффект присутствия на поле и более яркие эмоции, чем зрители на стадионе.



Рис. 2.2. Интерактивный игровой процесс с VR 360°

3GPP уже начал исследование текущей сети 5G [1] в контексте XR (расширенной реальности), в котором были определены многочисленные варианты использования, такие как иммерсивные онлайн-игры, обмен 3D-сообщениями, иммерсивная потоковая передача с шестью степенями свободы, 3D-общение в реальном времени, удаленные консультации

в промышленности и при покупках в интернете. Оценки показывают, что в течение следующего десятилетия все больше людей будут подписываться на новые приложения по мере появления более продвинутых XR-сервисов с эффектом глубокого погружения. В следующем разделе мы обсудим требования, которые эффект глубокого погружения предъявляет к пропускной способности и задержке мобильной передачи.

2.1.1. Требования к задержке передачи

Требования к задержке передачи в основном возникают из-за необходимости избегать ощущений укачивания и дезориентации, вызванных несоответствием между ощущениями, которые испытывает наш вестибулярный аппарат, и фактическим изображением, которое мы видим. Это несоответствие обычно вызвано задержкой в канале между пользователем, взаимодействующим с облачной системой виртуальной реальности, и системой, реагирующей на это взаимодействие с пользователем. 3GPP называет этот тип интерактивной задержки *лагом движение–фотон*¹⁰ (motion-to-photon, MTP) [2].

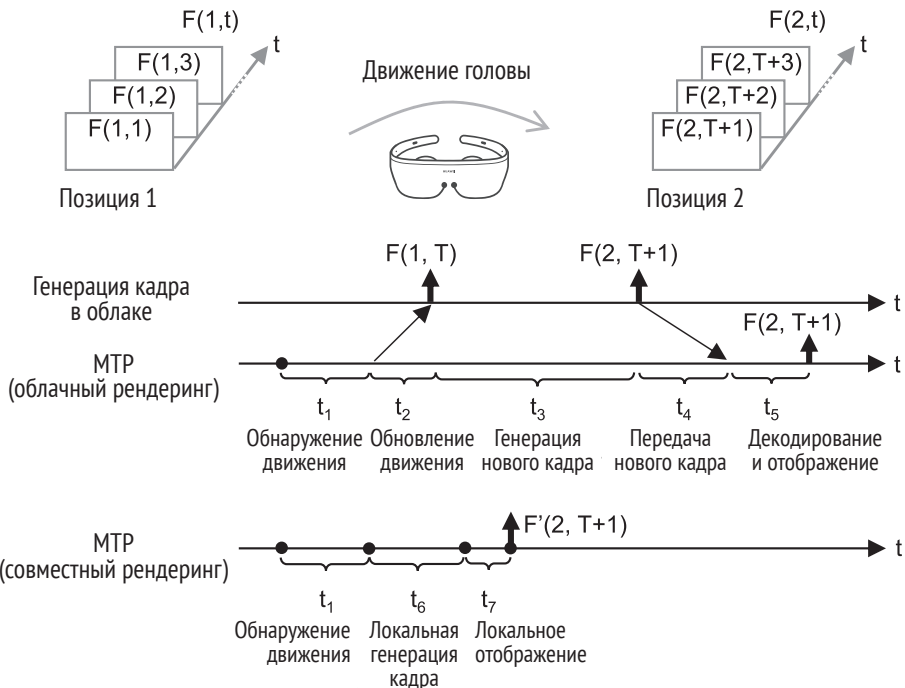


Рис. 2.3. Иллюстрация MTP с локальным рендерингом и без него

¹⁰ Этот немного странный термин ведет свое происхождение от шутивого жаргона компьютерных инженеров, которые «первым фотоном» называли момент начала прорисовки изображения лучом электронно-лучевой трубки. – Прим. перев.

Для достижения оптимального эффекта погружения целевая задержка МТР должна быть близка к пределу человеческого восприятия, что составляет примерно 10 мс [3, 4]. Это означает, что если задержка обновления изображения не превышает 10 мс, большинство людей будут испытывать комфортное погружение в виртуальную реальность.

Однако, как показано на рис. 2.3, между движением и «первым фоном» (появлением изображения на дисплее) существует множество этапов обработки и передачи сигнала. Задержка МТР будет зависеть от того, происходит ли на стороне устройства VR локальная обработка для генерации прогнозируемых кадров перед получением обновленных кадров из облака.

Система XR с полностью удаленным облачным рендерингом

Для работы системы XR с полностью удаленным облачным рендерингом между выполнением движения и отображением обновленного изображения необходимы следующие шаги:

- **обнаружение движения:** регистрация движения, например изменения положения головы, которая выполняется на стороне устройства, занимает время t_1 и зависит от чувствительности датчиков устройства и алгоритмов обнаружения;
- **обновление движения:** отправка данных с устройства в облако для обновления движения, которая занимает t_2 и связана с задержкой в сети;
- **генерация нового кадра в облаке:** формирование (включая рендеринг и кодирование) нового видеокadra после получения обновленной информации о движении, которое занимает t_3 и связано с эффективностью обработки в облаке и частотой кадров видео в секунду (frames per second, FPS);
- **передача нового кадра:** передача обновленного видеокadra из облака на устройство, которая занимает t_4 и связана с задержкой в сети;
- **декодирование и отображение** обновленного видеокadra на устройстве, которое занимает время t_5 и связано с быстродействием оборудования.

Следовательно, задержка МТР при полностью удаленном рендеринге может быть получена по следующей формуле:

$$\text{МТР}_{\text{remote}} = t_1 + t_2 + t_3 + t_4 + t_5 = (t_1 + t_5) + t_3 + (t_2 + t_4),$$

которую можно разделить на три части:

- $t_1 + t_5$: задержка обработки устройства;

- t_3 : задержка удаленной обработки;
- $t_2 + t_4$: время приема-передачи (round-trip time, RTT), то есть задержка передачи в сети, включая RTT в радиодоступе и базовых сетях.

При умеренной частоте кадров облачного видео, скажем 60 кадров в секунду, один только t_3 находится на уровне $1/60 = 16,7$ мс, что уже превышает целевой МТР в 10 мс. Увеличение частоты кадров видео, скажем с 60 до 120 кадров в секунду, поможет уменьшить $\text{МТР}_{\text{remote}}$ на 8,3 мс, хотя это будет происходить за счет увеличения требований к пропускной способности. В текущем стандарте 3GPP [2] требования смягчены, и целевая задержка МТР составляет 20 мс, оставляя больше времени для обработки и передачи. Кроме того, из-за более высокой обрабатывающей способности современных XR-устройств популярным методом является совместный локальный рендеринг в облачной системе и XR на стороне устройства, применяемый для исключения RTT из расчета задержки МТР.

По мере диверсификации устройств в будущем на рынке может остаться место только для устройств виртуальной реальности на основе полностью удаленного рендеринга, по крайней мере для очень легких носимых устройств с ограниченными возможностями обработки. Учитывая это, для достижения конечной цели задержки МТР в 10 мс все три части в $\text{МТР}_{\text{remote}}$ должны быть значительно сокращены, а именно:

- задержка обработки на стороне устройства ($t_1 + t_5$) может быть уменьшена просто вследствие действия закона Мура. Новые материалы и вычислительная архитектура наверняка помогут ускорить процесс, который будет обсуждаться в части IV книги;
- задержка удаленной обработки (t_3) может быть уменьшена за счет уменьшения степени сжатия и использования преимущества значительно увеличенной скорости передачи данных, ожидаемой от 6G;
- RTT ($t_2 + t_4$) можно уменьшить за счет распределенных архитектур граничных вычислений и радиоинтерфейсов с чрезвычайно малой задержкой в 6G.

Система XR с локальным рендерингом

Для облачной системы XR с локальным рендерингом и генерацией кадров интервал между движением и «первым фотоном» может быть сокращен следующим образом:

- **обнаружение движения:** то же, что и в предыдущем случае, занимает t_1 ;

- **генерация локального кадра:** локальная генерация кадра для обновления на основе алгоритмов локального рендеринга, занимает t_6 ;
- **локальное отображение:** отображает локально обновленный кадр без декодирования сигналов, как в предыдущем случае, и занимает время t_7 меньше, чем t_5 .

Следовательно, задержка MTP при локальном рендеринге может быть рассчитана по следующей формуле:

$$\text{MTP}_{\text{local}} = t_1 + t_6 + t_7.$$

Излишне говорить, что качество восприятия во многом зависит от качества локального рендеринга, и это решение нежизнеспособно в долгосрочной перспективе. Следовательно, мы можем рассчитывать только на обновление кадров в облаке. Но задержку приема-передачи, время удаленного рендеринга, а также задержку сжатия и декомпрессии можно значительно уменьшить. Например, $t_2 + t_4$ не включается в расчет $\text{MTP}_{\text{local}}$ для оптимального опыта XR с локальным рендерингом; следовательно, требование к RTT может быть облегчено с менее 2 мс до приблизительно 8 мс, как показано в табл. 2.1.

Таблица 2.1. Требования к пропускной способности и задержке для погружения VR 360°

Параметр	Текущая VR	Полная VR1	Полная VR2
Разрешение видеокadra	4K	24K	48K
Частота кадров	60	120	120
Глубина цвета (битов на пиксель)	24	36	36
Исходный поток данных	10,62 Гбит/с	1146,62 Гбит/с	2293,24 Гбит/с
Коэффициент сжатия	100:1	100:1	20:1
Поток после сжатия	0,1 Гбит/с	11,5 Гбит/с	114,66 Гбит/с
Задержка MTP	20 мс [2]	10 мс [2]	10 мс
Задержка RTT	< 20 мс [8]	< 8 мс [8]	< 2 мс

2.1.2. Требования к пропускной способности

Разрешение, глубина цвета и частота кадров видео – три ключевых фактора, определяющих требования к пропускной способности облачной системы VR. Мы можем рассчитать общую пропускную способность, просто перемножив три значения следующим образом:

пропускная способность = разрешение видеокadra × глубина цвета × частота кадров.

Для достижения оптимального эффекта погружения в виртуальную реальность все три параметра должны приблизиться к пределу человеческого восприятия. В частности, они следующие:

- **разрешение видеокадра:** это количество пикселей для отображения цифрового изображения. Требования к разрешению в основном обусловлены необходимой плотностью пикселей на единицу площади. Последний параметр обычно измеряется в пикселях на градус зрительного угла (pixels per angular degree, PPD). Ранние исследования [5] показали, что человеческий глаз имеет ограничения на минимальную степень разрешения на уровне 0,3 угловой минуты, где 1 угловая минута составляет $1/60$ градуса. Следовательно, предел человеческого восприятия равен $1 / [(1/60) \times 0,3] = 200$ PPD. С другой стороны, как сообщается в [2], общее горизонтальное поле зрения (field of view, FOV) для обычного человеческого глаза составляет примерно 120° , поэтому максимальное разрешение для нашего визуального восприятия составляет примерно $120^\circ \times 200 \text{ PPD} = 24\text{K}$ пикселей. Это означает, что изображения с разрешением выше 24K в большинстве случаев будут бесполезными. Обратите внимание, что для системы VR с полностью удаленной обработкой количество кадров или размер передаваемых видеок кадров необходимо удвоить по сравнению с локальной обработкой. Дело в том, что видеок кадры генерируются и передаются для *бинокулярного зрения* (два глаза), что удваивает размер видеок кадров для *монокулярного зрения* (один глаз). В системе VR с совместным локально-облачным рендерингом передаваемые видеок кадры предназначены только для монокулярного отображения, а рендеринг бинокулярного изображения на их основе может выполняться локально;
- **глубина цвета:** определяется как количество битов, используемых для обозначения цвета одного пикселя. Восьмибитная глубина цвета может обеспечить $2^8 = 256$ различных цветов, в то время как 24-битная глубина цвета (восемь битов на компонент цвета в цветовой структуре RGB) может обеспечить примерно 16 млн оттенков цвета. Последняя глубина называется *истинным цветом* (true colour) и широко применяется в сегодняшних системах виртуальной реальности. Хотя в некоторых случаях кажется, что истинный цвет позволяет отображать больше деталей, чем может воспринимать человеческий глаз (как было указано в [6], люди могут различать только до 10 млн цветов), существует риск возникновения заметной разницы между оттенками одного цвета, известный как *эффект цветowych полос* (color banding) [7]. Мы можем смягчить эту проблему, дополнительно увеличив глубину

цвета, например используя 36-битную глубину цвета с 12 битами на компонент цвета RGB;

- **частота кадров видео:** это скорость, с которой создаются видео-кадры, выражается в кадрах в секунду. Если отсутствует опыт интерактивного движения в виртуальной реальности (например, при просмотре фильма), 120 кадров в секунду (что равно 8,3 мс между двумя кадрами) является достаточно высоким показателем для иммерсивного опыта, поскольку он приближается к пределу человеческого восприятия (примерно 10 мс). Однако, как обсуждалось в разделе 2.1.1, более высокая частота кадров видео может помочь уменьшить МТР в интерактивных сценариях за счет более высокой пропускной способности. Например, видеопоток с разрешением 4K, частотой 60 кадров в секунду и 24-битной глубиной цвета приведет к генерации исходного потока данных 10,62 Гбит/с. Это число возрастает до 1,15 Тбит/с, если для всех параметров установлены значения, близкие к пределам человеческого восприятия, например разрешение 24K, 120 кадров в секунду и глубина цвета 36 бит.

2.1.3. Обзор основных требований для полной VR

На основе анализа, представленного в предыдущих разделах, в табл. 2.1 приведены требования к пропускной способности и задержке для полного погружения в виртуальную реальность в сравнении с сегодняшней системой VR. Как мы видим, увеличение разрешения, частоты кадров видео и глубины цвета коррелирует с более чем 100-кратным увеличением скорости необработанных данных, превышая 1 Тбит/с. Чтобы снизить нагрузку на сетевой канал, обычно применяется высокая степень сжатия (например, 133:1, как показано в [8]), когда возможен локальный рендеринг. Отсюда следует, что необходимо гарантировать скорость передачи данных выше 10 Гбит/с. Кроме того, когда установлено целевое ограничение МТР в 10 мс, требования к RTT должны быть пропорционально уменьшены. Например, в будущих системах виртуальной реальности с совместным локальным рендерингом требуется значение менее 8 мс [8].

С другой стороны, для устройств со строгими ограничениями по мощности и весу больше подходит архитектура полностью удаленного рендеринга. В этом случае при целевом значении $\text{МТР}_{\text{remote}} = 10$ мс, если степень сжатия слишком высока, для передачи не останется времени. Следовательно, предполагается, что используется малая степень сжатия (например, 20:1 в табл. 2.1), которая уменьшает $t_3 + t_5$, но RTT ($t_2 + t_4$) по-прежнему должен соответствовать очень строгому пределу, близкому к 1 мс (менее 2 мс по оценкам в табл. 2.1). Между тем из-за низкой

степени сжатия гарантированная скорость передачи данных в этом случае превышает 100 Гбит/с.

2.2. Тактильное и мультисенсорное общение

Пример 2.2 Кейт покупает тактильную одежду, чтобы носить ее во время игры в виртуальный футбол. Эта одежда обеспечивает мультисенсорную связь с игровым сервером, чтобы сделать игру более реалистичной, позволяя Кейт почувствовать текстуру, вес и давление виртуального мяча, когда она касается, удерживает и пинает его.

Кейт уехала в командировку. Хотя она не может лично встретиться со своим мужем и детьми во время поездки, сеть 6G позволяет ей виртуально встречаться с семьей и даже обнимать их через тактильную одежду почти так, как если бы она была рядом с ними.

Помимо аудио и видео, тактильные ощущения – это новое измерение ориентированного на человека контента, который будет передаваться по мобильным сетям. Тактильное общение включает в себя тактильную информацию в реальном времени, касающуюся текстуры поверхности, прикосновения, сдвига, движения, вибрации и силы нажатия, и все это передается по сети вместе с аудиовизуальной информацией. Эти возможности позволяют людям удаленно управлять машинами или роботами, или даже проецировать свое голографическое изображение через мобильные сети, чтобы выполнять сложные задачи в реальном времени, которые в противном случае были бы слишком опасными или дорогостоящими для личного участия, или просто чтобы обнять члена семьи, живущего в другом городе.

В некоторых приложениях, предлагающих погружение в виртуальную реальность, мультисенсорная обратная связь, такая как видео, аудио и тактильные ощущения, может работать одновременно. Например, в игровом приложении VR 360° геймеры могут одновременно видеть окружающую среду, слышать диалоги и воспринимать тактильные ощущения. Таким образом, для обеспечения максимально ориентированного на человека опыта – такого, который плавно соединяет физический и виртуальный миры, – потребуется мультисенсорное общение.

Человеческий мозг – сложная и мощная система, которая умеет строить ментальную модель физического мира на основе множества источников сенсорной информации, как показано на рис. 2.4. Общую задержку приема-передачи можно разделить на две составляющие: время передачи $T_{transmission}$ и время обработки $T_{processing}$ (которое включает передачу через нервную систему и обработку мозгом). Поскольку значение $T_{processing}$, как правило, одинаковое как для локальных, так и для дистанционных операций, и для компенсации изменений окружающей среды могут быть использованы прогнозы, целью тактильной и мультисенсорной связи является восприятие и реакция на изменения окружающей среды.

тисенсорной коммуникации является снижение $T_{transmission}$ для достижения опыта, сопоставимого с локальной обработкой.

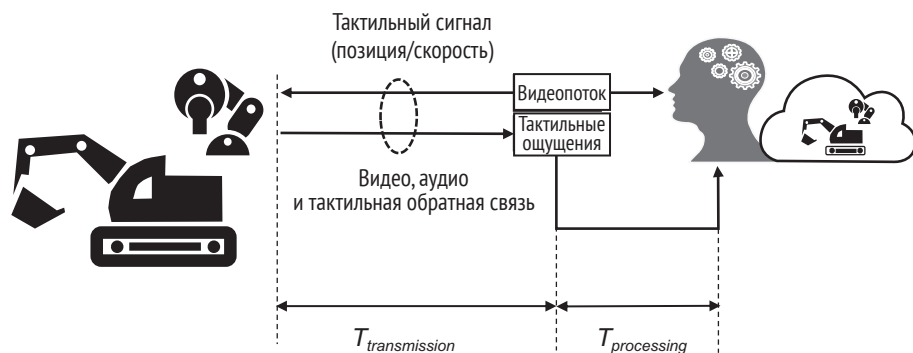


Рис. 2.4. Тактильное общение с мультисенсорной обратной связью

2.2.1. Дистанционная работа в высокодинамичных средах

Как показано на рис. 2.5, такие приложения, как интерактивная виртуальная реальность с тактильной обратной связью и дистанционное управление БПЛА – в дополнение к телеуправлению, – включают частые взаимодействия с целевыми объектами и требуют синхронизированной мультисенсорной обратной связи, представляющей высокодинамичную среду. Эти тактильные приложения являются наиболее сложными, потому что для обеспечения плавного и надежного соединения для них современные беспроводные сети необходимо оптимизировать или даже перепроектировать с учетом поведения человеческого мозга.

Тактильная обратная связь чрезвычайно важна в таких случаях телеуправления удаленными объектами, как телехирургия, теледиагностика и дистанционное управление движением. Это связано с тем, что тактильная обратная связь, например информация о силе сжатия или текстуре, может стимулировать человеческий мозг и помогать пользователям выбирать продолжительность воздействия, усилие, жесты и т. д.

Для тактильного управления и интерактивных тактильных приложений в высокодинамичных средах обычно требуется очень низкая задержка. Многие действия, выполняемые удаленно, такие как дистанционная хирургия и управление на промышленном уровне, требуют чрезвычайно точного управления движением. В таких случаях точность перемещения, которая не контролируется в пределах допуска 1 мм, может привести к сбою процесса, серьезно снизив эффективность работы или даже создав риск для жизни человека. Если мы движемся со

скоростью 1 м/с относительно объекта и хотим управлять движениями с допуском 5 мм, нам потребуется общая задержка менее 5 мс. Хотя мы можем в некоторой степени предсказать и компенсировать задержку передачи, для эффективной обработки изменений в высокودинамичных и интерактивных средах максимальная общая задержка все же не должна превышать 10 мс [9]. Стоит отметить, что указанная задержка представляет собой общую задержку, включая обнаружение движения, RTT передачи по сети и время интерактивной обратной связи. В некоторых случаях, таких как интерактивная дистанционная работа, требование к RTT для передачи по радиоинтерфейсу может составлять менее 1 мс или даже 0,1 мс.



Рис. 2.5. Типичные приложения с тактильной связью

Мультисенсорная обратная связь, в том числе тактильная, важна при телеуправлении. Например, чтобы управлять удаленной машиной точно и быстро, оператору необходимо одновременно воспринимать аудиовизуальную и тактильную информацию. Это означает, что относительная задержка передачи между аудио, видео и тактильной информацией должна лежать в пределах естественных порогов человеческого восприятия. Однако пороги восприятия для разных органов чувств не совпадают. Кроме того, пока не совсем понятно, как относительные пороги восприятия, определяемые как относительная задержка между различными органами чувств, влияют на одновременный прием мультисенсорной информации. Это все еще открытая тема для иссле-

дования, но в более раннем исследовании [10] были сделаны следующие выводы:

- индивидуальные различия в порогах восприятия выглядят значительными;
- связь между средними порогами восприятия звуковой, визуальной и тактильной информации можно описать соотношением *тактильный < аудио ≈ визуальный*, что указывает на то, что тактильная информация обладает минимальным, т. е. наиболее критичным к задержке, порогом из трех.

2.2.2. Основные требования к телеуправлению в высокодинамичных средах

Требования к телеуправлению в высокодинамичной среде (что является основным сценарием использования тактильной связи для 6G) описаны ниже и кратко изложены в табл. 2.2:

Таблица 2.2. Сводка требований к телеуправлению в высокодинамичной среде

Параметр	Тактильная информация	Прочая сенсорная информация
Общая задержка	$\leq 1\text{--}10$ мс	$\leq 10\text{--}20$ мс
Пропускная способность	1000–4000 пакетов/с	≤ 100 Мбит/с
Надежность	$\geq 99.999\%$	$\geq 99.999\%$

- **задержка:** как обсуждалось в предыдущем разделе, общая задержка, допустимая для высокодинамичного телеуправления, составляет от 1 до 10 мс. Для достижения этого показателя при условии обеспечения стабильного телеуправления нам требуется чрезвычайно высокая скорость передачи пакетов. Отправка 1000 пакетов в секунду соответствует задержке в 1 мс между двумя последовательными пакетами, т. е. для удовлетворения требований к общей задержке в 1 мс скорость передачи должна превышать 1000 (в идеале до 4000) пакетов в секунду [9];
- **относительная задержка:** важна синхронизация между видео, аудио и тактильной обратной связью. Некоторые эксперименты [10] показывают, что для обеспечения хорошего опыта взаимодействия с пользователем необходима относительная задержка менее 10–20 мс между любыми двумя типами сенсорной информации;
- **пропускная способность:** для тактильной информации пропускная способность определяется скоростью передачи и размером пакета. В большинстве случаев тактильная информация

состоит из нескольких байтов для каждой степени свободы (DoF) дистанционно управляемого движущегося объекта. Это означает, что размер пакета линейно зависит от количества DoF (количества точек тактильного сенсорного поля). Например, размер пакета, содержащего шесть DoF, обычно составляет 12–48 байт. Если предположить, что скорость передачи от 1000 до 4000 пакетов в секунду, тактильная пропускная способность составит от 96 кбит/с до 1,5 Мбит/с. Если количество DoF увеличивается до 100, размер пакета увеличивается до 200–800 байт, что означает смещение тактильной пропускной способности в диапазон от 1,6 до 25,6 Мбит/с. Для другой сенсорной информации пропускная способность зависит от качества звука и видео. Например, для передачи видео в формате 4K требуется около 100 Мбит/с, тогда как для варианта полной VR, обсуждаемого в разделе 2.1, требуется гораздо более высокая пропускная способность;

- **надежность:** сбой передачи может повлиять на качество работы в дистанционном режиме или даже создать риск для жизни человека, поэтому в некоторых высоконадежных системах телеуправления требуется надежность передачи 99,999 % [9].

2.3. Дисплеи 3D без стекла и голографические дисплеи

Пример 2.3 Джек навещает клиента в городе со сложным гористым рельефом. В центральной части города много высотных домов и путепроводов. Дорожная сеть похожа на трехмерный лабиринт, и люди могут легко заблудиться там, если они незнакомы с местностью. Однако на этот раз для Джека все было иначе. Перед поездкой Джек установил на свой смартфон приложение для 3D-навигации. Приложение использует методы трехмерного виртуального отображения всех зданий и дорог в реальном времени. С помощью этих методов трехмерного виртуального отображения Джек может легко перемещаться по сложной дорожной сети и вовремя прибывать в пункт назначения.

2.3.1. Глубина восприятия и 3D-дисплеи без стекла

Хотя виртуальная реальность формирует визуальный опыт, аналогичный реальному миру, она не может имитировать то, как человеческие глаза воспринимают глубину (то есть когда глаза фокусируются на близких объектах, удаленные объекты становятся размытыми, и наоборот). В VR пользователи всегда фокусируются на экране VR, независимо от того, находится ли объект близко или далеко. Это создает конфликт восприятия глубины, вызывая у пользователей головокружение или другие нежелательные эффекты. Ожидается, что для устранения этих

неприятных эффектов следующим поколением иммерсивных XR станут трехмерные дисплеи без стекла, основанные на визуальной аккомодации, а не на психологическом восприятии или параллаксе движения.

Концепция 3D-дисплеев без стекла существует уже давно. Разработаны различные конструкции с использованием методов светового поля и голографического отображения [11, 12]. Сегодня использование таких дисплеев в основном ограничивается локальными сценариями, например заранее записанные голографические концерты [13]. Эти варианты применения в настоящее время не требуют мобильной связи; следовательно, они не имеют прямого отношения к мобильным сетям. Однако в будущем, когда появятся новые приложения, такие как мобильная 3D-навигация, ожидается, что сети 6G должны будут передавать 3D-изображения или голограммы. Это предъявляет чрезвычайно высокие требования к производительности сетей 6G, включая такие факторы, как пропускная способность и задержка. В этом разделе мы сосредоточимся в основном на световом поле и методах голографического отображения, используемых в трехмерных дисплеях без стекла [12].

2.3.2. Методы реконструкции трехмерных изображений без использования стекла

Дисплеи светового поля собирают информацию о световом поле от источника, обычно с использованием массивов камер, и передают полученную информацию на экран устройства, где реконструируются трехмерные изображения. Процесс реконструкции основан на проецировании отдельных 2D-изображений с непрерывной информацией о параллаксе на несколько точек обзора. Этот процесс обычно основан на методах, которые могут разделять световые лучи, идущие от экрана, с помощью таких оптических компонентов, как решетки линз Френеля, решетки микролинз и оптика с направлением лучей. На рис. 2.6а показано, как пользователи просматривают разные параллаксные изображения в разных точках обзора, чтобы наблюдать трехмерные изображения. Каждая точка обзора получает набор световых лучей, испускаемых плоскостью экрана под разными углами.

Голографические дисплеи основаны на процессе реконструкции волнового фронта, в котором в качестве активных элементов управления обычно используется волновая оптика или пространственные модуляторы света (spatial light modulator, SLM). Как показано на рис. 2.6б, голографический дисплей обычно состоит из SLM и источника света, такого как лазер или светоизлучающий диод (light emitting diode, LED) (одноцветный или RGB) [14]. Изменение фазового распределения на SLM позволяет пользователям восстанавливать и воспринимать голографические изображения.

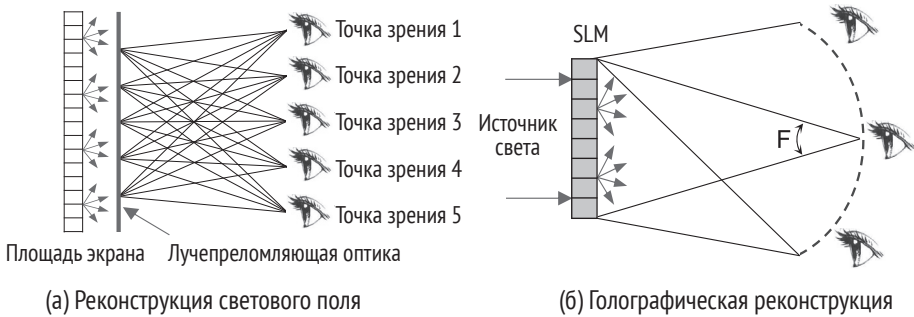


Рис. 2.6. Реконструкция трехмерного изображения без использования стекла на основе (а) направленных лучей из плоскости экрана и (б) пространственных модуляторов света

2.3.3. Требования к разрешению и задержке

Используя дисплеи светового поля, пользователи могут реконструировать 3D-изображения в различных точках обзора, наблюдая набор световых лучей, излучаемых из плоскости экрана. Поскольку для достижения хорошего восприятия трехмерного изображения требуется большее количество точек обзора, плотность пикселей на плоскости экрана может быть чрезвычайно высокой; следовательно, отображение объемного изображения с высоким разрешением требует передачи большого количества данных. Разрешение отображения светового поля можно рассчитать исходя из количества элементов отображения на плоскости экрана и количества лучей на элемент [12]. Его также можно рассчитать по количеству точек обзора и разрешению каждой точки обзора. В качестве примера возьмем изображение светового поля размером 6 дюймов ($133 \text{ мм} \times 75 \text{ мм}$): если количество точек обзора составляет 144 (т. е. 12×12), а разрешение каждой точки обзора равно 1080 пикселей (1920×1080), то размер пикселя $\sim 5,8 \text{ мкм}$, а общее разрешение составляет $3,0 \times 10^8$ пикселей.

Для голографических дисплеев на основе волнового фронта размер пикселя рассчитывается с использованием угла дифракции поля зрения голографического дисплея. Для поля зрения $\sim 30^\circ$ размер пикселя должен быть 1 мкм [15]. В качестве примера возьмем голографический дисплей размером 10 дюймов ($200 \text{ мм} \times 150 \text{ мм}$): согласно расчетам в [12], если размер пикселя трехмерного изображения равен 1 мкм, разрешение составляет $(200 \text{ мм} / 1 \text{ мкм}) \times (150 \text{ мм} / 1 \text{ мкм}) = 3,0 \times 10^{10}$ пикселей.

Поскольку разработка 3D-дисплеев без стекла все еще находится в зачаточном состоянии, еще не проводились углубленные исследования того, как частота обновления изображения влияет на качество отображения. Хорошей отправной точкой является использование час-

тоты обновления изображения 60 Гц для дисплеев светового поля [16] и 30 Гц для голографических дисплеев из-за сложного и трудоемкого процесса, связанного с созданием голограмм [12]. Однако для достижения максимального эффекта погружения, приближающегося к пределу человеческого восприятия, как обсуждалось в предыдущих разделах, частота обновления 30 или 60 Гц может оказаться недостаточной.

2.3.4. Основные требования к 3D-дисплеям без стекла

На основе анализа, представленного в предыдущих разделах, в табл. 2.3 перечислены типичные требования к скорости передачи исходных данных для различных методов трехмерного отображения. Скорость передачи несжатых данных для 3D-дисплеев без стекла чрезвычайно высока: около 0,4 Тбит/с для дисплея со световым полем размером с мобильный телефон и 184 Тбит/с для 50-дюймового голографического дисплея. Эффективное сжатие такого огромного количества данных – серьезная проблема. В [11] авторы предложили способ передачи сжатых данных 3D-объекта, а не данных 3D-голограммы. Этот метод направлен на снижение требований к пропускной способности, но его еще предстоит развить.

Таблица 2.3. Требования к скорости передачи исходных данных для различных типов 3D-дисплеев без стекла

Параметр	Световое поле	Голографический	Голографический
Размер изображения (дюймы)	6	10	50
Размер пикселя (мкм)	5,8	1	1
Частота обновления (Гц)	60	30	30
Поток пикселей (пикселей/с)	$1,79 \times 10^{10}$	$9,0 \times 10^{11}$	$2,3 \times 10^{13}$
Цветность	Полноцветный	Одноцветный	Одноцветный
Битов на пиксель	24	8	8
Поток несжатых данных (Тбит/с)	0,4	7,2	184

2.4. Источники

- [1] 3GPP, *Extended reality (XR) in 5G*, 3rd Generation Partnership Project (3GPP), Technical Report (TR) 26.928, 03 2020, version 16.0.0. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3534>.

- [2] 3GPP, *Virtual reality (VR) media services over 3GPP*, 3rd Generation Partnership Project (3GPP), Technical Report (TR) 26.918, 03 2020, version 16.0.0. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3053>.
- [3] P. Jombik, V. Bah`yl, *Short latency disconjugate vestibulo-ocular responses to transient stimuli in the audio frequency range*, Journal of Neurology, Neurosurgery & Psychiatry, vol. 76, no. 10, pp. 1398–1402, 2005.
- [4] M. S. Amin, *Vestibuloocular reflex testing*, Medscape Article Number 1836134, 2016. <https://emedicine.medscape.com/article/1836134-overview>.
- [5] H. R. Blackwell, *Contrast thresholds of the human eye*, Journal of the Optical Society of America, vol. 36, no. 11, pp. 624–643, 1946.
- [6] D. Judd, *Color in business, science, and industry*, 1975.
- [7] Wikipedia, *Colour banding*, 2020, https://en.wikipedia.org/wiki/Colour_banding.
- [8] Huawei Technologies Co., Ltd., Boe Technology Group Co., Ltd., and CAICT, *Ubiquitous display: Visual experience improvement drives explosive data growth*, 2020, https://www.huawei.com/minisite/static/Visual_Experience_White_Paper_en.pdf.
- [9] O. Holland, E. Steinbach, R. V. Prasad, Q. Liu, Z. Dawy, A. Aijaz, N. Pappas, K. Chandra, V. S. Rao, S. Oteafy et al., *The IEEE 1918.1 tactile internet standards working group and its standards*, Proceedings of the IEEE, vol. 107, no. 2, pp. 256–279, 2019.
- [10] D. L. Woods, J. M. Wyma, E. W. Yund, T. J. Herron, B. Reed, *Factors influencing the latency of simple reaction time*, Frontiers in Human Neuroscience, vol. 9, p. 131, 2015.
- [11] X. Xu, Y. Pan, P. P. M. Y. Lwin, X. Liang, *3D holographic display and its data transmission requirement*, in Proc. 2011 International Conference on Information Photonics and Optical Communications. IEEE, 2011, pp. 1–4.
- [12] M. Yamaguchi, *Light-field and holographic three-dimensional displays*, Journal of the Optical Society of America, vol. 33, no. 12, pp. 2348–2364, 2016.
- [13] P. Gallo, *Michael Jackson hologram rocks billboard music awards: Watch & go behind the scenes*, 2014. <https://www.billboard.com/articles/news/6092040/michael-jackson-hologram-billboard-music-awards/>.
- [14] L. Onural, F. Yara, s, H. Kang, *Digital holographic three-dimensional video displays*, Proceedings of the IEEE, vol. 99, no. 4, pp. 576–589, 2011.
- [15] Q. Jiang, G. Jin, L. Cao, *When metasurface meets hologram: Principle and advances*, Advances in Optics and Photonics, vol. 11, no. 3, pp. 518–576, 2019.
- [16] L. G. Factory, *Holographic displays*. <https://lookingglassfactory.com>.

Глава 3

Сканирование, локализация и визуализация

Как было сказано в главе 1, сканирование станет новой функцией, интегрированной в систему связи 6G. Новые возможности сканирования окружающей среды основаны на измерении и анализе параметров беспроводных сигналов, и этот подход значительно отличается от сетей беспроводных сенсоров, основанных на сборе и передаче данных с локальных чувствительных элементов. 6G-сканирование сможет заменить специализированное оборудование, такое как радар, лидар, а также профессиональные установки КТ и МРТ.

В системе ISAC технология сканирования может превратить базовые станции 6G, устройства 6G или даже всю сеть 6G в распределенные сенсоры. В качестве примеров применения 6G-сканирования можно назвать локализацию подвижных объектов, распознавание жестов и действий, а также съемку и картографирование.

В сети 6G более высокие диапазоны частот (миллиметровые волны и ТГц), более широкая полоса пропускания и массивные антенные решетки позволят создавать решения для сканирования и визуализации с очень высоким разрешением и точностью. Эти решения будут необходимым условием предоставления высококачественных услуг во многих областях, таких как общественная безопасность и защита критически важных активов, мониторинг здоровья и благополучия, интеллектуальный транспорт, умные дома и фабрики, распознавание жестов и движений, измерение качества воздуха и обнаружение газов / токсичных веществ. В этом контексте мы введем несколько новых KPI для технологии сканирования:

- **точность измерения:** разница между измеренными и реальными значениями в диапазоне, угле, скорости и т. д.;
- **разрешение обнаружения:** возможность различения объектов по дальности, углу, скорости и т. д.;

- **вероятности обнаружения / ложной тревоги:** вероятность того, что объект будет обнаружен, когда он присутствует/отсутствует.

В следующих разделах описаны четыре категории сценариев использования интегрированного сканирования и связи и проанализированы соответствующие требования KPI.

3.1. Высокоточная локализация

Пример 3.1 Джек – соучредитель стартапа по производству летающих автомобилей. На его заводе развернута высокоточная система трехмерной локализации с поддержкой 6G.

Система обеспечивает связь и сканирование, включая расширенную локализацию и навигацию, а также картографические сервисы для полной автоматизации роботов. Она служит киберфизическим мостом для физических объектов, материалов, обрабатываемых деталей и рабочего процесса цифрового производства. В рамках этой системы Джек и его деловые партнеры могут отслеживать, диагностировать и оптимизировать производственный процесс в любом месте и в любое время. Для экономии места мелкие детали хранятся на трехмерном координатном складе. Роботы-дроны могут легко извлекать товары со склада и загружать их в автоматизированное управляемое транспортное средство (automated guided vehicle, AGV), используя очень точную и мягкую вертикальную подачу, как показано на рис. 3.1. Процесс включает в себя точную трехмерную локализацию, которая необходима, чтобы не повредить доставляемые детали. Роботы AGV могут эффективно транспортировать детали на высокой скорости, поэтому им необходимо избегать столкновений друг с другом и препятствий, применяя локализацию в реальном времени. Задержка на уровне миллисекунд и точность на уровне миллиметра позволяют осуществлять динамическое планирование пути в реальном времени. Чтобы поднять турбовентиляторный двигатель из грузового отсека и установить его в заданном месте, роботы AGV должны будут тесно взаимодействовать со стационарными роботами. Для подъема, переноски, монтажа и сварки требуются высокоточные операции с использованием сочетания беспроводной локализации 6G и других датчиков ближнего действия.

Сеть 6G будет предоставлять услуги по локализации на основе устройств 6G и локализации объектов без устройств. Для локализации на основе устройств 6G информация о местоположении устройств выводится из принятых опорных сигналов или измерительных данных от целевых устройств. С другой стороны, локализация для объектов без устройств 6G больше похожа на работу радара. Она включает в себя обработку информации о задержке, доплеровском спектре и угловом спектре (используемой для отображения расстояния, скорости и углового положения объектов окружающей среды) от рассеянных и отраженных сигналов беспроводной связи. Путем дальнейшей обра-

ботки мы можем извлекать координаты, скорость позиционирования и другую геометрическую информацию в физическом трехмерном пространстве. Обратите внимание, что высокоточное измерение расстояния также может обеспечить синхронизацию часов, что важно для промышленных сетей.

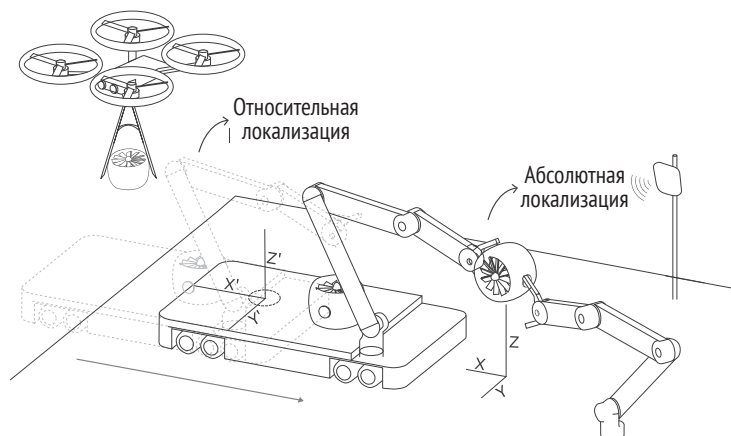


Рис. 3.1. Высокоточная локализация на фабрике летающих автомобилей из будущего

При использовании текущей сети 5G на открытом воздухе целевыми показателями являются ошибки горизонтальной и вертикальной локализации менее 10 м и 3 м соответственно, а также доступность 80 % для обоих измерений. При разворачивании внутри помещений наша цель состоит в том, чтобы ошибки локализации по горизонтали и вертикали (при 80 % доступности) составляли менее 3 м [1]. В благоприятных условиях сеть 6G обеспечит локализацию с точностью до сантиметрового уровня – на открытом воздухе с большой антенной решеткой, а внутри помещений со связью в ближнем поле. В целом внутренняя среда является более сложной и неоднородной, поскольку препятствия, например стены, мебель, оборудование, люди и т. д., сильно влияют на распространение сигналов 6G. Благодаря более высокой полосе пропускания, работе с несколькими диапазонами и увеличенной апертуре антенной решетки система ISAC в 6G может обеспечить отличные разрешающие способности для разделения многолучевого распространения и использования информации о многолучевом распространении для лучшей локализации и отслеживания.

3.1.1. Абсолютная локализация

От заводов до складов, от больниц до магазинов розничной торговли, от сельского хозяйства до горнодобывающей промышленности – вы-

сокоточная локализация и отслеживание обеспечивают связь между информацией кибермира и местоположением физических объектов. На 6G-сканирование не влияют неблагоприятные условия освещения, шум и механическая вибрация, а это означает, что данная технология обеспечивает дополнительные преимущества и надежность при определении местоположения, особенно по сравнению с традиционными световыми дальномерами на основе времени пролета или ультразвуковыми датчиками. Она позволяет оптимизировать производственные процессы, отслеживать работу в режиме реального времени, автоматизировать парковочные сервисы и т. д. Кроме того, локализация с низкой задержкой и высокой точностью может обеспечить навигацию без столкновений для грузовых автомобилей и вилочных погрузчиков. Уровень точности 10 см позволяет позиционировать устройства, а уровень точности 1 см может дополнительно обеспечить позиционирование в ограниченном пространстве на уровне модуля, обеспечивая эффективное хранение компонентов с малым форм-фактором, таких как микро-схемы и мелкие металлические детали [2].

3.1.2. Относительная локализация

Относительная локализация определяется как определение положения двух или более объектов, которые движутся друг относительно друга, или когда объекты имеют согласованные направление движения и скорость. Это крайне важно для автоматической стыковки и совместных действий, выполняемых несколькими роботами. Относительная локализация также может быть использована командой навигационных роботов в ситуациях, когда нам необходимо знать относительное местоположение судна по сравнению с его соседями. Относительная локализация необходима как условие функционирования на малых расстояниях, поскольку для каждого робота критически важное значение имеет точное определение его местоположения по отношению к общей системе координат. В качестве примера возьмем механическую деталь сложной формы длиной один метр. Если мы стремимся избежать потери баланса веса во время совместного подъема и переноски, учитывая, что максимальный допустимый наклон составляет три градуса, относительная ошибка локализации примерно в 5 см или даже меньше между двумя подвижными объектами может гарантировать, что объект перемещается с правильной ориентацией и балансом. Относительная локализация также может применяться к новым сценариям, например во время стыковки дрона с движущимся транспортным средством, обладающим чрезвычайно малым запасом места для посадки из-за ограниченной площади грузовой платформы.

3.1.3. Семантическая локализация

Благодаря повсеместному распространению ИИ будущие системы могут обеспечить возможность *семантической локализации* с учетом контекста. Существующие системы могут выполнять простую диспетчерскую локализацию, оперируя номером этажа, номером здания, названием улицы при обслуживании пользователей. Для поддержки будущих приложений умного дома / торгового центра / ресторана / отеля и автоматических заводских приложений объекты и компоненты должны иметь диспетчерскую локализацию, предоставляя такую информацию, как уровень полки, номер места, номер стола, номер секции и т. д. Но функциональность не ограничивается фиксированными адресами – всеобъемлющий ИИ также может поддерживать динамическую регистрацию адресов в соответствии с контекстом. Как показано на рис. 3.2, ресторан использует дроны в качестве официантов-роботов. С помощью ИИ эти роботы могут понимать контекст семантических инструкций (таких как доставка еды к нужному столу). Затем, благодаря возможности семантической локализации, они могут находить гостей и адресно доставлять еду. Действия этих дронов будут очень похожи на действия официантов-людей. Благодаря интеграции ИИ и локализации роботы идут еще дальше и подстраивают свое поведение в соответствии с характеристиками задачи, например при транспортировке хрупкие и твердые предметы будут обрабатываться с разным уровнем осторожности, включая точность определения местоположения и скорости.



Рис. 3.2. Семантическая локализация в ресторане будущего с роботами-официантами

3.2. Визуализация, картирование и локализация в реальном времени

Пример 3.2 Стартап Джека по производству летающих автомобилей расположен в центре футуристического города, в котором он живет. Важной особенностью футуристического города является реконструкция изображения в реальном времени, когда можно получить информацию об окружающей среде в реальном времени для нужд виртуального города и интеллектуальной сети, как показано на рис. 3.3. С помощью этой технологии он может замечать мелкие детали городской обстановки. Например, могут быть обнаружены и визуализированы заменяемые стеклянные навесные стены и рекламные щиты, а также потоки людей. Даже сезонные изменения листвы включены в фоновую информацию. Иногда, когда Джек возвращается с работы домой, он избегает определенных маршрутов из-за своей аллергии на иву, растущую в нескольких местах. Все эти объекты можно визуализировать и нанести на карту в реальном времени. Однажды ночью он едет в полностью автономном автомобиле по улице с ограниченным освещением, и через дорогу внезапно перебегает олень. Автономный автомобиль немедленно определяет ситуацию и применяет маневр торможения с уклонением, чтобы не травмировать ни оленя, ни пассажира.

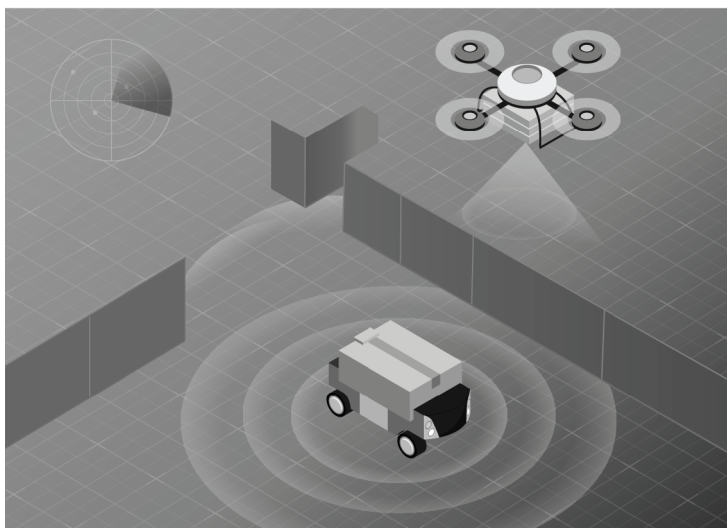


Рис. 3.3. Реконструкция и локализация виртуального образа окружающей среды в реальном времени

В предыдущем примере возможности восприятия окружающей обстановки тремя разными способами взаимно расширяются и дополняются. В частности, функция визуализации используется для получения цифровых образов окружающей среды, а функция локализации используется для получения местоположения окружающих объектов. Эти образы и/или местоположения затем используются функцией со-

поставления для построения карты. Наконец, карта помогает функции локализации улучшить вывод местоположений.

По сравнению с традиционным сканирующим радаром, ISAC в 6G будет использовать передовые алгоритмы, граничные вычисления и искусственный интеллект для создания карт сверхвысокого разрешения и хорошо распознаваемых образов. Функция сканирования в сети 6G обеспечит полностью функциональное решение для работы в течение всего дня, в отличие от существующих лидаров и оптических устройств, которые сильно зависят от света и атмосферных условий (например, тумана или облаков). Благодаря обширной сети транспортных средств, базовых станций и т. д., которые будут действовать как сенсоры, область съемки может быть значительно расширена. Более того, производительность значительно улучшится благодаря объединению результатов визуализации, которые передаются в глобальном масштабе через сеть, с облачными сервисами. Ниже мы обсудим примеры применения этих возможностей сканирования.

3.2.1. Одновременная локализация и картирование

Приложения одновременной локализации и картирования (simultaneous localization and mapping, SLAM) в миллиметровом или ТГц-диапазоне позволяют восстанавливать трехмерные карты окрестностей в неизвестных средах. SLAM основан на концепции, согласно которой обладающее сенсорами устройство, перемещаясь в неизвестной среде, распознает окружающие объекты (ориентиры) и впоследствии восстанавливает 2D- или 3D-карту окружающей среды для дальнейшего повышения точности локализации.

Текущая точная технология SLAM (для автоматического создания внутренних и внешних карт) требует как высокого разрешения при измерении расстояния (дальности), так и очень высоких угловых разрешений, которые традиционно достигаются с помощью лазерных и оптических технологий. Хотя лидары и оптические камеры могут обеспечивать высокое разрешение, они не могут работать в неблагоприятных погодных условиях (туман, дождь, облака и т. д.) и в условиях низкой внешней освещенности.

Чтобы решить предыдущие проблемы и одновременно избежать затрат на лидары и камеры, мы можем реализовать технологию SLAM, используя беспроводные сигналы 6G. Фактически из-за сложной природы окружающей среды (например, клиент находится внутри помещения с многочисленными комнатами и/или перегородками) большую часть целевой зоны обслуживания представляет собой покрытие вне зоны прямой видимости (non-line-of-sight, NLOS). В этом случае картографирование на основе 6G способно предоставить самые актуаль-

ные знания об окружающей среде, благодаря которым высокоточная локализация все еще возможна даже в сценариях NLOS. В частности, мы можем отображать расположение объектов в среде, чтобы предоставить дополнительную информацию о локализации. Затем уточненную информацию можно применить для определения точек многолучевого отражения с помощью методов трассировки лучей. По этой причине местоположение объектов может быть отслежено при помощи SLAM 6G как в условиях прямой видимости (line-of-sight, LOS), так и по NLOS. Большое количество информации о многолучевом распространении может косвенно увеличить вероятность обнаружения неоднородностей среды, тем самым улучшая производительность локализации. Мы считаем, что когда расстояние между сенсорным устройством и объектом находится в пределах обычного расстояния для помещений (примерно 10 метров), SLAM на основе 6G может точно определить местонахождение объекта на уровне сантиметра.

3.2.2. Визуализация и картирование внутри помещений

Сканирование на основе 6G открывает целый ряд возможностей в области трехмерной визуализации и картирования помещений, которые, в свою очередь, позволяют использовать различные приложения, такие как реконструкция сцены в помещении, пространственная локализация и внутренняя навигация. Такие приложения обычно требуют сверхвысокого разрешения и высокой точности. Благодаря тому факту, что рассеянные сигналы отражаются несколько раз, когда поверхности LOS действуют как зеркала, компенсированные изображения объектов NLOS могут быть восстановлены путем применения зеркального отражения.

Система визуализации, которая может распознавать объекты, расположенные за углом, будет очень ценной и применимой во многих новых сценариях использования. После реконструкции окружающей среды следующим шагом будет определение местоположения и визуализация целей NLOS. Целевые местоположения могут быть установлены с хорошей точностью, если известна предварительная информация о месте действия. Даже если предполагаемые границы объектов построены с некоторыми ошибками, полученные местоположения все равно будут скорректированы по истинным траекториям лучей. На рис. 3.4 показано, как локализацию цели можно использовать в сценарии NLOS с реконструкцией окружающей среды. Для точной локализации NLOS результаты реконструкции окружающей среды (расстояния между стенками блоков А и Б) имеют погрешность 5 %. Такая ошибка в предполагаемых расстояниях, хоть и небольшая, повлияет на некоторые целевые местоположения, но они все равно будут центрированы по истинным

траекториям. В этом случае во внутреннем коридоре, где расстояние между стенами предполагается равным 2 м, ошибка 5 % соответствует точности картографирования окружающей среды 10 см [3].



Рис. 3.4. Локализация в сценарии NLOS

3.2.3. Визуализация и картование на открытом воздухе

Датчики на беспилотных транспортных средствах обычно имеют ограниченный обзор и охват из-за погодных условий, препятствий и естественного лимита возможностей сенсоров. При этом близлежащие стационарные базовые станции могут иметь большее поле зрения, большее расстояние обнаружения и более высокое разрешение. Таким образом, транспортные средства могут достичь качественно нового уровня автономности, используя для планирования движения карты, реконструированные базовыми станциями. Кроме того, разрешение и точность сканирования значительно улучшаются благодаря объединению результатов визуализации, которые передаются в глобальном масштабе через сеть, с облачными сервисами. Для достижения высокого разрешения при автономном вождении от современных автомобильных лидаров ожидают разрешающую способность на уровне 3 см или меньше [4]. Плотнo распределенные базовые станции в городской зоне и ISAC делают возможной реконструкцию окружающей среды и трехмерную локализацию, которые, в свою очередь, образуют *виртуальный город*.

Поскольку расстояние между зданиями в подавляющем большинстве случаев больше одного метра, разрешение и точность на уровне метров позволят создавать визуализации и карты на уровне зданий, в то время как диапазон чувствительности в сотни метров подойдет для реконструкции городских территорий. Стоит отметить, что дорожная карта предоставляет исчерпывающую информацию о дорожном движении в режиме реального времени. Визуализацию на основе 6G можно

использовать для точного воссоздания картины дорожного движения, а также окружающей среды. Массив данных, извлеченных из снимка дороги, содержит миллионы отдельных точек отражения, которые соответствуют аварийным ограждениям, дорожным знакам и дорожной разметке. Реконструированную карту можно использовать для интеллектуального управления трафиком, такого как мониторинг транспортного потока, обнаружение пробок и аварий.

3.3. Расширенное человеческое восприятие

Пример 3.3 Рассмотрим медицинский центр будущего с поддержкой 6G, оснащенный множеством усовершенствованных устройств и оборудования для расширения сенсорных способностей человека. Доктор Кляйне во время операции надевает сенсорные очки со сверхвысоким разрешением. Это оборудование помогает доктору Кляйне видеть распределение кровеносных сосудов и лимфатической системы пациента с миллиметровым разрешением, тем самым снижая вероятность необратимого повреждения во время операции. Кроме того, эти очки могут сканировать толщу кожи пациента и определять его сердцебиение и кровоток в режиме реального времени, помогая доктору Кляйне отслеживать состояние организма и быстро принимать решения во время операции с помощью искусственного интеллекта. Во время операции доктор Кляйне извлекает ткани пациента для диагностических целей и отправляет патологический образец в центр медицинского обследования, где эксперт Томми исследует его с помощью сенсорных перчаток. Гибкие графеновые перчатки под названием «T-RAY» позволяют Томми исследовать спектрограмму молекулярных колебаний в реальном времени и быстро определять состояние образца, просто слегка прикоснувшись к нему. Схожим образом в этой лаборатории проводят фармакологические испытания, осторожно нанося порошок лекарственного препарата на сенсорные перчатки.

Расширение человеческого восприятия направлено на обеспечение безопасных и высокоточных функций сканирования и визуализации, которые превосходят возможности человека, в то же время обладая малым или сверхмалым энергопотреблением. Опираясь на ISAC в качестве базовой технологии, мы можем использовать для сканирования окружающей обстановки мобильное клиентское оборудование, а не громоздкие специализированные устройства. Сканирующие устройства могут быть мобильными телефонами с поддержкой 6G, носимыми устройствами или медицинским оборудованием, имплантированным под кожу человека. Научно-технический прогресс открывает перед нами перспективы расширения возможностей человеческого восприятия, интегрируя поток информации об окружающей обстановке непосредственно в сеть 6G.

Помимо больниц, другие возможные применения расширенного сканирования могут включать обнаружение объектов внутри упаков-

вок, водопроводных труб за стенами, дефектов продуктов, утечек в канализации и многое другое. Более того, чтобы получить информацию о состоянии кровеносных сосудов, состоянии органов и других жизненно важных функциях, мы сможем использовать носимые устройства, размещенные над или под кожей и работающие в наиболее высокочастотных диапазонах (миллиметровые волны, ТГц- или оптические диапазоны). Этот тип технологий позволит получать критически важную информацию более точно и эффективно, чем это возможно только благодаря человеческим способностям. Кроме того, точная информация о местоположении мобильных устройств и связанная с ними информация о собранных данных поможет своевременно предоставить пациентам медицинскую помощь во время чрезвычайных ситуаций.

3.3.1. За пределами возможностей глаза – сверхвысокое разрешение

В будущем методы сканирования будут поддерживать визуализацию и обнаружение с высоким разрешением, а с сетью 6G они откроют двери для множества применений, таких как дистанционная хирургия, диагностика рака, обнаружение повреждений упаковки и т. д. Как показано на рис. 3.5, хирург может проводить операцию удаленно, с помощью системы визуализации сверхвысокого разрешения и платформы телеуправления. Кроме того, интеллектуальные предприятия будут использовать эти превосходные сенсорные решения для реализации бесконтактного сверхточного наблюдения, трекинга и контроля качества.

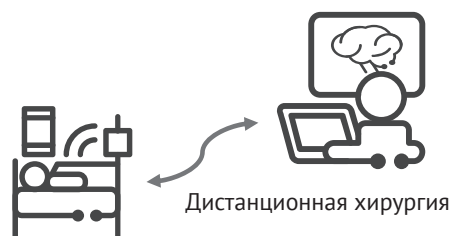


Рис. 3.5. Дистанционное хирургическое вмешательство с системой визуализации сверхвысокого разрешения

Для этого типа приложений требуются разрешение миллиметрового диапазона и сверхвысокое поперечное разрешение, которые нуждаются в более высокой полосе пропускания и увеличенной апертуре антенной решетки соответственно. Используя технологии связи 6G с частотой в ТГц-диапазоне и, соответственно, длиной волны менее 1 мм, эти расширения сенсорных возможностей человека можно реализовать в виде мобильных устройств.

3.3.2. За пределами возможностей глаза – сделать невидимое видимым

В то время как сценарии сверхвысокого разрешения требуют более широкой полосы пропускания и увеличенной апертуры антенны, следующее применение «расширенного зрения» предъявляет другие требования. Точнее говоря, для получения важной информации, которая может скрываться под кожей, за преградой или в темноте, необходимы частоты с относительно низкими потерями при проникновении. Другими словами, невидимое становится видимым.

Обычная технология оптической визуализации может обеспечить визуализацию LOS только на уровне, аналогичном возможностям человеческого глаза. Однако сканирование на основе 6G может реализовать потенциал визуализации NLOS, как показано на рис. 3.6. Будущая технология обнаружения скрытых объектов станет полагаться на портативные устройства с мощными возможностями визуализации и малой задержкой, в отличие от широко распространенных в настоящее время больших и громоздких машин. Иными словами, мы предполагаем, что для обнаружения замурованных в стены труб или для проверки содержимого коробок будут применяться мобильные телефоны. Мы можем эффективно реализовать эту функцию, используя характеристики проникновения электромагнитной волны. Проницаемость радиосигналов зависит от частоты и мощности передатчика. Неионизируемые материалы, такие как кожа (0,5–4 мм), подкожный жир (12–20 мм), чемоданы (0,5 см) и мебель (2 см), являются примерами того, где могут применяться проникающие радиоволны.



Рис. 3.6. Пример визуализации NLOS

Сканирование при помощи сетей 6G также может играть важную роль в медицинских технологиях, применяемых для диагностики, мониторинга и лечения. В настоящее время количество людей, живущих с хроническими заболеваниями, во всем мире стремительно растет, что создает нарастающую перегрузку системы здравоохранения. Больные некоторыми хроническими заболеваниями (астма, аритмия, гипогликемия, гипергликемия и т. д.) нуждаются в регулярном мони-

торинге состояния здоровья с помощью носимых устройств. Например, сердцебиение пациента, страдающего сердечным заболеванием, необходимо контролировать 24 часа в сутки, а уровень сахара в крови пациента, страдающего гипогликемией или гипергликемией, необходимо контролировать после каждого приема пищи. К счастью, технология сканирования 6G позволяет осуществлять атравматический мониторинг физиологических параметров организма. Например, мы можем не только считывать сердечные импульсы, но и получить полную картину состояния кровеносной системы организма. Очевидно, что эта технология помогает врачам мгновенно получать точную информацию о состоянии пациента. Технология атравматического мониторинга не причинит вреда или дискомфорта пациентам и отличается сверхвысокой надежностью и точностью, однако, например, мониторинг зоны *апикального импульса*¹¹ диаметром 1–2 см [5] потребует полосы пропускания от десятков до сотен ГГц. Благодаря этому типу расширения человеческого восприятия в будущем мы сможем корректировать, заменять и расширять возможности наших органов чувств с помощью новых интерфейсов «мозг–машина» и «мышцы–машина».

3.3.3. За пределами возможностей глаза – спектральное распознавание

Спектральное распознавание основано на идентификации целей посредством определения их электромагнитных или фотонных характеристик с помощью спектрограммы. Сюда входит анализ параметров поглощения, отражательной способности и диэлектрической проницаемости, помогающий различать качество материалов. В качестве примеров перспективных применений этой технологии можно назвать измерение загрязненности, подсчет калорий и управление качеством продукции.

Эти применения основаны на эффекте вибрации молекул, который связан с уникальной кривой поглощения материала. Коэффициент отражения материала зависит от частоты излучения. Например, электромагнитное излучение в частотном диапазоне ТГц и субТГц сильно поглощается жидкой водой, и кривая поглощения/отражения заметно меняется в зависимости от содержания воды в исследуемом объекте. Это свойство можно использовать для неразрушающего контроля промышленных продуктов и точного определения содержания воды в древесине, бумаге и т. д. Распределение и соотношение каждого компонента в смеси можно определить путем проведения теста на поглощение или отражение. Для считывания всех пиков поглощения радиоволн ко-

¹¹ Апикальный импульс (Ictus cordis) – это толчок, ощущаемый в точке максимального импульса на грудной клетке, где можно почувствовать биение сердца. – Прим. перев.

леблющимися молекулами требуется полоса пропускания от 2 ГГц до 8 ГГц. Кроме того, для получения читаемых спектрограмм требуется джиттер частоты менее ± 10 ГГц на частоте 1 ГГц [6].

Распознавание спектрограммы может пригодиться в приложениях, связанных с пищевыми продуктами. Состав пищи (тип и ингредиенты) можно определить по отражению сигналов ГГц-диапазона. Это поможет нам определить различные типы продуктов питания, калорийность, наличие загрязненных ингредиентов и т. д., как показано на рис. 3.7.



Рис. 3.7. Распознавание спектрограммы для определения калорийности пищи

При сканировании окружающей среды, которое включает в себя оценку качества воздуха и обнаружение загрязнений, спектр отклика загрязняющих частиц $PM_{2,5}^{12}$ имеет две отдельные полосы поглощения в диапазоне между 2,5 и 7,5 ТГц. Корреляции между полосами поглощения и кросс-пиками на синхронных и асинхронных графиках показывают, что поглощение оксидов металлов находится в диапазоне 2,5–7,5 ТГц. Эти результаты [7] подтверждают, что терагерцовый спектральный анализ загрязнений $PM_{2,5}$ является многообещающим инструментом, который можно использовать для понимания состава и массы загрязняющих веществ, а диапазон 0,1–7 ТГц является предпочтительным для получения полос поглощения.

3.4. Распознавание жестов и действий

По мере развития технологий возможности вычислений вышли за рамки привычных всем нам настольных компьютеров. Сообщество разработчиков интерфейсов человек–компьютер провело исследования множества альтернативных стратегий ввода. В последние годы наблюдается огромный интерес ко взаимодействию человека и компьютера посредством считывания движений всего тела или жестов рук. Распознавание жестов и действий без использования специальных устройств является ключом к реализации интерфейсов человек–компьютер, которые позволяют пользователям передавать команды и удобно взаимодействовать с устройствами с помощью жестов и действий. Мы можем

¹² Мелкодисперсные частицы и капли жидкости размером от 10 нм до 2,5 мкм. Частицы $PM_{2,5}$ легко проникают сквозь биологические барьеры и поэтому представляют наибольшую угрозу для организма. – Прим. перев.

легко интегрировать эту технологию в нашу повседневную жизнь, и как таковая она значительно изменит наш образ жизни. Можно с уверенностью сказать, что распознавание жестов и действий в радиочастотном спектре может применяться в качестве средства ввода для интерактивных приложений и интеллектуальных систем. Иными словами, эта технология открывает новые возможности использования радиочастотных сигналов в качестве среды для интерпретации жестов и действий без необходимости применения носимых устройств.

Извлечение признаков (feature extraction) – это главный этап процесса распознавания жестов и действий; признаки извлекаются из временной, частотной и пространственной областей или соответствующих пересечений областей. В 6G использование более высоких частот обеспечит более высокое разрешение и точность, а также сможет лучше фиксировать действия и жесты. Измерение доплеровских сдвигов, вызванных движением, лучше работает в высокочастотном диапазоне. Кроме того, массивные антенные решетки в 6G позволяют выполнять распознавание со значительно улучшенным пространственным разрешением и точностью. В отличие от распознавания на основе видео, сигналы 6G не зависят от условий освещения и, следовательно, могут обеспечить более надежное распознавание жестов и движений. Кроме того, благодаря характеристикам распространения сигналов 6G распознавание движений можно выполнять и вне зоны прямого распространения радиоволн. Еще одно важное преимущество распознавания жестов и движений в 6G заключается в том, что оно не связано с какими-либо рисками конфиденциальности (как видеонаблюдение), что может быть неприемлемым в определенных случаях (например, дома). В будущей системе распознавания жестов и действий, которая полностью использует пользовательское оборудование (например, смартфоны) и базовые станции сотовой связи, устройства будут совместно применяться для построения полной картины окружающей среды, и это станет возможным благодаря плотно распределенной сети 6G. Кроме того, обилие и доступность пользовательских устройств и базовых станций в сети означает, что дальность обнаружения и распознавания будет значительно расширена. Объединение собранных по всему миру данных сканирования, совместно используемых через сеть с облачными службами, значительно улучшит общую точность распознавания.

3.4.1. Бесконтактное управление на макроуровне

Пример 3.4 Будучи соучредителем стартапа по производству летающих автомобилей, Джек очень много работает ради своего бизнеса. Однако недавно он заболел из-за систематического переутомления и попал в замечательную современную больницу, где ему назначили операцию. На следующий день Джека прооперировали в умной опе-

рационной, и хирург использовал жесты, чтобы управлять медицинским оборудованием, не касаясь его. После операции Джеку пришлось пройти медицинскую реабилитацию, чтобы ускорить выздоровление. Распознавание движений гарантировало, что его реабилитация будет проходить без особых усилий, а распорядок дня будет выполняться правильно. В течение всего периода госпитализации за состоянием здоровья Джека постоянно следили, наблюдая, не падает ли он, не совершает ли какие-либо необычные движения. Все это было важно для его скорейшего выздоровления.

В обозримом будущем в умной больнице появятся расширенные функции распознавания жестов и движений. Система медицинской реабилитации в будущей умной больнице позволит организовать автоматическое наблюдение за пациентами. Например, система сможет следить за тем, что жесты и движения пациентов во время физиотерапии соответствуют стандартным требованиям для реабилитационных упражнений, и выдавать оперативные предупреждения о неправильных движениях или жестах, что значительно улучшит реабилитацию. Для распознавания жестов и движений во время физиотерапии требуется разрешение по скорости 0,01 м/с или лучше, поскольку движения выздоравливающих пациентов медленные и плавные. Кроме того, если пациент упадет во время тренировки или если обнаружится, что подозрительный человек вторгся в запретную зону, в центр управления больницы поступит сигнал тревоги. Чтобы реализовать функцию обнаружения падения или вторжения, когда смещение тела намного больше, чем жесты, требуется разрешение по дальности 2 см, разрешение по скорости 0,05 м/с и вероятность обнаружения более 99,9 %.

В умной больнице инфекционные заболевания в основном передаются через чихание и кашель. Поэтому крайне важно иметь возможность обнаруживать микрокапли, выделяемые при чихании и кашле. Соответствующая функция обнаружения пригодится для точного определения местоположения и отслеживания перемещений чихающих и кашляющих пациентов. Слизь и слюна могут вырываться изо рта человека со скоростью почти 40 м/с и разноситься на расстояние до восьми метров [8]. После того как человек выдохнет, микрокапли могут оставаться взвешенным в воздухе в течение нескольких минут, в зависимости от их размера. Чтобы точно обнаруживать и локализовать чихающего или кашляющего человека, требуются разрешение по дальности 5 см, разрешение по скорости 0,1 м/с и вероятность обнаружения более 99 %. Более того, система обнаружения и отслеживания чихания и кашля может определять размер аэрозольного облака, а также его траекторию (которая показывает, насколько далеко могут распространяться облака). Получив больше информации о чихании и кашле человека, исследователи надеются лучше контролировать и сдерживать распространение болезней.

3.4.2. Бесконтактное управление на микроуровне

Пример 3.5 После долгого дня на заводе по производству летающих автомобилей Джек возвращается домой. Войдя в гостиную, он просто делает жесты руками, чтобы включить свет и кондиционер, не касаясь выключателей. Затем он использует жесты рук для управления смарт-телевизором и взаимодействия с ним. После ужина он репетирует выученную вчера мелодию на виртуальном пианино.

Умный дом будущего будет оснащен продвинутой системой фиксации и распознавания жестов рук. Система позволит нам отслеживать трехмерное положение, вращение и жест руки. В результате, просто махнув рукой, мы можем включать или выключать свет в любом месте дома, не касаясь выключателя. Помимо освещения, с помощью жестов рук с использованием сигналов 6G мы также можем управлять, например, смарт-телевизором. Для основного взаимодействия со смарт-телевизором используется человеческое лицо и естественные жесты рук. Иными словами, система распознавания лиц применяется для идентификации пользователя, а система распознавания жестов рук управляет работой телевизора (каналы и громкость). Поскольку работа системы основана на сигналах 6G, пользователю не нужно находиться перед телевизором. Чтобы точно уловить жесты человека, требуется разрешение по дальности 1 см, разрешение по скорости 0,05 м/с, вероятность детектирования более 99 % и диапазон охвата до восьми метров (размер большой гостиной). Кроме того, для разделения жестов нескольких людей перед телевизором идеально подходит разрешение по диапазону не более 5 см.

Чтобы сделать еще один шаг вперед, можно было бы реализовать более сложные функции с помощью усовершенствованной системы захвата и распознавания жестов рук. Например, мы могли бы играть на виртуальном пианино в воздухе с помощью функции распознавания в сети 6G, обеспечивая полное погружение в любом месте и в любое время. Кинематика пальцев очень сложна, и есть исследования, показывающие, что средняя ширина указательного пальца составляет примерно 1,6–2 см для большинства взрослых [9]. Скорость пальцев во время выступления пианиста в медленном темпе имеет максимальное значение ± 2 м/с [10]. Следовательно, чтобы реализовать концепцию виртуального пианино, разрешение по диапазону и поперечному диапазону должно быть ниже 0,5 см с разрешением по скорости 0,01 м/с. Кроме того, вероятность распознавания выше 99 % гарантирует, что во время выступления виртуального пианиста не будет прерываний. Без сомнения, эта футуристическая концепция откроет широкие возможности для многих других инновационных приложений, связанных с высокоточным обнаружением и отслеживанием движений пальцев.

3.5. Источники

- [1] 3GPP, *Study on positioning use cases*, 3rd Generation Partnership Project (3GPP), Technical Report (TR) 22.872, 09 2018, version 16.1.0. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3280>.
- [2] S. Kumar, A. Majumder, S. Dutta, R. Raja, S. Jotawar, A. Kumar, M. Soni, V. Raju, O. Kundu, E. H. L. Behera et al., *Design and development of an automated robotic pick & stow system for an e-commerce warehouse*, arXiv preprint arXiv:1703.02340, 2017.
- [3] T. Johansson, Å. Andersson, M. Gustafsson, S. Nilsson, *Positioning of moving non-line-of-sight targets behind a corner*, in Proc. 2016 European Radar Conference (EuRAD). IEEE, 2016, pp. 181–184.
- [4] J. AmTechs Corporation, Tokyo, *VLP 16 puck*, 2020, <https://www.amtechs.co.jp/product/VLP-16-Puck.pdf>.
- [5] A. Chandrasekhar, *To evaluate apical impulse*. <http://www.meddean.luc.edu/lumen/meded/medicine/pulmonar/pd/pstep36.htm>.
- [6] J. Advantest Inc., Tokyo, *Terahertz spectroscopic system TAS7400 product specification*, 2019, https://www.advantest.com/documents/11348/146157/spec_TAS7400_EN.pdf.
- [7] H. Zhan, Q. Li, K. Zhao, L. Zhang, Z. Zhang, C. Zhang, and L. Xiao, *Evaluating PM2. 5 at a construction site using terahertz radiation*, IEEE Transactions on Terahertz Science and Technology, vol. 5, no. 6, pp. 1028–1034, 2015.
- [8] *See how a sneeze can launch germs much farther than 6 feet*, 2020, <https://www.nationalgeographic.com/science/2020/04/coronavirus-covid-sneeze-fluid-dynamics-in-photos/>.
- [9] *Finger-friendly design: Ideal mobile touchscreen target sizes*, 2020, <https://www.smashingmagazine.com/2012/02/finger-friendly-design-ideal-mobile-touchscreen-target-sizes/>.
- [10] S. D. Bella and C. Palmer, *Rate effects on timing, key velocity, and finger kinematics in piano performance*, PloS One, vol. 6, no. 6, p. e20518, 2011.

Глава 4

Полнофункциональная индустрия 4.0 и выше

Индустрия 4.0 относится к четвертой промышленной революции, которая автоматизирует традиционную промышленность с использованием современных интеллектуальных технологий [1]. По сравнению с индустрией 3.0, как показано на рис. 4.1, индустрия 4.0 нацелена на промышленное производство со значительно возросшей гибкостью, универсальностью, удобством использования и эффективностью на основе передовых киберфизических систем. Исходя из этого, мы можем предложить видение расширенной *индустрии 4.0+*, интеллектуального производства, которое станет следствием быстрого развития технологий беспроводной связи и автоматизации, а также искусственного интеллекта и машинного обучения в реальном времени.

Как показано на рис. 4.1, каждому поколению обрабатывающей промышленности требуются десятилетия для развития и совершенствования. Это сильно отличается от индустрии беспроводной связи, где каждое поколение имеет более короткий жизненный цикл. Для индустрии 4.0 ключевое значение имеют новые технологии, такие как беспроводная связь и интернет вещей на основе новейших достижений информационно-коммуникационных технологий.

Несмотря на то что на сегодняшних предприятиях используются беспроводные технологии, такие как Wi-Fi, LTE, Bluetooth и Zigbee, они не стали основными методами организации связи для использования в автоматизации производства. В основном это связано с ограничениями технологических характеристик. Тем не менее 5G имеет большой потенциал из-за того, что в нем уделяется большое внимание связи межмашинного типа, сверхнизкой задержке и высоконадежной связи.

Области применения беспроводной технологии можно разделить на пять категорий: автоматизация производства, автоматизация процессов, человеко-машинный интерфейс и производственные ИТ, логистика и складирование, а также мониторинг и профилактическое обслужи-

вание [2]. Типичные варианты использования, которые попадают в эти категории: управление движением, мобильная панель дистанционного управления, мобильный робот, обширная беспроводная сенсорная сеть, удаленный доступ и обслуживание, дополненная реальность, управление процессами с обратной связью, мониторинг процессов и управление складскими запасами [2]. Один вариант использования может относиться к одному или нескольким областям применения.

Несмотря на то что некоторые из перечисленных вариантов использования могут быть реализованы в 5G, в наиболее сложных случаях мы вынуждены ожидать появления систем связи следующего поколения. Например, некоторые варианты могут иметь экстремальные требования или быть размещены в суровых условиях с точки зрения сверхвысокой производительности сети, сложных условий распространения радиосигнала с потенциально высокими помехами, устаревших установок, которые могут потребовать бесшовной интеграции в новое производство, особых соображений безопасности и защиты данных. Во второй части этого раздела мы рассмотрим некоторые наиболее показательные примеры использования, которые вдохновляют на разработку новой системы 6G, направленной на обеспечение лучшей поддержки *вертикальных отраслей*¹³. С другой стороны, новые функции 6G могут также простимулировать активные инновации в области индустрии 4.0, и не только.



Рис. 4.1. Дорожная карта промышленной революции

¹³ Вертикальные отрасли (vertical industries) – это отрасли промышленности, работающие для узкого рынка B2B и продающие свою продукцию как часть цепочки поставок, т. е. вертикально, в отличие от горизонтального рынка, наполненного большим числом потребителей.

4.1. Фабрика будущего

Пример 4.1 Марк работает инженером на автомобильном заводе. В его обязанности входит надзор за производством автомобилей на заводе, построенном по технологии 6G. Вместо того чтобы находиться в цехе в шумной и иногда опасной обстановке, Марк теперь работает удаленно в уютной светлой диспетчерской, что намного безопаснее и экологичнее. Теперь он может сосредоточиться на контроле и управлении с помощью различных типов XR и оборудования для дистанционного управления со сверхбыстрым каналом, сверхнизкими задержками и сверхнадежными соединениями в сети 6G.

Однажды днем Марк получает новые заказы из интернета. Он дважды проверяет требования клиентов, а затем запускает ввод конфигураций в интеллектуальную систему управления на виртуальном планшете. Поскольку на заводе нет фиксированной производственной линии, интеллектуальная система сама находит лучший способ реорганизации производственной линии на основе всей собранной информации, а также своего собственного опыта и знаний. Таким образом, производственная линия автоматически реорганизуется для производства автомобилей с учетом требований заказчика.

Концепция сборочной линии по-прежнему используется на производственных предприятиях почти через 100 лет после изобретения Генри Фордом из-за присущей ей высокой эффективности. Однако такая эффективность лучше всего подходит для массового производства, что ставит под угрозу гибкость производства и возможность настройки. В результате обычная сборочная линия не может отвечать требованиям индивидуализации будущего производства [3].

Напротив, фабрика будущего стремится к полной революции, чтобы в конечном итоге реализовать производство «без освещения» с полной автоматизацией и гибкостью. Другими словами, это производство без людей на рабочем месте, что означает, что цеха не нуждаются в освещении. Фабрика будущего не только по-другому выглядит и работает, но и по-другому внедряет инновации.

- **Беспроводная связь как первый шаг к гибкости:** главный шаг, необходимый для создания гибкого заказного производства, заключается в освобождении машин от соединительных кабелей. Другими словами, это будут индивидуальные и гибкие производственные модули, которые мгновенно перемещаются для формирования сборочной линии и взаимодействуют с другими роботами, AGV, а также дронами посредством обмена информацией по сверхвысокопроизводительным радиоканалам. Таким образом, концепция сборочной линии может действительно трансформироваться из обычного «массового производства» в ориентированное на будущее «производство по индивидуальному заказу».

- **Быстрое распространение интеллекта среди роботов:** индивидуальная информация обо всех объектах на заводе будет собираться и выражаться в виде физических представлений в цифровом виртуальном мире, которые будут использоваться для проектирования, моделирования и оптимизации производственного процесса. Более того, ИИ получит широкое распространение в обрабатывающей промышленности. Например, машины и роботы смогут накапливать производственный опыт и знания, которые затем могут быть переданы другим машинам и роботам на той же производственной линии, на других производственных линиях или на разных заводах по всему миру с помощью передовых технологий мобильной связи. Таким образом, одна машина или робот может использовать интеллект всех других машин или роботов, что, в свою очередь, приводит к оптимизированному и быстро развивающемуся производственному процессу. Кроме того, в этих сценариях для совместной работы над задачей требуется несколько роботов, и поэтому их коммутации не будут основываться на простом подключении. Это типичный сценарий «подключенного интеллекта».
- **Обнаружение в реальном времени для профилактического обслуживания:** вся производственная среда будет контролироваться с помощью множества датчиков со сверхплотным подключением и интеллектуальной радиочастотной зондирующей системы, как описано в главе 3. Все это нацелено на обеспечение профилактического обслуживания всей производственной среды и технологического процесса, гарантируя как безопасность, так и эффективность производства. Даже незначительный производственный брак может быть обнаружен в режиме реального времени и мгновенно исправлен.
- **Низкий углеродный след и экологически чистый завод:** поскольку в большинстве цехов завода нет людей, нет необходимости держать свет включенным или обслуживать другие помещения для сотрудников. Например, весь мониторинг может выполняться через беспроводные сенсорные сети или функции RF-зондирования, интегрированные в систему связи 6G; следовательно, нет необходимости держать включенным свет и для камер наблюдения. В этом случае можно добиться гораздо меньших операционных затрат и выбросов углекислого газа.

Завод будущего предлагает широкий спектр сценариев использования и потребует задействовать как ОТ-, так и ИТ-технологии. В следующих разделах мы подробнее рассмотрим ряд сложных случаев использования.

4.2. Управление движением

Пример 4.2 На фабрике, где работает Марк, все машины и роботы подключены к беспроводной сети. Как и строительные блоки Lego, их можно легко перемещать и соединять с другими машинами и роботами в соответствии с требованиями к продукции.

Система управления движением – это мозг машины и жизненно важный компонент процесса автоматизации. Она широко используется во многих отраслях промышленности, включая производство, автомобилестроение и медицину. Она отвечает за управление, перемещение и/или вращение деталей машины четко определенным образом (например, в печатных машинах, станках или упаковочных машинах). В процессе управления с обратной связью контроллер движения отправляет команды одному или нескольким исполнительным механизмам строго циклически и детерминированно. После получения заданных значений от контроллера исполнительные механизмы выполняют соответствующие действия над одним или несколькими объектами. Например, они могут вращать руку робота или перемещать определенный компонент на машине. В то же время датчики определяют текущее состояние процессов (например, текущее положение и/или угол поворота одного или нескольких компонентов) и отправляют фактические значения обратно в контроллер движения [2].

Управление движением – один из самых сложных вариантов использования в области автоматизации; он требует сверхвысокой надежности и детерминированной связи с малой задержкой. Например, задержка канала E2E (включая радиоинтерфейс, ядро и транспортные сети), допустимая для приложений управления движением, таких как станки или упаковочные машины, может быть даже на уровне микросекунд при уровне надежности выше 99,9999%. Этот тип механизмов используется в современном производстве на протяжении десятилетий. Сегодня в системах управления движением широко используются проводные технологии, такие как промышленный Ethernet [2, 4].

Чтобы сместить парадигму от фиксированной к гибкой производственной линии, которая может быть собрана как строительные кубики Lego, первым и основополагающим шагом является замена проводной связи беспроводными каналами. Тем не менее, когда речь идет о той же пропускной способности, что и у промышленного Ethernet, это проще сказать, чем сделать. Несмотря на то что варианты использования 5G для управления движением уже обсуждаются, большинство из них могут быть реализованы только в 6G из-за ограничений пропускной способности. Таким образом, сверхнизкая задержка и сверхвысокая надежность связи с детерминированными функциями определенно станут проблемой при проектировании всей системы мобильной связи.

4.3. Совместная деятельность роботов в группе

Пример 4.3 На автомобильном заводе основную работу выполняют роботы, а не люди. В процессе производства сырье, детали и аксессуары для сборки автомобиля транспортируются со склада на производственную линию с помощью роботов, грузовых автомобилей или дронов. Крупные детали обычно доставляются совместно несколькими роботами. Локализация и синхронизация в реальном времени с помощью сети 6G помогают поддерживать уровень точности, необходимый для сложной совместной работы, например совместной доставки.

Сборочные работы выполняются группой совместно действующих роботов без какого-либо вмешательства человека. Локальное и совместное машинное обучение – это ключевой метод, который позволяет роботам лучше справляться с различными ситуациями. Если один робот сталкивается с неизвестной ситуацией и решает ее, соответствующие знания и опыт можно быстро передать другим роботам на том же заводе или между заводами через сеть 6G.

Благодаря развитию робототехники промышленные роботы нашли свой путь в индустриальный мир. Они могут выполнять сварку, покраску и пайку, а также множество других задач. Таким образом, они будут продолжать играть жизненно важную роль в промышленности будущего. Одним из вариантов использования является совместная транспортировка [2], когда большие или тяжелые детали доставляются в нужное место несколькими роботами. Безопасное и эффективное сотрудничество роботов достигается с помощью приложения киберфизического контроля, которое координирует их движения. Связь между роботами требует сверхвысокой надежности и доступности услуг, а также строгой синхронизации. Также стоит отметить, что команды управления передаются, а данные с датчиков принимаются посредством периодической детерминированной связи.

Коллективная рабочая сила может использоваться как для переноски жестких или хрупких деталей, которые требуют очень точной координации между роботами, так и более гибких или эластичных деталей, которые обеспечивают некоторую свободу в скоординированных движениях. В обоих случаях сети, которые работают лучше с точки зрения синхронизации, задержки и точности локализации, могут помочь повысить эффективность совместной работы. Как правило, задержка E2E, необходимая для коллективного функционирования роботов, может составлять приблизительно 1 мс, а надежность превышает 99,9999 %.

На фабрике будущего роботы будут работать в полной гармонии друг с другом и машинами. Это сотрудничество представляет из себя большее, чем просто заранее определенные программы. Другими словами, оно основано на роботах, воспринимающих окружающую среду, и интеллектуальной обработке полученной информации и знаний на основе ИИ.

4.4. От интеллектуальных коботов к киборгам

Пример 4.4 Внезапно срабатывает предупредительный сигнал, и Марк отправляется в производственную зону. Возможное местонахождение проблемы отображается его умными очками, которые мгновенно направляют его к нужному месту. Информация о потенциальной причине проблемы отображается на его планшете. Интеллектуальная система управления помогает Марку анализировать ситуацию и находить оптимальное решение по техническому обслуживанию.

Как только решение найдено, робот-помощник приносит все необходимые инструменты и останавливает производственную линию. Он легко взаимодействует с Марком и предлагает советы, если у него иное мнение по поводу определенных действий.

Коллега Марка, Бернд, ослеп из-за врожденного заболевания. Ему имплантировали устройство, помогающее восстановить зрение. Благодаря современным технологиям его инвалидность не помешала ему осуществить мечту стать автомобильным инженером.

В промышленности появилось составное слово *кобот* (cobot) – сочетание слов *collaboration* (сотрудничество) и *robot*. Это концепция, используемая для продвинутой индустрии 4.0, и не только. В отличие от упомянутого выше варианта использования, когда роботы работают друг с другом, кобот относится к роботам, которые могут тесно и интерактивно взаимодействовать с людьми как партнеры. Разница с обычными роботами в том, что коботы не отделены от людей традиционными защитными средствами, такими как экраны из плексигласа или ограждения. Коботы выполняют работу, которая для людей является слишком сложной или опасной.

Интеллект, точность, устойчивость и надежность являются ключами к обеспечению эффективного взаимодействия человека с машиной в повседневной жизни или при выполнении физически сложных работ в опасной среде. В отличие от обычных промышленных роботов, коботы не ограничиваются фиксированными задачами или областями. Их можно обучать посредством наблюдений за действиями и логическими выводами наставников-людей или электронных коллег. Для этого базовая инфраструктура, необходимая для поддержки мобильности и возможности подключения коботов, должна иметь пропускную способность промышленного уровня.

Оснащенные сложной системой сенсоров и быстродействующими средствами связи, коботы могут реагировать на людей и взаимодействовать с ними. Эта технология открывает широкие перспективы в промышленности, особенно с учетом того факта, что текущая глобальная тенденция заключается в старении населения. Пожилые сотрудники могут работать дольше, если коботы будут поддерживать их в физически сложных делах. Между тем коботы могут легко расширить свою профессиональную область применения от вертикальных отраслей

промышленности до ориентированных на клиента услуг в розничной торговле или общественном питании. Эти коботы могут расширить наши возможности как в профессиональной сфере, так и в сфере досуга с помощью механических, но интеллектуальных методов.

Если мы рассматриваем коботов как воплощение интеллектуальной кибернетики, представляющей слияние искусственного интеллекта, ИСТ и ОТ на высшем уровне, то следующим шагом будут *киборги*. Концепция киборга была сформулирована в 1960 году и первоначально определялась следующим образом: «киборг намеренно включает в свой состав экзогенные компоненты, расширяющие функцию саморегулирования организма, чтобы приспособить его к новым условиям» [5]. Иными словами, киборги – это кибернетические организмы, или, говоря иначе, это человеческие тела, усиленные машинами. Благодаря последним передовым разработкам в области медицины, механики и взаимодействия человека с машиной, таким как нейроинтерфейсы, в ближайшем будущем киборги перейдут из научной фантастики в реальность. Это принесет невероятную пользу, особенно людям с ограниченными физическими возможностями. Например, эти люди смогут видеть и перемещать предметы и, как таковые, будут на равных со всеми в рабочей среде. Даже люди без физических проблем смогут защитить себя в опасной профессиональной среде, повысив при этом свои знания и осведомленность об окружающей среде. Подводя итог, можно сказать, что нейробиология и робототехника будут развиваться в тандеме с технологиями мобильной связи, сами люди могут служить новыми мобильными устройствами, а 6G станет ключом к взаимодействию.

4.5. Источники

- [1] Wikipedia, «Fourth industrial revolution», 2020, accessed Sept. 21, 2020. [Online]. Available: https://en.wikipedia.org/wiki/Fourth_Industrial_Revolution.
- [2] 3GPP, «Service requirements for cyber-physical control applications in vertical domains», 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 22.104, 12 2019, version 17.2.0. [Online]. Available: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3528>.
- [3] Y. Koren, The global manufacturing revolution: product–process–business integration and reconfigurable systems. John Wiley & Sons, 2010.
- [4] 5GACIA 5G alliance for connected industries and automation, «White paper, 5G for connected industries & automation (Second Edition)», 2018, accessed Sept. 20, 2020. [Online]. Available: <https://www.5g-acia.org>.
- [5] M. E. Clynes, «Cyborgs and space», *Astronautics*, vol. 14, pp. 74–75, 1960.

Глава 5

Умный город и умная жизнь

Понятия «умный город» и «умная жизнь» охватывают очень широкий круг вариантов использования сетей 6G, которые преобразуют все аспекты городской жизни. Со стратегической точки зрения, эти варианты направлены на обеспечение эффективного муниципального управления, высококачественных государственных услуг и устойчивого экономического развития. Эта трансформация началась в эпоху 4G с поддержки IoT в мобильной связи и эволюционировала до 5G как mMTC. Следующий этап предполагает переход к 6G, что подразумевает более строгие требования к системам мобильной связи.

5.1. Умный транспорт

Пример 5.1 Начинаются летние каникулы, и Марк, его жена Дженнифер и трое детей отправляются в поездку. Им потребуется 12 часов, чтобы добраться до места назначения, и поэтому поездка на дальние расстояния была бы утомительной и опасной, но не в этот раз. Их новый автомобиль оснащен системой беспилотного управления 5-го уровня, что делает путешествие увлекательным, расслабляющим и приятным. Интеллектуальный автомобиль полностью берет на себя управление, и теперь Марк и Дженнифер могут наслаждаться красивыми пейзажами, а также играть в игры со своими детьми.

Несмотря на то что во время летних каникул толпы людей путешествуют на своих машинах, Марку и Дженнифер не приходится застревать в дорожных заторах. Это связано с тем, что «летающие автомобили» превращают дорогу из плоской в трехмерную, как показано на рис. 5.1, значительно улучшая ее пропускную способность. Дополнительным преимуществом является то, что им не нужно учиться управлять автомобилем. Благодаря наличию системы беспилотного управления 5-го уровня все автомобили в системе трехмерного движения хорошо управляются и движутся к месту назначения по своим оптимальным маршрутам.

Использование мобильной связи для подключения транспортных средств (vehicle-to-everything, V2X) началось с 4G LTE и изменило всю транспортную экосистему. Подключение не только повышает комфортность вождения, но также обеспечивает большую безопасность и более

высокую эффективность использования транспортных ресурсов. Сотовая связь V2X имеет очень широкую сферу применения, которая включает в себя: безопасность, оптимизацию эксплуатации транспортных средств, беспилотное управление, повышение эффективности дорожного движения, экологичность, а также взаимодействие с обществом и сообществами [1]. Среди них наиболее сложным является беспилотное управление, которое и принесет нам больше всего инноваций.



Рис. 5.1. Летающие автомобили с беспилотным управлением 5-го уровня

Общество автомобильных инженеров (society of automotive engineers, SAE) установило несколько уровней автоматизации транспортных средств, где уровень 0 соответствует транспортным средствам с полностью ручным управлением, а уровень 5 соответствует полностью автоматизированным транспортным средствам [2]. Сканирование, подключенность и связь являются фундаментальными факторами, способствующими развитию беспилотных транспортных средств.

Первым по-настоящему рыночным вариантом беспилотного вождения может стать телеуправление. В качестве примеров применения телеуправления можно назвать такие области, как промышленная добыча полезных ископаемых, карьерные разработки, строительство и сельское хозяйство. Благодаря высокому уровню автоматизации всеми тяжелыми машинами (например, кранами и буровыми установками) в опасных зонах можно будет управлять удаленно, что приведет к появлению новых профессий. Наличие в повседневной жизни потребителей беспилотных транспортных средств или дронов 5-го уровня изменит наше понимание автомобильных путешествий. Беспилотное вождение

5-го уровня без пробок в час пик и необходимости заранее планировать маршрут повлияет на нашу жизнь так же, как переход от лошадиной тяги к паровому двигателю. Для беспилотного вождения 5-го уровня наиболее сложными факторами являются непредсказуемые природные явления на пути следования. В свете этого услуги сканирования и искусственного интеллекта, предоставляемые системами мобильной связи, помогут транспортным средствам получать точную информацию для принятия решений. Эффективность вождения (например, сочетание таких факторов, как скорость автомобиля и плотность движения) неразрывно связана с пропускной способностью систем связи. Само собой разумеется, что требования к уровню обслуживания требуют сверхмалой задержки (порядка одной миллисекунды или ниже) и высокой надежности (99,9999 % или выше) связи, а также точной локализации (например, для определения расстояния между транспортными средствами). Разумеется, безопасность и надежность также относятся к перечню критически важных параметров. Последствия взлома злоумышленниками системы управления беспилотным транспортом могут быть настолько ужасными, что их трудно вообразить.

Нужно учитывать некоторые практические вопросы и с точки зрения архитектуры. Одна из проблем связана с обновлением платформы граничных (периферийных) вычислений, в которой задействованы несколько операторов и поставщиков, чтобы устранить узкие места в быстродействии, особенно для высокоскоростных оконечных устройств. Помимо скоростных беспроводных соединений, необходим системный интеллект. Он не только включает адаптивные решения, основанные на продолжительности ситуаций, но также предсказывает опасные ситуации и предлагает соответствующие действия или решения.

Транспортные средства представляют собой чрезвычайно своеобразное пользовательское оборудование – мобильный интеллектуальный объект без заметных энергетических ограничений – и поэтому сильно отличаются от обычных мобильных устройств, таких как смартфоны. Они генерируют огромные объемы данных и собирают данные от ближайших коллег; следовательно, они могут действовать как подвижные грани фиксированной инфраструктуры, что порождает интересные, но сложные вопросы, касающиеся проектирования системы.

5.2. Умное здание

Пример 5.2 Марк и Дженнифер прибывают в свой отель, где уже зарезервирована парковка и завершена регистрация. Робот поджидает их, помогает нести багаж и проводит семью в комнату. В течение следующих нескольких дней робот также служит гидом семьи во время экскурсий по городу с учетом их предпочтений и увлечений.

Строительная отрасль представляет собой замкнутую экосистему с разнообразными вариантами использования. Как и другие отрасли бизнеса, она также находится в процессе цифровой трансформации. Технология умных зданий включает в себя не только установку в зданиях новейших электронных продуктов (например, лифтов и телевизоров), но также и управление зданиями как интеллектуальными объектами с непрерывным обменом информацией между электронными устройствами, интеллектуальными материалами и системами управления и безопасности зданий. Технология умных зданий предусматривает еще гибкое создание многоцелевых зданий со значительно повышенной энергоэффективностью и интеллектом.

Первая тенденция – это интеграция и автоматизация систем здания, когда одно здание может содержать до 10 различных подсистем, связанных между собой с использованием гетерогенных каналов. Эти подсистемы могут включать, например, систему камер наблюдения, систему управления лифтами, систему пожарной сигнализации, систему управления парковкой, систему кондиционирования воздуха, систему электроснабжения и отопления, а также систему управления автоматизацией здания. Содержание и эксплуатация таких зданий могут быть весьма затратными, но при этом неэффективными. Ключом к внедрению технологии мобильной связи в индустрию умного домостроения является использование ее в качестве общей инфраструктуры для объединения различных подсистем, которая имеет строгую изоляцию ресурсов и высокую собственную надежность. Индустрия умного домостроения также будет широко применять ИСТ – от мобильной связи до интернета вещей. Например, в современном коммерческом здании количество датчиков (для мониторинга окружающей среды, мониторинга инфраструктуры, управления персоналом и т. д.) может достигать уровня 10^5 и выше. Следовательно, очень важно, чтобы для индустрии умного домостроения предоставлялась доступная услуга массового подключения. Кроме того, к некоторым датчикам, используемым для мониторинга окружающей среды (например, детекторам дыма), обычно предъявляются очень высокие требования с точки зрения энергопотребления. Батарея обычно должна прослужить не менее двух лет, а в идеале – более 20 лет. Кроме того, здания также будут оснащены умными материалами и объектами, такими как роботы и коботы, которые нуждаются в общении и являются интерактивными.

Вторая тенденция – взаимодействие между зданиями в разных регионах. В будущем здания больше не будут изолированы; вместо этого они будут связаны между собой внутри одного региона или в разных регионах. Поэтому ожидается, что инфраструктура мобильной связи обеспечит цифровую основу для надежного взаимодействия между платформами, объектами и регионами.

5.3. Умное здравоохранение

Пример 5.3 Из-за того, что Лиза, мать Марка, живет одна и у нее несколько хронических заболеваний, Марк оснастил ее дом умным медицинским оборудованием, включая систему датчиков для наблюдения за ее состоянием здоровья, и терминалом для трехканального голографического телеприсутствия (изображение, звук, тактильные ощущения). Это гарантирует, что в случае возникновения чрезвычайной ситуации врачи получат жизненно важные данные, а Марк будет немедленно об этом проинформирован. Лиза чувствует, что рядом с ней ее сын Марк и личный врач, и она может немедленно получить медицинскую консультацию.

Для поддержания здоровья Лизе необходимо регулярно заниматься аэробикой и растяжкой. У нее есть костюм из умной тактильной ткани, который может отслеживать ее движения и обеспечивать как звуковую, так и тактильную обратную связь, словно голос и руки обычного тренера. Лиза может загружать в этот костюм соответствующие программы для различных упражнений, и, при необходимости, настоящий тренер, одетый в аналогичный костюм, может передавать тактильную информацию в реальном времени, чтобы помочь ей выучить новые упражнения.

Всеобъемлющее и индивидуализированное предоставление медицинских услуг без каких-либо географических ограничений – это перспектива развития электронного здравоохранения на следующее десятилетие. Реализация этого видения во многом зависит от развития систем мобильной связи, к которым будут предъявлять все более жесткие требования относительно надежности, доступности, безопасности и конфиденциальности.

В сочетании с искусственным интеллектом анализ данных, полученных от пациентов в режиме реального времени, может помочь в прогнозировании серьезных заболеваний. В этом отношении передовая инфраструктура с возможностями зондирования будет иметь большое значение для помощи в борьбе с крупными пандемиями, такими как COVID-19. Опираясь на передовые видео-, голографические и тактильные технологии, они выведут профессиональное взаимодействие и опыт работы с эффектом присутствия на новый уровень. (Соответствующие требования KPI для 6G были описаны в главе 1.) Можно смело предположить, что дистанционная диагностика, дистанционная хирургия и динамический мониторинг, а также голографическое медицинское обучение станут преобладающими технологиями в будущем, как показано на рис. 5.2. Такие сценарии использования, как дистанционная хирургия, могут потребовать сверхмалой задержки и сверхвысокой надежности связи на межконтинентальном расстоянии (например, для врача в Китае, выполняющего операцию на сердце пациенту в Африке). Это будет иметь важное значение для снижения остроты проблемы ста-

рения общества, особенно в регионах с ограниченными медицинскими ресурсами.



Рис. 5.2. Возможные варианты использования умного здравоохранения

5.4. Интеллектуальные сервисы на базе БПЛА

Пример 5.4 Лиза, мать Марка, не может поехать в семейный отпуск из-за несчастного случая, из-за которого она потеряла подвижность. Марк использует дрон, чтобы совершать XR-звонки Лизе, и управляет дроном, чтобы запечатлеть Альпы с неба. Это позволяет Лизе полностью погрузиться в атмосферу, и ей кажется, что она путешествовала с семьей.

Во время отпуска Марк покупает для Лизы сувенир в местном магазине. Это потрясающе красивое кольцо, которое очень понравилось Лизе, и она с нетерпением ждет, когда сможет надеть его. Чтобы как можно скорее доставить кольцо Лизе, Марк вызывает дрона-курьера, который немедленно взлетает, как посыльный голубь в древние времена, как показано на рис. 5.3. Когда на пути к Лизе заканчивается заряд батареи, дрон-курьер летит к ближайшей зарядной станции, что позволяет ему преодолевать очень большие расстояния.

БПЛА, также называемые дронами, представляют собой летательный аппарат без пилота-человека на борту. Ими можно управлять дистанционно с помощью пульта управления, и они имеют ряд частично автономных полетных функций, которые в будущем станут полностью автономными. БПЛА бывают самых разных размеров и грузоподъемности и могут использоваться в разнообразных сферах бизнеса. Например, как показано на рис. 5.4, у них могут быть коммерческие и промышленные применения, такие как автоматическая беспилотная инспекция, экстренное реагирование для умного и безопасного города, мониторинг окружающей среды и интеллектуальная логистика, как сказано в примере 5.4 и показано на рис. 5.3.



Рис. 5.3. Доставка посылок БПЛА

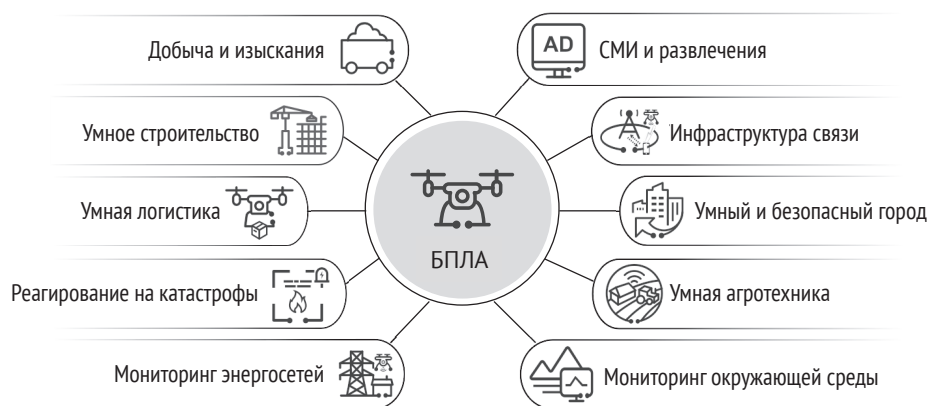


Рис. 5.4. Области применения и варианты использования БПЛА

Дроны также могут быть использованы для расширения границ вышеуказанных отраслей за счет новых вариантов использования. По данным Keystone, рынок технологий БПЛА достигнет 41–114 млрд долларов США во всем мире в течение следующих пяти лет [3].

Независимо от отрасли промышленности, главным свойством беспилотных летательных аппаратов является воздушная мобильность. Они служат удобной платформой для перемещения различных датчиков и камер высокого разрешения. По этой причине беспилотные летательные аппараты будут играть важную роль в повышении сенсорных возможностей 6G. БПЛА не обязательно должны ограничиваться ролью оконечного устройства; они могут применяться в качестве ретранслятора или точки доступа для формирования временной сети и расширения зоны действия мобильной связи, особенно в исключительных сценариях (например, при авариях в горах или стихийных бедствиях). Иными словами, гибкость и маневренность БПЛА могут сделать их хо-

рошей альтернативой в критических ситуациях при отсутствии фиксированной инфраструктуры.

БПЛА являются действующим вариантом использования 5G, и они будут развиваться в направлении 6G с точки зрения более высоких требований KPI, включая сверхвысокую скорость передачи данных (например, видеопоток с обзором 360° с пропускной способностью более 10 Гбит/с), сверхвысокую надежность и сверхнизкую задержку (для плавного управления или полностью автономного полета). БПЛА нуждаются в высокоточном позиционировании, и показательный пример заключается в том, что для прямой видеотрансляции 8K, лазерного картографирования/патрулирования и серийных фотоснимков может потребоваться определение местоположения с точностью до одного сантиметра. Помимо требований KPI, для БПЛА будет обязательной надежность промышленного уровня. Она востребована в сценариях, где, например, БПЛА становятся источником частных данных. В этом контексте разработки новых систем мобильной связи также будут учитывать возможность обеспечения суверенитета данных.

Однако ограниченная емкость аккумулятора не позволяет беспилотным летательным аппаратам работать в течение продолжительных периодов времени. Следовательно, «зеленая связь» определенно пригодится для организации энергосберегающей связи с БПЛА.

5.5. Источники

- [1] 5GAA, *Whitepaper, C-V2X use cases, methodology, examples and service level requirements*, 2019, https://5gaa.org/wpcontent/uploads/2019/07/5GAA_191906_WP_CV2X_UCs_v1-3-1.pdf.
- [2] SAE On-Road Automated Vehicle Standards Committee et al., *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles*, SAE International: Warrendale, PA, USA, 2018.
- [3] HUAWEI Technologies (UK) Co., Ltd., *White paper: Connected drones. A new perspective on the digital economy*, 2017, https://www-file.huawei.com/-/media/corporate/pdf/x-lab/connected_drones_a_new_perspective_on_the_digital_economy_en.pdf?la.

Глава 6

Глобальное покрытие сети мобильного доступа

От беспроводной сети следующего поколения мы ожидаем глобального покрытия поверхности Земли, чтобы люди и подключенные вещи могли получить доступ к интернету в любом месте и в любое время. В 6G наземные сети смогут объединяться с узлами неназемного доступа, такими как спутники, дроны, *высотные платформы* (high-altitude platform, HAP) и летающие автомобили. Это объединение сформирует глобальную интегрированную сеть связи, которая обеспечит бесшовное покрытие, охватывающее сушу, море и небо, как показано на рис. 6.1. Кроме того, эта интеграция позволит пользователям подключаться к сети 6G независимо от того, идут ли они пешком, едут на автомобиле или летят на борту самолета, и обеспечит бесперебойное обслуживание, даже если какая-то часть инфраструктуры неисправна или повреждена.

В 6G наземные и неназемные сети будут полностью интегрированы на системном уровне, обеспечивая конвергенцию услуг, радиоинтерфейсов, сетей и пользовательских устройств. Органически объединяя эти две среды доступа в одну конвергентную многоуровневую гетерогенную сеть, охватывающую весь земной шар, 6G предоставит пользователям единообразное обслуживание. Обеспечение глобальной доставки мобильных услуг будет важным аспектом развития сети 6G.

Интегрированная сеть 6G расширит спектр предоставляемых услуг. Например, за счет интеграции наземных и неназемных сетей 6G обеспечит широкополосную связь и услуги IoT в регионах и областях, где отсутствует покрытие наземной сети, включая удаленные фермы, и даже на борту кораблей и самолетов. Кроме того, 6G будет способствовать появлению новых применений, таких как высокоточное позиционирование объектов комбинированными средствами спутниковых и наземных систем и высокоточная съемка земной поверхности в реальном времени для сельского хозяйства.

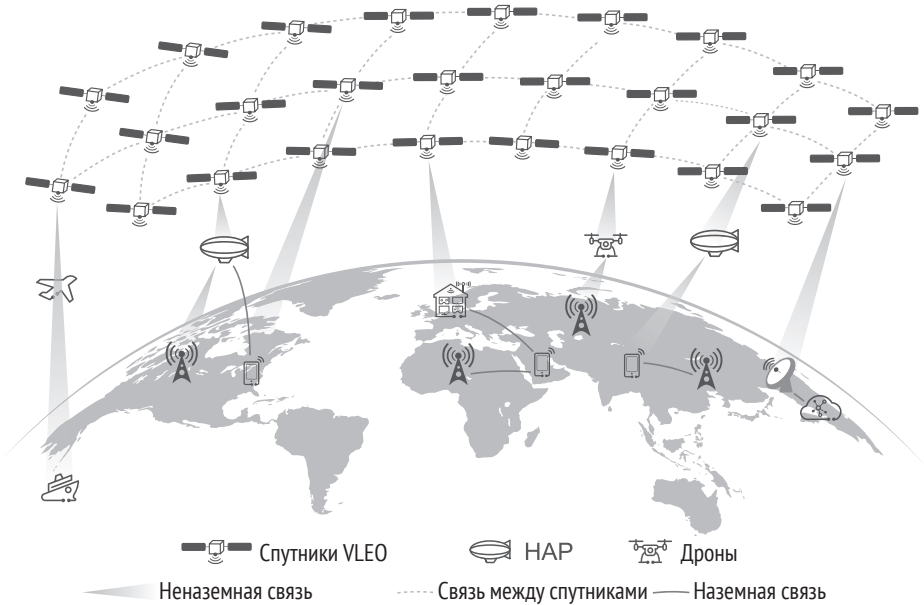


Рис. 6.1. Глобальное интегрированное покрытие 6G

6.1. Широкополосный беспроводной доступ для всех

Интегрированная сеть 6G обеспечит 3D-покрытие, устраняя все пробелы в покрытии по всему миру. Например, это позволит людям, проживающим в отдаленных и безлюдных районах, на борту транспортных средств, таких как корабли и самолеты, а также в море на нефтедобывающих платформах, подключаться к интернету через высокоскоростные беспроводные соединения. Люди смогут подключаться к сети 6G через наземные и неназемные каналы, используя один и тот же мобильный телефон или портативное устройство в любом месте и в любое время. Кроме того, интегрированная сеть 6G будет устойчивой к стихийным бедствиям, а это означает, что она может обеспечить подключение для служб быстрого реагирования и оказания помощи при бедствиях.

6.1.1. Мобильная широкополосная связь в любом месте

Пример 6.1 Карл владеет фермой далеко от центра города. Покрытие сотовой связи на ферме слабое, иногда случаются перебои в обслуживании. Оператор электросвязи сообщил Карлу, что теперь ему доступны услуги 6G с высококачественным широкополосным доступом независимо от его местонахождения. Карл заказывает услугу 6G и покупает новый мобильный телефон с поддержкой 6G, который объединяет наземный и неназемный доступ. В центре города мобильный телефон с поддержкой 6G может

подключаться к наземной сети, как и старый телефон. На ферме Карл теперь может прозрачно подключаться как к наземным, так и к неназемным сетям, стабильно получая широкополосный доступ независимо от своего местоположения.

Сегодня примерно 3,7 млрд человек все еще не имеют даже базового доступа к интернету [1]. Большинство из них проживают в сельских и отдаленных районах, где наземные и неназемные сети могли бы работать вместе для предоставления широкополосного доступа, как показано на рис. 6.2. Для организации широкополосного доступа на море будет применяться интеграция широкополосных спутников, корабельных и наземных станций.

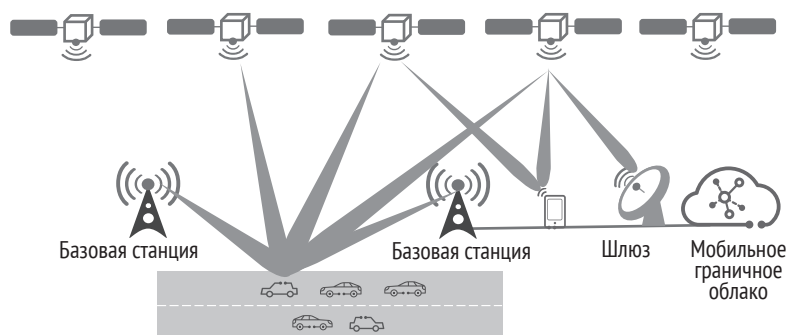


Рис. 6.2. Интегрированный сервис для пешеходов и автомобилей

Помимо использования фиксированной и подвижной ретрансляции, привлекательной перспективой выглядит прямое соединение между неназемными станциями и мобильными телефонами. В течение многих лет такое соединение было чрезвычайно дорогим и предлагало низкую скорость передачи данных. Кроме того, абонентам приходилось иметь два разных телефона: один для доступа к спутниковой сети, а другой – для доступа к сотовой сети. В будущем 6G позволит интегрировать все типы доступа в один мобильный телефон, обеспечивая плавное переключение между услугами.

Для обеспечения стабильного и высококачественного обслуживания в сельской местности интегрированная сеть 6G должна обеспечивать высокую скорость передачи данных, аналогичную той, которая предлагается через сельские базовые станции 5G, в частности 50 Мбит/с в нисходящем канале (загрузка) и 25 Мбит/с в восходящем канале (выгрузка) в пересчете на пользователя [2].

6.1.2. Широкополосное соединение в движении

Пример 6.2 Селия – исследователь пингвинов. Последние шесть месяцев она находилась в Антарктиде. Завершив исследования, она готовится отправиться домой, для чего

ей необходимо сесть на ледокол, следующий в Буэнос-Айрес, а затем на самолет. На ледоколе она загружает видео с пингвинами в социальные сети через широкополосную спутниковую службу ледокола. Прибыв в Буэнос-Айрес, она садится в самолет до дома и, используя бортовую систему связи 6G, делает видеозвонок своей семье в середине полета.

Предоставление людям доступа к интернет-услугам в любое время через одно и то же устройство независимо от их местонахождения, будь то дома или в самолете, является одной из конечных целей создания беспроводных сетей. Сегодня транспортные средства можно разделить на четыре основных типа: автомобили, поезда, самолеты и корабли (рис. 6.3). Если взять в качестве примера самолет, то в 2019 году этим типом транспорта воспользовались более 4 млрд человек [3]. Большинство из них не имели доступа к интернету во время полетов; для тех, у кого он был, соединение было медленным и дорогим. В будущем интегрированные системы 6G должны обеспечивать мобильное широкополосное подключение для всех пассажиров самолетов. Скорость передачи данных для каждого пользователя должна превышать 15 Мбит/с и 7,5 Мбит/с для загрузки и выгрузки соответственно [2], при условии 20%-го коэффициента активности и 400 пользователей на самолет. В этом случае общая скорость передачи данных для одного самолета должна быть не менее 1,2 Гбит/с и 600 Мбит/с для загрузки и выгрузки соответственно.

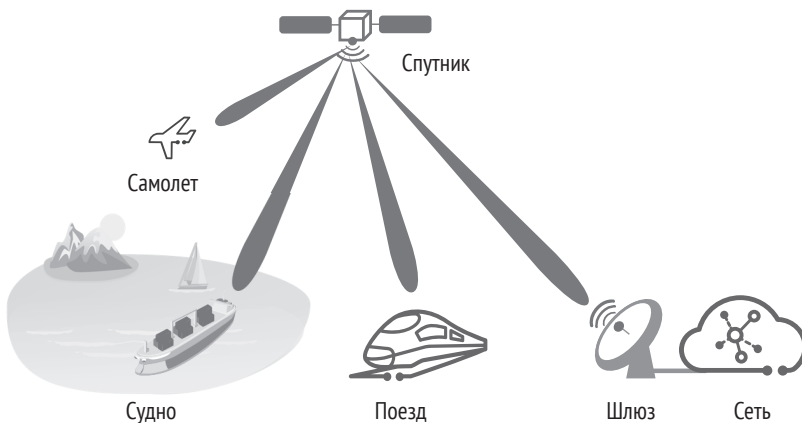


Рис. 6.3. Широкополосное соединение в движении

6.1.3. Связь служб быстрого реагирования и помощь при стихийных бедствиях

Пример 6.3 Андреа работает в Национальном центре реагирования на чрезвычайные ситуации, где она отвечает за оповещение о стихийных бедствиях и организацию ава-

рийно-спасательных операций в случае стихийных бедствий. В сезон торнадо Андреа передает информацию о торнадо рыбакам и судам через спутниковые сети. Как правило, в случае землетрясения часть оборудования наземной сотовой сети повреждается, а это означает, что Андреа должна использовать неназемные сети (включая спутники, НАР и дроны) для связи с местными спасательными командами. 6G расширяет возможности аварийно-спасательных служб за счет интеграции неповрежденных наземных и неназемных сетей. Интеграция также позволяет создавать ячеистые сети (меш-сети), чтобы члены спасательных команд могли общаться друг с другом и с центром реагирования на чрезвычайные ситуации.

Надежная и бесперебойная система экстренной связи имеет решающее значение в сценариях реагирования на бедствия, поскольку она обеспечивает прогнозирование бедствий, оповещение, реагирование на чрезвычайные ситуации и экстренную связь. В некоторых случаях стихийные бедствия могут нарушить работу наземных сетей; однако 6G объединяет наземные и неназемные сети для обеспечения непрерывного обслуживания и поддержки управления в чрезвычайных ситуациях. Система связи для управления в чрезвычайных ситуациях включает в себя несколько систем, таких как проводная и беспроводная передача голоса, сеть передачи данных и видеосистемы. С помощью программного обеспечения для управления и диспетчеризации в чрезвычайных ситуациях можно быстро связаться с командным центром, соответствующими подразделениями, группами экспертов и местными спасательными командами.

6.2. Широкий спектр услуг интернета вещей в труднодоступных местах

Пример 6.4 Джон – морской биолог, изучающий влияние морской среды на рыбу. Он сбросил много автономных буев, чтобы собрать информацию об океане. Эти буи питаются от солнечных батарей и могут подключаться к сети 6G. Джон может собирать информацию с буев, такую как загрязненность морской воды, ее температура, высота волн и скорость ветра, а также точное местоположение через сеть 6G. Он может скачать обновленную информацию на свой офисный компьютер, чтобы изучать изменения в водной среде.

В настоящее время связь для интернета вещей основана на покрытии земной поверхности наземной сетью. Однако во многих случаях этот подход не может гарантировать бесперебойное подключение. Например, сбор информации с буев в океане или передача информации из контейнеров во время их транспортировки по морю могут прерываться. В будущем устройства интернета вещей должны иметь возможность

подключаться к 6G и передавать информацию в любом месте и в любое время. Широкий спектр услуг интернета вещей охватит неподключенные места, такие как удаленные районы, океаны и т. д.

В удаленных районах с ограниченной сотовой связью может не хватать ресурсов для массового подключения устройств интернета вещей. Во многих ненаселенных районах, таких как Антарктида и пустыни, люди размещают устройства интернета вещей для сбора информации. Однако эти области не покрываются наземными сетями, а это означает, что возможности сбора информации ограничены. 6G, интегрируя наземные и неназемные сети, обеспечит доступ для устройств IoT в ненаселенных районах, что упростит сбор такой информации, как перемещения пингвинов или белых медведей, и мониторинг посевов на удаленных фермах.

В некоторых случаях людям необходимо разместить в океане буи для измерения высоты волн, температуры воды и скорости ветра. Такая информация может помочь морякам избежать участков с сильным ветром и волнами или принять необходимые меры предосторожности. Кроме того, важно сообщать информацию о состоянии контейнеров при морских перевозках на большие расстояния. Если людям будет доступна информация о каждом контейнере в режиме реального времени, они смогут проверять температуру, влажность и местонахождение контейнеров в течение всего процесса транспортировки.

6.3. Высокоточное позиционирование и навигация

Пример 6.5 После приземления в аэропорту Селия решает взять напрокат беспилотный автомобиль. Автомобиль, использующий точное позиционирование и прогнозирование дорожной ситуации, останавливается прямо перед ней, когда она выходит из терминала. Селия садится в машину, которая с помощью точного позиционирования и навигации везет ее домой, пока она читает новости в интернете. Селия живет далеко от центра города, поэтому, когда автомобиль теряет наземное покрытие, он динамически переходит на неназемную сеть 6G, чтобы продолжить получать услуги навигации и связи. Спустя час машина благополучно прибывает к дому Селии, где ее ждет семья.

Современные спутниковые навигационные системы, широко применяемые в быту, обеспечивают позиционирование с точностью около 10 м. Глобальная навигационная спутниковая система (global navigation satellite system, GNSS) и система интегрированного позиционирования на основе низкоорбитального созвездия (low-earth orbit, LEO) могут обеспечивать высокоточное позиционирование и навигацию, достигая точности около 10 см на открытом воздухе. Благодаря интеграции

наземных и неназемных сетей высокоточное позиционирование и навигация могут быть реализованы по всему миру, независимо от того, используются ли они в центре города или в отдаленных районах. Это упростит предоставление многих новых услуг, таких как высокоточные службы определения местоположения пользователей и службы навигации для беспилотных автомобилей, точные сельскохозяйственные и строительные работы.

Высокоточная навигация может заметно повысить эффективность сельскохозяйственных работ, например за счет повышения точности позиционирования машин в поле. Усовершенствованные технологии позиционирования и навигации обеспечат надежную техническую поддержку для таких операций, как точная подготовка земли, посев, культивация, внесение удобрений, защита растений, сбор урожая, опрыскивание и многое другое.

В будущем большинство автомобилей смогут подключаться к сети 6G. Наземная сеть может предоставлять высококачественные услуги V2X для транспортных средств в городе, в то время как неназемная сеть может предоставлять услуги точного (в пределах нескольких десятков сантиметров) позиционирования и навигации для транспортных средств в отдаленных и ненаселенных районах с помощью мегагруппировок LEO, как показано на рис. 6.2. Кроме того, для повышения точности навигационных услуг сети 6G будут широко использоваться БПЛА.

6.4. Мониторинг земной поверхности в реальном времени

Пример 6.6 Во время поездки в национальный парк Селия и ее семья стали свидетелями аварии, когда бензовоз опрокинулся в кювет и загорелся. Система мониторинга земной поверхности с помощью 6G обнаруживает эту ситуацию и отправляет данные в центр управления в режиме реального времени. Центр управления анализирует состав топлива и обнаруживает, что возникла опасность взрыва цистерны. Центр немедленно информирует Селию о риске через навигационную систему автомобиля, рекомендуя держаться на безопасном расстоянии от зоны происшествия. Соответственно, Селия выбирает новый маршрут и благополучно прибывает в национальный парк. Система мониторинга Земли не только уведомляет Селию, но и отправляет сигналы тревоги дорожной полиции и пожарным, помогая избежать взрыва.

Одним из типичных вариантов использования сегодняшней спутниковой системы является дистанционное сканирование и наблюдение поверхности Земли с помощью методов многоканальной съемки. Такие методы включают в себя оптическую съемку, при которой изображения получают с помощью камер видимого света и частично инфракрасного

диапазона, и радиочастотное сканирование, при котором изображение формируют путем записи и анализа радиоволн, отраженных от земной поверхности, с помощью радара с синтезированной апертурой. Однако из-за ограниченных возможностей каналов связи пройдет время, прежде чем люди смогут массово использовать эти изображения. В будущем интегрированная сеть 6G позволит одновременно осуществлять мониторинг Земли и связь в рамках единой системы.

6.5. Источники

- [1] ITU Broadband Commission for Sustainable Development, *State of broadband report 2019*, 2020. https://www.itu.int/dms_pub/itu-s/opb/pol/S-POL-BROADBAND.20-2019-PDF-E.pdf.
- [2] 3GPP, *Service requirements for the 5G system*, 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 22.261, 07 2020, version 17.3.0. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3107>.
- [3] E. Mazareanu, *Air transportation – statistics & facts*, 2020. <https://www.statista.com/topics/1707/air-transportation>.

Глава 7

Подключенное машинное обучение и сетевой ИИ

С 2012 года прорывы в технологиях, связанных с *машинным обучением* (machine learning, ML) на основе *глубоких нейронных сетей* (deep neural network, DNN), достижения в области суперскалярных вычислений и современного интернета, сделавшего большие данные доступными, привели к появлению искусственного интеллекта. Революция передовых технологий опирается на искусственный интеллект, который достиг, а в некоторых случаях превзошел человеческие способности. Например, ИИ превосходит людей в сложных играх, таких как шахматы и го, и даже в таких задачах, как распознавание изображений и голоса.

Недавнее исследование [1] предсказывает, что ИИ сможет выполнять широкий спектр задач в течение срока службы 6G (то есть с 2030 по 2040 год). Сценарии использования 6G для предоставления услуг искусственного интеллекта предсказать непросто, но пространство для инноваций будет огромным благодаря объединению возможностей 6G и искусственного интеллекта с машинным обучением. Это означает, что ИИ может оказать значительное влияние на общество и стать наиболее важной услугой и областью применения сетей 6G.

С математической точки зрения мы можем рассматривать методы машинного обучения в ИИ как реализацию серии сложных алгоритмов оптимизации. Ожидается, что с развитием технологий (особенно DNN), ростом вычислительных возможностей и появлением массивных данных реального времени, циркулирующих в мобильных сетях, ИИ с машинным обучением решит проблемы оптимизации более высокого уровня, которые не могут быть решены с помощью традиционных алгоритмов.

Как было сказано в главе 1, 6G – это ключевой фактор, позволяющий сделать ИИ доступным каждому человеку, в любом месте и в любое время. Следовательно, сеть 6G – это не просто «соединительный кабель» для служб и приложений искусственного интеллекта; скорее, она будет

разработана с нуля для более эффективной оптимизации ИИ. Передовые технологии мобильной передачи, а также вычислительные мощности, предлагаемые сетевыми узлами, будут использоваться для дополнения крупномасштабного интеллекта широко распространенным распределенным обучением. Это позволит обществу воспользоваться возможностями ИИ в самых разнообразных отраслях. Знания будут извлекаться из массивных данных, считываемых в целевой среде, а затем использоваться для принятия решений и гибкого интеллектуального управления объектами. Поэтому мы верим, что с помощью ИИ и 6G в ближайшем будущем произойдет *всемирная революция в познании*.

Как показано на рис. 7.1, ИИ будет играть важную роль как в оптимизации систем связи, так и в реализации «умных» вариантов использования (умные вещи, умный дом, умный город и т. д.). В следующих разделах основное внимание уделяется двум категориям базовых вариантов использования, а именно услугам и операциям 6G с поддержкой ИИ и услугам ИИ с поддержкой 6G.



Рис. 7.1. Нативный ИИ в системах связи и «умных» сценариях использования

7.1. Услуги и операции 6G, дополненные искусственным интеллектом

Существующие в настоящее время сети поставщиков услуг, особенно беспроводные сети, нуждаются в большом количестве сотрудников для управления, обслуживания и эксплуатации. Для более чем 250 операторов по всему миру их расходы на сетевые операции, администрирование и обслуживание (operations, administration, and maintenance, OA&M) являются серьезным финансовым бременем. Есть надежда, что скоро

мы перестанем нуждаться в персонале для поддержания функционирования и технического обслуживания сетей – это будет идеальная полностью автоматическая сеть.

7.1.1. Качество услуг сети 6G, дополненной искусственным интеллектом

Пример 7.1 Том отвечает за операции на умном предприятии, на котором работают всего несколько человек. Искусственный интеллект в 6G помогает сети обеспечивать чрезвычайно надежную и быструю связь между роботами индустрии 4.0. Во время перерыва на обед Том виртуально играет в футбол с друзьями. Нативный искусственный интеллект обеспечивает минимальную задержку в 6G-сети, позволяя Тому наслаждаться погружением в виртуальную реальность с обзором 360° без какого-либо дискомфорта, например укачивания. Ощувив жажду, Том дает устные распоряжения роботу, который понимает живую речь благодаря встроенному в сеть 6G искусственному интеллекту и доставляет напиток Тому, как официант. После работы Том едет в другой город, чтобы встретиться с друзьями. Нативный ИИ помогает сети обслуживать умный транспорт с беспилотным вождением пятого уровня, что гарантирует безопасную и расслабляющую поездку. Во время поездки ИИ динамически адаптирует покрытие своей сети 6G в соответствии с местоположением абонента и потребностями в услугах.

К 2030 году умные устройства по всему миру будут включать в себя постоянно растущее число личных и домашних устройств, различные городские датчики, беспилотные автомобили, умных роботов и многое другое. Этим умным устройствам потребуется беспроводная связь, чтобы облегчить полную координацию и взаимодействие при решении задач автоматизации. От беспроводной связи потребуется более высокая пропускная способность, более высокая надежность, а также более низкие задержки и джиттер, чем могут обеспечить существующие системы связи. Благодаря интеграции коммуникационных и вычислительных платформ мы сможем в полной мере использовать возможности сетей к восприятию и обучению. Системы связи будут поддерживать автоматическое конфигурирование ресурсов, предоставлять пользователям высокоинтеллектуальные и персонализированные услуги и обеспечивать почти оптимальную производительность.

В примере 7.1 ИИ может адаптироваться к изменениям окружающей среды и находить оптимальное решение в потенциальных чрезвычайных ситуациях, помогая обеспечить соответствие сети строгим требованиям к надежности и задержкам индустрии 4.0 для обмена информацией между роботами и центральной системой. В случае иммерсивной виртуальной реальности с обзором 360°, как было сказано в разделе 4.1, сжатие и локальный рендеринг помогают частично снизить строгие требования к пропускной способности и задержке E2E в сети 6G. Кроме

того, ИИ работает как механизм оптимизации, разумно оптимизируя как коммуникационные, так и вычислительные ресурсы. Возможности искусственного интеллекта в сетях 6G в этих областях напрямую повлияют на качество обслуживания.

ИИ, используемый в сетях 6G, может наделять роботов расширенными семантическими способностями, позволяя им понимать контекст семантических инструкций. Это поможет реализовать более человекоподобный интерфейс для услуг зондирования и локализации. Кроме того, на основе сетевого ИИ можно организовать понятное человеку семантическое общение между роботами, в отличие от сегодняшнего общения на уровне машинного кода.

ИИ также может применяться для синтеза сетевых структур с целью создания гетерогенных сетей, таких как интегрированные неназемные и наземные сети. Искусственный интеллект в 6G поможет координировать сложные многоуровневые гетерогенные сети и в конечном итоге предоставить пользователям оптимальное покрытие.

7.1.2. Управление сетью, дополненной ИИ

Пример 7.2 Эстер работает у оператора мобильной связи инженером по планированию сети. Ей поручили подготовиться к концерту, который, как ожидается, привлечет много людей. Чтобы приспособить сеть к резкому увеличению трафика во время концерта, Эстер планирует добавить несколько узлов БПЛА или низкоорбитальных спутников. Подобные задачи в прошлом требовали настройки множества параметров и проведения длительных тестов. Но теперь, имея доступ к возможностям 6G с нативным ИИ, она может просто перетащить узлы в желаемое место на экране планшета, связанного с интеллектуальной системой сетевого планирования, после чего последующая работа будет выполнена автоматически.

Поскольку частью инфраструктуры 6G станет большое количество БПЛА, HAPS и VLEO, управление такой многоуровневой системой может оказаться чрезвычайно сложной задачей. ИИ станет жизненно важным инструментом для решения таких проблем, как интеллектуальный гетерогенный доступ, оптимизация энергопотребления и десятков тысяч других параметров и повышение производительности системы. Например, агент ИИ будет обучаться оптимальному использованию спектра, мощности, антенн и энергии. Это поможет обеспечить эффективное взаимодействие различных узлов гетерогенной сети, таких как спутники, дроны и наземные станции.

Более того, расширенный интеллект позволит приблизить некоторые варианты использования 5G к уровню 6G. Но основной целью является применение ИИ для автоматизации управления сетью и в конечном итоге полного устранения вмешательства человека в работу, управление и обслуживание.

Другой многообещающий пример – автоматическое создание специализированных сегментов сети в соответствии с характеристиками услуг. Это отвечает потребностям операторов в гибком конфигурировании сети, помогая им превратиться в своего рода «фабрики» заказных сегментов.

7.2. Сервисы искусственного интеллекта с поддержкой 6G

Во многих областях ИИ достиг человеческого уровня. Однако с точки зрения вычислений он не так эффективен, как человеческий мозг; еще предстоит ликвидировать значительный пробел. В соответствии с законом Мура скорость вычислений нынешней компьютерной платформы удваивается каждые 18 месяцев [2]. Однако спрос на вычисления с искусственным интеллектом за последние 8–10 лет растет намного быстрее, примерно в 40–100 раз быстрее, чем закон Мура [3].

Недавнее исследование вычислительных ограничений ИИ предполагает, что использование машинного обучения для масштабных задач будет дорогостоящим и неэкологичным [4]. Полное использование возможностей глобального обучения с помощью высокоэффективной сети, такой как 6G, устранил дублирование вычислений машинного обучения и сведет к минимуму передачу больших данных, необходимых для машинного обучения, по беспроводной сети. Это будет ключевым фактором, определяющим устройство 6G. Ожидается, что 6G с расширенными возможностями связи и распределенных вычислений поможет сократить разрыв между вычислениями ИИ и быстродействием человеческого мозга. Для максимального повышения эффективности вычислений требуется распределенное и совместное машинное обучение, обеспеченное связью с малой задержкой и высокой пропускной способностью. В следующих разделах мы приводим два примера, демонстрирующих нативную поддержку различных услуг искусственного интеллекта в системах 6G.

7.2.1. Совместный анализ и управление в реальном времени

Пример 7.3 Том работает в здании, когда происходит землетрясение, и он оказывается в ловушке внутри. На помощь отправляется флот мобильных роботов, чтобы найти его и, в идеале, помочь ему выбраться наружу. Эти роботы наделены возможностями совместного использования искусственного интеллекта с сетью 6G. По отдельности они ищут все возможные маршруты, которые могут привести к местонахождению Тома. Каждый робот может видеть, слышать и ощущать свое окружение на предмет признаков жизни, огня, ядовитых газов и многого другого. Оперативно формируется локальная модель

ИИ, обучаемая на данных, полученных из окружающей среды. В то же время модели, обученные на месте, вкупе с некоторыми необработанными данными сканирования (при необходимости), полученными от всего парка роботов, передаются в сеть. Сетевые узлы используют свою глобальную информацию и расширенные вычислительные возможности для дальнейшего совершенствования глобальной модели ИИ. Затем эта глобальная модель возвращается и распространяется на каждого робота для улучшения локальных выводов, тем самым улучшая и синхронизируя интеллект всех роботов. Роботы в конечном итоге совместными усилиями находят Тома и показывают ему, как добраться до безопасного места, определяя лучший локальный маршрут и обрабатывая глобальную информацию от всех роботов. Несмотря на свой крошечный размер, роботы работают вместе, используя приемы совместной переноски предметов, чтобы преодолевать некоторые более крупные препятствия на пути.

В этом случае некоторые конкретные требования к сети 6G можно резюмировать с учетом трех аспектов:

- **низкая задержка и массивный доступ к восходящей линии связи:** в совместном ИИ каждый робот передает свою локально обученную модель ИИ (и, возможно, некоторые необработанные данные) в сеть, что способствует развитию глобального интеллекта. Сеть 6G должна поддерживать низкую задержку и массивные восходящие соединения, каждому из которых требуется пропускная способность, достигающая нескольких Гбит/с;
- **интеллект граничных узлов:** Том может беспокоиться о конфиденциальности данных (например, компрометирующих изображений и записей голоса). Ввиду критического характера ситуации важно обеспечить связь между узлами сети и роботами с малой задержкой. Чтобы решить проблемы Тома, обеспечивая при этом низкую задержку, вместо отправки всей информации в облако используется *локальный граничный узел* (edge node). Этот узел физически расположен близко к месту событий и предоставляет определенный набор обязательных услуг;
- **распределенное машинное обучение:** передача всех необработанных данных в центральное облако занимает значительную часть пропускной способности сети, а обработка этого большого объема данных создает нагрузку на вычислительные возможности центрального узла. Организуя распределенное выполнение вычислительных задач, совместный ИИ и машинное обучение частично компенсируют эти нагрузки;
- **сжатие данных машинного обучения для радиointерфейсов:** передача большого количества данных, необходимых для обучения глубоких нейросетей, по беспроводным каналам связи является дорогостоящей и неэффективной. Новые исследования

в области теории информации, такие как обнаружение «узких мест» информационных потоков, имеют решающее значение для изучения того, как сжимать большой объем данных, необходимых для обучения, поскольку методика сжатия обучающих данных сильно отличается от существующих технологий кодирования звука и изображения, которые были оптимизированы под особенности человеческого восприятия.

7.2.2. 6G для крупномасштабного интеллекта

Пример 7.4 Леса – сложная экосистема. Начиная с 1930-х годов ученым, изучающим леса, потребовалось 60 лет, чтобы правильно понять исходные данные и осознать, что естественные лесные пожары помогают поддерживать здоровье лесов. Впоследствии в 1990-х годах было принято решение прекратить вмешательство в естественные лесные пожары¹⁴. Однако леса, заново засаженные людьми, более уязвимы для природных пожаров, поскольку в среднем они в 2–3 раза плотнее естественных лесов. В таких лесах пожар распространяется слишком быстро и широко, выходя из-под контроля. Это считается основной причиной частых эпидемических лесных пожаров с 2015 года, когда большинство деревьев, высаженных после 1990-х годов, превратились во взрослый лес. Нам потребовалось еще 25 лет, чтобы понять, что повторно засаженные леса слишком густые.

В традиционном мире сбор необходимого объема данных, анализ и понимание данных и выработка на основе этого понимания оптимальной политики занимают длительное время. Такая политика зачастую устаревает еще до того, как ее можно будет реализовать.

Кибернетический мир с поддержкой 6G решит эти проблемы. Благодаря трехмерному полному покрытию и возможностям зондирования сетей 6G в лесах можно будет собирать самые разные данные в режиме реального времени. Данные о лесном хозяйстве, такие как температура, скорость ветра, влажность почвы, высота верхушек деревьев и т. д., будут собираться непрерывно. Затем ИИ можно использовать для поиска корреляции и периодичности, распознавания закономерностей, выполнения моделирования и сравнения данных с другими лесами и историческими данными. В кибермире можно смоделировать и оценить различные политики, что позволяет минимизировать ущерб физической природе.

В этом случае покрытие леса беспроводной сетью является фундаментальным требованием. Однако, поскольку деятельность человека в лесу ограничена, покрывать лес только наземной системой мобильной связи экономически нецелесообразно. С другой стороны, спутниковые услуги в настоящее время слишком дороги для реализации сканирования и сбора больших данных, достаточных для построения модели в кибермире. Высокая потребляемая мощность и потенциально корот-

¹⁴ Здесь и далее авторы говорят о Китае. – Прим. перев.

кое время автономной работы спутниковых оконечных устройств также препятствуют широкомасштабному развертыванию беспроводного доступа в лесу. Здесь не обойтись без новых возможностей 6G.

6G обеспечит трехмерное полное покрытие за счет интеграции наземных и неназемных сетей. По мере развития инфраструктуры VLEO, БПЛА и НАР вместе с всеобъемлющими интеллектуальными и сенсорными возможностями сетей 6G станет возможным воспроизведение физического мира (то есть леса) в кибермире с последующей выработкой оптимальной политики сбережения окружающей среды.

В этом случае большие объемы данных собираются по лесам всего мира. Эти данные передаются в распределенные локальные узлы граничного уровня для выполнения машинного обучения, а затем эти узлы обмениваются между собой параметрами обученной нейронной сети для формирования совместного ИИ. Обширные данные значительно повысят уровень осведомленности о проблемах лесов. Результаты прогнозирования различных политик будут более точными и эффективными, без длительных периодов сбора данных.

7.3. Источники

- [1] K. Grace, J. Salvatier, A. Dafoe, B. Zhang, O. Evans, *When will AI exceed human performance? Evidence from AI experts*, Journal of Artificial Intelligence Research, vol. 62, pp. 729–754, 2018. arXiv preprint arXiv: 1705.08807, 2018.
- [2] Wikipedia, *Moore's law*, [https://en.wikipedia.org/wiki/Moore %27s_law](https://en.wikipedia.org/wiki/Moore%27s_law).
- [3] R. Perrault, Y. Shoham, E. Brynjolfsson, J. Clark, J. Etchemendy, B. Grosz, T. Lyons, J. Manyika, S. Mishra, J. C. Niebles, *The AI index 2019 annual report*, AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, 2019.
- [4] N. C. Thompson, K. Greenewald, K. Lee, G. F. Manso, *The computational limits of deep learning*, arXiv preprint arXiv: 2007.05558, 2020.

Резюме части II

6G откроет множество новых возможностей для применения беспроводных технологий. Шесть категорий, о которых мы рассказали в этой части, охватывают будущие сценарии использования 6G, очевидные для нас *в настоящее время*. Однако с точки зрения всех возможных применений 6G эти варианты использования 6G являются просто отдельными примерами того, что ждет нас в будущем.

Численный анализ требований к производительности в этих типичных случаях использования приводит нас к набору целевых KPI, представленных в главе 1, особенно с точки зрения радиодоступа. Конкретные значения KPI установлены на основании самых строгих требований среди шести категорий. Например, с точки зрения скорости передачи данных, самые высокие требования предъявляет трехмерная голографическая связь. Что касается задержки передачи, то самые высокие требования предъявляют дистанционная работа в высокочастотных средах и промышленное управление движением. Чрезвычайно жесткие требования к джиттеру и надежности предъявляют сценарии использования детерминированной связи. В будущих умных городах и умной жизни варианты использования IoT определяют требования к плотности подключения и времени автономной работы сенсора. Что еще более важно, новые варианты использования мобильных сетей, такие как зондирование и искусственный интеллект, влекут за собой новые показатели качества услуги – разрешение зондирования и вероятность вывода. Для некоторых новых применений могут потребоваться дальнейшие исследования связанных с ними показателей, например гибкости и масштабируемости для поддержки собственных сервисов AI и уровня надежности сети.

Часть III



Теоретические основы

Введение в часть III

Беспроводные сети быстро превращаются в чрезвычайно сложные системы с экспоненциально увеличивающимся числом устройств, требующих еще большего количества подключений при все более жестких требованиях. В частности, сетям 6G будут присущи следующие характерные свойства.

- **Беспрецедентный размер и плотность сети:** в 2020 году на беспроводную связь приходилось 78 % глобального IP-трафика и было на 10 млрд подключенных устройств больше, чем в 2015 году. В перспективе сетевая инфраструктура будет полностью интегрирована с окружающей средой через встроенные в стены *транسمиттеры* (передатчики), кеширование данных и беспроводные датчики, расположенные повсюду вокруг нас. Коммуникационная среда превратится в интеллектуальную беспроводную сеть, где абсолютно все объекты, такие как здания, стены, автомобили и дорожные знаки, будут оснащены *интеллектуальными поверхностями*, способными усиливать входящие электромагнитные сигналы, выполнять вычисления и хранить данные.
- **Беспрецедентный уровень стохастичности:** поскольку динамика сети и потребности в подключении в общем случае не поддаются точному прогнозированию, сетевые ресурсы будут развертываться динамически – в зависимости от эволюции условий трафика и мобильности пользователей, при этом сама сетевая инфраструктура также станет мобильной. Это открывает новые возможности для обработки и хранения данных, которые теперь можно реализовать и на границе сети с помощью новых технологий, таких как мобильные граничные вычисления и граничный ИИ.
- **Беспрецедентная неоднородность с точки зрения типов услуг и требований QoS:** будущие беспроводные сети должны будут одновременно обеспечивать скорость передачи 1 Тбит/с для услуг MBW, задержку радиointерфейса 1 мс для URLLC, точность локализации 10 см для автомобильных приложений, способность обрабатывать 100 млн конечных устройств на км² в массовых сценариях IoT, увеличение покрытия на 10 дБ по сравнению с 5G и энергоэффективность в 10 раз выше, чем у существующих сетей.

Чтобы идти в ногу с быстро развивающимися беспроводными технологиями, нам не обойтись простым повышением быстродействия каналов передачи данных. Реализация концепции 6G требует изменения парадигмы в проектировании беспроводных сетей. Приходится признать, что научных знаний, которыми мы располагаем сегодня, просто недостаточно, и нам нужны новые фундаментальные знания, чтобы понять, как различные части такой сложной системы могут работать вместе, и достичь максимальной производительности сети с точки зрения пропускной способности, энергоэффективности, коэффициента битовых ошибок, точности локализации и задержки обработки/связи.

Инструменты, в которых мы сейчас так нуждаемся, представляют собой комбинацию теоретических математических и физических методов моделирования и анализа взаимосвязи между очень большим количеством узлов и потреблением энергии, вызванным вычислениями, необходимыми для обработки гигантского объема информации. На этом фундаменте мы и будем строить и развивать сеть 6G с нативным ИИ.

Глава 8

Теоретические основы искусственного интеллекта и машинного обучения

8.1. Фундаментальная теория искусственного интеллекта

Технология ИИ развивается быстрыми темпами, особенно с тех пор, как нейронные сети получили в свое распоряжение высокопроизводительные графические процессоры. В первой части этого раздела мы рассмотрим базовые определения в области машинного обучения, а затем более подробно опишем основы теории искусственного интеллекта.

8.1.1. Определения

Искусственная нейронная сеть

Искусственная нейронная сеть (artificial neural network, ANN) состоит из связанных между собой элементарных процессоров, называемых нейронами, которые обычно организованы в несколько последовательных слоев. Входной слой получает исходные данные, которые затем обрабатываются одним или несколькими скрытыми слоями и, наконец, выходным слоем, который предоставляет выходные данные ANN для дальнейшего использования. Если ANN содержит только один скрытый слой, она называется *мелкой*, а если присутствует более одного скрытого слоя, она называется *глубокой*. Обычно на практике используют глубокие ANN, так как установлено, что они требуют меньше нейронов, чем мелкие сети при равной функциональности [1]. В глубоком обучении каждый нейрон выполняет три операции:

- вычисляет аффинную комбинацию входных данных;
- вычисляет функцию активации (обычно нелинейную), вход которой является результатом ранее вычисленной аффинной комбинации (например, спрямленных функций, сигмоидальных функций и гиперболических тангенсов);
- передает результат нейронам следующего слоя.

Хотя операции, выполняемые каждым нейроном, довольно просты, объединяя достаточное количество нейронов, можно решать очень сложные задачи. Было показано, что ANN являются универсальными *аппроксиматорами функций*, благодаря чему их отношения ввода-вывода могут воспроизводить любые функции, при условии что веса и смещения аффинных комбинаций правильно настроены [2]. К сожалению, эта возможность бесполезна, если мы не знаем, как настроить веса и смещения для выполнения данной задачи. Процедура поиска весов и значений называется *обучением ANN* и представляет собой подачу набора данных с известными парами ввода-вывода в ANN, а затем использование хорошо зарекомендовавших себя алгоритмов обучения и стохастического градиентного спуска (SGD, *stochastic gradient descent*) [3, 4] для экстраполяции подходящей конфигурации весов и смещений.

Глубокая нейронная сеть

Глубокие нейронные сети (deep neural network, DNN) образуют подмножество ANN. Глубокая нейросеть состоит из нескольких слоев, каждый из которых содержит несколько нейронов. Нейроны одного слоя связаны с нейронами следующего слоя, и каждая связь имеет обучаемый вес. Каждый нейрон объединяет все свои входы в один выход. Эта объединяющая функция является нелинейной и чаще всего похожа на сигмовидную или спрямленную линейную функцию. Сегодняшние DNN имеют несколько заметных ограничений, которые следует тщательно учитывать:

- **обобщение:** на практике каналы мобильной связи меняются со временем. Если DNN включает в себя канал с изменяющимся во времени затуханием, это повлияет на возможность обобщения DNN;
- **сложность:** минимальное количество нейронов, используемых для кодировщика и декодера, определяет их сложность;
- **размер обучающих данных:** точность DNN зависит от обучающего набора данных. Способы формирования и уменьшения размера обучающего набора данных в настоящее время являются предметом интенсивных исследований;

- **продолжительность обучения:** процесс обучения DNN, как правило, является долгим и переменчивым. Это зависит от обучающих данных, что представляет проблему для его применения в системах реального времени.

Сверточная нейронная сеть

Среди всех типов DNN наиболее популярной является *сверточная нейронная сеть* (convolutional neural network, CNN); в ней реализованы двумерная сверточная фильтрация и максимальное объединение в дополнение к полной связности между слоями. Посредством сверточных слоев и слоев с максимальным объединением входные данные высокой размерности могут быть сведены к представлению меньшей размерности, что позволяет избежать возможных проблем оптимизации, вызванных размерностью. Низкоразмерное представление называется *скрытым слоем* в терминах DNN и содержит наиболее важные признаки, текстуры или семантику соответствующих многомерных входных данных, в зависимости от конкретного предмета обучения.

Благодаря такой многоуровневой архитектуре оптимизация Лагранжа для стохастического градиентного спуска может быть реализована с помощью цепочки дифференцируемых функций. Используя правило цепочки, CNN выполняет обратное распространение за эпоху данных, чтобы настроить все нейроны на их обучающие цели, применения или задачи. Типичные цели (т. е. критерии качества) обучения включают среднеквадратическую ошибку, максимальную вероятность и минимальную ошибку классификации. В некоторых случаях для управления процессом обучения (оптимизации) используется более одной цели обучения.

Рекуррентная нейронная сеть

В человеческих (семантических) языках предложение можно рассматривать как цепь Маркова, при этом каждое слово более или менее зависит от предшествующих ему слов. Точно так же радиоканал может быть упрощен до цепи Маркова. Однако, хотя традиционные DNN не могут учитывать предыдущие события, *рекуррентные нейронные сети* (recurrent neural network, RNN) используют для решения этой проблемы петли памяти, в которых сохраняется определенный набор предшествующих данных. Концепция RNN имеет большое значение для семантической коммуникации, которая будет обсуждаться в главе 10. *Долговременная краткосрочная память* (long short-term memory, LSTM) является примером рекуррентной нейросети для *обработки естественного языка* (natural language processing, NLP) [5]. Ее также предложили использовать в механизме выбора беспроводных каналов [6].



Рис. 8.1. Обобщенное представление машинного обучения как карты функций

8.1.2. Таксономия машинного обучения

С математической точки зрения ИИ может в общих чертах сводиться к вопросу поиска неизвестного отображения функций $f: X \rightarrow Y$, где X – это входное пространство признаков, представляющее точки данных, а Y – пространство меток, представляющее выходные данные знаний. Исходя из различий в имеющемся опыте, целевой функции и конкретных алгоритмах обучения, как показано на рис. 8.1, задачи ИИ можно разделить на три категории: обучение с учителем, обучение без учителя, обучение с подкреплением. В качестве усовершенствования разрабатывается самообучающееся (частично контролируемое) обучение, чтобы снизить затраты на ручную разметку обучающих данных. Кроме того, генеративные модели предназначены для непосредственного моделирования данных или установления условных распределений вероятностей между переменными.

Обучение с учителем

В модели *обучения с учителем* (supervised learning, контролируемое обучение) набор обучающих данных предварительно размечен, т. е. содержит пары ввод–вывод, на которых обучается DNN. Соответственно, чем больше обучающих данных, тем выше точность предсказаний нейросети. Обучение с учителем генерирует предсказатель на основе размеченных данных и предполагает хорошее обобщение неизвестных данных. Это наиболее распространенный метод обучения моделей, свя-

занных с проблемами классификации, регрессии и ранжирования. Хотя традиционные варианты, такие как метод опорных векторов и методы ядра, по-прежнему широко используются, обучение с учителем является наиболее эффективным вариантом для больших данных, особенно для задач компьютерного зрения и NLP. Успех обучения с учителем обусловлен удивительно мощной способностью DNN извлекать признаки из массива данных (так называемое обучение представлением). Средство извлечения структуры данных – ключевой компонент контролируемого обучения; например, свертка чрезвычайно важна при работе с изображениями. Однако разные структуры данных требуют своих собственных эквивалентов «свертки».

Обучение без учителя

В модели обучения без учителя (*unsupervised learning*, неконтролируемое обучение) набор обучающих данных не размечен, и DNN должна научиться самостоятельно извлекать признаки, шаблоны, семантику или текстуры. Следовательно, модель, обученная без учителя, делает прогнозы для неизвестных данных на основе опыта с неразмеченными данными. С помощью этого подхода трудно точно оценить способность обучаемой модели к обобщению. Модель должна изучить внутренние структуры и представления неразмеченных данных и сделать прогнозы на основе структур и представлений. Наиболее распространенными задачами, связанными с обучением без учителя, являются кластеризация и уменьшение размерности. *Автоматический кодировщик* (*auto encoder*, автоэнкодер) – широко используемый инструмент обучения без учителя.

Самоконтролируемое (или частично контролируемое, *semi-supervised*) обучение – это потенциально мощный метод обучения представлению, который автоматически генерирует какой-то контролируемый сигнал для решения конкретной задачи (например, для изучения представлений данных или автоматической разметки набора данных).

Обучение с подкреплением

В течение последних 20 лет обучение с подкреплением (*reinforcement learning*, RL) привлекает большое внимание и остается одним из важнейших направлений исследований машинного обучения. Агент в процессе обучения с подкреплением периодически наблюдает за состоянием среды, принимает решения, получает результаты и корректирует свою политику для достижения оптимальной точности. Одна проблема, которая препятствует широкому использованию RL на практике, заключается в том, что агенту требуется значительное время, чтобы изучить все возможные состояния, прежде чем перейти к наилучшей политике.

Глубокое обучение с подкреплением

Быстрое развитие глубокого обучения [1] в последние годы создало новые возможности в развитии RL, поскольку использование глубокого обучения, особенно глубоких нейросетей, может значительно ускорить обучение и вывод RL. Новая технология *глубокого обучения с подкреплением* (deep reinforcement learning, DRL) использует преимущества алгоритмов глубоких нейросетей и специального оборудования, такого как графические и математические процессоры. DRL также обеспечивает более высокую скорость обучения и лучшую производительность. Более того, когда группа автономных и взаимодействующих объектов существует в общей среде, каждый объект может быть оснащен агентом DRL, помогающим в сотрудничестве, конкуренции и координации с другими объектами, за счет принятия правильных решений для достижения глобальных целей.

Большинство проблем принятия решений в неопределенных и стохастических средах можно смоделировать с помощью так называемого *марковского процесса принятия решений* (Markov decision process, MDP) [7], который обычно реализуется с помощью динамического программирования [8]. Однако с увеличением масштаба системы вычислительная сложность быстро становится непреодолимой. Более того, точное моделирование системы иногда нецелесообразно. В результате DRL предоставляет альтернативное решение проблемы. На рис. 8.2 схематически представлены основные различия между RL и DRL.

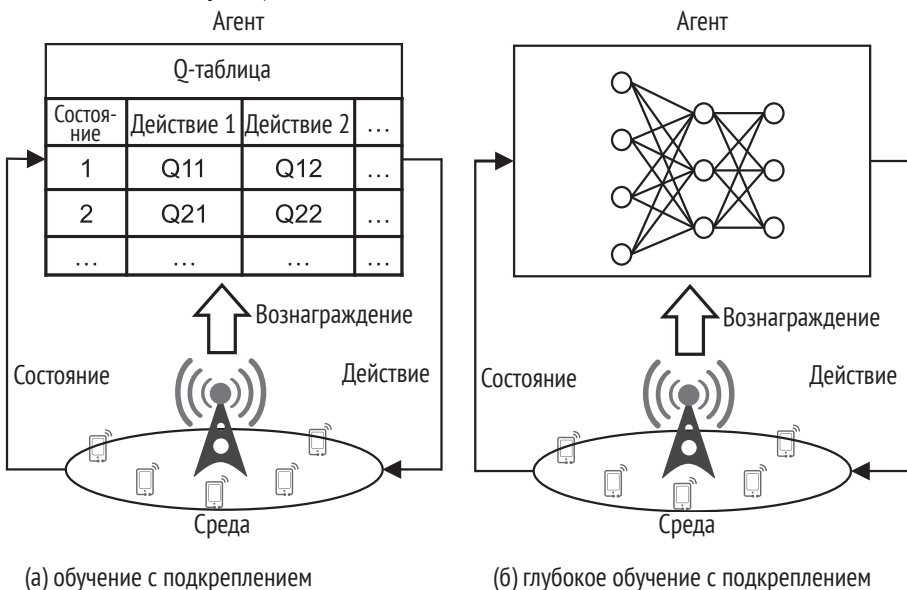


Рис. 8.2. Структуры RL и DRL в контексте беспроводных сетей

Вышеупомянутые категории ИИ или машинного обучения должны быть реализованы с использованием конкретной структуры нейронной сети, такой как ANN, DNN или CNN и т. д. Далее мы обсудим сначала теоретико-информационные принципы DNN, а затем опишем несколько классов реализации DNN.

8.1.3. Теоретико-информационный принцип DNN

Для интерпретации DNN с акцентом на технологии глубокого обучения используется множество различных теорий, таких как теория оптимизации, алгоритмы максимизации ожидания, теория топологии, теория графов, семантическая теория, теория динамических систем и теория приближений. Выбор подходящей теории для интерпретации DNN зависит от реальной проблемы. Одними из наиболее подходящих теорий в контексте беспроводной связи являются теория информации и теория информационных узких мест, в частности, поскольку они исходят из той же точки зрения на теорию информации, что и беспроводные системы. Далее мы тоже будем придерживаться этой точки зрения.

Одним из показателей эффективности коммуникаций является теория *информационных узких мест* [9]. Эта теория находится на стыке машинного обучения и прогнозирования, статистики и теории информации.

Шеннон определил информацию, которая представляет собой строительные блоки для теории глубокого обучения, как неопределенность с вероятностной точки зрения.

Во-первых, *взаимная информация* $I(X; Y)$ отражает степень вероятностного взаимодействия между двумя случайными величинами: X и Y . Если изменяется X , то в некоторой степени будет изменяться Y , и наоборот: $I(X; Y) = I(Y; X)$. Глубокую нейросеть можно смоделировать как поток взаимной информации от входа к выходу.

Во-вторых, *состязательные отношения* взаимной информации $I(X; Y) = I(P_X; P_{Y|X})$ включают в себя вогнутую функцию $I(X; Y)$ от переменного P_X и фиксированного $P_{Y|X}$, а также выпуклую функцию $I(X; Y)$ переменного $P_{Y|X}$ и фиксированного P_X . Отсюда вытекает необходимость кодирования источника информации, чтобы минимизировать $I(X; Y)$, и кодирования канала, чтобы максимизировать $I(X; Y)$.

Шеннон утверждал, что надежная система связи должна включать и то, и другое. Точно так же DNN совмещает два состязательных шага. В то время как в процессе обучения в одни эпохи преобладает понижение размерности вектора признаков, а в другие – повышение, оба этих фактора необходимы для обучения. В конце обучения достигается равновесие.

Для объяснения того, что происходит в DNN с точки зрения теории информации, в [9] предлагается теория информационного узкого мес-

та с использованием потока взаимной информации и состязательного характера повышения и понижения размерности.

Обучая DNN, мы ожидаем, что выходной слой Y будет максимально похож на входной слой X , а скрытый слой Z будет максимально отличаться от входного слоя X , насколько это возможно. С точки зрения теории информации это означает максимизацию $I(X; Y)$ при минимизации $I(X; Z)$. При движении от X к Z происходит сжатие (кодирование) информации¹⁵, а при движении от Z к Y происходит декодирование¹⁶. Предположим, что в скрытом слое Z присутствует информационное узкое место γ : если энтропия Z меньше, чем γ , то вся информация Z может пройти; в противном случае может пройти только самая существенная γ -часть Z .

Взаимная информация между X и Z определяется следующим образом:

$$I(X; Z) = \sum_x \sum_z P_{Z,X} \cdot \log \frac{P_{Z,X}}{P_Z \cdot P_X} = \sum_x \sum_z P_X \cdot P_{Z|X} \cdot \log \frac{P_{Z|X}}{P_Z}.$$

Перекрыстная энтропия в Y равна:

$$H_{\text{cross}}(Z, Y) = - \sum_z P_Z \cdot \log P_{Y|Z} = E_{x \sim P_X} \left[- \sum_z P_{Z|X} \cdot \log P_{Y|Z} \right].$$

С учетом информационного узкого места γ окончательная функция потерь DNN равна:

$$\begin{aligned} f_{\text{loss}} = E_{x \sim P_X} \left[- \sum_z P_{Z|X} \cdot \log P_{Y|Z} \right] \\ + \beta \sum_x \sum_z P_X \cdot P_{Z|X} \cdot \max \left(\log \frac{P_{Z|X}}{P_Z} - \gamma, 0 \right). \end{aligned}$$

Поскольку математически сложно минимизировать компонент

$$\sum_x \sum_z P_X \cdot P_{Z|X} \cdot \max \left(\log \frac{P_{Z|X}}{P_Z} - \gamma, 0 \right),$$

альтернативой является минимизация его верхней границы $E_{x \sim P_X} [D(P_{Z|X} | Q_Z)]$, где $D(P_{Z|X} | Q_Z)$ – дивергенция (расхождение) Кульбака–Лейблера $P_{Z|X}$ по Q_Z . В итоге функция потерь принимает следующий вид:

¹⁵ В привычных терминах нейросетей это соответствует уменьшению размерности вектора признаков при переходе к Z -слою. – *Прим. перев.*

¹⁶ Соответственно, здесь происходит увеличение размерности вектора, причем размерность X в общем случае не совпадает с размерностью Y . – *Прим. перев.*

$$f_{loss} = E_{x \sim P_X} \left[- \sum_z P_{Z|X} \cdot \log P_{Y|Z} \right] + \beta \cdot E_{x \sim P_X} \left[D(P_{Z|X} \| Q_Z) \right].$$

Функция потерь определяет информационное узкое место [10] и выявляет глубинную взаимосвязь между целью обучения, архитектурой нейронной сети и распределением входных данных.

Некоторые методы обучения могут использовать теорию информационных узких мест для оптимизации своей архитектуры. Например, вариационный вывод строит нижнюю границу цели информационного узкого места и обучается на собранных данных без предварительного знания об их распределении [11]. Эта структура может быть расширена на случаи с различными диспропорциями, возникающими в задачах высокого уровня, или на случаи с многозадачными проблемами [12].

С точки зрения архитектуры нейронной сети и метода обучения проще всего разделить модель нейронной сети, предварительно обученную для задач высокого уровня, на две части: одну, используемую на стороне передатчика, и одну, используемую на стороне приемника. Затем можно вставить нейронные слои между двумя частями. Вся модель обучается, а затем настраивается с учетом существующих каналов [13].

8.1.4. Реализации DNN

Автоэнкодер

Если сжатие ($E_x \sim P_X [D(P_{Z|X} \| Q_Z)]$) неявно реализовано в виде DNN, оно выполняется *автоэнкодером* (auto encoder, AE) [14].

Коэффициент сжатия $|Z|/|X|$ соответствует выборке данных X , а цель Y априори неизвестна. Когда $Y = X$, автоэнкодер становится источником на основе обучения для блочного кода фиксированной длины с самым низким коэффициентом сжатия $|Z|/|X|$, который в целом ограничен снизу $H(X)$. Если размер X недостаточно велик (маленький блок), автоэнкодер может помочь найти исходный код со степенью сжатия $|Z|/|X|$ ниже, чем $H(X)$.

Вариационный автокодер

Вариационный автоэнкодер (variational auto encoder, VAE) [15] вводит в скрытый слой априорное распределение (Q_Z), и это распределение является показателем восприятия поверх скорости и сжатия. Типичный вариационный автоэнкодер показан на рис. 8.3. Различные априорные распределения (Q_Z) приведут к разным коэффициентам DNN. Если Q_Z является однородным многомерным распределением Гаусса, проекция по каждому измерению представляет собой оценку по этому измерению. Теоретически Q_Z может быть любым другим распределением, и,

следовательно, неоднородные подканалы беспроводной системы также могут быть смоделированы как Q_z , что позволяет настраивать VAE для конкретного ранее известного распределения.

В отличие от автоэнкодеров, которые только максимизируют $I(X; Y)$, вариационные автоэнкодеры работают по двум конкурирующим направлениям, одновременно минимизируя $I(X; Z)$ и максимизируя $I(X; g_Y(Z; \theta_g))$. В конце обучения $I(X; g_Y(Z; \theta_g))$ представляет собой минимальную взаимную информацию (Z), которая присутствует в слое узкого места. В главе 23 книги мы объясним, как вариационные автоэнкодеры помогают генерировать схему кодирования общего источника.

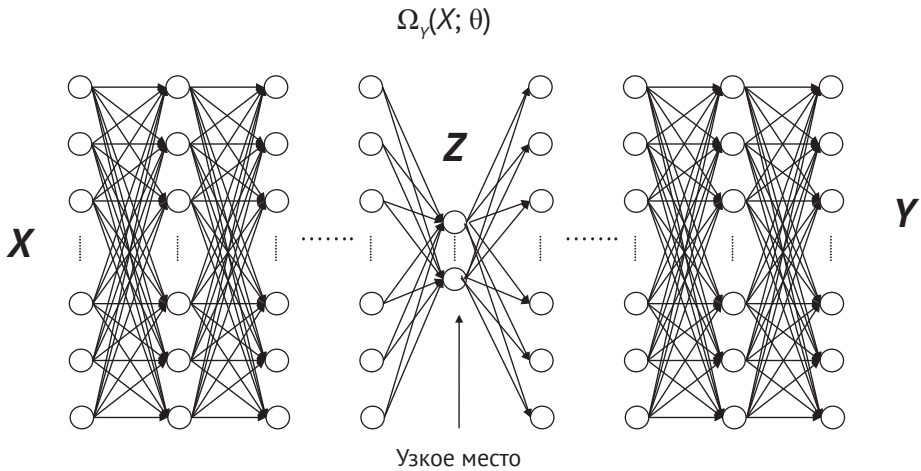


Рис. 8.3. Информационное узкое место автоэнкодера, такое что $|X| > |Z|$ и $|Y| > |Z|$, т. е. скрытый слой Z является узким местом

Генеративно-сопоставительная сеть

Генеративно-сопоставительная сеть (generative adversarial network, GAN) – это большой шаг вперед по сравнению с VAE [16]. В типичной GAN Вассерштейна априорные распределения (Q_z) реальны и не являются гауссовыми образами, а метрика больше не является взаимной информацией – теперь это расстояние Вассерштейна. Хотя GAN – одна из лучших генеративных DNN, доступных в настоящее время, VAE по-прежнему предпочтительнее в системах беспроводной связи из-за наличия скрытого гауссова уровня – над гауссовыми распределениями можно выполнять линейные операции.

8.2. Теория распределенного ИИ

На сегодняшний день глубокое обучение используется только в традиционных централизованных архитектурах, тогда как его исполь-

зование в распределенных архитектурах редко становится объектом исследований. Это критический вопрос с точки зрения перспектив создания сетей 6G, поскольку распределенные архитектуры необходимы для обеспечения масштабируемости и гибкости системы связи. Однако применение глубокого обучения в распределенной архитектуре вызовет несколько фундаментальных проблем. Например, у каждого беспроводного узла будет своя собственная ANN, которая будет обучена с использованием собственного набора данных, полученного на основе локальных измерений и опыта. В результате разные беспроводные узлы должны научиться вести себя на основе наборов данных, которые могут отличаться как по количеству (разные узлы могут иметь разные возможности измерения и хранения), так и по качеству (разные узлы могут испытывать разные искажения данных из-за различий в датчиках). Это потенциально может привести к нестабильности и, в худшем случае, к сбою беспроводной сети. Кроме того, в распределенной архитектуре каждый узел будет пытаться оптимизировать свою собственную производительность, а не общесистемную полезность, тем самым рассматривая другие узлы как возможных противников. Это может вызвать нежелательные или опасные ситуации, в которых беспроводной узел может научиться жульничать, чтобы максимизировать свою производительность. Следовательно, крайне важно получить твердые знания обо всех этих проблемах, чтобы гарантировать правильное поведение узлов в беспроводных сетях на основе ANN. Наконец, большинство результатов глубокого обучения не предусматривают получение распределенных масштабируемых решений, учитывающих большие размеры и непредсказуемую эволюцию с течением времени в сценариях беспроводной связи 6G. В результате создание подходящих обучающих наборов становится очень сложной задачей, поскольку общие подходы к обучению с учителем предполагают, что ANN будут работать в условиях, аналогичных тем, которые были представлены на этапе обучения. В реальности внезапные изменения трафика и/или топологии могут сделать доступные обучающие наборы бесполезными. Следовательно, срочно требуются новые методики обучения ANN, способные предвидеть или, по крайней мере, приспосабливаться к внезапным изменениям в беспроводной сети.

Ниже представлены две такие методики: федеративное обучение (federated learning) и многоагентное обучение с подкреплением (multi-agent reinforcement learning, MARL).

Федеративное обучение

Федеративное обучение – это подход к машинному обучению, при котором центральная модель обучается децентрализованно [17, 18] с использованием данных на узлах. Суть процесса заключается в следую-

щем: в каждом раунде каждый узел независимо вычисляет обновление для центральной модели, используя свои сохраненные данные, а затем передает это обновление (но не данные) на центральный сервер. Центральный сервер объединяет обновления, отправленные с разных узлов, а затем вычисляет улучшенную глобальную модель.

Современные мобильные устройства хранят в своей памяти множество данных, и эти данные можно использовать для обучения моделей. Хотя такие данные могут значительно улучшить модели, они часто состоят из конфиденциальной информации и поэтому находятся в частной собственности. Владельцы обычно не хотят делиться такими данными с центральным сервером в целях обучения. Чтобы обойти эту проблему, сервер может отправить центральную модель узлам для обучения на местах с использованием локальных данных. После создания обновленной модели узлы отправляют ее на центральный сервер. Таким образом, центральный сервер будет получать большое количество обновлений модели от многих узлов, которые затем объединяются в улучшенную глобальную модель. Затем начинается еще один раунд процесса обучения: центральный сервер отправляет улучшенную модель новым узлам, а новые узлы обучают ее дальше, используя свои личные данные. После многих раундов параметры модели могут сходиться, в результате чего получается хорошо обученная модель.

Основная цель федеративного обучения – отделить обучение модели от прямого доступа к необработанным обучающим данным, что позволяет использовать огромные объемы данных, хранящихся на узлах. Ниже приводится общее описание процесса федеративного обучения, которое в основном основано на [17]. Также существует множество других дополнений, например [19, 20].

Предположим, что обучаемая модель является нейронной сетью, а $f_i(w)$ является функцией потерь прогноза в примере (x_i, y_i) с параметром w . Также предположим, что существует K пользователей, по которым разделены данные, P_k – это множество индексов точек данных клиента k , а $n_k = |P_k|$. Тогда минимизируемая целевая функция может быть выражена как

$$f(w) = \sum_{k=1}^K \frac{n_k}{n} F_k(w),$$

где

$$F_k(w) = \frac{1}{n_k} \sum_{i \in P_k} f_i(w).$$

Пусть C будет долей узлов, которые выбирают в каждом раунде для вычисления обновлений и градиентов функции потерь. Если $C = 1$, дан-

ные на всех узлах используются для вычисления полного (нестохастического) градиентного спуска.

Типичная реализация федеративного обучения выбирает t случайных устройств в каждом раунде. Центральный сервер отправляет каждому узлу параметр сети w_t ; затем каждый пользователь k вычисляет $g_k = \nabla F_k(w_t)$, который является средним градиентом его локальных данных в текущей модели w_t , а потом обновляет параметр $w_t^k \leftarrow (w_t^k - ag_k)$, (где a – скорость обучения). Этот шаг повторяется E раз на устройствах пользователей, чтобы гарантировать отправку наиболее надежного w_t^k обратно на центральный сервер. Затем центральный сервер объединяет эти w_t^k и применяет обновление $w(t+1) \leftarrow \sum_{k=1}^K \frac{n_k}{n} w_t^k$. Объем вычислений определяется тремя ключевыми параметрами: C – доля узлов, участвующих в вычислениях в каждом раунде; E – количество проходов обучения, которые каждый клиент выполняет над своим локальным набором данных в каждом раунде, и B – размер локального мини-пакета, используемого для обновлений клиента. Если $B = \infty$, полный локальный набор данных рассматривается как один мини-пакет. Для пользователя с n_k локальными примерами за каждый раунд будет выполняться $u_k = E \frac{n_k}{B}$ локальных обновлений.

Федеративное обучение дает ряд преимуществ, включая следующие:

- **для обучения модели можно использовать большой объем реальных данных:** поскольку узлов много и каждый узел содержит определенный объем данных, агрегирование всех данных увеличивает возможность оптимального обучения модели;
- **обучение выполняется на узлах децентрализованно**, что снижает вычислительную нагрузку на центральный сервер;
- **соблюдаются требования к конфиденциальности** [20], поскольку пользователи загружают на сервер только обновления параметров модели, а не свои собственные данные.
- С другой стороны, есть и недостатки:
- **центральный сервер все же нуждается в определенном доверии**, поскольку процесс проверки обновлений модели, отправленных пользователем, иногда может приводить к извлечению некоторой информации о пользователе;
- **стоимость связи является существенной проблемой**, поскольку ресурсы связи потребляются дважды на пользователя: один раз для отправки параметров пользователю и второй раз для возврата обновлений на сервер. Случайный характер беспроводной связи делает такие обновления довольно сложными;

- **дополнительные трудности** связаны с тем фактом, что обучающие данные не берутся из независимых и одинаково распределенных источников, поскольку каждый пользователь имеет свой собственный набор данных, который не может представлять распределение популяции. Отдельные пользователи также обладают разными объемами локальных обучающих данных, и эти различия могут повлиять на сходимость алгоритма обучения;
- **децентрализованное обучение сложнее централизованного** с точки зрения продолжительности и настройки. Следует заранее тщательно выбирать значения гиперпараметров и архитектуру нейронной сети;
- **эффективность федеративного обучения** ниже, чем у централизованного метода.

Многоагентное обучение с подкреплением (MARL)

В обучении с подкреплением [21] *решатель* (т. е. агент) стремится достичь оптимального поведения в присутствии неопределенностей путем взаимодействия с окружающей средой, что обычно моделируется как марковский процесс принятия решений. С развитием глубокого обучения [22] было показано, что обучение с подкреплением обеспечивает выдающуюся производительность в таких областях, как настольные игры [23, 24], автономное вождение [25] и робототехника [26]. В публикации [27] представлен обзор последних достижений в области глубокого обучения с подкреплением. В большинстве приложений агент учится реагировать на неопределенную среду или на одного соперника, который реагирует в ответ. Однако во многих важных приложениях друг с другом взаимодействуют большое количество агентов; некоторые из них могут быть устройствами, реализующими одинаковые или разные стратегии для принятия решений о своих действиях. MARL обладает потенциалом для систематического анализа среды с помощью стратегически взаимодействующих агентов. Изучение MARL сочетает в себе обучение с подкреплением и методы MDP с теорией игр, особенно с разделом теории игр, известным как игры среднего поля [28]. Он стремится предоставить алгоритмы, способствующие изучению стабильных политик наилучшего реагирования в многоагентной системе с достаточно большим числом агентов, что остается серьезной проблемой в большинстве случаев, поскольку с увеличением числа агентов растет сложность [29].

Сложность распространения классических алгоритмов обучения (таких как Q-обучение с подкреплением, адаптивный эвристический критик или многорукий бандит с минимизацией сожаления) на многопользовательские сценарии проистекает из баланса, который необ-

ходимо поддерживать между использованием информации, полученной во время обучения, и адекватным определением области поиска. Например, MARL среднего поля может быть очень эффективным при большом количестве взаимодействующих устройств.

Существуют различные примеры приложений MARL в беспроводных сетях. Например, проблема полностью децентрализованного MARL рассматривается в [30], где агенты расположены в узлах изменяющейся со временем сети связи. Авторы предлагают два децентрализованных алгоритма «субъект–критик» с аппроксимацией функций, которые применимы к крупномасштабному MARL, где как количество состояний, так и количество агентов чрезвычайно велико. В децентрализованной структуре этап действия выполняется каждым агентом индивидуально, без необходимости делать выводы о политике других агентов. В то же время для критического шага принимается согласованное обновление по сети. Оба алгоритма являются полностью инкрементными и могут быть реализованы онлайн. Для этих алгоритмов также доступен анализ сходимости, когда функции значений истинности аппроксимируются в классе линейных функций.

8.3. Теория динамической байесовской сети

Своевременная и экономичная обработка сигналов и информации в сетях радиодоступа требует использования пространственно-временных зависимостей между сетевыми переменными. Эти зависимости возникают в нескольких пространственно-временных областях, например:

- **области активации пользователей**, которые включают устройства МТС/IoT, сообщающие/реагирующие на коррелированные события;
- **области каналов**, когда определенные пользователи имеют схожие характеристики распространения;
- **области контента данных**, когда пользователи в социальной структуре разделяют или запрашивают одинаковый контент.

Эти зависимости открывают возможности для использования прироста производительности в беспроводных сетях. Поэтому изучение зависимостей между сетевыми переменными имеет важное значение для предоставления расширенной информации и алгоритмов обработки сигналов для улучшенного радиодоступа.

Динамический байесовский вывод – это естественная структура, используемая для изучения зависимостей между переменными во времени, включая помехи и шаблоны трафика, структуру канала и обновление активности пользователя/данных. Выученная изменяющаяся

во времени структура может затем использоваться для динамической кластеризации в бессотовой массивной MIMO, например обеспечивая улучшенные решения радиодоступа. Многие такие проблемы могут быть сформулированы в терминах оценки разреженной изменяющейся во времени обратной ковариационной матрицы, которая выявляет динамическую сеть взаимозависимостей между объектами. Однако по сравнению с инвариантным во времени выводом обучение переменной по времени сети требует больших вычислительных ресурсов, поскольку существует больше параметров и требуется дополнительная связь. Кроме того, стандартные методы плохо масштабируются для более крупных сетей, где предпочтительны методы невысокой сложности и приближительного вывода. В идеале эти методы должны также должным образом учитывать накладные расходы на обучение путем компромисса между ресурсами канала и производительностью. Кроме того, алгоритмы обучения должны работать с неполными нелинейными измерениями; например, когда данные объединяются с использованием каналов с ограниченной пропускной способностью, как в случае распределенного MIMO.

Чтобы получить масштабируемые решения, мы будем использовать приближенные методы вывода (такие как передача сообщений) и их распределенные (многопроцессорные) версии. Передача сообщений и приближенная передача сообщений недавно продемонстрировали масштабируемую производительность в обобщенных обратных линейных задачах и будут служить отправной точкой для применяемых здесь методов. Существующие решения для статических сценариев, не зависящих от времени, будут расширены для учета изменчивости по времени. Далее мы рассмотрим типичные теоретические инструменты.

Байесовская сеть

Байесовская сеть (Bayesian network, BN) – это тип графовой модели, представленной в виде ориентированного ациклического графа. Узлы – это графовые представления объектов и событий, существующих в реальном мире, обычно называемые *переменными* или *состояниями*. Для каждой переменной в ориентированном ациклическом графе существует функция распределения вероятностей, размеры и определение которой зависят от ребер, ведущих к переменной.

Байесовские сети можно определить как частный случай в более общем классе, известном как графовая модель, в которой узлы представляют случайные величины, а отсутствие дуг – предположения об условной независимости между переменными. Если рассматривать граф в целом, условные вероятности, структуру BN и совместное распределение вероятностей можно использовать для определения предельной

вероятности или правдоподобия каждого узла. Эта процедура известна как *маргинализация*. Расчет степени доверия в BN выполняется всякий раз, когда изменяется одна из этих предельных вероятностей. Эффект наблюдения распространяется по сети, и на каждом шаге распространения обновляются вероятности различных узлов. Согласно [31], в простых сетях предельная вероятность, или правдоподобие каждого состояния, может быть вычислена на основе знания совместных распределений согласно теореме Байеса.

Динамическая байесовская сеть

Большинство событий в нашей повседневной жизни не обнаруживаются в конкретный момент времени и могут быть описаны с помощью нескольких состояний наблюдений, которые позволяют судить об одном законченном конечном событии. Модель *динамической байесовской сети* (dynamic Bayesian network, DBN) описывает систему, которая динамически изменяется или развивается с течением времени. Эта модель позволяет пользователям постоянно отслеживать и обновлять систему и даже прогнозировать ее поведение. DBN обычно определяется как частный случай односвязной BN, специально предназначенный для моделирования временных рядов, как указано в [32]. Все узлы, ребра и вероятности, которые формируют статическую интерпретацию системы, идентичны таковым в BN. Переменные здесь могут быть обозначены как состояния DBN, поскольку они включают временное измерение, а состояние системы, описываемое DBN, удовлетворяет марковскому условию, как показано на рис. 8.4.

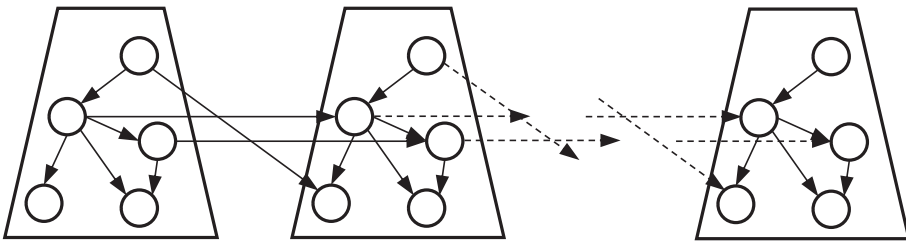


Рис. 8.4. Временной интервал, представляющий моментальный снимок развивающегося во времени процесса

Скрытая марковская модель

Скрытая марковская модель (Hidden Markov Model, HMM) – это стохастический конечный автомат, в котором каждое состояние генерирует (излучает) наблюдение. Мы будем использовать X_t для обозначения скрытого состояния и Y_t для обозначения наблюдения. Если существует K возможных состояний, то $X_t \in \{1, K\}$. Параметры модели включают

в себя распределение начального состояния, модель перехода и модель наблюдения. Модель перехода обычно характеризуется условным полиномиальным распределением $A(i; j) = P(X_t = j | X_{t-1} = i)$, где A – стохастическая матрица. Стохастическая матрица A часто разреженная, и ее структура обычно изображается в виде графа, где узлы представляют состояния, а стрелки – допустимые переходы (то есть переходы с ненулевой вероятностью).

Технические характеристики DBN можно разделить на три категории:

- **вывод:** поскольку в каждом временном интервале можно наблюдать только подмножество состояний, нам необходимо вычислить все неизвестные состояния в DBN. Это достигается с помощью процедуры, известной как вывод, в соответствии с которой различные типы DBN требуют различных видов оценок и расчетов на основе их конкретных структур. В определенных обстоятельствах может оказаться более подходящим не оценивать условную функцию плотности вероятности, а вместо этого оценивать статистику достаточности функции плотности вероятности. Если имеется сильное сходство между DBN и односвязными BN, можно использовать эффективный алгоритм прямого и обратного хода. Также для вычисления логического вывода в DBN используются алгоритмы сглаживания и прогнозирования;
- **декодирование последовательности:** другая проблема, связанная с DBN, – это поиск наиболее вероятной последовательности скрытых переменных с учетом наблюдений. Поскольку узлы DBN могут иметь более одного состояния, необходимо определить последовательность скрытых состояний с наивысшими вероятностями. Эта процедура обычно обозначается как декодирование последовательности и может быть достигнута с использованием *алгоритма динамического программирования Витерби*. Следует отметить, что если мы намерены декодировать последовательность скрытых состояний с помощью алгоритма Витерби, необходимо иметь полный набор наблюдений. Вариант алгоритма, известный как усеченный алгоритм Витерби, может использоваться после получения фиксированного количества наблюдений;
- **обучение:** представление реальных проблем в структуре DBN часто требует введения нескольких узлов, для которых невозможно точно определить условные вероятности. Даже экспертные знания не могут предложить решения для некоторых условных отношений в определенных областях. В таких обстоятельствах становится необходимо изучить эти конкретные распределения вероятностей. Этот процесс обучения сложен и основан на *алго-*

ритме максимизации ожидания или общей максимизации ожидания (general expectation maximization, GEM).

В табл. 8.1 приведены наиболее эффективные методы в зависимости от того, известна ли структура. Каждый метод также имеет уникальные преимущества и недостатки, перечисленные в табл. 8.2.

Таблица 8.1. Методы для различных состояний структуры

Состояние структуры	Наблюдаемость	Метод
Известное	Полная	Простая статистика
Известное	Частичная	Динамическое программирование
Неизвестное	Полная	Максимальное правдоподобие

Таблица 8.2. Преимущества и недостатки различных методов

Метод	Преимущества	Недостатки
Простая статистика	Простые и недорогие вычисления	Требуется полная структура и наблюдаемость, что вряд ли возможно на практике
Динамическое программирование	Известный алгоритм: Витерби	Опасность ловушки локального оптимума
Максимальное правдоподобие	Алгоритм EM/GEM, хорошее поведение	Опасность ловушки локального оптимума

В сетях 6G существует множество потенциальных приложений для DBN. Например, изученные зависимости между сетевыми переменными могут быть использованы для потенциального увеличения производительности в различных беспроводных услугах. В частности, изученная изменяющаяся во времени структура может использоваться для динамической кластеризации пользователей или сети (обычно массивная MIMO без сот) в различных областях (канал, активность и контент) с прямым воздействием на приложения радиодоступа. Ярким примером является изменяющаяся во времени корреляционная структура беспроводного канала, которая, если ее изучить, может быть использована для повышения производительности RAN. Ожидается огромный прирост производительности в отношении массового MIMO, поскольку знание структуры канала (в частности, канала MIMO) предоставит возможность динамически кластеризовать пользователей в RAN на основе условий их канала. Структура канала и кластеризация пользователей облегчают предоставление схем оценки канала с меньшими издержками за счет уменьшения как количества ресурсов канала для обучения, так и количества обратной связи, связанной с оценками канала. Байесовская структура будет чрезвычайно эффективной в этом сценарии,

поскольку она дает убеждения, касающиеся качества задействованных оценок. Эти убеждения можно оценить как локально на стороне пользователя и на границе сети, так и глобально на центральном сервере и макробазовой станции, чтобы облегчить принятие решений относительно оценки канала и декодирования данных. В контексте массовых MIMO, распределенных MIMO или бессотовых сетей этот подход приведет к новым схемам оценки канала и обратной связи. В частности, динамическая кластеризация пользователей будет влиять на структуру и длину пилотных последовательностей, используемых для обучения (как в восходящей, так и в нисходящей линии связи) сотовой системы, в которой базовые станции оснащены большим количеством антенн. Это эффективно снизит накладные расходы на обучение, высвободив ресурсы канала для передачи данных. В частности, структура обратной связи CSI может быть оптимизирована на основе кластеризации пользователей и мягкой информации (убеждений), выводимой процедурой байесовского вывода. Теперь можно исследовать новые предложения, основанные на квантованной явной обратной связи.

Кроме того, байесовская структура объединяет обнаружение активности пользователей, включая модель статистики активации пользователей (которую также можно изучить). Она распространяется на неадаптивные и адаптивные методы вероятностного группового тестирования, чтобы включить зависимости между пользователями и распределения вероятностей, которым они следуют. Это особенно актуально в массовых сценариях МТС/IoT, где вопросы оценки канала и обнаружения активности пользователя должны решаться совместно, чтобы уменьшить как служебные данные, так и время ожидания, связанные с процедурой произвольного доступа. В частности, в приложениях с массовым подключением с двухфазной передачей (начальный доступ с последующей передачей данных) байесовская структура оценивает оптимальное количество ресурсов, выделенных для начального доступа (первая фаза), по сравнению с ресурсами, выделенными для передачи данных (вторая фаза), с немедленным перераспределением ресурсов на этапе передачи данных. В общем, ожидается, что в перегруженных системах, где количество активных пользователей намного превышает количество антенн на базовой станции, планирование пользователей с помощью байесовского вывода может значительно улучшить общую эффективность использования спектра. Это повлияет на структуры как плановых, так и безгрантовых (grant-free, GF) схем, которые нацелены на приложения МТС/IoT с передачей малых пакетов и объединяют начальный доступ и передачу данных.

Байесовский подход также может применяться для локализации и прогнозирования траектории. Кроме того, распределенное прогнозирование состояний каналов и местоположений пользователей рассмат-

ривается как ключевой компонент при разработке надежных систем беспроводной связи. Например, прогнозируя (краткосрочное) местоположение пользователей или эволюцию их углового спектра мощности (т. е. функцию, показывающую среднюю энергию на угловую единицу в антенной решетке), планировщики получают полезную информацию о предварительном кодировании обучающих символов, чтобы, опираясь на один и тот же пилотный сигнал, можно было извлечь информацию о состоянии канала для нескольких пользователей.

DBN могут повлиять на направление развития многих технологий, включая обнаружение активности пользователя, оценку канала и отслеживание мобильности.

8.4. Источники

- [1] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, Vol. 1. MIT Press, Cambridge, 2016.
- [2] K. Hornik, M. Stinchcombe, H. White et al., *Multilayer feedforward networks are universal approximators*. Neural Networks, vol. 2, no. 5, pp. 359–366, 1989.
- [3] J. Duchi, E. Hazan, and Y. Singer, *Adaptive subgradient methods for online learning and stochastic optimization*, Journal of Machine Learning Research, vol. 12, no. 7, 2011.
- [4] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, arXiv preprint arXiv:1412.6980, 2014.
- [5] S. Fernández, A. Graves, and J. Schmidhuber, *An application of recurrent neural networks to discriminative keyword spotting*, in Proc. International Conference on Artificial Neural Networks. Springer, 2007, pp. 220–229.
- [6] D. Madhubabu and A. Thakre, *Long-short term memory based channel prediction for siso system*, in Proc. 2019 International Conference on Communication and Electronics Systems (ICCES). IEEE, 2019, pp. 1–5.
- [7] M. L. Puterman, *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons, 2014.
- [8] D. P. Bertsekas, *Dynamic programming and optimal control*, Vol. 1, no. 2. Athena Scientific, Belmont, MA, 1995.
- [9] N. Tishby, F. C. Pereira, and W. Bialek, *The information bottleneck method*, arXiv preprint physics/0004057, 2000.
- [10] R. Shwartz-Ziv and N. Tishby, *Opening the black box of deep neural networks via information*, arXiv preprint arXiv:1703.00810, 2017.
- [11] A. A. Alemi, I. Fischer, J. V. Dillon, and K. Murphy, *Deep variational information bottleneck*, in Proc. International Conference on Learning Representations, 2016.

- [12] I. Estella-Aguerri and A. Zaidi, *Distributed variational representation learning*, IEEE Transactions on Pattern Analysis and Machine Intelligence, to be published.
- [13] A. E. Eshratifar, A. Esmaili, and M. Pedram, *Bottleneck: A deep learning architecture for intelligent mobile cloud computing services*, in Proc. 2019 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED). IEEE, 2019, pp. 1–6.
- [14] G. E. Hinton and R. S. Zemel, *Autoencoders, minimum description length and Helmholtz free energy*, in Proc. Conference on Advances in Neural Information Processing Systems, 1994, pp. 3–10.
- [15] D. P. Kingma and M. Welling, *An introduction to variational autoencoders*, Foundations and Trends in Machine Learning, vol. 12, no. 4, 2019.
- [16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative adversarial nets*, in Proc. Conference on Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.
- [17] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, *Communication efficient learning of deep networks from decentralized data*, in Proc. Conference on Artificial Intelligence and Statistics. PMLR, 2017, pp. 1273–1282.
- [18] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, *Federated learning: Strategies for improving communication efficiency*, in Proc. NIPS Workshop on Private Multi-Party Machine Learning, 2016.
- [19] B. Hitaj, G. Ateniese, and F. Perez-Cruz, *Deep models under the gan: Information leakage from collaborative deep learning*, in Proc. the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 603–618.
- [20] V. Smith, C.-K. Chiang, M. Sanjabi, and A. S. Talwalkar, *Federated multi-task learning*, in Proc. Conference on Advances in Neural Information Processing Systems, 2017, pp. 4424–4434.
- [21] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT Press, 2018.
- [22] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, Vol. 1 MIT Press, 2016.
- [23] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot et al., *Mastering the game of Go with deep neural networks and tree search*, Nature, vol. 529, no. 7587, pp. 484–489, 2016.
- [24] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton et al., *Mastering the game*

- of Go without human knowledge*, Nature, vol. 550, no. 7676, pp. 354–359, 2017.
- [25] S. Shalev-Shwartz, S. Shammah, and A. Shashua, *Safe, multi-agent, reinforcement learning for autonomous driving*, arXiv preprint arXiv:1610.03295, 2016.
- [26] J. Kober, J. A. Bagnell, and J. Peters, *Reinforcement learning in robotics: A survey*, International Journal of Robotics Research, vol. 32, no. 11, pp. 1238–1274, 2013.
- [27] Y. Li, *Deep reinforcement learning: An overview*, arXiv preprint arXiv:1701.07274, 2017.
- [28] B. Jovanovic and R. W. Rosenthal, *Anonymous sequential games*, Journal of Mathematical Economics, vol. 17, no. 1, pp. 77–87, 1988.
- [29] Y. Shoham and K. Leyton-Brown, *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.
- [30] H. Tembine, R. Tempone, and P. Vilanova, *Mean-field learning: A survey*, arXiv preprint arXiv:1210.4657, 2012.
- [31] R. Sterritt, A. H. Marshall, C. M. Shapcott, and S. I. McClean, *Exploring dynamic Bayesian belief networks for intelligent fault management systems*, in Proc. 2000 Conference on Systems, Man and Cybernetics, vol. 5. IEEE, 2000, pp. 3646–3652.
- [32] K. P. Murphy, *Dynamic Bayesian networks: Representation, inference and learning*, Ph. D. thesis, 2002.

Теоретические основы беспроводных сетей большой емкости

9.1. Теория электромагнитной информации

Это направление исследований связано с формулировкой фундаментальных проблем беспроводной связи и устройства антенн, которые возникают на пересечении хорошо известных областей электромагнитной теории Максвелла и теории информации Шеннона. Междисциплинарное поле, образованное такими теоретическими проблемами волн и информации, и их решения можно описательно назвать *теорией электромагнитной информации* (electromagnetic information theory, EIT).

Работа в этой области, сочетающей волновую физику с теорией информации, имеет долгую историю, восходящую к истокам теории информации [1] и, в частности, к новаторским работам по свету и информации [2, 3]. Для обозначения количества эффективных измерений коммуникации было введено понятие *степеней свободы* (degrees of freedom, DoF). Строгие методики определения числа степеней свободы с использованием волновой теории были разработаны еще в 1960-х годах, и с тех пор дальнейший прогресс в этой области был весьма неравномерным.

Пропускная способность канала определяется как максимальная взаимная информация:

$$C = \max_{p(x)} \{I(x, y)\},$$

где x и y – векторы Tx и Rx соответственно. Чтобы учесть влияние законов электромагнетизма на пропускную способность канала, в [4] определяется пространственная пропускная способность S как максимальная

взаимная информация между векторами T_x и R_x , а также информация о канале (при условии что идеальная информация о состоянии канала известна в R_x):

$$S = \max_{p(x), \mathbf{E}} \{I(x, \{y, \mathbf{G}(\mathbf{E})\})\}.$$

Ограничения:

$$\langle x + x \rangle \leq P_T, \quad \nabla^2 \mathbf{E} - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t^2} = 0, \quad \mathbf{E} = \mathbf{E}_0, \quad \forall \{\mathbf{r}, t\} \in B,$$

где $\mathbf{E}$ – электрическое поле, используемое для передачи данных, B – граничное условие (в зависимости от среды рассеяния), $\mathbf{G}$ – матрица канала. Количество степеней свободы электромагнитного поля необходимо для понимания физических ограничений беспроводных (излучающих) систем связи.



Рис. 9.1. Принципиальная схема обобщенного беспроводного канала

Как показано на рис. 9.1, обобщенный беспроводной канал состоит из массива T_x , канала распространения (среды) и массива R_x . Решетки T_x и R_x влияют, по крайней мере, на диаграмму направленности и взаимное влияние каналов. Сложный обобщенный беспроводной канал $\mathbf{G}(\mathbf{E})$ обычно характеризуется как свертка векторов управления решеткой и многолучевых физических каналов распространения:

$$\mathbf{G}(\mathbf{E}) = \mathbf{H} = f(\{\varphi_n, \theta_n\}) = \sum_n, \quad h_n \mathbf{a}_{\theta_n} \mathbf{a}_{\varphi_n}^H,$$

где n – количество многолучевых распространений, φ_n и θ_n – угол прихода (AoA) и угол выхода (AoD) волны физических каналов распространения соответственно, $\mathbf{a}_\theta$ и $\mathbf{a}_\varphi$ – управляющие векторы передающей и приемной решеток, а h_n – реакция на влияние i -го физического канала с многолучевым распространением.

После успешного развития связи ММО в 2000-х годах интерес к ЕИТ неуклонно рос и привел к появлению концепции канала ММО, не зависящего от конфигурации антенны [5–8]. В любом случае оптимальное количество антенн и пропускная способность ММО ограничены для данного размера апертуры, независимо от сценария (например, богатое или разреженное рассеяние) [4]. В настоящее время массивные ММО-системы – это зрелая технология, ключевые компоненты которой включены в стандарт 5G NR. Преимущества массивных ММО с точки зрения эффективности использования спектра и энергоэффективности хорошо поняты и признаны, и возникают естественные вопросы: «Возможно ли приблизиться к пределам бесконечного количества антенн?» и «Чего нам ждать дальше?» [9]. Ответы можно найти в ЕИТ, поскольку данная теория обеспечивает общую основу для проектирования беспроводной сети и анализа производительности. Первоначальные исследования в этом направлении были проведены в [10–12].

Потенциальные технические исследования, направленные на увеличение количества степеней свободы обобщенного беспроводного канала $\mathbf{G}(\mathbf{E})$, можно разделить на три категории: (1) физический канал распространения, (2) диаграмма направленности антенной решетки и взаимное влияние, (3) электромагнитные физические характеристики.

9.1.1. Физический канал распространения

От неконтролируемой среды физического распространения к контролируемой

Соответствующие дисциплины включают *реконфигурируемые интеллектуальные поверхности* (reconfigurable intelligent surfaces, RIS) [13] и *интеллектуальные отражающие поверхности* (intelligent reflecting surfaces, IRS) [14], и в результате этих исследований все большее внимание привлекает концепция *интеллектуальной среды*. Ее главное преимущество – возможность активного построения полноранговой матрицы каналов. На данный момент, с учетом ряда ограничений, которые включают разрешение оборудования, накладные расходы обратной связи и мобильность клиентских устройств, теоретические пределы пропускной способности интеллектуальных сред неизвестны. ЕИТ может сыграть ключевую роль в соответствующих исследованиях.

От необратимого поведения траектории распространения электромагнитного излучения к обратимому

Обращение времени (TR) [15] является типичной дисциплиной в этой области; оно зависит от временной симметрии решения уравнения электромагнитной волны (также известного как принцип обратимости оптического пути). Ожидается, что при использовании принципа обращения времени теория электромагнитной информации будет играть ключевую роль в достижении сверхвысокого разрешения пространственного и временного измерений в сверхширокополосных (ultra-wideband, UWB) системах.

От электромагнитного дальнего поля до ближнего поля и даже поверхностных волн

Соответствующие дисциплины включают *большую интеллектуальную поверхность* (large intelligent surface, LIS) [16] и *связь на поверхностных волнах* (surface wave communication, SWC) [17]. LIS должна учитывать пространственную нестационарность сферических волн и алгоритм выравнивания¹⁷ приемника, в то время как SWC использует преимущества давно игнорируемых *волн Ценнека*, которые плотно связаны около поверхности среды и не могут распространяться на окружающую среду. И LIS, и SWC нуждаются в математическом инструменте EIT для разработки конструкции антенны и обработки микронаноструктур.

9.1.2. Диаграммы направленности антенных решеток и эффект взаимного влияния

От инвариантной (физической или виртуальной) геометрической топологии антенных решеток к вариативной

Помимо обычных реконфигурируемых антенн, в последнее время интерес представляют *информационные метаповерхности* [18]. Электромагнитная модуляция в точных пространственных измерениях может быть реализована на основе гибкого управления метаматериалами с помощью индексной модуляции [19]. Хотя не считается, что чистая индексная модуляция может улучшить NDF каналов, общепризнано, что дальнейшее развитие EIT может открыть нам множество других технических идей, которые могут быть реализованы на основе метаматериалов.

От полуволновых решеток к сверхкомпактным

Наиболее очевидным предметом изучения в этой области является сверхнаправленность [20]. В отличие от традиционной антенной решетки, которая сводит к минимуму взаимную связь M элементов для поддер-

¹⁷ Выравнивание сигнала во временной и частотной областях на выходе многолучевого канала.

жания усиления формирования луча на уровне M за счет максимального коэффициента передачи, сверхнаправленная антенная решетка создает сильную взаимную связь за счет разнесения антенн менее чем на половину длины волны, что приводит к усилению формирования луча больше M . С математической точки зрения подпространство матрицы импеданса включает собственные векторы с небольшими собственными значениями, представляющие часть векторов распространения. С физической точки зрения спектр плоских волн имеет большое количество суперволновых компонентов, и волны исчезают в конечном направлении. Исчезающие волны затухают по экспоненте. ЕИТ может решить проблему сильной взаимной связи с точки зрения теории электромагнетизма Максвелла и теории цепей Кирхгофа [20, 21].

От малых апертурных решеток к сверхбольшим

Типичные направления исследований здесь включают большие интеллектуальные поверхности [16] и *сверхбольшие апертурные решетки* (extremely large aperture arrays, ELAA) [22], последние из которых формируют нестационарные пространственные поля вдоль решеток. Чтобы уменьшить вычислительную нагрузку схемы ELAA–MIMO, рекомендуется разделить обработку сигналов полной решетки на более мелкие непересекающиеся подрешетки [22]. Ожидается, что теория электромагнитной информации предоставит новую основу для анализа, позволяющую лучше рассмотреть свойства нестационарного пространственного поля, возникающего по разным причинам.

9.1.3. Электромагнитные физические характеристики

От одинарной (или двойной) к тройной (или мульти-) поляризации

Хотя некоторые исследователи проанализировали эффект усиления DoF в элементах с тройной поляризацией в начале 2011 г. [23], как результат теоретических выкладок он был известен еще раньше, но использование трехмерных (3D) антенных решеток с тройной поляризацией по ряду причин не получило распространения. Благодаря увеличению пропускной способности MIMO на базе *четырёхзаходной спиральной антенны* (quadrifilar helix antenna, QHA) трехмерные антенные решетки и связанные с ними измерения многополяризованных каналов теперь привлекают повышенное внимание [24]. В эпоху 6G измерения и моделирование мультиполяризованных каналов – это вопросы, которые волнуют исследователей, и можно сосредоточить внимание на некоторых ранее игнорировавшихся характеристиках каналов, чтобы увеличить DoF. Теория электромагнитной информации будет играть важную роль в обобщенном моделировании каналов.

От линейного момента (энергии) к угловому

Типичным предметом работ в этой области является *орбитальный угловой момент* [25]. Дальнейшие исследования показали, что орбитальный угловой момент вряд ли станет новой независимой степенью свободы (DoF), а вместо этого является лишь частным случаем ММО [26]. Однако в некоторых квазипараксиальных средах и средах ближнего поля орбитальный угловой момент может дополнять ММО для достижения некоторых полезных эффектов. Остается без ответа вопрос о том, можно ли под руководством ЕИТ найти новые электромагнитные характеристики для беспроводной связи.

Традиционный анализ степени свободы электромагнитных волн в пространстве основан на том, является ли свободное пространство безграничным (или имеет бесконечно большую границу). Сложный волновой фронт электромагнитной волны сначала разлагается на члены мультипольного разложения со стандартными сферическими гармониками, а затем получается матрица отклика беспроводного канала посредством линейной суперпозиции функции Грина, распространяемой электромагнитной волной. Наконец, определяется независимый ортогональный базис в собственном подпространстве [27, 28]. В эпоху 6G новые тенденции в области антенных решеток и канальных сред побуждают нас находить более мощные инструменты анализа электромагнитных явлений, чтобы справиться со сложностью электромагнитной среды, порождаемой сверхнаправленными антенными решетками, ELAA и интеллектуальными метаповерхностями.

Вычислительный электромагнетизм

Вычислительный электромагнетизм (computational electromagnetics, CEM) – это новая смежная область, в которой ищут решения сложных задач теории электромагнетизма и инженерии [29]. Она охватывает все области электромагнетизма и взаимосвязана с теорией и практикой использования электромагнитного поля. Для разработки электромагнитных полей CEM используется для решения задач моделирования, оптимизации и проектирования электромагнитных полей со все более сложными фактическими сценариями. Для теории электромагнитного поля CEM может предоставить методы, средства и результаты для сложных численных и аналитических расчетов, а также исследовать законы и математические уравнения электромагнитных полей. Общие методы CEM включают метод конечных элементов [30], метод конечных разностей во временной области [31], матрицы линий передачи [32], метод моментов [33] и различные методы ускорения вычислений. Необходимо обратить внимание на развитие теории CEM и инструментов моделирования для автоматизированного проектирования, чтобы улучшить

наше понимание и применение функции Максвелла Грина и матрицы отклика беспроводного канала в сложных электромагнитных средах 6G. Как мы обсудим далее в главе 13, СЕМ требуется при моделировании канала 6G для новых полос спектра и сценариев использования.

По мере развития 6G начинают появляться различные беспроводные технологии, которые могут повлиять на развитие теории электромагнитной информации. К ним относятся: LIS, RIS, цифровой управляемый рассеиватель, орбитальный угловой момент, голографические MIMO, обращение времени, связь поверхностных волн, сверхмассивные MIMO и ELAA.

9.2. Теория крупномасштабной коммуникации

Во многих современных инженерных областях системы имеют тенденцию организовываться во все более крупные сети (такие как сверхмассивные MIMO и плотные сети). Кроме того, все более важной становится системная динамика, поскольку теперь требуется быстрая обработка в быстро меняющихся средах. Выступая своего рода общим знаменателем, такие сети теперь сталкиваются с проблемой, связанной с необходимостью одновременно справляться с растущим размером системы (больше узлов в сетях и больше субъектов, взаимодействующих с ними), повышенным уровнем стохастичности (случайные изменения в топологии и случайное развитие среды), а также с потребностью в децентрализованной и самоорганизованной обработке (сетевые узлы должны быть автономными и способны динамически адаптироваться к изменяющейся среде). Эти три аспекта представляют собой основные компоненты, необходимые для моделирования, анализа и оптимизации производительности будущих сетей. Традиционные методы, используемые для решения подобных проблем в больших сложных системах, включают прямое моделирование (метод Монте-Карло) или эвристические инструменты, такие как нейронные сети или рои частиц. Однако они имеют ряд серьезных недостатков, например они не являются математически стабильными и плохо поддаются простой интерпретации, в результате чего направленные улучшения и оптимизации системы либо основаны на пробах и ошибках, либо просто невозможны. В результате эти методы часто рассматриваются как резервные решения в тех случаях, когда математические модели не имеют прочной основы.

Следовательно, требуется разработка новых теоретических инструментов, позволяющих проводить систематический, надежный и интерпретируемый анализ больших случайных сетей. Эти инструменты должны быть достаточно общими, чтобы охватывать широкий спектр сетевых моделей, но в то же время обеспечивать легкую специализацию для практических сетей. Как будет более подробно показано ниже, для

решения части проблем, присущих этим большим системам, уже были определены различные современные математические инструменты в качестве потенциальных кандидатов. Однако большинство этих инструментов все еще находятся на ранних стадиях математической проработки или только начали внедряться в инженерные изыскания. Из этих новых инструментов наиболее многообещающими оказались следующие три:

- **теория случайных матриц**, в частности теория случайных матриц большой размерности [34], технически представляет собой исследование крупномерных эрмитовых матриц со случайными элементами. В начале 2000-х годов в этой области наблюдался всплеск интереса к беспроводной связи в ходе анализа новых технологий той эпохи – множественных антенн и методов расширения спектра, математические модели которых были основаны на случайных матричных каналах связи. В то время как научный прогресс в этой области продолжался с 2000 по 2010 год (на основе математических результатов, полученных к 2000 году), мало кто из исследователей развивал математические основы теории случайных матриц дальше, чем это было необходимо для решения сложных проблем. Однако совсем недавно в контексте массового ММО и сверхмассивного ММО появились новые теоретические инструменты, связанные с анализом сетей беспроводной связи. Теория случайных матриц направлена на изучение спектральных свойств (собственных значений и собственных подпространств) матриц со случайными элементами, размеры которых считаются очень большими. Подобно тому, как случайные переменные или случайные векторы конечного размера демонстрируют детерминированные пределы при достижении большого набора выборок этих переменных (в основном на основе закона больших чисел или центральной предельной теоремы), спектральные меры некоторых моделей случайных матриц демонстрируют детерминированное поведение, когда размеры строки и столбца растут до бесконечности. Долгое время эта теория оставалась прерогативой математиков, однако в начале 2000-х стало очевидным, что случайные матрицы могут адекватно моделировать каналы беспроводной связи, которые являются стохастическими и быстро меняются по своей природе [35]. После нескольких лет работы исследователям беспроводной связи стало ясно, что большинство моделей случайных матриц, предложенных математиками, либо неадекватны, либо слишком просты для эффективного изучения наиболее реалистичных каналов связи. Вторая волна исследований началась примерно в 2010 году и преследовала цель дальней-

шего изучения математического инструмента теории случайных матриц для конкретных потребностей инженеров беспроводной связи. Последние достижения в этой области принесли важные новые результаты в моделировании, анализе и оптимизации производительности многочисленных беспроводных каналов, таких как каналы ММО точка-точка с райсовским замиранием [36], многолучевые каналы ММО [37], каналы ММО множественного доступа [38], линейно предкодированные широкополосные каналы [39], каналы связи с унитарными прекодерами [40], многозвенные каналы [41] и сети с одной/несколькими ячейками [42, 43]. Хотя эти результаты иногда основаны на очень сложных системных моделях, они представлены в относительно упрощенной форме и хорошо поддаются оптимизации. Существенным ингредиентом, вводимым теорией случайных матриц, является приближение большой размерности, которое позволяет эффективно анализировать большие сложные стохастические системы с детерминированными величинами, известными как детерминированные эквиваленты [44]. Компоненты крупноразмерной системы измеряются на основе количества антенн в беспроводных устройствах, количества пользователей в одной ячейке или количества ячеек в данной сети. Таким образом, теория случайных матриц признана эффективным инструментом для анализа многих локальных систем беспроводной связи, где «локальный» понимается как «одномасштабный» (велико либо количество антенн, либо количество пользователей). Ожидается, что благодаря дополнительным обобщенным инструментам, таким как итерационные детерминированные эквиваленты [40], теория случайных матриц станет адекватным инструментом для моделирования сложных телекоммуникационных систем, в которых учитываются многие крупноразмерные компоненты, что позволяет представить себе совместный анализ на макроскопическом и микроскопическом уровнях. Кроме того, инструменты теории случайных матриц также произвели впечатление на сообщество обработки сигналов, предложив инновационные схемы детектирования и оценки (статистического вывода), такие как радарная технология ММО, для обработки больших массивов сигналов [45]. Наконец, были проведены исследования по разработке децентрализованных алгоритмов для оптимизации производительности многоячеечных сетей ММО [46];

- **децентрализованная стохастическая оптимизация.** Второй интересный инструмент – это недавно предложенная структура, объединяющая стохастические аппроксимации и алгоритмы

сплетен (gossip algorithm) для децентрализованной стохастической оптимизации. Происхождение децентрализованной стохастической оптимизации можно проследить до работы [47], в которой получены методы выполнения децентрализованных сложных вычислений на потенциально больших кластерах процессоров. Однако первоначальная работа была ограничена тем фактом, что межпроцессорная связь должна быть недорогой и независимой от расстояния между процессорами и, следовательно, независимой от базовой топологии сети. Параллельно с этим были разработаны децентрализованные консенсусные алгоритмы для решения совершенно другой проблемы: достижения общего решения [48] на взаимосвязанном графе датчиков с низкими возможностями обработки и памяти. Классическим примером децентрализованных алгоритмов консенсуса является алгоритм достижения консенсуса в распределенной системе ММО. То есть, исходя из соображений, что каждая точка доступа в кластере выполняет начальное измерение, постоянно сохраняет одно значение в памяти и может отправлять это значение только своему ближайшему соседу, должны быть разработаны быстрые алгоритмы, чтобы помочь кластеру в конечном итоге достичь общего решения по среднему значению исходных измерений.

В данном случае топологические соображения важны для свойств выполнимости и сходимости алгоритмов. В работе [49] предложена децентрализованная обработка в больших взаимосвязанных сетях с ограниченными коммуникационными возможностями. Этот подход направлен на решение общесетевой проблемы, которую можно разделить на несколько независимых подзадач. Каждый сетевой узел оптимизирует свою собственную подзадачу, а затем отправляет результат своему ближайшему соседу (детерминированно или случайным образом), таким образом достигается консенсус, в котором каждый узел в конечном итоге знает о решении проблемы. Среди различных расширений этих схем предлагаются некоторые алгоритмы, основанные на стохастической аппроксимации [50], которая предполагает, что каждый сетевой узел обладает лишь частичным знанием информации о подзадаче. Децентрализованная обработка уже была предметом практических исследований в многопользовательских сетях [51]. В децентрализованной ММО различные проблемы, такие как максимизация общей пропускной способности сети, требуют решения, в котором центральный объект имеет полное представление о сети. Из-за ограниченных возможностей вычислений и памяти, а также коротких диапазонов связи точек доступа де-

централизованная обработка и алгоритмы сплетен также являются важными ингредиентами для достижения максимальной производительности плотных сетей. В частности, ожидается, что такие децентрализованные алгоритмы обработки и сплетен станут адекватной альтернативой теоретико-игровым и обучающим подходам, которые давно рассматриваются в беспроводной связи, несмотря на то что исследователи редко приводят убедительные аргументы в пользу их использования в самоорганизованных сетях. Следует отметить, что теоретико-игровые соображения, в первую очередь основанные на последовательности испытаний и обратной связи, часто требуют от сетевых участников принятия неправильных решений до того, как будет достигнуто ожидаемое равновесие по Нэшу. Если решаемые задачи оптимизации ставят строгие ограничения между узлами, теоретико-игровые подходы часто приводят к алгоритмам, результаты которых не могут изначально удовлетворять ограничениям. Это не относится к децентрализованной обработке и методам сплетен, когда информация, которой обмениваются узлы, не используется до объединения;

- **тензорная алгебра и разложение тензора низкого ранга:** в простейшем воплощении тензор может быть истолкован как D -мерная структура данных. Это похоже на обобщение векторов и матриц до более чем двух измерений (в классической линейной алгебре векторы являются одномерными структурами данных, тогда как матрицы имеют два измерения, а именно строки и столбцы). Результирующий D -мерный массив можно интерпретировать либо как представление полилинейного приложения (как в ньютоновской физике, где бескоординатные тензорные представления используются для моделирования физических законов), либо как структуру данных, которая естественным образом индексируется D -измерениями (или модами) [52]. В последнем случае понятие тензорного ранга играет решающую роль. Тензор ранга 1 порядка D определяется как внешнее произведение векторов D соответствующих размерностей. Это аналогично матрице ранга 1, которая определяется внешним произведением только двух векторов: вектора-столбца и вектора-строки. Минимальное количество членов ранга 1, необходимое для восстановления данного тензора, известно как ранг.

Проблема тензорной декомпозиции была впервые выявлена в конце 1920-х годов [53] и глубоко укоренилась в экспериментальной науке. В частности, разложение заданного тензора на сумму компонентов ранга 1 (известное как каноническая полиадическая декомпозиция) имеет практическое значение во многих

приложениях, поскольку оно раскрывает внутреннюю структуру данных. В 1970-х годах тензорная декомпозиция низкого ранга была применена к области психометрии, где она получила название PARAFAC [54]. Позднее она была распространена на область хемометрики в 1980-х годах [55].

В основе тензорной алгебры лежит глубокая математическая теория. Несмотря на относительно знакомую схему, обусловленную аналогией с матричным случаем, свойства канонической полиадической декомпозиции во многом отклоняются от интуитивного представления, как в случае матриц [56]. В частности, тензоры высокого порядка ($D > 3$) могут иметь высокий ранг даже при умеренных размерах тензора. В отличие от матриц, где ранг ограничен минимумом по размерам строки и столбца, ожидаемый ранг общего тензора масштабируется намного быстрее [57]. Кроме того, для тензоров подобного ранга (т. е. тех, которые содержат данные, полученные из структурированной модели) каноническая полиадическая декомпозиция по существу уникальна при мягких условиях. Опять же, это контрастирует с матрицами. Например, если мы принимаем во внимание QR или разложение по сингулярным числам, условия ортогональности, налагаемые на факторы, которые делают разложение матриц уникальным, являются чисто техническими и не всегда соответствуют значимому ограничению в рассматриваемой задаче. Эти уникальные свойства разложения тензора низкого ранга можно использовать в беспроводной связи несколькими способами. Например, если тензорная модель низкого ранга возникает из модели зеркального распространения радиочастоты в сочетании с антенной решеткой, тензорное разложение позволяет нам выполнять слепое разделение источников и оценку направления с более мягкими допущениями и большим количеством источников по сравнению с классической множественной классификацией сигналов (multiple signal classification, MUSIC) или оценкой параметров сигнала с помощью методов инвариантности вращения [58]. Такие смягченные предположения могут включать использование гораздо более коротких выборок данных, основанных на статистике высокого порядка. Возможность слепого разделения источников также лежит в основе подхода некогерентного, неортогонального многопользовательского разделения, предложенного в [59] в контексте массового доступа, который будет обсуждаться в главе 24.

Современные расширения тензорной теории нацелены на решение все более сложных задач, таких как связанная каноническая полиадическая декомпозиция [60] (где связь возникает из рассма-

триваемого приложения) и разложение на последовательность тензорных цепочек с эффективным хранением [61], применимое к большим размерам, с большим объемом памяти. Недавние теоретические разработки также включают в себя изучение случайных тензоров, в частности с акцентом на модели с пиками. В этом контексте цель состоит в том, чтобы аналитически охарактеризовать условия, при которых информативный компонент низкого ранга может быть надежно отделен от аддитивного шума измерения с помощью приближения тензора низкого ранга, и спрогнозировать достижимую точность. Один из подходов – использовать инструменты статистической физики [62]. Предпринимаются и другие попытки разработать спектральную теорию случайных тензоров с чистого листа – печально известная трудная проблема [63]. Кроме того, возможность обобщения тензорного разложения низкого ранга изучается в контексте формул «недостающих данных» [64].

По мере того как 6G продолжает развиваться, начинают появляться различные технологии, которые тесно связаны с теорией крупномасштабных коммуникаций, в том числе плотная сеть, федеративное обучение, бессотовый массивный MIMO, ориентированная на пользователя связь, граничная связь и распределенный MIMO.

9.3. Источники

- [1] C. E. Shannon, *A mathematical theory of communication*, ACM SIGMOBILE Mobile Computing and Communications Review, vol. 5, no. 1, pp. 3–55, 2001.
- [2] D. Gabor, *CIII. Communication theory and physics*, London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, vol. 41, no. 322, pp. 1161–1187, 1950.
- [3] D. Gabor, *IV light and information*, in Progress in optics, Vol. 1. Elsevier, 1961, pp. 109–153.
- [4] S. Loyka, *Information theory and electromagnetism: Are they related?* in Proc. 2004 10th International Symposium on Antenna Technology and Applied Electromagnetics and URSI Conference. IEEE, 2004, pp. 1–5.
- [5] J. Y. Hui, C. Bi, and H. Sun, *Spatial communication capacity based on electromagnetic wave equations*, in Proc. 2001 IEEE International Symposium on Information Theory. IEEE, 2001, p. 337.
- [6] J. W. Wallace and M. A. Jensen, «*Intrinsic capacity of the MIMO wireless channel*», in Proc. IEEE 56th Vehicular Technology Conference, vol. 2. IEEE, 2002, pp. 701–705.

- [7] M. A. Jensen and J. W. Wallace, *Capacity of the continuous-space electromagnetic channel*, IEEE Transactions on Antennas and Propagation, vol. 56, no. 2, pp. 524–531, 2008.
- [8] F. K. Gruber and E. A. Marengo, *New aspects of electromagnetic information theory for wireless and antenna systems*, IEEE Transactions on Antennas and Propagation, vol. 56, no. 11, pp. 3470–3484, 2008.
- [9] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, *Massive MIMO is a reality. What is next? Five promising research directions for antenna arrays*, Digital Signal Processing, vol. 94, pp. 3–20, 2019.
- [10] T. L. Marzetta, *Spatially-stationary propagating random field model for massive MIMO small-scale fading*, in Proc. 2018 IEEE International Symposium on Information Theory (ISIT). IEEE, 2018, pp. 391–395.
- [11] A. Pizzo, T. L. Marzetta, and L. Sanguinetti, *Spatial characterization of holographic MIMO channels*, arXiv preprint arXiv:1911.04853, 2019.
- [12] A. Pizzo, T. L. Marzetta, and L. Sanguinetti, *Degrees of freedom of holographic MIMO channels*, in Proc. 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2020, pp. 1–5.
- [13] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, *Wireless communications through reconfigurable intelligent surfaces*, IEEE Access, vol. 7, pp. 116 753–116 773, 2019.
- [14] Q. Wu and R. Zhang, *Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming*, IEEE Transactions on Wireless Communications, vol. 18, no. 11, pp. 5394–5409, 2019.
- [15] I. H. Naqvi, G. El Zein, G. Lerosey, J. de Rosny, P. Besnier, A. Tourin, and M. Fink, *Experimental validation of time reversal ultra wide-band communication system for high data rates*, IET Microwaves, Antennas & Propagation, vol. 4, no. 5, pp. 643–650, 2010.
- [16] S. Hu, F. Rusek, and O. Edfors, *Beyond massive MIMO: The potential of data transmission with large intelligent surfaces*, IEEE Transactions on Signal Processing, vol. 66, no. 10, pp. 2746–2758, 2018.
- [17] K. K. Wong, K. F. Tong, Z. Chu, and Y. Zhang, *A vision to smart radio environment: Surface wave communication superhighways*, arXiv preprint arXiv: 2005.14082, 2020.
- [18] H. Wu, G. D. Bai, S. Liu, L. Li, X. Wan, Q. Cheng, and T. J. Cui, *Information theory of metasurfaces*, National Science Review, vol. 7, no. 3, pp. 561–571, 2020.
- [19] J. A. Hodge, K. V. Mishra, and A. I. Zaghloul, *Reconfigurable metasurfaces for index modulation in 5G wireless communications*, in Proc. 2019 Inter-

- national Applied Computational Electromagnetics Society Symposium (ACES). IEEE, 2019, pp. 1–2.
- [20] T. L. Marzetta, *Super-directive antenna arrays: Fundamentals and new perspectives*, in Proc. 2019 53rd Asilomar Conference on Signals, Systems, and Computers. IEEE, 2019, pp. 1–4.
 - [21] M. T. Ivrlač and J. A. Nossek, *The multiport communication theory*, IEEE Circuits and Systems Magazine, vol. 14, no. 3, pp. 27–44, 2014.
 - [22] A. Amiri, M. Angjelichinoski, E. De Carvalho, and R. W. Heath, *Extremely large aperture massive MIMO: Low complexity receiver architectures*, in Proc. 2018 IEEE Globecom Workshops. IEEE, 2018, pp. 1–6.
 - [23] A. S. Poon and N. David, *Degree-of-freedom gain from using polarimetric antenna elements*, IEEE Transactions on Information Theory, vol. 57, no. 9, pp. 5695–5709, 2011.
 - [24] B. Yang, P. Zhang, H. Wang, and W. Hong, *Electromagnetic vector antenna arraybased multi-dimensional parameter estimation for radio propagation measurement*, IEEE Wireless Communications Letters, vol. 8, no. 6, pp. 1608–1611, 2019.
 - [25] B. Thidé, H. Then, J. Sjöholm, K. Palmer, J. Bergman, T. Carozzi, Y. N. Istomin, N. Ibragimov, and R. Khamitova, *Utilization of photon orbital angular momentum in the low-frequency radio domain*, Physical Review Letters, vol. 99, no. 8, p. 087701, 2007.
 - [26] O. Edfors and A. J. Johansson, *Is orbital angular momentum (OAM) based radio communication an unexploited area?* IEEE Transactions on Antennas and Propagation, vol. 60, no. 2, pp. 1126–1131, 2011.
 - [27] M. R. Andrews, P. P. Mitra, and R. DeCarvalho, *Tripling the capacity of wireless communications using electromagnetic polarization*, Nature, vol. 409, no. 6818, pp. 316–318, 2001.
 - [28] A. S. Poon and N. David, *Degree-of-freedom gain from using polarimetric antenna elements*, IEEE Transactions on Information Theory, vol. 57, no. 9, pp. 5695–5709, 2011.
 - [29] D. B. Davidson, *Computational electromagnetics for RF and microwave engineering*. Cambridge University Press, 2010.
 - [30] O. C. Zienkiewicz and P. Morice, *The finite element method in engineering science*. McGraw-Hill, London, 1971.
 - [31] Y. Liu, R. Mittra, T. Su, X. Yang, and W. Yu, *Parallel finite-difference time-domain method*. Artech, 2006.
 - [32] C. Christopoulos, *The transmission-line modeling (TLM) method in electromagnetics*, Synthesis Lectures on Computational Electromagnetics, vol. 1, no. 1, pp. 1–132, 2005.

- [33] W. C. Gibson, *The method of moments in electromagnetics*. CRC Press, 2014.
- [34] Z. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*. Springer, 2010.
- [35] D. N. C. Tse and S. V. Hanly, *Linear multiuser receivers: Effective interference, effective bandwidth and user capacity*, IEEE Transactions on Information Theory, vol. 45, no. 2, pp. 641–657, 1999.
- [36] W. Hachem, P. Loubaton, J. Najim et al., *Deterministic equivalents for certain functionals of large random matrices*, Annals of Applied Probability, vol. 17, no. 3, pp. 875–930, 2007.
- [37] F. Dupuy and P. Loubaton, *On the capacity achieving covariance matrix for frequency selective MIMO channels using the asymptotic approach*, IEEE Transactions on Information Theory, vol. 57, no. 9, pp. 5737–5753, 2011.
- [38] R. Couillet, M. Debbah, and J. W. Silverstein, *A deterministic equivalent for the analysis of correlated MIMO multiple access channels*, IEEE Transactions on Information Theory, vol. 57, no. 6, pp. 3493–3514, 2011.
- [39] S. Wagner, R. Couillet, and M. Debbah, *Large system analysis of linear precoding in MISO broadcast channels with limited feedback*, arXiv preprint arXiv:0906.3682, 2009.
- [40] J. Hoydis, R. Couillet, and M. Debbah, *Deterministic equivalents for the performance analysis of isometric random precoded systems*, in Proc. 2011 IEEE International Conference on Communications (ICC). IEEE, 2011, pp. 1–5.
- [41] N. Fawaz, K. Zarifi, M. Debbah, and D. Gesbert, *Asymptotic capacity and optimal precoding in MIMO multi-hop relay networks*, IEEE Transactions on Information Theory, vol. 57, no. 4, pp. 2050–2069, 2011.
- [42] J. Hoydis, M. Kobayashi, and M. Debbah, *Optimal channel training in uplink network MIMO systems*, IEEE Transactions on Signal Processing, vol. 59, no. 6, pp. 2824–2833, 2011.
- [43] H. Sifaou, A. Kammoun, L. Sanguinetti, M. Debbah, and M.-S. Alouini, *Max–min SINR in large-scale single-cell MU–MIMO: Asymptotic analysis and low-complexity transceivers*, IEEE Transactions on Signal Processing, vol. 65, no. 7, pp. 1841–1854, 2016.
- [44] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*. Cambridge University Press, 2011.
- [45] X. Mestre and M. Á. Lagunas, *Modified subspace algorithms for DoA estimation with large arrays*, IEEE Transactions on Signal Processing, vol. 56, no. 2, pp. 598–614, 2008.

- [46] H. Asgharimoghaddam, A. Tölili, L. Sanguinetti, and M. Debbah, *Decentralizing multicell beamforming via deterministic equivalents*, IEEE Transactions on Communications, vol. 67, no. 3, pp. 1894–1909, 2018.
- [47] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: numerical methods*. Prentice Hall, 1989.
- [48] F. Benaych-Georges and R. R. Nadakuditi, *The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices*, Advances in Mathematics, vol. 227, no. 1, pp. 494–521, 2011.
- [49] S. S. Ram, A. Nedić, and V. V. Veeravalli, *Distributed stochastic subgradient projection algorithms for convex optimization*, Journal of Optimization Theory and Applications, vol. 147, no. 3, pp. 516–545, 2010.
- [50] C. Kubrusly and J. Gravier, *Stochastic approximation algorithms and applications*, in Proc. 1973 IEEE Conference on Decision and Control Including the 12th Symposium on Adaptive Processes. IEEE, 1973, pp. 763–766.
- [51] J. J. P. Bianchi, *Distributed stochastic approximation for constrained and unconstrained optimization*, arXiv preprint arXiv: 1104.2773, 2011.
- [52] P. Comon, *Tensors: A brief introduction*, IEEE Signal Processing Magazine, vol. 31, no. 3, pp. 44–53, 2014.
- [53] F. L. Hitchcock, *The expression of a tensor or a polyadic as a sum of products*, Journal of Mathematics and Physics, vol. 6, no. 1–4, pp. 164–189, 1927.
- [54] R. A. Harshman et al., *Foundations of the PARAFAC procedure: Models and conditions for an ‘explanatory’ multimodal factor analysis*, UCLA Working Papers in Phonetics, vol. 16, pp. 1–84, Dec. 1970.
- [55] C. J. Appellof and E. R. Davidson, *Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents*, Analytical Chemistry, vol. 53, no. 13, pp. 2053–2056, 1981.
- [56] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Review, vol. 51, no. 3, pp. 455–500, 2009.
- [57] L. Chiantini, G. Ottaviani, and N. Vannieuwenhoven, *An algorithm for generic and lowrank specific identifiability of complex tensors*, SIAM Journal on Matrix Analysis and Applications, vol. 35, no. 4, pp. 1265–1287, 2014.
- [58] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, *Parallel factor analysis in sensor array processing*, IEEE Transactions on Signal Processing, vol. 48, no. 8, pp. 2377–2388, 2000.
- [59] A. Decurninge, I. Land, and M. Guillaud, *Tensor-based modulation for unsourced massive random access*, arXiv preprint arXiv:2006.06797, 2020.

- [60] M. Sørensen, I. Domanov, and L. De Lathauwer, *Coupled canonical polyadic decompositions and multiple shift invariance in array processing*, IEEE Transactions on Signal Processing, vol. 66, no. 14, pp. 3665–3680, 2018.
- [61] I. V. Oseledets, *Tensor-train decomposition*, SIAM Journal on Scientific Computing, vol. 33, no. 5, pp. 2295–2317, 2011.
- [62] T. Lesieur, L. Miolane, M. Lelarge, F. Krzakala, and L. Zdeborová, *Statistical and computational phase transitions in spiked tensor estimation*, in Proc. 2017 IEEE International Symposium on Information Theory (ISIT). IEEE, 2017, pp. 511–515.
- [63] L. Qi and Z. Luo, *Tensor analysis: spectral theory and special tensors*. SIAM, 2017.
- [64] M. Nickel and V. Tresp, *An analysis of tensor models for learning on structured data*, in Proc. Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, 2013, pp. 272–287.

Глава 10

Теоретические основы беспроводных сетей большой емкости

10.1. Теория семантической коммуникации

Слово «семантика» (значение) происходит от языков (естественных или формальных) и от концепции композиционности, согласно которой значение предложения определяется тремя составляющими:

- правило составления предложений (синтаксис);
- значение (семантика) каждого компонента;
- контекст.

Чтобы дать формальное определение семантической коммуникации, было предложено множество понятий семантики, но до сих пор ни одно из них не было удовлетворительным. Отсутствие общепринятого математического определения семантики сильно ограничивает дальнейшее развитие в этой области.

Однако такая теория семантической информации становится необходимой, когда мы ожидаем, что форма коммуникации не предполагает вмешательства человека. В традиционной теории информации мы не рассматриваем *значение* того, что мы передаем, поскольку всегда будет «человеческий мозг», чтобы интерпретировать это. Изображения, видео, текст, речь, аудиозаписи – вся эта информация интерпретируется нашим мозгом. Если мы слышим лай собаки, мы понимаем, что «есть СОБАКА, и она ЛАЕТ», потому что наш мозг способен интерпретировать звуки. Когда интеллектуальным машинам необходимо общаться без какого-либо вмешательства человека, они отправляют информацию, закодированную на их собственном внутреннем языке (как люди, использующие естественный язык).

Важным новым вариантом использования 6G может быть связь между интеллектуальными машинами. Однако ожидается, что 6G не появится в ближайшие 10 лет, и к этому времени машины приобретут более продвинутые возможности, особенно с точки зрения композиционности.

Семантика естественного языка является *порождающей*. Компетентные пользователи естественного языка могут понимать и генерировать (порождать) неопределенное количество различных предложений, даже если словарный запас языка ограничен. Применяя конечное число синтаксических правил к ограниченному словарю, мы можем генерировать и понимать почти бесконечное количество значимых предложений. Эту замечательную способность мы имеем в виду, когда говорим, что человеческий естественный язык является порождающим. Семантика естественного языка также является композиционной, поскольку значимые выражения строятся из других значимых выражений. Мы можем проанализировать значение такого предложения, как «Боб – учитель и скрипач», отметив, что выражение содержит два составляющих предложения: «Боб – учитель» и «Боб – скрипач». Точно так же значение сложного предложения, такого как «Если Джон покупает билеты, то либо Бен, либо Мэри заплатят за попкорн», частично определяется значениями составных предложений «Джон покупает билеты», «Бен заплатит за попкорн» и «Мэри заплатит за попкорн». В лингвистике *принцип композиционности* гласит, что значение сложного выражения определяется его структурой и значениями составляющих его частей.

Логика высказываний, как и исчисление предикатов, была разработана специально для моделирования такого рода композиционной семантики и предполагает очевидную взаимосвязь между логикой и языком: логические системы могут представлять или моделировать важные структурные особенности естественного языка, такие как генеративность и композиционность. Конечно, мы всегда можем задать вопрос, действительно ли естественный язык является генеративным и композиционным в той мере и теми способами, как мы это предполагаем. Об этом размышляли многие исследователи, но это эмпирический вопрос, который может быть решен только путем эмпирического исследования. (Естественный язык может не оказаться строго композиционным, но есть веские аргументы в пользу того, что он в целом композиционен.)

Какова нынешняя ситуация с семантической теорией информации? Ниже приводится отрывок из монографии Шеннона и Уивера 1949 года [1]: «Если рассуждать о коммуникации в широком смысле, то нам видятся проблемы на трех уровнях. Отсюда возникают три последовательных вопроса:

- **Уровень А.** Насколько точно могут передаваться символы коммуникации? (Техническая проблема).
- **Уровень В.** Насколько точно передаваемые символы передают желаемое значение? (Семантическая проблема).
- **Уровень С.** Насколько эффективно полученное значение влияет на поведение желаемым образом? (Проблема эффективности)».

Хотя можно восхищаться попытками Шеннона и Уивера представить теорию информации как зонтик, охватывающий синтаксис, семантику и прагматику, общие результаты выглядят разочаровывающими. Тем, кто считает уравнения Шеннона техническими моделями передачи сигналов, расширение Уивера на вопросы значения и эффективности должно показаться неуместным, поскольку Шеннон не проявлял особого интереса к семантическому значению сообщения или его прагматическому влиянию на слушателя. Подобно производителям современных устройств обработки звука, его не волновало, транслировал ли канал Бетховена или Моцарта, или предпочитает ли слушатель ритм рока, свинг джаза или контрапункт Баха. Вместо этого его теория была направлена на решение технических проблем передачи звука с высокой точностью воспроизведения. Шеннон несколько настороженно относился к массовому применению своих математических уравнений к семантическим и прагматическим вопросам межличностного общения. Но его сомнения не разделял Уоррен Уивер, руководитель Фонда Рокфеллера и Института исследования рака Слоуна-Кеттеринга, а также консультант ряда частных научных фондов. Опубликованная теория Шеннона сочеталась с интерпретирующим эссе Уивера, в котором теория информации представлена как «чрезвычайно общая по своему охвату, фундаментальная касательно проблем, которые она рассматривает, и классическая по простоте и силе получаемых ею результатов». В эссе говорилось, что какой бы ни была проблема с общением, ее решением является сокращение потерь информации. Однако у Шеннона есть техническое определение слова «информация», которое не приравнивает *информацию* к *значению*. Он подчеркивает, что «семантические аспекты коммуникации не имеют отношения к инженерным аспектам». С точки зрения Шеннона информация относится к возможности уменьшить неопределенность. Это дает нам шанс уменьшить энтропию.

Начиная с работ Карнапа, Бар Хиллеля и др. [2] и до современных концепций, таких как семантика вопросов Гренендейка и Стохофа [3] или информационный поток Барвайса, Селигмана и др. [4], дедуктивный вывод обычно квалифицируют как аналитический. В контексте информации эту характеристику можно сформулировать так: информация, содержащаяся в заключении, является подмножеством информации, содержащейся в предпосылке. Другими словами, информация, содержа-

щаяся в заключении, уже содержится в предпосылке. В контексте функции истинности – функции, которая принимает значения истинности в качестве входных данных и выдает уникальное значение истинности в качестве выходных данных, – обычно используется утверждение, что действительный дедуктивный вывод сохраняет истинное значение, содержащееся в предпосылках, или что множество возможных слов, которое проверяет набор (или конъюнкцию) всех предпосылок, является подмножеством множества возможных слов, проверяющих заключение. Примирение между двумя точками зрения может быть достигнуто путем семантического подхода к теории информации. Точнее, обе точки зрения предполагают, что каждое значащее предложение является объединением непересекающихся элементарных предложений Z и что они определяют содержание предложения P как объединение всех Z , так что отрицание Z подразумевается предложением P .

Таким же образом можно рассматривать и нейронные сети. Ответы на заданные вопросы видны на выходном уровне обученной нейронной сети, а логическое содержание максимизируется на этом выходном слое. Однако даже если «обычное» искажение (не учитывающее значение) на выходном слое очень велико, выход сети дает нам то, что нам нужно.

- **Пример:** предположим, что нейросеть обучена распознавать кошку на изображении.

Логика, связанная с этой нейросетью, очень проста: один тип (Кошка), отсутствие композиционности и одно предложение:

$$\{x, T x \in \text{Кошка}\},$$

где x – входное изображение, а T означает True (истина). Итак, запись $\{x, T x \in \text{Кошка}\}$ означает, что изображение x представляет кошку (изображение имеет тип Кошка). Конечно, нейросеть была обучена на наборе данных, помеченном людьми, и те, кто пометил набор данных для обучения, являются контекстом. Истина относительна. Семантическое сообщение будет передавать только один бит, соответствующий утверждению «это кошка, Истина (1) или Ложь (0)». Таким образом, DNN можно рассматривать здесь как кодировщик семантического источника, который кодирует одно изображение в один бит.

Если говорить о более сложных ситуациях, нейронная сеть может отвечать на многие вопросы, например используемый язык может содержать несколько предложений, несколько типов, конъюнкций и дизъюнкций. Нейронная сеть будет сжимать данные и преобразовывать их в предложения, требующие гораздо меньшего количества бит. Совместное кодирование источника и канала

на основе машинного обучения в качестве наглядного примера семантической коммуникации будет обсуждаться в главе 23.

- **Математический подход:** очевидно, что семантическую информацию получить нелегко, в первую очередь из-за отсутствия соответствующих математических инструментов. Однако некоторые перспективные направления исследований все же существуют. Мы не будем здесь вдаваться в подробности, лишь отметим, что Бодо и Беннекен в [5] пересмотрели информационные меры Шеннона с категориальной/топологической точки зрения.

Для определения информационных мер уровня A Шеннон использовал аксиоматический подход. Он сказал: «Эта теорема и предположения, необходимые для ее доказательства, никоим образом не являются необходимыми для данной теории. Они даются в основном для того, чтобы придать определенную правдоподобность некоторым из наших более поздних определений. Однако подлинное обоснование этих определений будет заключаться в выводах, которые из них следуют». Выводы показали, что он был прав, и в данном случае скалярные величины, такие как энтропия и взаимная информация, достаточны и успешны.

Распространение этого аксиоматического подхода на семантический случай до сих пор не было плодотворным, и похоже, что связи между различными точками зрения на энтропию – алгебраическими, вероятностными, комбинаторными или динамическими – все еще плохо изучены. В работе [5] это было основной мотивацией для введения категориальной структуры как формализма, достаточно общего для интеграции различных теорий, и мы считаем, что это может привести к теории семантической информации. Используя такую структуру, в [6] обнаружили нешенноновские информационные меры, такие как энтропия Тсаллиса, квантовая энтропия или комбинаторная теория информации, где вероятности заменены частотами.

Отправной точкой является определение *информационной структуры*. Это пара (S, E) , где S – категория [7], представляющая переменные (морфизмы S задаются обуславливанием), а E – ковариантный функтор из S в категорию измеримых пространств. Затем между информационными структурами определяют морфизмы, чтобы создать категорию таких структур. В современной топологии пространства изучаются с помощью их (ко)гомологии, которая в некоторой степени кодирует их формы. Здесь естественным понятием пространства является понятие, предложенное Гротендиком; топосом [8] и элементарными топосами здесь являются категории контравариантных функторов

множеств в категории S для данной информационной структуры и самой грубой топологии Гротендика (все предпучки являются пучками). Эти предпучки и естественные преобразования между ними образуют абелеву категорию, что влечет за собой возможность определения когомологических функторов.

Гротендик обнаружил, что энтропия Шеннона – это когомология топосов. Кроме того, мы можем заменить S на другую, более общую категорию. (Исследования в настоящее время проводятся в синтаксической категории, соответствующей данному языку, которая зависит от вероятностей предложений.) Кроме того, в семантическом контексте информационные меры могут быть не скалярами, а фактически более крупными объектами, такими как сомножества и множества.

И наконец, следует отметить очень сильную связь, которая возникает между геометрией и логикой, поскольку у каждого топоса есть свой внутренний язык.

6G будет стимулировать возникновение и развитие многих технологий в тандеме с теорией семантической коммуникации, таких как совместное кодирование канала источника и сжатие источника.

10.2. Теория сверхвысокого разрешения

Теория сверхвысокого разрешения относится к представлению исходной информации таким образом, чтобы ее можно было восстановить как можно ближе к исходному формату [9]. Такие события, как выбросы и точки разрыва, могут быть сверхразрешены с бесконечной точностью всего по нескольким низкочастотным выборкам путем решения удобных выпуклых планов (выпуклого программирования¹⁸). Очевидным вариантом использования сверхвысокого разрешения является дальнейшее повышение разрешения, выходящее за рамки физических ограничений сенсорной системы (например, микроскопа). Сверхразрешение также можно использовать в других случаях, когда желательно экстраполировать мелкомасштабные детали из данных с низким разрешением или разрешить детали субпикселей (например, при зондировании).

В беспроводной связи передаваемые сигналы обычно достигают приемника несколькими путями из-за отражения от объектов (например, зданий). Предположим, что $h(t)$ – передаваемый сигнал, $z(t)$ – характеристика канала, а d_i и t_i – комплексная амплитуда и задержка в i -м тракте; тогда полученный сигнал во временной области можно представить следующим образом:

¹⁸ Выпуклое программирование – это подобласть математической оптимизации, которая изучает задачу минимизации выпуклых функций на выпуклых множествах.

$$x(t) = z(t) * h(t) = \sum_{i=1}^r d_i h(t - t_i),$$

где r – количество путей, которые в дальнейшем мы будем называть *модами* (mode). Используя преобразование Фурье, мы можем представить сигнал в спектральной области следующим образом:

$$\hat{x}(f) = \hat{z}(f) * \hat{h}(f) = \sum_{i=1}^r d_i \hat{h}(f) e^{j2\pi f t_i}.$$

Поскольку передаваемый сигнал $\hat{h}(f)$ известен априори, наблюдаемые данные могут быть выражены как

$$\frac{\hat{x}(f)}{\hat{z}(f)} = \sum_{i=1}^r d_i e^{j2\pi f t_i} \quad \forall d : \hat{h}(f) \neq 0.$$

Наконец, мы можем получить классическую модель оценки параметров теории сверхвысокого разрешения, которая представляет собой смесь сигналов r мод:

$$x[t] = \sum_{i=1}^r d_i \psi(t; f_i) \quad t \in \mathbb{Z},$$

где d_i – амплитуда, f_i – частота, ψ – (известная) модельная функция (например, $\psi(t; f_i) = e^{j2\pi f_i t}$, а r – количество мод или порядок модели. Наша цель – оценить $2r$ неизвестных параметров $\{d_i\}$ и $\{f_i\}$. Простой пример представлен на рис. 10.1.

Первой широко используемой теорией разрешения была теорема выборки Шеннона–Найквиста [10], которая требует, чтобы частота дискретизации как минимум вдвое превышала максимальную частоту исходной информации. Этот критерий известен как *частота дискретизации Найквиста* и гарантирует идентичность и целостность восстановленной информации. Эта теорема основана на преобразованиях Фурье, которые линейно преобразуют исходную информацию в новую систему координат, содержащую бесконечное количество косинусоидальных волновых оснований. С помощью преобразования Фурье последовательный во времени сигнал или последовательность могут быть представлены суммированием проекций (внутренним произведением) по бесконечному числу оснований. Частота дискретизации Найквиста обеспечивает одинаковое разрешение для всех проекций.

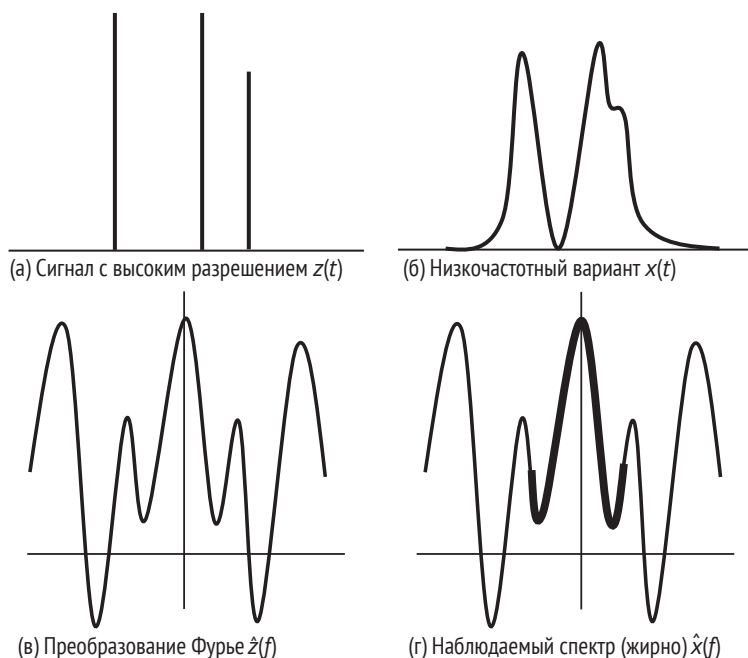


Рис. 10.1. Суперразрешение, которое экстраполирует спектр верхнего уровня (мелкие детали) из спектра нижнего уровня (данные с низким разрешением)

Теорема выборки Шеннона–Найквиста длительное время доминировала в разработке систем коммуникации. Она указывала нам, какая полоса пропускания требуется для надежного трансивера или для кодеков, применяемых к аудио, видео, изображениям и аналоговым сигналам. Однако за последние четыре десятилетия мы стали свидетелями огромного увеличения занимаемой полосы частот радиосвязи благодаря новым поколениям беспроводных систем. Полоса пропускания становится все дороже, а несущие частоты – все выше, что еще больше стимулирует изучение теории сверхвысокого разрешения.

Если говорить об источнике сигнала, то может возникнуть вопрос, действительно ли необходимо единообразно представлять сигнал или информацию в соответствии с его самой высокочастотной составляющей, и этот вопрос можно разделить на два подвопроса: (1) требуется ли равномерное разрешение и (2) необходимо ли постоянно придерживаться частоты дискретизации Найквиста?

Равномерное или неравномерное разрешение: чтобы ответить на этот вопрос, мы должны рассмотреть источник или сигнал, приемник или декодер которого выделяют и используют только некоторые из его частотных компонентов. Примером этой концепции служит видеопоток, когда человеческое зрение уделяет больше внимания переднему

плану, чем фону. Соответственно, передний план состоит из высокочастотных компонентов, а низкочастотные относятся к фону. В этом сценарии логичнее и эффективнее использовать более высокое разрешение для переднего плана, чем реализовывать равномерное разрешение.

Это соображение дает начало второй группе широко используемых теорий: метод главных компонент [11], оконное преобразование Фурье и вейвлеты. Метод главных компонент использует разложение по сингулярным векторам для разделения компонентов. Однако, в отличие от преобразования Фурье, в методе главных компонент в качестве новой координаты используется левая или правая собственная матрица, а не базис Фурье. Кодировщик JPEG применяет метод главных компонент для разделения и сохранения основных компонентов, в то время как приемопередатчики MIMO также используют метод основных компонент для детектирования основных слоев. Однако метод главных компонент теряет некоторую информацию из-за равномерного разрешения в соответствии с принципом неопределенности Гейзенберга. В случае широкополосного сигнала равномерное разрешение дает преимущество одним частям, но ухудшает другие. Чтобы решить эту проблему, Габор сначала предложил оконное преобразование Фурье – а затем и вейвлеты – для реализации множественного разрешения. Если мы преобразуем вейвлет-сигнал обратно в исходное окружающее пространство (в качестве декодера), вполне вероятно, что разрешение на самом деле в общем случае не будет однородным. JPEG-2000 использует вейвлеты для сжатия изображений. Некоторые исследователи даже предложили заменить OFDM вейвлет-преобразованием для широкополосных сигналов.

Равномерная или неравномерная частота выборки: мы также должны учитывать тот факт, что высокочастотные компоненты исходного сигнала могут игнорироваться приемником или не иметь для него большого значения. Наиболее очевидный пример – оцифровка звука для передачи. Поскольку человеческое ухо может воспринимать звук только с частотой до 20 кГц, все, что превышает этот порог, не имеет смысла для людей. Таким образом, бессмысленно учитывать аудиосигнал высокой четкости за пределами этого диапазона. В связи с этим для оцифровки голоса с хорошим качеством достаточно частоты дискретизации 8 кГц.

Это соображение порождает третью широко используемую теорию – *разреженной выборки* [12], которая основана на разреженности естественной информации. Например, исходная информация x , хотя и не является разреженной в своем пространстве, может быть линейно преобразована в разреженное представление z только с K ненулевыми элементами в новой системе координат. Преобразование может быть

быстрым преобразованием Фурье (БПФ), дискретным преобразованием Фурье (ДПФ) или вейвлетами, как упомянуто выше. Для этой разреженной информации x случайная матрица размером $M \times N$ (известная как матрица измерений) определяет $M(K < M \ll N)$ случайных точек выборки, чтобы выполнить выборку x в вектор y размером $M \times 1$. Очевидно, y дискретизируется неравномерно, и его частота дискретизации намного ниже, чем удвоенная максимальная частота. На стороне декодирования сжатая выборка использует норму $L1$, а не норму $L2$, чтобы отбросить выбросы, а для аппроксимации разреженного вектора можно использовать аппроксимацию с преследованием (matching pursuit, MP) или ортогональную аппроксимацию с преследованием (orthogonal matching pursuit, OMP) [13–16]. Сжатая выборка устраняет необходимость в частоте дискретизации Найквиста для выборки разреженной по своей природе информации или сигналов. В 6G датчики могут использовать схемы неоднородной случайной выборки для значительной экономии полосы пропускания радиосвязи.

Сверхразрешение для 6G: за последние 40 лет мы стали свидетелями того, что неравномерное разрешение во многих отношениях лучше, чем равномерное, а неравномерная выборка лучше, чем равномерная в таких случаях, как кодирование источника или сжатие изображения.

Теория сверхвысокого разрешения объединяет как неоднородное разрешение, так и неоднородную выборку [9] и может использоваться во многих областях 6G.

Простым примером является кодирование на стороне источника для машин. Дело не только в большом количестве машинных приемников, подключенных к системам 6G; машинные приемники работают не так, как люди. К фону может применяться более высокое разрешение, чем в алгоритмах, ориентированных на человека, а для сохранения полосы пропускания передачи стратегия кодирования источника с несколькими разрешениями может гибко адаптироваться к различным свойствам объекта.

Последний пример в этой главе – неравномерная выборка с низким энергопотреблением. Системы 6G с широкой полосой частот будут создавать проблемы для динамического квантования, а соблюдение критерия дискретизации Найквиста приведет к огромному энергопотреблению. Если исходная информация является разреженной по своей природе, то становится возможной неравномерная выборка с низким энергопотреблением для снижения энергопотребления.

Теория сверхвысокого разрешения может повлиять на направления развития многих технологий, включая алгоритмы локализации, сжатую выборку и алгоритмы беспроводной визуализации.

10.3. Источники

- [1] C. E. Shannon and W. Weaver, *The mathematical theory of communication*. University of Illinois Press, 1949.
- [2] R. Carnap, Y. Bar-Hillel et al., *An outline of a theory of semantic information*, Technical Report no. 247, Research Laboratory of Electronics, MIT, 1952.
- [3] J. A. G. Groenendijk and M. J. B. Stokhof, *Studies on the semantics of questions and the pragmatics of answers*, joint Ph.D. dissertation, University of Amsterdam, 1984.
- [4] J. Barwise, J. Seligman et al., *Information flow: The logic of distributed systems*. Cambridge University Press, 1997.
- [5] P. Baudot and D. Bennequin, *The homological nature of entropy*, *Entropy*, vol. 17, no. 5, pp. 3253–3318, 2015.
- [6] J.-P. Vigneaux-aritzia, *Topology of statistical systems: A cohomological approach to information theory*, thesis, Max Planck Institute for Mathematics in the Sciences, Max Planck Society, 2019.
- [7] S. Awodey, *Category theory*. Oxford University Press, 2010.
- [8] S. MacLane and I. Moerdijk, *Sheaves in geometry and logic: A first introduction to topos theory*. Springer Science & Business Media, 2012.
- [9] E. J. Candès and C. Fernandez-Granda, *Towards a mathematical theory of superresolution*, *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.
- [10] H. Nyquist, *Transmission systems for communications*, *AT&T Technical Journal*, vol. 2, p. 26, 1959.
- [11] I. T. Jolliffe, *Principal component analysis* (Second Edition). Springer, 2002.
- [12] D. L. Donoho, *Compressed sensing*, *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [13] E. J. Candes, J. K. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [14] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [15] E. J. Candes and T. Tao, *Near-optimal signal recovery from random projections: Universal encoding strategies?* *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [16] R. G. Baraniuk, *Compressive sensing [lecture notes]*, *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, 2007.

Глава 11

Теоретические основы энергоэффективных систем

11.1. Энергоэффективные коммуникации и теория вычислений

Проектирование беспроводных сетей на основе показателей энергоэффективности – чрезвычайно сложная задача, поскольку традиционные схемы распределения мощности приводят к непрерывной передаче с использованием максимальной доступной мощности. Однако в последнее время эта точка зрения начала меняться, так как ключевым показателем эффективности для сетей 5G стала энергоэффективность в битах на джоуль, определяемая как количество информации, которое может быть надежно передано на один джоуль потребляемой энергии [1]. В работе [2] обсуждаются систематические подходы к решению задач максимизации энергоэффективности. В соответствии с этими подходами структура из [3] используется в [4] и [5] для разработки энергоэффективных алгоритмов управления мощностью для массивных систем MIMO, а в [6] – для энергоэффективных сетей с малыми сотами. Однако все доступные исследования рассматривают два основных пути расходования энергии, а именно:

- **энергия передачи:** энергия, выделяемая каждой передающей антенной;
- **статическая энергия:** энергия, потребляемая всеми аппаратными блоками в каждом передатчике и приемнике (например, аналого-цифровым преобразователем, цифроаналоговым преобразователем, аналоговой фильтрацией, резервным аккумулятором и системой охлаждения).

Фактически анализ энергопотребления, необходимого для выполнения вычислений при обработке информации в цифровой области (например, для вычислений при цифровом предварительном кодиро-

вании и цифровом приеме), плохо изучен. В этом контексте доступны лишь некоторые результаты, в которых используются принципы фундаментальной физики/термодинамики, чтобы для начала понять: действительно ли цифровые вычисления должны потреблять энергии?

Удивительно, но теоретически ответ отрицательный: ни один физический принцип не предписывает потреблять энергию для выполнения вычислений в цифровой области! Действительно, если выполняемое вычисление не уничтожает никакую информацию, то это обратимое преобразование, которое, согласно второму закону термодинамики, не увеличивает энтропию. В результате обратимые вычисления теоретически могут выполняться без каких-либо затрат энергии, как впервые было замечено в [7] и [8]. Модели машин, которые реализуют эту обратимую логику, появились в [9] и [10] и работают, сохраняя как выходные, так и входные данные каждого вычисления, без стирания каких-либо битов из памяти компьютера. Таким образом, всегда можно вернуть машину в исходное состояние. Тогда возникает другой вопрос: почему современные компьютеры основаны на булевой логике, которая, как правило, необратима (например, не всегда возможно вывести входные данные операции «И» или «ИЛИ» из ее вывода)?

Первое практическое соображение состоит в том, что даже если были предложены модели обратимых логических машин, они носят чисто теоретический характер. На практике обработка данных выполняется макроскопическими устройствами, которые рассеивают макроскопические количества энергии. Второе соображение более фундаментальное – даже если бы мы могли построить обратимую логическую машину, у нее были бы невыполнимые требования к памяти, так как ей нужно было бы хранить абсолютно все выходные и входные данные каждой выполняемой операции без удаления каких-либо битов из памяти. Однако поскольку объем данных, которые могут быть сохранены в компьютере, ограничен объемом банка памяти [11], даже обратимые логические машины в конечном итоге должны будут перезаписывать некоторые ячейки памяти, таким образом потребляя энергию.

Аналогичная ситуация возникает при обнаружении или исправлении ошибок передачи информации по каналу связи. Обнаружение ошибки в теории не влечет за собой рассеивания энергии, но ее исправление приводит к потере информации (в окружающую среду), генерируя энтропию. Это приводит к нереалистичной ситуации, когда во избежание генерации энтропии система связи должна также хранить полученные сообщения, содержащие ошибки. В итоге мы приходим к тому, что необходимо учитывать необратимые операции, но очень мало известно об их фундаментальных ограничениях, касающихся рассеивания энергии в системах связи. Краеугольный камень в этой области – работа [12];

однако в этой статье рассматривается только изолированный компьютер, а не полная система связи. Примеры [12] применительно к системам связи немногочисленны.

Как сказано в предыдущих главах, к энергоэффективности 6G будут предъявляться очень строгие требования. Следовательно, чрезвычайно важно изучить стоимость вычислений в нейросетях с точки зрения энергопотребления. Кроме того, необходимо изучить энергетический аспект в связи с требованиями к задержке обработки для 6G, которые также очень строги и требуют быстрых (и энергоемких) вычислений. Понимание фундаментальных энергетических ограничений вычислений, связанных с обработкой информации, является важным шагом на пути к успешному развертыванию беспроводных сетей 6G. Как было сказано ранее, если предполагается бесконечное пространство памяти, то теоретически для выполнения обратимых вычислений не требуется энергия. Однако мы также увидели, что физически невозможно построить сеть, основанную только на обратимых операциях. Следовательно, с инженерной точки зрения принципиально важно понимать, какого минимального предела мы можем достичь, выполняя необратимые операции.

Поскольку современные компьютеры основаны на необратимой (булевой) логике, необходимо заполнить теоретический пробел между обратимыми и необратимыми вычислениями, чтобы понять фундаментальные энергетические ограничения обработки информации. Эти исследования выглядят проблематичными из-за того, что беспроводные сети становятся все сложнее, и в будущем станут намного более сложными, чем все прошлые и нынешние сети связи, что требует разработки специальных теоретических моделей для оценки объема информации, которая должна быть обработана для поддержания желаемой функциональности сети. Помимо исследования энергии, необходимой для обработки информации, также важно понимать, как фундаментальные ограничения энергии обработки информации связаны с фундаментальными ограничениями производительности связи. Например, каков оптимальный компромисс между производительностью связи и потреблением энергии в сети? Этот вопрос исследовался только в отношении энергии, используемой для связи, тогда как почти ничего неизвестно об энергии, используемой для вычислений, связанных с обработкой информации. Кроме того, во всех предыдущих исследованиях рассматривались традиционные сетевые архитектуры и топологии, без учета влияния использования глубокого обучения и нейросетей.

Термодинамика в приложении к коммуникациям и теории вычислений способна помочь в разработке новых алгоритмов для энергоэффективных сетей 6G.

11.2. Теория зеленого ИИ

В будущем ИИ претерпит невероятную эволюцию и проникнет во все слои общества. Во время этого процесса ИИ и связь будут сближаться, при этом либо сеть связи будет обслуживать и обеспечивать конвейеры данных для приложений ИИ, либо ИИ повысит эффективность передачи данных в сети связи. Однако потребление электроэнергии и выбросы углерода в результате использования искусственным интеллектом больших вычислительных ресурсов являются весьма неприятным фактором как с точки зрения защиты окружающей среды, так и с точки зрения экономических выгод. По этой причине зеленый ИИ привлекает все большее внимание в научном и инженерном сообществах [15, 16], и переход от красного ИИ к зеленому существенно повлияет на принципы проектирования сетей беспроводной связи следующего поколения.

Красный ИИ использует модель, которая обучается с помощью обучающего набора данных и оценивается с помощью тестового набора данных. Чтобы разработать красную модель ИИ, нам обычно требуется использовать набор обучающих данных для итеративной корректировки гиперпараметров. Соответственно, зеленый ИИ относится к решению, в котором модели, алгоритмы и оборудование дают новые результаты без увеличения вычислительных затрат и фактически снижают их в идеальных сценариях [15].

Как правило, стоимость вычислений модели ИИ пропорциональна произведению трех ключевых факторов: стоимости выполнения модели на одной выборке, размера обучающего набора данных и количества экспериментов с гиперпараметрами. Количество операций с плавающей запятой, вследствие использования разных алгоритмов для реализации одной и той же функции ИИ, уменьшилось в 44 раза с 2012 по 2019 год, что эквивалентно удвоению эффективности алгоритма каждые 16 месяцев в течение семи лет [17]. Между тем из-за быстрого расширения наборов обучающих данных и гиперпараметров и, в частности, из-за чрезмерного стремления к повышению точности обучения на основе огромных вычислительных ресурсов потребление электроэнергии ИИ увеличилось в 300 000 раз с 2012 по 2018 год, что эквивалентно удвоению потребления ресурсов каждые 3,4 месяца [18]. Эмма Струбелл и др. предложили оценку потребления энергии искусственным интеллектом (p_t) и выбросов CO_2 (CO_{2e}) [16]:

$$p_t = \frac{PUE \times t \times (p_c + p_r + gp_g)}{1000} \quad (\text{кВтч}),$$

$$\text{CO}_{2e} = \text{CUP} \times p_t \quad (\text{кг}),$$

где t – ожидаемое общее время, затраченное на обучение модели; p_c , p_r и p_g представляют собой среднее энергопотребление (в ваттах) от всех сокетов ЦП, от всех сокетов DRAM (основной памяти) и от каждого графического процессора соответственно во время обучения; g – количество графических процессоров, используемых для ускорения обучения. *PUE* – это сокращение от *power usage effectiveness*, означающее эффективность использования энергии с учетом дополнительной энергии, необходимой для поддержки вычислительной инфраструктуры (в первую очередь охлаждения). Средний мировой показатель *PUE* для центров обработки данных составляет 1,58 [19]. *CUP* – это сокращение от *carbon dioxide per unit power* (количество углекислого газа на единицу мощности), которое зависит от уровня развития местной энергетики. Агентство по охране окружающей среды США предоставило средний коэффициент производства CO_2 0,433 кг/кВтч [20].

В работе [16] на примере модели NLP (обработка естественного языка) показано, что потребление энергии и выбросы углерода, соответствующие различным моделям ИИ и аппаратному обеспечению, нельзя игнорировать с экономической точки зрения. Средняя стоимость электроэнергии в США в 2018 году составила 0,12 доллара США за кВтч без учета потерь выработки. В частности, любые модели, чувствительные к наборам данных и гиперпараметрам, будут очень дорогими, поскольку они могут нуждаться в переобучении каждый раз, когда возникает новая ситуация.

В последние годы появилось огромное количество научных статей по ИИ в различных областях применения, включая ИИ, используемый в сетях беспроводной связи. Однако мало внимания уделяется вытекающим из этого проблемам энергопотребления и выбросов углерода, вызванным огромным объемом прямых неоптимизированных вычислений во время внедрения ИИ. Сообщество ИИ пытается привлечь внимание исследователей к этим проблемам и в настоящее время выступает за следующее: (1) сообщать о времени обучения и чувствительности к гиперпараметрам в публикациях по ИИ, (2) предоставлять равноправный доступ к вычислительным ресурсам во время реализации ИИ и (3) уделять приоритетное внимание вычислительной эффективности аппаратного обеспечения и алгоритмов искусственного интеллекта.

Чтобы измерить эффективность, Рой Шварц и др. предложили сообщать объем работы, необходимой для получения результата (для обучения модели и настройки гиперпараметров) в приложениях ИИ [15]. При составлении отчета об объеме работы, выполненной приложением ИИ, необходимо измерить определенные показатели, чтобы обеспечить справедливое сравнение между различными моделями. Речь идет о следующих показателях:

- **выбросы углерода:** количество, которое зеленый ИИ стремится минимизировать;
- **энергопотребление:** связано с выбросами углерода, но не зависит от времени и места;
- **затраченное реальное время:** общее время работы для создания результата ИИ, которое является естественным показателем эффективности;
- **количество параметров:** количество параметров (обучаемых или общих), используемых моделью ИИ, которое является еще одним распространенным показателем эффективности;
- **операции с плавающей запятой:** конкретная мера эффективности в процессе генерации результата ИИ.

В [21] была предложена концепция зеленого ИИ под названием Once-for-All (раз и навсегда), согласно которой мы обучаем одну сеть, а затем специализируем ее при развертывании, чтобы обеспечить эффективный логический вывод на многих устройствах и уложиться в различные ограничения ресурсов. Традиционные подходы сводятся либо к проектированию вручную, либо к поиску нейронной архитектуры, чтобы найти специализированную нейронную сеть и обучить ее с нуля для каждого случая, что недопустимо с вычислительной точки зрения и генерирует большое количество выбросов CO_2 . И наоборот, универсальная сеть поддерживает различные архитектурные настройки за счет разделения обучения и поиска, и можно быстро получить специализированную подсеть без дополнительного обучения.

Возвращаясь к разговору о проектировании беспроводных коммуникационных сетей следующего поколения, нужно отметить, что традиционные алгоритмы оптимизации ИИ (такие как федеративное обучение) обычно рассматривают пропускную способность или задержку беспроводных соединений как вес для распределенного мультипроцессорного обмена данными, без учета предельных значений энергии или разницы в стоимости мощности между различными устройствами в различных регионах. Это неоднозначное рассмотрение ограничений энергии ИИ или затрат на электроэнергию может привести к большому расхождению между проектом беспроводной сети и фактическим развертыванием ИИ в будущем. По этой причине мы выступаем за необходимость придавать одинаковое значение зеленому ИИ и зеленым коммуникациям. С самого начала проектирования архитектуры следует в полной мере учитывать влияние моделей, алгоритмов и оборудования искусственного интеллекта на энергопотребление, чтобы обеспечить экономическую выгоду клиентам с соответствующими операционными расходами системы.

В эпоху 6G возникнет необходимость в архитектуре энергоэффективных вычислений, которые должны учитывать энергетические ограничения ИИ. Впрочем, это важно уже сегодня, поскольку такие соображения одинаково актуальны в нынешних сетях 4G и 5G.

11.3. Источники

- [1] S. Buzzi, I. Chih-Lin, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, *A survey of energy-efficient techniques for 5G networks and challenges ahead*, IEEE Journal on Selected Areas in Communications, vol. 34, no. 4, pp. 697–709, 2016.
- [2] A. Zappone and E. Jorswieck, *Energy efficiency in wireless networks via fractional programming theory*, Foundations and Trends in Communications and Information Theory, vol. 11, no. 3–4, pp. 185–396, 2015.
- [3] A. Zappone, L. Sanguinetti, G. Bacci, E. Jorswieck, and M. Debbah, *Energy-efficient power control: A look at 5G wireless technologies*, IEEE Transactions on Signal Processing, vol. 64, no. 7, pp. 1668–1683, 2015.
- [4] A. Zappone, E. Björnson, L. Sanguinetti, and E. Jorswieck, *Globally optimal energy efficient power control and receiver design in wireless networks*, IEEE Transactions on Signal Processing, vol. 65, no. 11, pp. 2844–2859, 2017.
- [5] M. Di Renzo, A. Zappone, T. T. Lam, and M. Debbah, *System-level modeling and optimization of the energy efficiency in cellular networks: A stochastic geometry framework*, IEEE Transactions on Wireless Communications, vol. 17, no. 4, pp. 2539–2556, 2018.
- [6] C. H. Bennett and R. Landauer, *The fundamental physical limits of computation*, Scientific American, vol. 253, no. 1, pp. 48–57, 1985.
- [7] D. Deutsch, *Is there a fundamental bound on the rate at which information can be processed?* Physical Review Letters, vol. 48, no. 4, p. 286, 1982.
- [8] Y. Lecerf, *Machines de Turing reversibles-recursive insolubilité en $n \in \mathbb{N}$ de l'équation $u = \theta nu$, où θ est un isomorphisme de codes*, Comptes Rendus hebdomadaires des séances de l'académie des sciences, vol. 257, pp. 2597–2600, 1963.
- [9] C. H. Bennett, *Logical reversibility of computation*, IBM Journal of Research and Development, vol. 17, no. 6, pp. 525–532, 1973.
- [10] S. Lloyd, V. Giovannetti, and L. Maccone, *Physical limits to communication*, Physical Review Letters, vol. 93, no. 10, p. 100501, 2004.
- [11] R. Landauer, *Irreversibility and heat generation in the computing process*, IBM Journal of Research and Development, vol. 5, no. 3, pp. 183–191, 1961.

- [12] J. Izydorczyk and L. Cionaka, *A practical low limit on energy spent on processing of one bit of data*, in Proc. 2008 IEEE 8th International Conference on Computer and Information Technology Workshops. IEEE, 2008, pp. 509–514.
- [13] B. Perabathini, V. S. Varma, M. Debbah, M. Kountouris, and A. Conte, *Physical limits of point-to-point communication systems*, in Proc. 2014 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt). IEEE, 2014, pp. 604–610.
- [14] E. D. Demaine, J. Lynch, G. J. Mirano, and N. Tyagi, *Energy-efficient algorithms*, in Proc. 2016 ACM Conference on Innovations in Theoretical Computer Science, 2016, pp. 321–332.
- [15] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, *Green AI*, arXiv preprint arXiv:1907.10597, 2019.
- [16] E. Strubell, A. Ganesh, and A. McCallum, *Energy and policy considerations for deep learning in NLP*, arXiv preprint arXiv:1906.02243, 2019.
- [17] D. Hernandez and T. B. Brown, *Measuring the algorithmic efficiency of neural networks*, arXiv preprint arXiv:2005.04305, 2020.
- [18] D. Amodei and D. Hernandez, *AI and compute*, Heruntergeladen von <https://blog.openai.com/aiand-compute>, 2018.
- [19] R. Ascierto, *Uptime institute global data center survey*, Technical Report, Uptime Institute, 2018.
- [20] US Environmental Protection Agency, *Emissions and generation resource integrated database (eGRID2007, version 1.1)*, 2008.
- [21] H. Cai, C. Gan, T. Wang, Z. Zhang, and S. Han, *Once-for-all: Train one network and specialize it for efficient deployment*, arXiv preprint arXiv:1908.09791, 2019.

Резюме части III

Беспроводные сети быстро развиваются по сценарию, согласно которому инфраструктура связи будет сливаться с окружающей средой. Если сети 5G движутся в направлении плотного развертывания базовых станций и антенных решеток, то 6G перейдет на следующий уровень, оснастив объекты в области связи, такие как здания, стены, автомобили и дорожные знаки, интеллектуальными поверхностями, способными усиливать электромагнитные сигналы, выполнять вычисления и сохранять данные. В связи с экспоненциальным увеличением количества подключенных устройств, требующих более высокопроизводительных услуг, коммуникационная среда будет преобразована в интеллектуальную беспроводную сеть, характеризующуюся беспрецедентным уровнем плотности, стохастичности и неоднородности, а также очень большими размерами. В таком сложном сценарии традиционные сетевые архитектуры утратят работоспособность, и потребует смена парадигмы. В части III мы изложили теоретические основы коммуникаций в интеллектуальной беспроводной сети 6G, установив базис для вычисления теоретических пределов производительности, а также разработки прикладных алгоритмов для приближения к этим пределам. Мы исходим из предположения, что сетевая инфраструктура состоит из нескольких распределенных сегментов, каждый из которых наделен возможностями искусственного интеллекта и действует как подсистема, способная принимать независимые решения. Это видение может быть реализовано посредством взаимодействия теоретических инструментов из множества дисциплин, берущих начало в фундаментальной математике и физике. Однако важно отметить, что некоторые пробелы в знаниях остаются, и их необходимо устранить, чтобы эффективно создать единую структуру. На сегодняшний день можно отметить следующие пробелы в знаниях.

Пробел в знаниях 1:

В целом машинное обучение для беспроводных сетей сегодня адаптировано только к централизованным сетевым архитектурам, в которых центральный контроллер настраивает единую нейросеть для определения политики во всей сети. Однако в распределенной сети каждый узел будет обучать свою собственную ИНС на основе различных алгоритмов и локальных наборов обучающих данных. Отсутствие надлежащего понимания и контроля может вызвать нестабильность системы, что в свою очередь может привести к сбою системы и/или кибератакам.

Пробел в знаниях 2:

Традиционные методы обучения предполагают, что условия, в которых будут работать нейросети, аналогичны тем, которые были испытаны на этапе обучения. Однако это трудно обеспечить в больших и стохастических сетях из-за их быстро меняющейся во времени и неоднородной природы.

Пробел в знаниях 3:

Ни одна из описанных выше математических/физических структур никогда не использовалась в распределенных беспроводных сетях. Имеющиеся методологии, основанные на теории случайных матриц и вычислительной физике, не применимы к распределенным беспроводным сетям на основе ИИ, в которых каждый узел учится вести себя, независимо обрабатывая свой собственный набор данных без какого-либо явного централизованного управления.

Пробел в знаниях 4:

Каждая из математических/физических структур, описанных выше, способна охватить определенный аспект сложных беспроводных сетей, но для того, чтобы полностью охарактеризовать большие и стохастические сети нативного ИИ, все эти структуры должны использоваться в тандеме. В настоящее время нет известных результатов совместного использования такого количества сложных структур в беспроводных сетях.

Пробел в знаниях 5:

Вычисления требуют энергии; нейросети обучаются и эксплуатируются путем численной обработки больших наборов данных. Поэтому ожидается, что вычисления станут основным препятствием на пути развития зеленых и устойчивых сетей. Фундаментальные ограничения энергопотребления обработки информации доступны только для P2P-систем, а о многопользовательских беспроводных сетях ничего не известно.

Пробел в знаниях 6:

Чтобы достичь наилучшего компромисса между достижением оптимальной связи и обеспечением устойчивости беспроводной экосистемы, необходимо исследовать компромисс между фундаментальной производительностью и энергетическими пределами вычислений и связи.

Пробел в знаниях 7:

Использование сложных теоретических инструментов может привести к сложным решениям, которые трудно реализовать в практических системах. Например, может быть очень сложно решить различные

уравнения, возникающие в результате анализа EIT или распределенной стохастической оптимизации. Вычислительная структура, способная реализовать теоретические решения, в настоящее время отсутствует.

Эти пробелы в знаниях представляют собой лишь некоторые из препятствий, которые необходимо будет преодолеть, если мы хотим достичь единой фундаментальной теории сетей 6G. С теоретической точки зрения, мы приближаемся к концу эпохи, которой руководствовались информационно-коммуникационные технологии на протяжении последнего столетия. Многие из выдающихся инженерных достижений в области коммуникации (знаменитая «эпоха G») и вычислений (знаменитая «эпоха Мура») были основаны на довольно старых фундаментальных принципах: теорема выборки Найквиста, восходящая к 1924 году, закон Шеннона, датированный 1948 годом, и архитектура фон Неймана 1946 года. Сегодня, приблизившись к пределам этих принципов, мы обнаружили, что нам не хватает фундаментальных знаний, необходимых для разработки новых инженерных решений, и существует потребность в обновлении и развитии основ, изложенных в части III, чтобы дать толчок развитию новой научно-технической эры.

Часть IV



Новые элементы

Введение в часть IV

Ожидается, что 6G будет содержать новые возможности и предоставлять новые услуги с использованием новых беспроводных технологий. В дополнение к разработкам, сделанным в области беспроводной передачи, система 6G будет включать в себя множество новых элементов, таких как новый спектр, новые каналы, новые материалы, новые антенны, новые вычислительные технологии и новые оконечные устройства. Эту часть книги мы начнем с представления идеи нового спектра, а затем обсудим возможности и проблемы использования ТГц-диапазона как для связи, так и для сканирования. Мы также описываем моделирование и измерение новых каналов наряду с высокочастотными диапазонами, для таких сценариев, как экстремально большие апертурные решетки, неназемные сети и сканирование в миллиметровых и ТГц-диапазонах. Учитывая, что связь и сканирование в новых высокочастотных диапазонах требуют новых материалов и антенн, мы обсудим применение кремниевой фотоники, фотонных кристаллов, гетерогенных, реконфигурируемых, фотоэлектрических и плазмонных материалов. Кроме того, мы обсуждаем несколько новых типов антенн для диапазонов частот ТГц. Это особенно важно, поскольку из-за значительных потерь при передаче в ТГц-диапазоне антенны существенно отличаются от обычных, которые подключаются к радиочастотным системам через коаксиальные кабели или микрополосковые линии. И поскольку ожидается, что закон Мура вскоре утратит свою актуальность, в этой части мы рассмотрим новые вычислительные технологии, такие как вычисления на основе структур, подобных мозгу, и квантовые вычисления. Наконец, мы дадим прогноз относительно тенденций развития будущих оконечных устройств и обсудим их новые возможности.

Глава 12

Новый спектр

Подходящий спектр – это главное условие успешного развития для каждого поколения беспроводных систем. Во-первых, всегда желательно располагать более широким спектром, чтобы поддерживать более высокие скорости передачи данных и пропускную способность сети. Во-вторых, глобально унифицированный спектр позволит добиться большей экономии за счет масштабов как инфраструктуры, так и пользовательского оборудования. Поэтому очень важно определить и согласовать спектр во всем мире – то, что отрасль беспроводной связи успешно сделала через ITU-R и Всемирную конференцию радиосвязи (world radiocommunication conference, WRC). В-третьих, многополосные радиотехнологии позволяют нам и дальше использовать существующие и новые диапазоны. В-четвертых, глобальный роуминг и стандартизация технологий имеют решающее значение для повсеместного предоставления услуг и доступности приложений. И в-пятых, критически важны единые правила распределения и регулирования спектра во всем мире – это уникальная проблема, с которой мы сталкиваемся, используя радиочастотный спектр.

В инфраструктуре беспроводной связи спектр является жизненно важным компонентом. Для освоения радиочастотного спектра, как правило, требуется два поколения связи, и по мере развития технологий беспроводного поколения использование спектра продолжает расширяться до более высоких частотных диапазонов. Например, полосы ниже 1 ГГц использовались в первом поколении (1G), тогда как в 2G и 3G спектр, используемый для систем международной подвижной электросвязи (international mobile telecommunications, IMT), был расширен до диапазона ниже 3 ГГц. В поколении 5G С-диапазон является наиболее широко используемым диапазоном во всем мире для начального развертывания сети, а поддержка диапазона миллиметровых волн с частотой до 52,6 ГГц заявлена в релизе 3GPP 15/16 (2020). Ожидается, что в будущих версиях 5G эта поддержка будет расширена примерно до 100 ГГц. Что касается 6G, то мы предвидим расширение радиоспектра до еще более высоких частотных диапазонов, включая терагерцовый,

так как этого требуют более высокие скорости передачи данных и новые услуги. Следующее поколение беспроводных систем, занимающих более высокочастотные полосы спектра, откроет новые возможности как для более качественных услуг связи, так и для новых услуг, помимо связи.

В этой главе мы обсуждаем потенциальный спектр для 6G, но начнем с определения спектра 5G. Затем мы проанализируем потребности в спектре 6G и определим диапазоны частот-кандидатов в соответствии с требованиями, которые будут предъявляться новыми услугами 6G. В частности, мы обращаем внимание читателей на необходимость расширения спектра в сторону значительно более высоких частот и важность использования интегрированного многоуровневого спектра со всеми диапазонами, включая нижний, средний, миллиметровый и терагерцовый.

12.1. Глобальное распределение спектра 5G до 2020 года

ITU-R предполагает, что сеть 5G должна включать три репрезентативных сценария использования (т. е. EMBB, mMTC и URLLC) и обеспечивать скорость передачи данных около 20 Гбит/с, задержку на уровне миллисекунд и сверхплотную связь, что представляет собой значительные улучшения по сравнению с предыдущими поколениями систем мобильной связи [1].

С этой целью в сети 5G применяется многоуровневое распределение частотных полос, как показано на рис. 12.1. При таком подходе мобильная сеть может получить доступ к «высоким диапазонам» (диапазон миллиметровых волн от 24,25 до 71 ГГц), «средним диапазонам» (от 2 до 6 ГГц) и «низким диапазонам» (ниже 2 ГГц) по запросу, поддерживая разнообразные сценарии использования.

Среди этих частотных диапазонов наиболее важную роль для приложений 5G играют средние полосы (от 2 до 6 ГГц), поскольку они предлагают лучший компромисс между покрытием и пропускной способностью для рентабельной реализации, а также поддерживают большинство сценариев использования 5G в широкомасштабных зонах покрытия. По состоянию на январь 2020 года многие страны начали выделять средние полосы частот для первой волны коммерческого развертывания 5G [2].

Ключевыми средними полосами являются полосы *дуплексной связи с временным разделением* (time division duplex, TDD) на частотах 3300–3800, 2600 и 2300 МГц. Доступность этих диапазонов увеличивается во всем мире, поскольку операторы сетей считают их основными

полосами частот для крупномасштабного коммерческого развертывания 5G. Если эти полосы недоступны в краткосрочной перспективе или недостаточны для удовлетворения спроса, некоторые страны используют альтернативные полосы TDD на частотах 4800–4990 МГц (например, Россия, Китай, Бразилия и Южная Африка) и 3800–4200 МГц (например, Япония, США, Великобритания и Южная Корея).

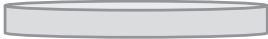
	Классификация	Показатели
eMBB, URLLC 	Высокие частоты (24,25–71 ГГц) Верхний уровень данных	• 24,25–27,5, 37–43,5 и 66–71 ГГц (глобально) • 45,5–47 и 47,2–48,2 ГГц (многие страны)
eMBB, URLLC, mMTC 	Средние частоты (2–6 ГГц) Уровень покрытия и связи	• TDD 3300–3800, 2600 и 2300 МГц • TDD 4800–4990 и 3800–4200 МГц (альтернативы)
eMBB, URLLC, mMTC 	Низкие частоты (ниже 2 ГГц) Уровень покрытия	• FDD 600 и 700 МГц • FDD 800, 900, 1800 и 2100 МГц

Рис. 12.1. Многоуровневая структура и конкретные варианты частотных диапазонов для сценариев использования в 5G

Низкие диапазоны обычно используются в сочетании со средними диапазонами, чтобы обеспечить более широкое и глубокое покрытие в помещении. Некоторые сетевые операторы используют эти низкие диапазоны для независимого быстрого и экономичного развертывания, если повышенная пропускная способность не является основной задачей. Ресурсы в этом частотном диапазоне в основном состоят из двух частей: (а) новые полосы *дуплексной связи с частотным разделением* (frequency division duplex, FDD) (например, 600 и 700 МГц), распределенные на WRC в 2015 году (WRC-15) [3], и (б) существующие полосы 2G/3G/4G FDD (например, 800, 900, 1800 и 2100 МГц), которые будут совместно использоваться с 5G посредством рефрейминга спектра или динамического совместного использования спектра (dynamic spectrum sharing, DSS).

Средние полосы могут быть объединены с несколькими низкими полосами для достижения более широкой пропускной способности сети или дополнены низкими полосами для улучшения высокоскоростных восходящих каналов и достижения более широкого покрытия. Такое новаторское использование диапазонов называется *дополнительными восходящими каналами связи* (supplementary uplinks, SUL).

Высокие диапазоны, в частности миллиметровые волны, впервые включены в спектр ИМТ начиная с 5G. Эти полосы в основном исполь-

зуются для поддержки дополнительной пропускной способности и высоких скоростей передачи данных, необходимых для некоторых приложений в определенных местах, например они могут быть использованы для городских точек доступа или фиксированного беспроводного доступа. Однако из-за высоких потерь при распространении радиоволн и ограничений современных радиотехнологий использование миллиметровых волн для бесшовных, широкозонных и мобильных приложений по-прежнему сталкивается с рядом технологических проблем. На WRC-19 верхние полосы с общей шириной полосы 14,75 ГГц (например, сумма 24,25–27,5, 37–43,5 и 66–71 ГГц) и 2,5 ГГц (например, сумма 45,5–47 и 47,2–48,2 ГГц) полосы частот миллиметрового диапазона были определены для IMT как полосы глобального и регионального согласования соответственно [4]. Благодаря использованию технологии агрегации несущих сочетание миллиметрового диапазона и средних или низких частот может способствовать достижению как более широкого покрытия, так и более высоких скоростей передачи данных.

Распределение спектра 5G ускоряется во всем мире в связи с ускоренным развертыванием 5G. К 2018 году только несколько стран-первопроходцев, таких как Китай, Южная Корея и США, завершили или объявили о планах распределения спектра 5G, но к январю 2020 года это число выросло до более чем 60 [2]. Ожидается, что большинство стран и регионов зарегистрируют свой собственный спектр 5G не позднее 2021 или 2022 года. Это означает, что большая часть упомянутого ранее спектра будет доступна для использования во всем мире с 2023 по 2025 год.

12.2. Требования к спектру 6G

Повестка дня ITU-R уже включает обсуждение потенциально доступного спектра для 6G [5].

В целом распределение спектра тесно связано со сценариями использования и ключевыми показателями эффективности сети. Следовательно, многие аспекты использования спектра 5G (как обсуждалось в части II, такие как eMBB, mMTC и URLLC) получают продолжение в 6G. Кроме того, распределение спектра подчеркивает преемственность политики и правил в области радиосвязи, поскольку жизненный цикл выделенного спектра охватывает десятилетия. Это означает, что концепция использования многоуровневой структуры частотных диапазонов в 5G будет по-прежнему применяться для 6G.

Кроме того, по мере появления новых требовательных к полосе пропускания приложений, таких как голографическая связь, наряду

с новыми услугами и функциями, такими как сканирование с высоким разрешением, ожидается, что 6G будет использовать гораздо более широкую полосу пропускания – выше, чем даже полоса миллиметрового диапазона, и более высокую частоту – вплоть до ТГц-диапазона или даже спектра видимого света. Что касается конкретных вариантов использования 6G (см. часть II), мы считаем, что необходимо внимательно рассмотреть требования к спектру, предъявляемые новыми технологиями, услугами и функциями в 6G.

В рамках этих требований необходимы дополнительные средние полосы, чтобы обеспечить непрерывное и рентабельное улучшение пропускной способности и покрытия при развертывании 6G. Кроме того, прорыв в области радиотехнологий и их инновационное использование при развертывании макросот в диапазоне миллиметровых волн станут ключом к предоставлению новых услуг и приложений 6G, таких как определение местоположения и сканирование с высоким разрешением, а также для удовлетворения более высоких требований к пропускной способности в 2030-х годах. Терагерцовый участок спектра также должен быть включен в сферу влияния 6G, поскольку он имеет большой потенциал использования для широкого перечня приложений связи и сканирования. На рис. 12.2 показаны потенциальные потребности в спектре и возможности 6G.



* Потенциальные кандидаты на ИМТ в 2030-е годы

Рис. 12.2. Потенциальные потребности в спектре и возможности 6G

12.3. Средние диапазоны – наиболее экономичный способ широкого охвата

По сравнению с более ранними системами ИМТ, требования 5G более разнообразны. Например, необходимо обеспечить как комфортную для пользователя скорость передачи данных (100 Мбит/с), так и плотность трафика в зоне (10 Мбит/(с·м²)). Поэтому сетевые операторы должны учитывать рентабельность широкого покрытия сети при поддержке большинства сценариев целевых приложений. Средние диапазоны частот от 3 до 5 ГГц играют решающую роль в 5G и, как ожидается, также будут иметь жизненно важное значение для 6G.

Стимулы использования дополнительного среднего диапазона

В системе 5G полностью осознается важность средних диапазонов (от 3 до 5 ГГц) и предлагается множество возможных частотных диапазонов, таких как 3300–3800 МГц, 2,6 ГГц и 2,3 ГГц. Однако из-за различий в правилах использования и распределения спектра в разных странах глобально в средних диапазонах могут быть выделены полосы менее 500 МГц. Следовательно, в случае сосуществования нескольких операторов один оператор может получить в среднем максимум 100 МГц непрерывного спектра в среднем диапазоне. Этого может хватить для начальной фазы развертывания 5G, но к 2030-м годам этого будет недостаточно, поскольку ожидается, что трафик вырастет в десятки или даже сотни раз. Согласно отчету [6], ежемесячные данные об использовании (data of usage, DoU) на пользователя на некоторых развитых рынках к 2025 году превысят 150 ГБ, что обусловлено ростом популярности видео сверхвысокой четкости (UHD) и приложений AR/VR. Ожидается, что DoU вырастет в 30 раз с 2019 по 2025 год (рис. 12.3) и продолжит расти до 2030-х годов.

Это означает, что потребуются дополнительный средний диапазон частот для поддержки устойчивого роста трафика для 6G к 2030 году. Исторически сложилось так, что каждое новое поколение системы ИМТ обеспечивает как минимум в 10 раз лучшую производительность, чем предыдущее поколение. Исходя из этого, мы определяем, что каждое новое поколение требует примерно в пять раз большей пропускной способности, чем предыдущее. Например, в 2G нужно было менее 1 МГц, но требуемый спектр для систем ИМТ увеличился до 5 МГц в 3G и достиг примерно 20 МГц в 4G. В 5G это значение увеличивается до 100 МГц в соответствии с «правилом пятикратного роста». Следуя этой тенденции, естественно предположить, что каждому оператору сети 6G потребуется средний диапазон частот 500 МГц, чтобы обеспечить устойчивую

бизнес-деятельность. В случае сосуществования нескольких операторов это означает, что дополнительный средний диапазон частот шириной как минимум 1–1,5 ГГц является оптимальной целью для достижения наилучшего компромисса между покрытием и пропускной способностью в 2030-х годах.

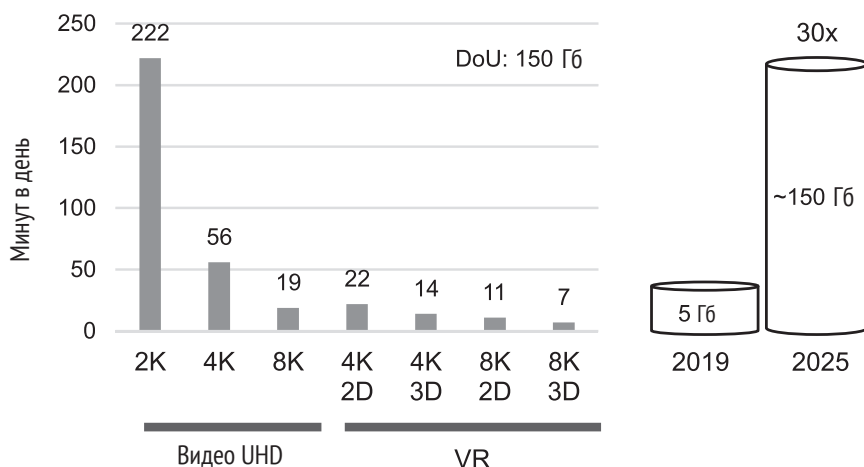


Рис. 12.3. Прогноз ежемесячного объема использования беспроводной сети к 2025 году. Источник данных [6]

Потенциальные кандидаты на дополнение среднего диапазона

Полосы 6 ГГц (т. е. 5925–7125 МГц) и 10 ГГц (т. е. 10–13,25 ГГц) являются убедительными кандидатами на дополнение среднего диапазона.

С точки зрения характеристик спектра, эти полосы частот могут поддерживать большие непрерывные блоки (теоретически доступная полоса пропускания составляет не менее 1 ГГц). Хотя затухание при распространении сигналов будет немного увеличиваться с переходом в диапазон более высоких частот, такое увеличение является приемлемым, особенно когда речь идет о передовых технологиях крупномасштабных массивов (например, массивный MIMO) наряду с более широким использованием промежуточных и радиочастотных (IF/RF) компонентов.

В табл. 12.1 показано увеличение затухания сигнала по сравнению с С-диапазоном 5G (т. е. 3,5 ГГц). Мы можем видеть, что увеличение потерь на трассе составляет примерно 5 и 9 дБ для диапазонов 6 и 10 ГГц соответственно по сравнению с диапазоном 3,5 ГГц в сценарии городских микроячеек (где радиус соты составляет 100 м), независимо от

того, рассматриваем мы LOS или NLOS. То же самое верно для сценария городской макросоты (где радиус соты составляет 500 м).

Таблица 12.1. Сравнение интервалов потерь на трассе между предполагаемыми средними полосами и 3,5 ГГц в сценариях городских макро- и микросот

Сценарии	6 ГГц		10 ГГц	
	Затухание LOS	Затухание NLOS	Затухание LOS	Затухание NLOS
Макросоты (радиус 500 м)	4,7	5,0	9,1	9,7
Микросоты (радиус 100 м)	4,7	5,0	9,1	9,7

С точки зрения политики регулирования радиосвязи эти полосы уже распределены подвижным службам на первичной или равной первичной основе в Регламенте радиосвязи [7], как показано в табл. 12.2. Действительно, ITU-R представил новый пункт повестки дня – пункт 1.2 повестки дня WRC-23 – для проведения технико-экономического обоснования некоторых из упомянутых ранее полос (например, 6425–7125 МГц и 10–10,5 ГГц) для будущей идентификации IMT [8].

Таблица 12.2. Распределение спектра мобильных служб в соответствии с правилами радиосвязи в диапазоне 6–11 ГГц

Полосы (ГГц)		Регионы ITU-R			Пункт повестки дня WRC-23 Полосы (ГГц) 1 2 3 для идентификации IMT
		1	2	3	
Полосы 6G	5,825–7,25	×	×	×	6,425–7,025 (R1)
	7,25–8,5	×	×	×	7,025–7,125 (глобальная)
Полосы 10G	10–10,45	×		×	10–10,5 (R2)
	10,5–10,68	×	×	×	
	10,7–11,7	×	×	×	
	11,7–12,2	×		×	
	12,2–12,5	×	×	×	
	12,5–12,75		×	×	
	12,75–13,25	×	×	×	

Следует также рассмотреть другие возможности использования спектра, помимо 6 ГГц и 10 ГГц в средних диапазонах, поскольку изуче-

ние диапазонов 3–11 ГГц для использования в приложениях 6G имеет важное значение.

12.4. Полосы миллиметрового диапазона в эпоху 6G

По сравнению с полосами нижнего диапазона, полосы миллиметрового диапазона имеют более жесткие характеристики распространения радиоволн. Например, миллиметровым волнам присущи более сильные потери в тракте (уменьшение дальности передачи), более сильные потери при проникновении (уменьшение мощности сигнала в помещениях) и разреженные кластеры (уменьшение ранга канала и, как следствие, снижение мультиплексирования многопользовательского MIMO). Таким образом, практическая реализация связи в миллиметровом диапазоне сталкивается с большими проблемами в области покрытия большой площади из-за ограничений покрытия и мобильности, вызванных блокирующим эффектом распространения миллиметровых волн. Обычно считается, что диапазон миллиметровых волн не является рентабельным для широкого покрытия eMBB – скорее, он лучше подходит для дополнения емкости в сценариях ограниченных приложений, таких как городские точки доступа, с экстремальной плотностью трафика и сверхвысокими скоростями передачи данных.

Однако разработка и внедрение диапазонов миллиметровых волн для 5G приведет к тому, что миллиметровая технология и связанная с ней экосистема станут более зрелыми. Что касается дальности передачи и энергопотребления базовых станций, а также времени автономной работы пользовательского оборудования, коммерциализация и широкое использование технологий миллиметрового диапазона в конечном итоге обеспечат развертывание макроячеек в эпоху 6G.

Новые стимулы внедрения миллиметровых волн

Помимо обеспечения сверхвысоких скоростей передачи данных для многих приложений 6G, миллиметровые волны будут играть ключевую роль в интегрированном сканировании и передаче данных для 6G.

С точки зрения связи, как обсуждалось в главе 1, 6G будет первой системой мобильной связи, которая достигает пиковых скоростей передачи данных в диапазоне Тбит/с. Однако даже увеличенного спектра в районе низких частот (например, в средних диапазонах) будет недостаточно для достижения этой цели. Рассматривая текущий спектр, выделенный для мобильной связи в Европе, как показано на рис. 12.4, мы видим, что существующая полоса пропускания унаследованных диапазонов 5G (то есть низких и средних диапазонов) составляет около 1 ГГц.

Даже с дополнительным среднечастотным спектром 1,5 ГГц к 2030 году будет доступно только 2,5 ГГц спектра. При этом если диапазоны миллиметровых волн, определенные на WRC-19, постепенно начнут использоваться для IMT-2030 (на рис. 12.4 мы предполагаем, что используется примерно половина общей полосы 14,75 ГГц), прирост доступной полосы пропускания будет значительным – в семь раз больше, чем в 5G. Следовательно, полосы миллиметрового диапазона будут играть важную роль в заполнении таких пробелов в пропускной способности.

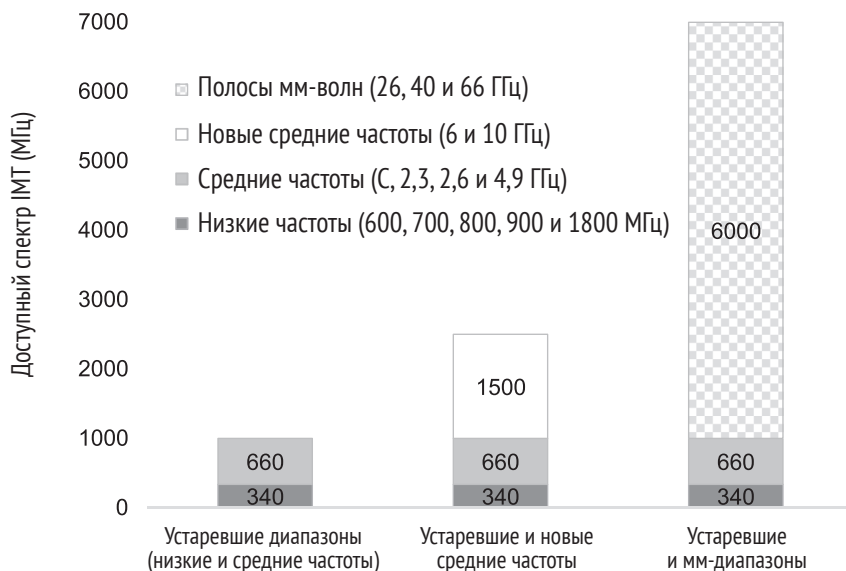


Рис. 12.4. Увеличение спектра подвижной связи к 2030 году

Новой возможностью 6G является сканирование окружающей среды (включая точное позиционирование, визуализацию и т. д.), и диапазон миллиметровых волн обеспечивает достаточный спектр для достижения сантиметрового разрешения сканирования. Согласно теории электромагнитного изображения, три аспекта разрешения сканирования – разрешение по дальности, угловое разрешение и разрешение смещения (cross-range) – можно рассчитать с помощью следующих формул [9]:

$$\text{Разрешение по дальности} = \frac{\text{Скорость света}}{2 \times \text{Полоса спектра}};$$

$$\text{Угловое разрешение} = 1,22 \times \frac{\text{Длина волны}}{\text{Апертура антенны}}.$$

Затем разрешение поперечного смещения (поперечное разрешение) можно рассчитать как

$$2d \times \tan(0,5 \times \text{Угловое разрешение}),$$

где d – расстояние сканирования.

Используя предыдущие формулы, мы можем сделать вывод, что для достижения разрешения в 10-сантиметровом диапазоне требуется непрерывная полоса пропускания не менее 1,5 ГГц (т. е. $3 \times 10^8 / 0,5 \times 10 \times 10^{-2}$). Для еще более высокого разрешения требуется еще большая полоса пропускания, которую средний диапазон не может обеспечить.

Чтобы достичь поперечного разрешения 10-сантиметрового уровня, угловое разрешение должно составлять $0,01^\circ$ для расстояния сканирования 10 м. При использовании миллиметровых волн с частотой около 60 ГГц апертура антенны, необходимая для достижения такого углового разрешения, составляет 0,6 м. И наоборот, использование средних частот, таких как 6 ГГц, приводит к увеличению необходимой апертуры антенны до 6 м. На практике это сложно реализовать.

Новые технологии для спектра миллиметровых волн

В эпоху 6G технические проблемы, выявленные в 5G, такие как мобильность и покрытие, а также проблемы реализации оборудования, будут решаться с помощью более совершенных методов. Например, в 6G потребуются моделирование каналов с более высокой точностью, чтобы облегчить оптимальное проектирование радиотехнологий (дополнительные подробности представлены в главе 13). Кроме того, для улучшения характеристик передачи миллиметровых волн необходимы усовершенствования в технологии формирования луча, в частности в формировании луча с помощью сканирования (более подробная информация представлена в главе 25). Обнадешивает то, что использование спектра миллиметрового диапазона будет значительно улучшено по мере того, как отрасли и технологии, связанные с материалами, радиочастотными компонентами и обработкой сигналов, станут более зрелыми. Таким образом, мы сможем достичь сверхвысокой скорости передачи данных и высокоточного разрешения сканирования.

Полосы-кандидаты для спектра миллиметровых волн

На WRC-19 полосы частот 14,75 ГГц (24,25–27,5, 37–43,5 и 66–71 ГГц) и 2,5 ГГц (45,5–47 и 47,2–48,2 ГГц) были определены для ИМТ в качестве полос глобального и регионального согласования соответственно. Однако следует отметить, что эти полосы не являются смежными; скорее, они разбросаны по нескольким областям в диапазоне от 24 до 71 ГГц. Совместное использование нескольких миллиметровых диапазонов, охватывающих десятки ГГц (например, 26 и 39 ГГц, 39 и 66 ГГц), будет важным, но чрезвычайно сложным, если мы хотим достичь скоростей передачи данных в диапазоне Тбит/с или сканирования с высоким разрешением сантиметровой точности.

Следовательно, на будущих собраниях WRC может возникнуть необходимость определить новые полосы миллиметрового диапазона, поддерживающие большие непрерывные блоки. Например, полосы Е-диапазона (71–76 ГГц и 81–86 ГГц) являются основными кандидатами, поскольку обеспечивают 10 ГГц непрерывного спектра при совместном использовании диапазона 66–71 ГГц (который уже был выделен). В настоящее время полоса Е 71–76 ГГц выделена мобильным службам на равной первичной основе с фиксированными службами или транзитными линиями связи. По мере развития технологий интегрированного доступа и транзитного соединения (integrated access and backhaul, IAB) мы ожидаем, что проблемы с разделением спектра между этими двумя услугами будут надлежащим образом решены.

12.5. Новые возможности для сканирования и связи в ТГц-диапазоне

ТГц-диапазон – это следующая возможность использования спектра для беспроводной связи 6G. По сравнению с низкочастотными диапазонами, ТГц-диапазон имеет очевидные преимущества для связи с очень высокой скоростью передачи данных и сканирования со сверхвысоким разрешением.

Одной из наиболее примечательных особенностей ТГц-диапазона является его способность обеспечивать сверхширокую полосу пропускания и достаточный ресурс спектра для предстоящего 6G. На WRC-19 подвижной и фиксированной службам было выделено в общей сложности 137 ГГц в разных полосах частот в ТГц-диапазоне от 275 до 450 ГГц (т. е. 275–296, 306–313, 318–333 и 356–450 ГГц) [4]. Таким образом, совокупный объем спектра, выделенного мобильным службам в ТГц-диапазоне от 100 до 450 ГГц, достиг более чем 230 ГГц с учетом частот, выделенных на предыдущих WRC. В табл. 12.3 показано, как был распределен спектр. Обеспечивая полосу пропускания, превышающую десятки или даже сотни ГГц, можно будет достичь широкого диапазона, достаточного для пиковых скоростей передачи данных и приложений, чувствительных к задержкам.

Таблица 12.3. Спектр подвижной связи, распределенный в ТГц-диапазоне 100–450 ГГц

Полоса частот (ГГц)	Ширина полосы (ГГц)	Полоса частот (ГГц)	Ширина полосы (ГГц)
102–109.5	7,5	252–275	23
141–148.5	7,5	275–296*	21
151.5–164	12,5	306–313*	7

Полоса частот (ГГц)	Ширина полосы (ГГц)	Полоса частот (ГГц)	Ширина полосы (ГГц)
167–174.8	7,8	318–333*	15
191.8–200	8,2	356–450*	94
209–226	17		

Примечания: 1. *Новые полосы, определенные в пункте 1.15 повестки WRC-19.

2. Фрагментарные полосы частот шириной менее 5 ГГц не показаны.

Возможность связи в ТГц-диапазоне

Хотя сигналы в ТГц-диапазоне затухают сильнее, чем миллиметровые волны, все же есть возможности найти подходящие диапазоны для передачи. Мы оценили теоретическую пропускную способность канала передачи в ТГц-диапазоне, принимая во внимание мощность передачи и шумы компонентов современных технологий, а также влияние атмосферного затухания [10]. Результаты представлены на рис. 12.5.

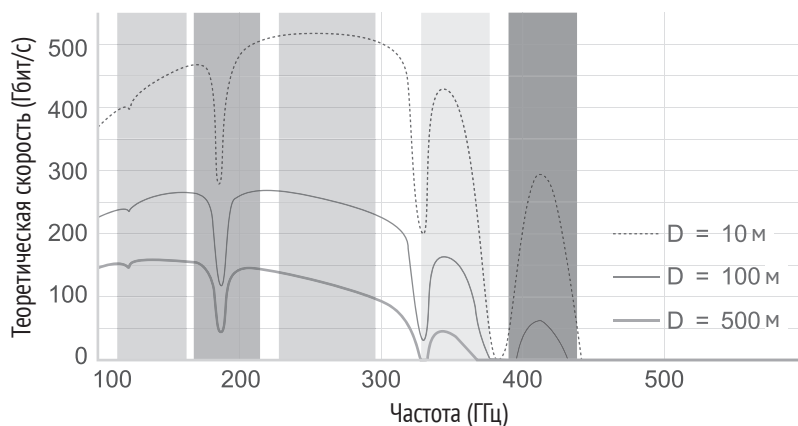


Рис. 12.5. Оценка теоретической пропускной способности канала передачи в ТГц-диапазоне (D – расстояние передачи)

Из этих результатов можно сделать вывод, что количество потенциальных окон частотных диапазонов с лучшими характеристиками распространения относительно велико (например, 140, 220 и 300 ГГц). Эти окна частотных диапазонов можно использовать для передачи на средние (например, 200 м) или короткие (менее 10 м) расстояния, избегая частотных областей с высоким атмосферным поглощением. Кроме того, длины волн в ТГц-диапазоне намного меньше, чем в диапазоне миллиметровых волн, и в одной и той же области на кристалле может быть размещено больше антенн, чтобы преодолеть затухание при распространении, тем самым улучшая покрытие.

Возможность интегрированного сканирования сверхвысокого разрешения

Как упоминалось ранее в этой главе, сканирование в ТГц-диапазоне может обеспечить гораздо более высокую точность и разрешение благодаря сверхширокой полосе пропускания. Кроме того, изображение, полученное в ТГц-диапазоне, демонстрирует улучшенное пространственное разрешение по сравнению с более низкочастотными полосами благодаря более коротким длинам волн и меньшему размеру оборудования. Это значительно расширяет область применения терагерцовых технологий в устройствах мобильной связи. Например, в будущем интеллектуальные устройства, поддерживающие технологию сканирования в ТГц-диапазоне, будут способны формировать электромагнитные изображения для получения такой информации, как количество калорий в пище, или для обнаружения скрытых объектов (более подробная информация представлена в главе 3).

В отличие от других диапазонов электромагнитного излучения, ТГц-диапазон обладает уникальными возможностями менее опасного сканирования, поскольку это неионизирующее излучение с малой глубиной проникновения [11]. ТГц-спектрометрия в 6G имеет множество потенциальных применений в области здравоохранения, промышленности, контроля качества пищевых продуктов и окружающей среды, поскольку частоты вибрации и вращения большинства молекул лежат в ТГц-диапазоне. ТГц-спектрометрия в будущем вызовет значительный интерес благодаря своей способности предоставлять непрерывную информацию в реальном времени посредством динамических, неинвазивных, пассивных и бесконтактных измерений. Особо следует отметить, что ТГц-спектрометрия дает результаты, сопоставимые с профессиональными аппаратами КТ или МРТ, но она намного безопаснее и портативнее, как показано на рис. 12.6.

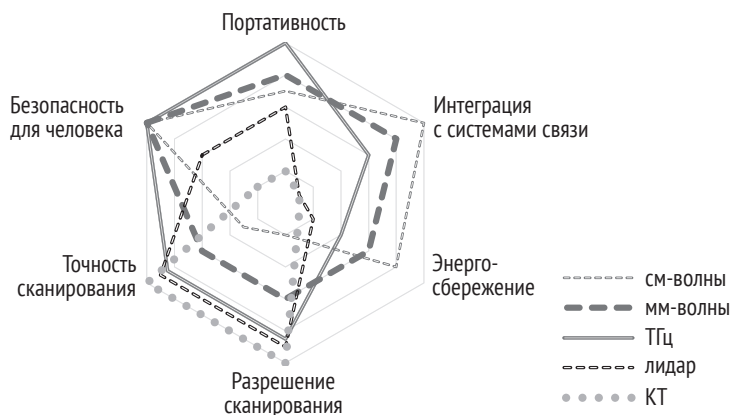


Рис. 12.6. Сравнение возможностей сканирования

Хотя ТГц-диапазон обеспечивает более широкую полосу пропускания и лучшие условия апертуры антенны, диапазоны миллиметровых волн более приспособлены к несовершенным условиям окружающей среды, таким как пыль, туман и т. д., что делает их подходящим вариантом для приложений сканирования окружающей среды вне помещений, таких как картографирование и электромагнитная реконструкция изображений.

Проблемы и перспективы

Для поддержки быстрого развития технологий связи в ТГц-диапазоне необходимы исследования и разработки, ведущие к прорыву в создании новых устройств, такие как проектирование электронных, фотонных и гибридных приемопередатчиков, крупногабаритные антенные решетки, решетки на кристалле или на корпусе и новые технологии материалов для изготовления решеток. В то же время необходимы дальнейшие исследования в области проектирования мощных высокочастотных устройств, новых материалов для антенн и РЧ-транзисторов, архитектур приемопередатчиков, моделирования каналов, обработки сигналов решетки и вопросов энергоэффективности.

Хотя ITU-R выделил полосы свыше 230 ГГц для подвижных служб в диапазоне частот 100–450 ГГц, правила и политика для отрасли ИМТ не ясны и еще не унифицированы на международном уровне. Собрания на уровне ITU и WRC должны работать вместе, чтобы способствовать достижению консенсуса.

12.6. Источники

- [1] ITU-R, *IMT Vision – framework and overall objectives of the future development of IMT for 2020 and beyond*, Recommendation ITU-R M. 2083-0, Sept. 2015.
- [2] GSA Report, *5G spectrum for terrestrial networks: Licensing developments worldwide*, 2020, accessed Feb. 2020. <https://gsacom.com/paper/5g-spectrumreport-for-terrestrial-networks-executive-summary-feb-2020/>.
- [3] ITU. World Radiocommunication Conference, *Final Acts WRC-15*, 2015. <https://www.itu.int/pub/R-ACT-WRC.12-2015>.
- [4] ITU. World Radiocommunication Conference, *Final Acts WRC-19*, 2019. <https://www.itu.int/pub/R-ACT-WRC.14-2019>.
- [5] ITU-R, *IMT future technology trends towards 2030*, ITU-R, M. [Future technology trends] (ongoing developing), 2020.
- [6] White paper, *5G spectrum: Public policy position*, 2020. https://www-file.huawei.com/-/media/corporate/pdf/public-policy/public_policy_position_5g_spectrum_2020_v2.pdf?la=en-gb.

- [7] *Radio Regulations (Edition of 2016)*. ITU-R publications, International Telecommunications Union, Geneva, Switzerland, 2016.
- [8] *Results of the First Session of the Conference Preparatory Meeting for WRC-23 (CPM23-1)*. BR Administrative Circular CA/251, Geneva, Switzerland, 2019.
- [9] J. Hasch, E. Topak, R. Schnabel, T. Zwick, R. Weigel, and C. Waldschmidt, *Millimeterwave technology for automotive radar sensors in the 77 GHz frequencyband*, IEEE Transactions on Microwave Theory and Techniques, vol. 60, no. 3, pp. 845–860, 2012.
- [10] ITU-R, *Attenuation by atmospheric gases and related effects*, Recommendation ITU-R, p. 672–12, Aug. 2019.
- [11] I. F. Akyildiz, J. M. Jornet, and C. Han, *Terahertz band: Next frontier for wireless communications*, Physical Communication, vol. 12, pp. 16–32, 2014.

Глава 13

Новые каналы

Распространение радиоволн – это основа основ беспроводной связи. Прежде чем строить и эксплуатировать реальные системы, мы должны понять принципы распространения радиоволн и разработать соответствующие модели каналов. Эти модели отражают ключевые процессы распространения и позволяют надежно оценивать и сравнивать различные системы. Каждое поколение беспроводной системы использует разные полосы частот для коммерческих целей. Следовательно, модель канала, связанная именно с этими полосами, является неотъемлемой частью развития технологий. До 5G диапазон частот составлял от 0,4 ГГц до 100 ГГц [1, 2].

Стандарты 3GPP TR 38.901 содержат важные методы моделирования каналов, удовлетворяющие требованиям, которые возникли с появлением новых сценариев и полос спектра в 5G. Например, бюджет канала является важным показателем для оценки потерь на трассе в полосах частот выше 6 ГГц. По этой причине в модели потерь на трассе за пределами 6 ГГц мы принимаем во внимание частотную зависимость [3]. Другой пример – технология массивного MIMO, которая требует целенаправленных усилий по формированию луча, поскольку это становится ключевым методом 5G. Пространственная согласованность определена как дополнительный параметр для работы с изменением угла в модели канала при формировании диаграммы направленности и трекинге [4, 5]. Стандарты также определяют и моделируют множество других новых характеристик распространения, включая перегораживание пути, потери проникновения из помещения в помещение [6], зависимость от полосы пропускания [7] и потери на поглощение кислородом и другими молекулами воздуха. Использование этих новых характеристик канала гарантирует, что мы можем всесторонне и точно оценить влияние пространственного распространения при разработке новых технологий 5G.

Кроме того, при проектировании системы и технологии 6G нам необходимо использовать правильную модель канала. Это связано с тем, что при моделировании каналов 6G неизбежно возникают серъ-

езные проблемы, связанные с новыми элементами системы, такими как новый спектр, сценарии использования и антенны, которые будут обсуждаться в других главах. Как упоминалось в статье [8], в ТГц-диапазоне применяется другой коэффициент потерь в тракте, чем в миллиметровом диапазоне. Также было введено понятие *пространственного нестационарного канала*, ориентированное на решетку с очень большой апертурой (ELAA) [9]. С развитием новых технологий частью модели распространения стала модель фазовращателя с угловой зависимостью, включенная в *реконфигурируемую интеллектуальную поверхность* (reconfigurable intelligent surface, RIS) [10]. Кроме того, в сценариях сканирования, таких как сканирование окружающей среды с помощью отражения луча, вместо формулы распространения, используемой в традиционной связи, должна быть включена формула распространения отраженного сигнала на основе принципов радара. Эти проблемы влияют на моделирование каналов и требуют не только модификации формул и параметров. Кроме того, традиционные методологии моделирования могут не соответствовать новым характеристикам каналов.

13.1. Новые требования к моделированию канала 6G

Если взглянуть на модель канала с исторической точки зрения, то модели физического канала, которые используют многолучевое двуправленное распространение электромагнитных волн для извлечения характеристик физической среды, можно разделить на три типа: детерминированная модель, стохастическая модель на основе геометрии (geometry-based stochastic model, GBSM) и негеометрическая стохастическая модель [11]. В детерминированной модели физические параметры распространения полностью фиксированы. Следовательно, в конкретных сценариях могут быть реконструированы реальные физические каналы. Детерминированные модели включают СЕМ, модели трассировки лучей и измерения. GBSM – это беспроводной канал, моделируемый посредством распределения рассеивающих кластеров, которые генерируются случайным образом с помощью заданной функции плотности вероятности. Благодаря своему стохастическому характеру GBSM лучше описывает распространение в группе физических сред, чем детерминированные модели, и вполне подходит для статистической оценки, такой как системное моделирование. Модель 3GPP – это типичный пример GBSM. Квазидетерминированная модель строится с использованием комбинации детерминированных моделей и стохастических моделей. Доминирующие пути рассчитываются с помощью

детерминированной модели, а пути рассеяния генерируются с помощью стохастических моделей.

На основе измерений и трассировки лучей в полосах частот от 6 ГГц до 100 ГГц исследователи в области 5G сформировали модели в полосах более высоких частот до 100 ГГц, как показано в табл. 13.1. Основными методиками моделирования каналов в этих стандартах являются стохастическое и квазидетерминированное моделирование. Эти методики отражают суть статистических моделей замираний, а также некоторые из фактически измеренных характеристик распространения для определенных конструкций антенн, таких как кросс-поляризационные.

Таблица 13.1. Методики стандартизации каналов 5G

Модель канала	Организация	Методика
3GPP 38.901	3GPP	Стохастическая модель
METIS	METIS	Стохастическая, на основе карты или гибридная модель
MiWEBA	MiWEBA	Квазидетерминированная модель канала
ITU-R M	ITU	Стохастическая модель
COST2100	COST	Стохастическая модель
IEEE 802.11	IEEE	Квазидетерминированная или трассировка лучей
NYU WIRELESS	NYU	Стохастическая
QuaDRiGa	Fraunhofer HHI	Стохастическая или квазидетерминированная
5G mmWave	NIST	Стохастическая
Channel Model Alliance		
mmMAGIC	mmMAGIC	Стохастическая или строгая проверка
IMT-2020	IMT-2020	Стохастическая

Статистические модели могут описывать канал более просто и эффективно и полезны для крупномасштабного моделирования из-за их низкой вычислительной сложности. Мы также обнаружили, что стохастические модели не могут выразить детерминированные параметры, относящиеся к конкретной системе или сценарию, такие как геометрическая информация, относящаяся к параметрам многолучевого канала или расположению устройств связи и рассеивателей. Однако детерминированное выражение канала не было требованием для проектирования каналов и систем на протяжении последних десятилетий. Потребность в таком моделировании в основном возникает при планировании сети, когда геометрическая схема размещения сотовых узлов влияет на конфигурации базовых станций.

В новой сети 6G определенные методы (такие как RIS, локализация и визуализация) тесно связаны с конкретными средами, которые не могут быть воссозданы с помощью стохастических моделей. Поэтому мы ожидаем появления детерминированных методик моделирования каналов, которые могут привести к более точной оценке.

Например, в связи с разработкой новых антенн и технологий интеграции мы предполагаем, что большие антенные решетки, такие как EAA, значительно повлияют на моделирование каналов и оценку производительности. Как показано в некоторых статьях [12], большие антенные решетки создают новые проблемы при моделировании, такие как сферические волны ближнего поля и нестационарные каналы [13]. Раньше мы просто моделировали дальнее поле, которое можно было аппроксимировать с помощью полных волн. Однако ближнее поле теперь слишком велико, чтобы его можно было игнорировать, и нам нужно учитывать сферические волны. Эти особенности распространения открывают новые возможности для улучшения пропускной способности каналов. Пространственно-временные характеристики каналов с множеством антенн являются ключевыми факторами, определяющими производительность пространственно-временной обработки. Следовательно, исследования каналов с большой антенной решеткой должны быть сосредоточены на пространственно-временных характеристиках каналов с множеством антенн. Следует отметить, что эти характеристики сильно зависят от окружающей среды, и это особенно верно для распределений рассеяния, которые трудно описать с помощью стохастических моделей. Кроме того, эти характеристики значительно увеличивают сложность моделирования. В свете этого будущие исследования должны быть сосредоточены на том, чтобы сначала определить, как эти характеристики канала влияют на производительность связи, после чего можно будет определить степень, в которой эти характеристики следует моделировать в канале.

Кроме того, как описано в главе 3, существуют новые сценарии использования 6G (например, сканирование). Алгоритм и производительность таких сценариев сильно зависят от местоположения цели и окружающей среды. Следовательно, предпочтение отдается детерминированным моделям, связанным с географическим положением. Но это еще не все; типичное применение сканирования и визуализации должно учитывать эффекты распространения (такие как дифракция), когда размер объекта приблизительно равен длине волны, и это трудно смоделировать с помощью обычного метода геометрической оптики. Как показано на рис. 13.1, ожидается, что методология вычислительного электромагнетизма (computational electromagnetic, CEM) будет описывать физические явления, такие как дифракция. Некоторые технологии, используемые в CEM, такие как метод конечных элементов, метод

конечных разностей во временной области и метод моментов, описаны в главе 8.

Из предыдущих рассуждений следует, что одна схема сама по себе может не соответствовать требованиям оценки всех приложений. То есть модели с более высокой точностью и детерминированными уровнями привнесут более высокую вычислительную сложность, которая, в свою очередь, намного превзойдет возможности оценки системы. Поэтому мы представляем эволюцию моделей распространения в соответствующих стандартах от 1G до 5G. Как показано на рис. 13.2, исследования модели канала исторически имели тенденцию фокусироваться на улучшении детерминированных уровней в условиях ограничения сложности. Следовательно, это может быть потенциальным направлением развития гибридной модели, в которую включено множество механизмов, но веса механизмов варьируются в соответствии с различными сценариями применения и критериями оценки.

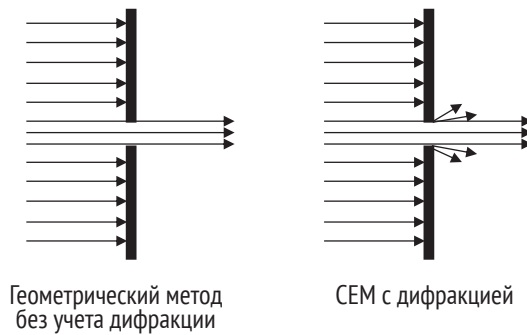


Рис. 13.1. Сравнение детерминистических моделей дифракции

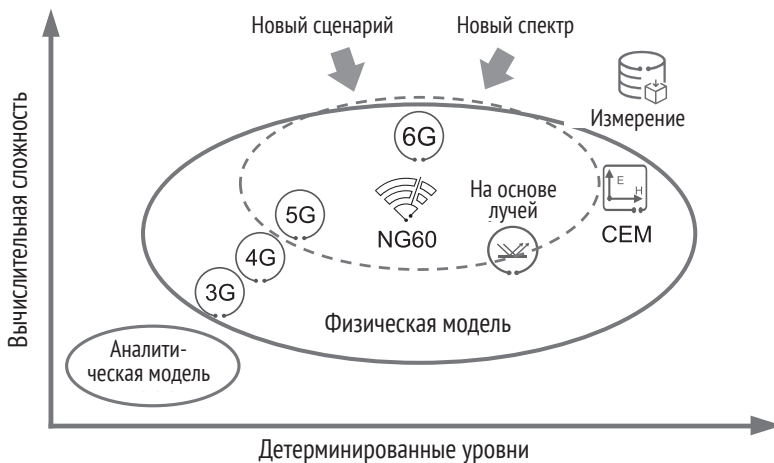


Рис. 13.2. Эволюция методологий моделирования каналов.
Пунктирный эллипс относится к 6G

13.2. Канальные измерения в 6G

13.2.1. Канальные измерения в новом спектре

По сравнению с полосой частот ниже 100 ГГц, терагерцевой полосе присущи более высокие потери на трассе луча в свободном пространстве. Кроме того, ТГц-волны возбуждают молекулы газа в атмосфере. Часть мощности сигнала будет преобразована в кинетическую энергию молекул газа, и это называется *молекулярным поглощением*. Оно обладает высокой частотной избирательностью, поскольку разные молекулы газа имеют разные резонансные частоты, которые соответствуют разным пикам молекулярного поглощения, как показано на рис. 13.3. В ТГц-диапазоне преобладает эффект поглощения водяного пара, тогда как эффект поглощения кислорода преобладает в диапазоне миллиметровых волн. Согласно закону Бера–Ламберта, потери на молекулярное поглощение экспоненциально возрастают с увеличением расстояния передачи. В результате эффект молекулярного поглощения приводит к выраженным частотно-избирательным и зависящим от расстояния потерям на молекулярное поглощение в каналах ТГц-диапазона.

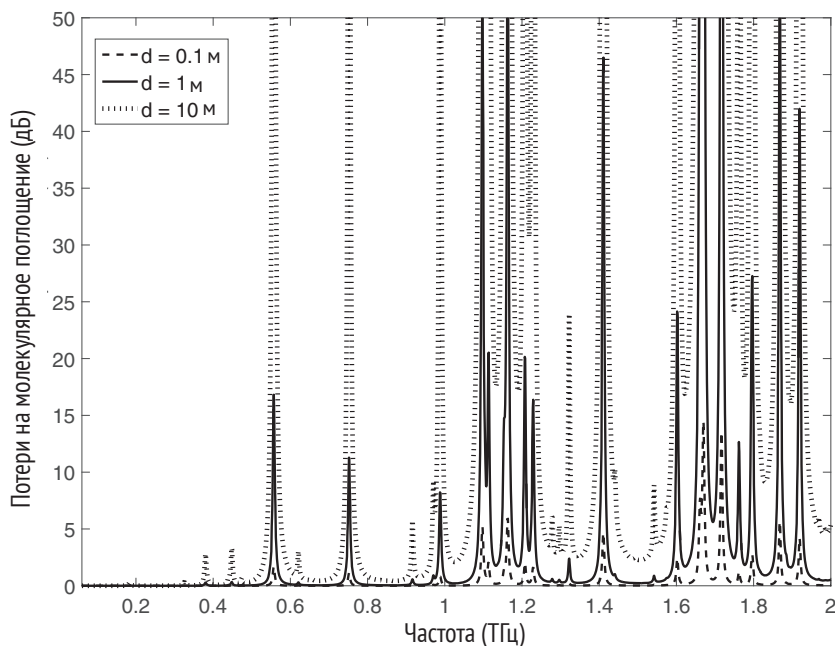
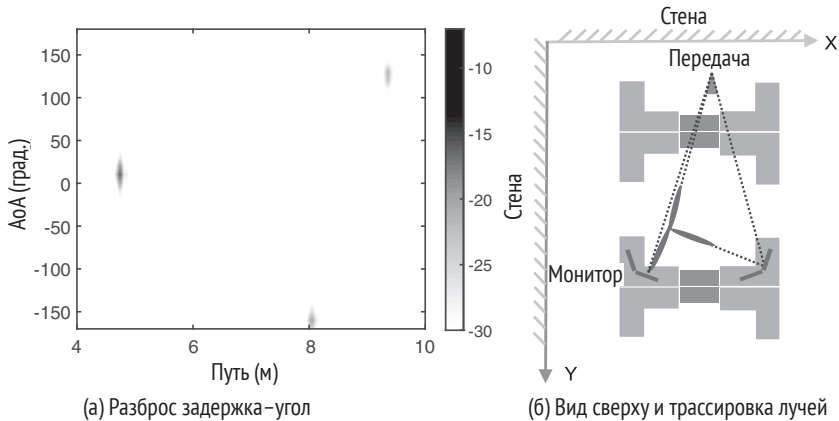


Рис. 13.3. Потери на молекулярное поглощение в ТГц-диапазоне

С другой стороны, канал ТГц-диапазона также демонстрирует характеристики распространения, отличные от канала миллиметрового диапазона. Обычно компонент многолучевого распространения состоит из

путей прямого распространения, путей зеркального отражения и путей рассеяния. Высокий К-фактор (в среднем 13 дБ в конференц-зале на частоте 140 ГГц согласно измерениям Huawei [8]) показывает соотношение мощностей между путями прямого и непрямого распространения. Измерения терагерцовых каналов показывают, что в них более выражено прямое распространение. Поверхность, которая является гладкой на более низких частотах, становится шероховатой в ТГц-диапазоне, поскольку волны ТГц-диапазона имеют длину, сопоставимую с высотой микронеровностей поверхности. Эта шероховатость приводит к потерям при отражении, а также рассеивает мощность по путям рассеяния, что в конечном итоге ослабляет многолучевое распространение и делает канал разреженным. Тем не менее некоторые измерения показывают, что в нижней части ТГц-диапазона, например 140 ГГц и 220 ГГц, многолучевое распространение все еще присутствует в помещении.



(в) Сценарий эксперимента

Рис. 13.4. Результаты и сценарий измерения многолучевого распространения на частоте 140 ГГц

Например, в офисном помещении на частоте 140 ГГц принимаемая мощность отраженного сигнала при многолучевом распространении всего на 6–7 дБ ниже, чем мощность прямого сигнала из-за сильного отражения от настольных жидкокристаллических мониторов (как показано на рис. 13.4).

Более того, различные распределения многолучевого распространения приносят новые факторы в сильно- и малозначащие параметры модели в ТГц-диапазоне. Например, для оценки модели потерь в трассе были проведены измерения распространения сигнала типичной внутренней точки доступа (indoor hotspot, InH), расположенной в коридоре внутри офиса, на частоте 140 ГГц, как показано на рис. 13.5 и в табл. 13.2. Было проведено сравнение со значением, найденным по формуле потерь в тракте 3GPP TR 38.901 InH, расширенной до 140 ГГц. Измерение показало степень потерь в тракте (path loss exponent, PLE) немного выше, чем в модели 3GPP, и ниже, чем PLE при распространении в свободном пространстве. Это означает, что слабый волноводный эффект можно наблюдать и в субТГц-диапазоне. Для малозначащих параметров, полученных в типичных сценариях InH, измерения также указывают на заметное отличие от существующих стандартов модели канала [14], как показано в табл. 13.3. Поэтому более точные и репрезентативные параметризованные модели каналов для нового спектра выше 100 ГГц еще предстоит глубоко исследовать и смоделировать.

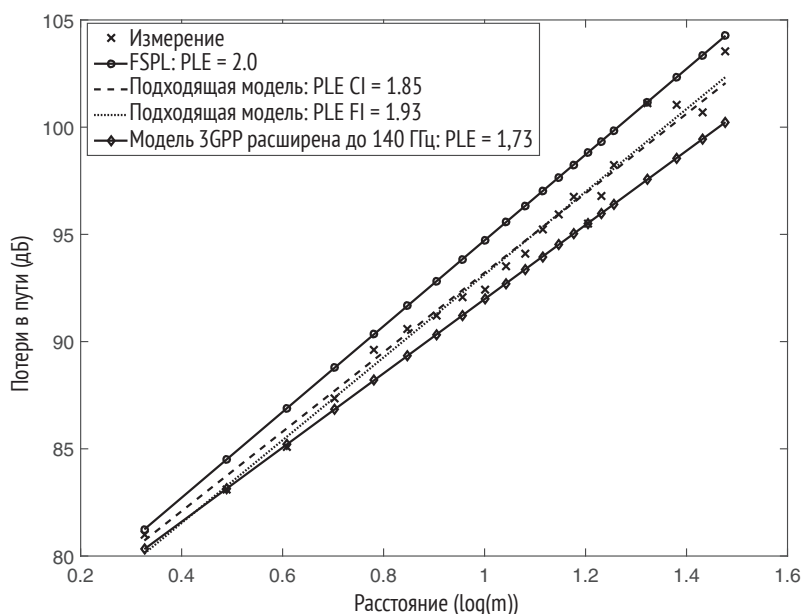


Рис. 13.5. Сравнение результатов измерений модели потерь на трассе, в свободном пространстве и модели 3GPP TR 38.901

Таблица 13.2. Потери в тракте на частоте 140 ГГц в сценариях InH

Потери при распространении (дБ)		Затухание в тени (дБ)
Измерения Huawei	$PL_{InH140} = 32,4 + 18,5 \log_{10}(d_{3D}) + 20\log(f_{140})$	$\delta_{SF} = 0,8$
Модель 38.901, расширенная до 140 ГГц	$PL_{InH140} = 32,4 + 17,3 \log_{10}(d_{3D}) + 20\log(f_{140})$	$\delta_{SF} = 3$

Таблица 13.3. Малозначащие параметры на частоте 140 ГГц в сценариях InH

Сценарий		Прямое распространение в офисе	
		Модель 3GPP, расширенная до 140 ГГц	Измерение
Разброс задержки (DS)	μ_{lgDS}	–7,71	–8,70
$lgDS = \log_{10}(DS/1s)$	δ_{lgDS}	0,18	0,50
Разброс AOA (ASA)	μ_{lgDS}	1,37	1,29
$lgASA = \log_{10}(ASA/1^\circ)$	δ_{lgDS}	0,37	0,30
Разброс ZOA (ZSA)	μ_{lgZSA}	0,88	0,73
$lgZSA = \log_{10}(ZSA/1^\circ)$	δ_{lgZSA}	0,18	0,15

13.2.2. Канальные измерения в новых сценариях

Как было сказано в разделе 13.1, новые требования к мобильной связи 6G привели к исследованию новых технологий, таких как ELAA, RIS, связь устройство–устройство в миллиметровом диапазоне, интегрированная неназемная связь в миллиметровом диапазоне и сканирование в миллиметровом и ТГц-диапазоне.

Моделирование каналов с ELAA

Естественно, новые сценарии применения и архитектуры антенны также создают новые проблемы для моделирования каналов. Например, ELAA может эффективно увеличить пропускную способность канала и пиковую скорость пользователя. Однако по мере увеличения апертуры антенной решетки увеличивается и расстояние ближней зоны, и это нельзя игнорировать, поскольку теперь некоторые пользователи будут находиться в области ближнего поля. По этой причине модель канала с плоской волной, используемая в 5G, может оказаться неприменимой. Таким образом, нам необходимо всесторонне смоделировать и изучить характеристики распространения нового канала со сферическими волнами и метод оценки канала.

Другой характерной особенностью канала распространения ЕЛАА является нестационарность в пространственных областях. Другими словами, по мере увеличения апертуры антенных решеток разные под-решетки могут иметь разные характеристики канала распространения, такие как мощность, кластеры, угол прихода и ранг.

Моделирование каналов неназемных сетей

Некоторые новые сценарии применения, такие как интегрированная неназемная связь, в свою очередь, создали большие проблемы для систем тестирования каналов. Показательным примером является тот факт, что обычное измерение канала требует высокого отношения сигнал/шум и высокой точности синхронизации, чего трудно достичь в случае неназемных сетей. Первое требуется, потому что низкое отношение сигнал–шум затрудняет получение информации о пространственном многолучевом распространении. Кроме того, системы тестирования интегрированных неназемных каналов также должны учитывать влияние атмосферных условий, таких как ветер, облака, дождь и снег.

Сканирование в миллиметровом и ТГц-диапазоне

Помимо применения в области связи, важным направлением исследований частотного диапазона 6G являются также различные применения сканирования. Благодаря применению более широкого спектра устройства связи могут использовать более широкую полосу пропускания и более интегрированные антенны. В результате эти устройства могут обеспечивать сверхточное пространственное и временное разрешение, дополнительно улучшая возможности выбора каналов. Особенно это актуально в ТГц-диапазоне частот. Другими словами, благодаря маленькой длине волны он может обеспечивать визуализацию на миллиметровом уровне. При этом мы по-прежнему не можем использовать текущую методологию моделирования каналов для точной оценки качества изображения. Например, как показано на рис. 13.6, для небольшого отверстия, размер которого сопоставим с длиной волны, нельзя использовать традиционный метод геометрической оптики для моделирования явления дифракции электромагнитной волны и, следовательно, невозможно точно оценить характеристики изображения. В этом контексте потенциальным направлением исследований, на котором мы можем сосредоточиться, является методология СЕМ, которая может эффективно моделировать распределение поля электромагнитных волн и восстанавливать среду распространения канала с высокой точностью. В конечном итоге это поможет повысить точность формирования изображения. Как показано на рис. 13.6, измерительное оборудование с отраженным каналом 140 ГГц было использовано для

обнаружения металлических объектов с небольшими отверстиями (минимальный размер составляет 3,2 мм). Для получения электромагнитного изображения использовался алгоритм обратной задачи рассеяния на основе СЕМ. Полученное изображение в основном соответствует оптическому изображению, как показано в правой части рисунка. Минимальное разрешение – 3,2 мм. Эти результаты визуализации подтверждают возможность получения изображений миллиметрового масштаба в субТГц-диапазоне.

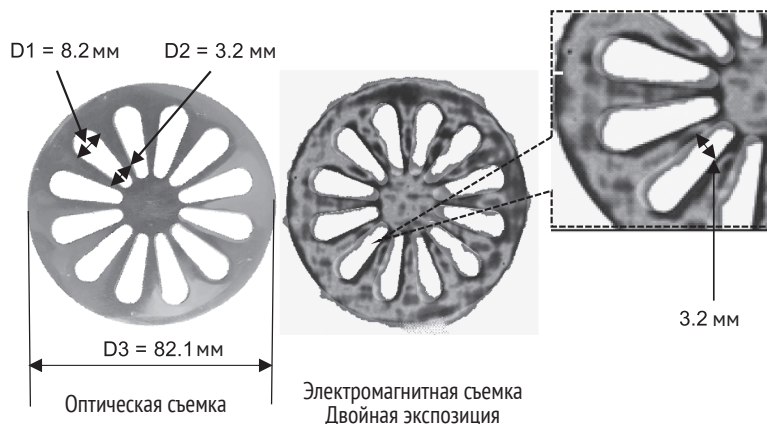


Рис. 13.6. Пример изображения миллиметрового уровня в субТГц-диапазоне, включающего небольшое отверстие с минимальным размером 3,2 мм

13.3. Источники

- [1] K. Haneda et al., *5G channel model for bands up to 100 GHz*, 2015, accessed Dec. 6, 2015. [http://www.5gworkshops.com/2015/5G_Channel_Model_for_bands_up_to100_GHz\(2015-12-6\).pdf](http://www.5gworkshops.com/2015/5G_Channel_Model_for_bands_up_to100_GHz(2015-12-6).pdf).
- [2] K. Haneda, J. Zhang, L. Tan, G. Liu, Y. Zheng, H. Asplund, J. Li, Y. Wang, D. Steer, C. Li et al., *5G 3GPP-like channel models for outdoor urban microcellular and macrocellular environments*, in Proc. 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring). IEEE, 2016, pp. 1–7.
- [3] H. Yan, Z. Yu, Y. Du, J. He, X. Zou, D. Steer, and G. Wang, *Comparison of large scale parameters of mmwave wireless channel in 3 frequency bands*, in Proc. 2016 International Symposium on Antennas and Propagation (ISAP). IEEE, 2016, pp. 606–607.
- [4] Y. Wang, Z. Shi, M. Du, and W. Tong, *A millimeter wave spatial channel model with variant angles and variant path loss*, in Proc. 2016 IEEE Wireless Communications and Networking Conference. IEEE, 2016, pp. 1–6.

- [5] K. Zeng, Z. Yu, J. He, G. Wang, Y. Xin, and W. Tong, *Mutual interference measurement for millimeter-wave D2D communications in indoor office environment*, in Proc. 2017 IEEE Globecom Workshops. IEEE, 2017, pp. 1–6.
- [6] Y. Du, C. Cao, X. Zou, J. He, H. Yan, G. Wang, and D. Steer, *Measurement and modeling of penetration loss in the range from 2 GHz to 74 GHz*, in Proc. 2016 IEEE Globecom Workshops. IEEE, 2016, pp. 1–6.
- [7] N. Iqbal, J. Luo, C. Schneider, D. Dupleich, S. Haefner, R. Müller, and R. S. Thomas, *Frequency and bandwidth dependence of millimeter wave ultra-wide-band channels*, in Proc. 2017 11th European Conference on Antennas and Propagation (EUCAP). IEEE, 2017, pp. 141–145.
- [8] Z. Yu, Y. Chen, G. Wang, W. Gao, and C. Han, *Wideband channel measurements and temporal-spatial analysis for terahertz indoor communications*, in Proc. 2020 IEEE International Conference on Communications Workshops. IEEE, 2020, pp. 1–6.
- [9] S. Wu, C.-X. Wang, H. Haas, M. M. Alwakeel, B. Ai et al., *A non-stationary wideband channel model for massive MIMO communication systems*, IEEE Transactions on Wireless Communications, vol. 14, no. 3, pp. 1434–1446, 2014.
- [10] W. Chen, L. Bai, W. Tang, S. Jin, W. X. Jiang, and T. J. Cui, *Angle-dependent phase shifter model for reconfigurable intelligent surfaces: Does the angle-reciprocity hold?* IEEE Communications Letters, 2020.
- [11] P. Almers, E. Bonek, A. Burr, N. Czink, M. Debbah, V. Degli-Esposti, H. Hofstetter, P. Kyösti, D. Laurenson, G. Matz et al., *Survey of channel and radio propagation models for wireless MIMO systems*, EURASIP Journal on Wireless Communications and Networking, vol. 2007, no. 1, p. 019070, 2007.
- [12] S. Wu, C.-X. Wang, H. Haas, M. M. Alwakeel, B. Ai et al., *A non-stationary wideband channel model for massive MIMO communication systems*, IEEE Transactions on Wireless Communications, vol. 14, no. 3, pp. 1434–1446, 2014.
- [13] J. Chen, X. Yin, X. Cai, and S. Wang, *Measurement-based massive MIMO channel modeling for outdoor LoS and NLoS environments*, IEEE Access, vol. 5, pp. 2126–2140, 2017.
- [14] 3GPP, «Study on channel model for frequencies from 0.5 to 100 GHz», 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.901, 12 2019, version 16.1.0. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3173>.

Глава 14

Новые материалы

Огромную эволюцию, наблюдаемую в цифровой связи на протяжении многих лет, можно охарактеризовать как результат значительного прогресса в полупроводниковых технологиях. Поскольку 6G больше не является вопросом далекого будущего, технологии получения новых материалов будут способствовать скорейшему движению в этом направлении. В данной главе мы рассмотрим несколько технологий изготовления материалов, которые являются ключевыми для 6G. Во-первых, мы кратко остановимся на эволюции кремния и материалов типа III–V¹⁹ как основы технологий для нескольких применений, связанных с терагерцовыми частотами. Затем мы опишем новые реконфигурируемые материалы, которые позволяют гибко управлять устройствами. По мере того как в рамках ТГц-диапазона мы приближаемся к более высокочастотной части спектра, в игру вступают современные фотонные материалы. В частности, фотонные кристаллы позволяют изготавливать оптические компоненты с низкими потерями. Кроме того, фотоэлектрические материалы преобразуют свет в электрические сигналы и позволяют фотодетекторам служить ключевым компонентом фотонной системы ТГц-диапазона. И последнее, но не менее важное: плазмонные материалы поддерживают поверхностные плазмоны, которые способны увеличивать взаимодействие между светом и материалами. Это особенно полезно, когда нам нужно улучшить работу фотонной/оптоэлектронной системы ТГц-диапазона.

14.1. Развитие кремниевой платформы

К существующей зрелой кремниевой платформе, которая по своей сути имеет низкую стоимость, высокий выход, небольшую геометрию и низкое энергопотребление, добавляются новые технологические решения, позволяющие реализовать новые возможности. Хорошим примером является процесс разработки SiGe-BiCMOS, который начался в 1990-х годах.

¹⁹ Полупроводниковые материалы типа A^{III}B^V, в число которых входят антимонид алюминия (AlSb), арсенид алюминия (AlAs), нитрид алюминия (AlN), фосфид алюминия (AlP), нитрид бора (BN), фосфид бора (BP), арсенид бора (Bas), антимонид галлия (GaSb) и др. – *Прим. перев.*

Сегодня платформа SiGe-BiCMOS успешно работает во многих применениях, таких как формирование изображений, спектроскопия и обмен данными, которые невозможно реализовать на стандартной платформе CMOS.

Для эффективного построения изображений объектов в ТГц-диапазоне требуется более быстрое получение данных, и мы можем добиться этого, используя антенные решетки в фокальной плоскости. Для этих целей были продемонстрированы многопиксельные решетки с использованием технологии SiGe [1]. Также была продемонстрирована компактная система ТГц-спектроскопии с использованием технологии SiGe BiCMOS до 1 ТГц [2]. Кроме того, для получения изображений сверхвысокого разрешения (изображений за пределами дифракционного предела) в [3] был разработан датчик изображения ближнего поля с использованием процесса SiGe BiCMOS, и было достигнуто пространственное разрешение 10 микрометров. Также было продемонстрировано применение технологий CMOS и SiGe во многих схемах связи ТГц-диапазона.

Развитие кремниевой платформы послужило толчком к развитию фотонных систем [4, 5]. Определенно стоит упомянуть прогресс, достигнутый кремниевой фотоникой за последние несколько десятилетий. Это стало возможным благодаря оптимизации стандартного процесса CMOS, например добавлению шагов наращивания Ge, чтобы обеспечить интеграцию фотодетектора и модулятора помимо стандартных технологических шагов CMOS. Для интеграции фотонных и электронных структур были предложены и продемонстрированы как плоская, так и многослойная архитектуры. Кроме того, были разработаны различные оптические межсоединения, обеспечивающие низкие потери и эффективное использование пространства. Многослойная интеграция предпочтительнее по сравнению со сложными планарными волноводами, поскольку она позволяет более эффективно использовать пространство и снижает сложность. По этим причинам она является многообещающей перспективной архитектурой для многих оптоэлектронных приложений.

Как упоминалось ранее, кремниевые технологии использовались для развития применений следующего поколения в области связи, обработки изображений, вычислений и т. д. Современные технологические процессы позволяют более эффективно и компактно интегрировать как фотонные, так и электронные компоненты на одном кремниевом кристалле. Ожидается, что прогресс, достигнутый в области кремниевой платформы, приведет к тому, что в ближайшем будущем мы станем свидетелями более совершенных компактных и недорогих фотонных и электронных устройств.

14.2. Гетерогенная платформа на материалах типа III–V

Электронные и фотонные материалы стали следствием прогресса в различных областях. По мере стремительного развития производственных процессов растут и число применений современных материалов. В последние десятилетия мы повышали сложность и производительность кремниевых устройств в соответствии с законом Мура. Однако кремний имеет фундаментальные ограничения использования в фотонной электронике из-за его не прямой запрещенной зоны. Для решения этой проблемы были предложены полупроводники типа III–V с прямой запрещенной зоной. Некоторые материалы III–V, например фосфид индия (InP) и арсенид галлия (GaAs), были испытаны в терагерцовом диапазоне частот (> 1 ТГц) в составе фотонных интегральных схем и показали удовлетворительные результаты. К сожалению, это дорогостоящая технология, и поэтому она не получила широкого распространения на рынке. Тем не менее интерес к фотонике не пропал, и сегодня она нашла применение в технологии сканирования окружающей среды, радарах и связи.

Фотонные интегральные схемы могут получить существенный выигрыш от использования гетерогенных платформ. В частности, соединение разных слоев материалов III–V на одной и той же матрице позволяет оптимизировать строение каждого компонента. Иными словами, для каждого компонента можно выбрать свои оптимальные материалы и конструкцию, а затем интегрировать их на одном кремниевом кристалле. В работе [7] была продемонстрирована гетерогенная система «сеть на кристалле», которая включает полный набор компонентов фотонной связи, в том числе лазеры, модуляторы и усилители на различных стеках слоев III–V.

Используя подход гетерогенной интеграции, мы также можем добавить в систему на кристалле другие материалы, не относящиеся к группе III–V. Было показано, что в кремниевый кристалл можно интегрировать невзаимные магнитные материалы, такие как Ce:YIG [8], тем самым формируя в фотонной интегральной схеме более сложные блоки, такие как изоляторы и циркуляторы.

14.3. Реконфигурируемый материал

Используя материалы с электрически изменяемыми свойствами, можно создавать устройства с большей функциональностью, меньшими размерами и меньшими затратами. По существу, различные настраиваемые материалы используют для гибкого и динамического управления свойствами системы. Например, нанеся реконфигурируемые материа-

лы на интеллектуальную поверхность и соединив ее с цифровой схемой управления, можно реализовать программируемую RIS.

Графен – это двумерный углеродный материал, который может поддерживать распространение поверхностных плазмон-поляритонов (surface plasmon polaritons, SPP). В зависимости от величины приложенного напряжения графен может демонстрировать разные уровни проводимости и, следовательно, управляемое электромагнитное поведение. К настоящему времени были предложены различные графеновые устройства, использующие это свойство переменной проводимости, включая переключатели, фазовращатели и антенны. В статье [9] представлен исчерпывающий обзор потенциальных применений графена в ТГц-коммуникациях, биологическом зондировании и многих других областях. В статьях [10, 11] показано, что передающая графеновая антенна может управлять лучом. В реконфигурируемых графеновых отражательных решетках [12] путем приложения различных напряжений смещения были достигнуты изменяемые фазы отражения. В [13] был представлен графеновый переключатель, который работает путем смещения сегментов графеновых секций под воздействием различного напряжения. Стандартные недорогие и неуправляемые реакторы для осаждения графена стоят примерно 1 млн долларов США.

Жидкие кристаллы проявляют свойства как жидкости, так и кристалла. Нематический жидкий кристалл может изменять свойства падающей волны, когда она проходит через жидкость, и этот эффект зависит от интенсивности приложенных электрических полей. Нематический жидкий кристалл можно эффективно использовать в реконфигурируемых электронных системах. Так, в работах [14, 15] предложили реконфигурируемую отражательную решетку и метаповерхность с использованием жидкого кристалла для управления лучом. В статье [16] представлены фазовращатель и фазированная антенная решетка на основе жидких кристаллов. Там же обсуждалась возможность коммерциализации подобных разработок.

Материалы с фазовым переходом (phase change material, PCM) реконфигурируют путем нагревания и охлаждения, в результате чего происходит переключение между аморфным (изолирующим) и кристаллическим (проводящим) состояниями. Переключатели PCM быстрые, компактные и имеют низкие вносимые потери. По сравнению с микроэлектромеханическими системами (MEMS), PCM показали лучшую надежность и быстродействие на частотах миллиметрового диапазона. В работах [17–19] впервые были продемонстрированы ВЧ-переключатели PCM GeTe для миллиметровых волн и фазовращатель с истинной задержкой по времени. В другой разработке [20] был представлен монолитно интегрированный фазовращатель на основе PCM для миниатюр-

ных реконфигурируемых приложений управления лучом с фазированной антенной решеткой миллиметрового диапазона.

14.4. Фотонный кристалл

Фотонный кристалл – это искусственная периодическая структура, полученная путем выстраивания различных диэлектрических материалов в повторяющуюся последовательность в виде сверхрешетки. Благодаря наличию локализованных дефектов периодической структуры фотонный кристалл может управлять потоком света, выступая как ловушка или как волновод. По сравнению со своими металлическими аналогами фотонно-кристаллические волноводы изготавливаются исключительно из диэлектрических материалов, что означает минимизацию омических потерь. Кроме того, фотонно-кристаллическая структура может быть эффективно интегрирована на кремниевую платформу при сохранении низкой стоимости компонента.

К настоящему времени разработаны различные компоненты на фотонных кристаллах [21] для использования в таких устройствах, как модуляторы, фильтры высокого разрешения, датчики и лазеры. К новым перспективным применениям фотонных кристаллов можно отнести биохимические датчики, лидары и связь в ТГц-диапазоне. В [22] представлен фотонно-кристаллический нанолазер для биомолекулярного и химического тестирования. Фазированные фотонные решетки на основе фотонных кристаллов также успешно продемонстрировали свои возможности в LiDAR-зондировании [22]. Более того, фотонные кристаллы показали свой потенциал в области связи в ТГц-диапазоне. В статье [23] представлена платформа связи ТГц-диапазона на базе фотонных кристаллов на кремнии. Разработчики продемонстрировали высокую скорость передачи данных с помощью встроенного резонансно-туннельного диодного детектора и фотонно-кристаллического волновода.

Совсем недавно топологический фотонный кристалл был исследован как новый метод управления светом [24]. Благодаря наличию топологических фаз материи мы можем направлять свет в заранее определенном направлении даже при наличии дефектов, поскольку существуют топологически защищенные граничные состояния. Кроме того, были исследованы различные топологические платформы для устойчивого однонаправленного распространения света, такие как нелинейные фотонно-кристаллические системы [25], неэрмитовы системы [26] и топологические системы более высокого порядка [27]. Эти системы будут способствовать защите распространения света в наноразмерных фотонных устройствах.

14.5. Фотоэлектрические материалы и фотодетектор

Когда падающее фотонное излучение достигает фотоэлектрического материала, в нем генерируются электронные токи. Некоторые обычные фотоэлектрические материалы представляют собой кристаллические структуры на основе кремния. С развитием нанотехнологий на сцену вышли новые многообещающие фотоэлектрические материалы, такие как наноразмерные квантовые точки, углеродные нанотрубки и графен, обладающие множеством перспектив применения.

Наиболее прямым и очевидным применением фотоэлектрической технологии, вероятно, можно считать солнечные элементы, но мы не должны игнорировать CMOS-сенсоры, которые повсеместно используются в цифровых камерах и многих других приложениях. Традиционно для производства солнечных элементов использовались полупроводниковые материалы, такие как кремний. Впоследствии для повышения эффективности фотопреобразования в состав структуры был добавлен металл, что привело к разработке чувствительных фотодетекторов, например детектора Шоттки. Кроме того, было доказано, что применение наноразмерных светоулавливающих слоев, таких как металлические наночастицы, увеличивает концентрацию света и рассеяние, что, в свою очередь, увеличивает фотоэлектрический эффект. Более того, металлические решетки также могут способствовать возбуждению поверхностных плазмонных волн, тем самым увеличивая взаимодействие между фотонами и электронами.

Фотодетекторы на основе фотоэлектрических материалов имеют множество применений. Например, квантовые точки использовались для увеличения поглощения света при мониторинге состояния здоровья [28]. Было показано, что матрица фотоприемников на основе углеродных нанотрубок [29] может создавать изображения с высоким разрешением. Был изготовлен и продемонстрирован браслет на основе массива графеновых фотодетекторов для мониторинга состояния здоровья [30]. Поскольку потребность в фотонных системах возрастает, использование фотоэлектрических материалов будет по-прежнему приводить к созданию компактных, а также высокопроизводительных фотонных устройств.

14.6. Плазмонный материал

Плазмонные материалы используют энергию света для генерации резонансных колебаний носителей заряда – плазмонов. Как упоминалось ранее, плазмонные наноструктуры использовались в фотоэлектрических устройствах для повышения эффективности фотопреобразования. Двумя распространенными плазмонными структурами являются

металлические наночастицы и металлические электроды с решетчатым рисунком. Первая структура увеличивает рассеяние света за счет возбуждения резонансов локализованных поверхностных плазмон-поляритонов (surface plasmon polariton, SPP), а вторая увеличивает длину пути света, связанного с SPP [31]. Стоит отметить, что обычно в обеих структурах используются золото и серебро. Сообщалось об увеличении эффективности фотопреобразования до 30 % с наночастицами золота [32]. С другой стороны, в [33] сообщается об увеличении эффективности на 19 % при использовании электрода с решеткой из серебра.

Теоретическая физика твердого тела изучала плазмонные материалы с 1970-х годов [34], но сейчас они создаются в лабораториях, а технология плазмонных метаматериалов за последнее десятилетие совершила большой рывок. Метаматериалы можно использовать для получения материалов с желаемыми свойствами, манипулируя периодически распределенными искусственными структурами. Первоначально они использовались в микроволновом диапазоне. Для работы в оптическом диапазоне нужны структуры нанометрового размера, а их трудно изготовить. В статье [35] был дан подробный обзор того, как реализовать отрицательное преломление на оптических частотах. В [36] сетчатая структура, состоящая из двух слоев металлических сеток с диэлектрической прокладкой между ними, была предложена как пример компонента с отрицательной диэлектрической проницаемостью, прозрачного в ближнем инфракрасном диапазоне. Также был продемонстрирован пакетный волновод металл–диэлектрик–металл с отрицательным показателем преломления на видимых частотах [37].

Плазмонные метаповерхности состоят из плазмонных структур нанометрового размера с различной геометрией и разделителями, напечатанными на тонкой пластине. Управляя фазовым профилем наноструктур, метаповерхности могут использовать различные конфигурации падающего света для генерации желаемых откликов. Недавно было доказано, что помимо более общих функций, таких как управление лучом, генерация луча с орбитальным угловым моментом (orbital angular momentum, OAM) и управление отраженным лучом, плазмонные метаповерхности способны управлять оптическими поверхностными волнами в ближней области поля [38]. Когда узорчатые наноструктуры взаимодействуют со светом, часть повторно излучаемой энергии преобразуется в SPP в ближнем поле. Путем тщательного проектирования формы и ориентации наноструктур можно управлять возбуждением SPP в соответствии с такими задачами, как зондирование в ближнем поле и биохимическое обнаружение.

Поверхностные плазмонные поляритоны также могут возникать в графене. Как упоминалось ранее, графен может работать как реконфигурируемый материал благодаря электрически настраиваемым

свойствам. Более того, графен может проявлять плазмонное поведение в ТГц- и инфракрасном диапазоне частот. В графене возбуждение SPP ограничено поперечным размером. Следовательно, графеновые плазмоны могут взаимодействовать с волнами значительно меньшей длины и обеспечивать более сильное взаимодействие света с веществом по сравнению с металлическими аналогами. Кроме того, используя гибридизацию графена и других обычных плазмонных наноструктур или метаматериалов, можно создавать еще более продвинутые устройства на основе графена. Вот почему это многообещающее направление будущего с разных точек зрения.

Плазмонные материалы, взаимодействующие с волновым излучением, вызывают интерес во многих областях. Например, графеновые антенны показали, что их уникальное настраиваемое поведение и потенциал плазмонного резонанса можно использовать во многих устройствах ТГц-диапазона. Используя эффект возбуждения SPP на плазмонных метаповерхностях, интеллектуальные многофункциональные поверхности могут выполнять не только управление лучом, но и служить биосенсорами. По мере достижения более глубокого понимания природы SPP, а также поведения материалов, поддерживающих SPP, для плазмонных материалов найдутся новые интересные применения.

14.7. Источники

- [1] U. R. Pfeiffer, R. Jain, J. Grzyb, S. Malz, P. Hillger, and P. Rodríguez-Vázquez, *Current status of terahertz integrated circuits-from components to systems*, in Proc. 2018 IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium (BCICTS). IEEE, 2018, pp. 1–7.
- [2] K. Statnikov, J. Grzyb, B. Heinemann, and U. R. Pfeiffer, *160-GHz to 1-THz multi-color active imaging with a lens-coupled SiGe HBT chip-set*, IEEE Transactions on Microwave Theory and Techniques, vol. 63, no. 2, pp. 520–532, 2015.
- [3] P. Hillger, R. Jain, J. Grzyb, W. Förster, B. Heinemann, G. MacGrogan, P. Mounaix, T. Zimmer, and U. R. Pfeiffer, *A 128-pixel system-on-a-chip for real-time superresolution terahertz near-field imaging*, IEEE Journal of Solid-State Circuits, vol. 53, no. 12, pp. 3599–3612, 2018.
- [4] M. U. Khan, Y. Xing, Y. Ye, and W. Bogaerts, *Photonic integrated circuit design in a foundry+ fabless ecosystem*, IEEE Journal of Selected Topics in Quantum Electronics, Article no. 8201014, 2019.
- [5] R. Helkey, A. A. Saleh, J. Buckwalter, and J. E. Bowers, *High-performance photonic integrated circuits on silicon*, IEEE Journal of Selected Topics in Quantum Electronics, Article no. 8300215, 2019.

- [6] T. Komljenovic, D. Huang, P. Pintus, M. A. Tran, M. L. Davenport, and J. E. Bowers, *Photonic integrated circuits using heterogeneous integration on silicon*, Proceedings of the IEEE, vol. 106, no. 12, pp. 2246–2257, 2018.
- [7] C. Zhang, S. Zhang, J. D. Peters, and J. E. Bowers, *8×8×40 Gbps fully integrated silicon photonic network on chip*, Optica, vol. 3, no. 7, pp. 785–786, 2016.
- [8] T. Shintaku and T. Uno, *Optical waveguide isolator based on nonreciprocal radiation*, Journal of Applied Physics, vol. 76, no. 12, pp. 8155–8159, 1994.
- [9] D. Correias-Serrano and J. S. Gomez-Diaz, *Graphene-based antennas for terahertz systems: A review*, arXiv preprint arXiv:1704.00371, 2017.
- [10] J. Gómez-Díaz, M. Esquius-Morote, and J. Perruisseau-Carrier, *Plane wave excitation–detection of non-resonant plasmons along finite-width graphene strips*, Optics Express, vol. 21, no. 21, pp. 24 856–24 872, 2013.
- [11] M. Esquius-Morote, J. S. Gómez-Díaz, J. Perruisseau-Carrier et al., *Sinusoidally modulated graphene leaky-wave antenna for electronic beamscanning at THz*, IEEE Transactions on Terahertz Science and Technology, vol. 4, no. 1, pp. 116–122, 2014.
- [12] E. Carrasco, M. Tamagnone, and J. Perruisseau-Carrier, *Tunable graphene reflective cells for THz reflectarrays and generalized law of reflection*, Applied Physics Letters, vol. 102, no. 10, p. 104103, 2013.
- [13] J.-S. Gómez-Díaz and J. Perruisseau-Carrier, *Graphene-based plasmonic switches at nearinfrared frequencies*, Optics Express, vol. 21, no. 13, pp. 15 490–15 504, 2013.
- [14] S. Bildik, S. Dieter, C. Fritzsche, W. Menzel, and R. Jakoby, *Reconfigurable folded reflectarray antenna based upon liquid crystal technology*, IEEE Transactions on Antennas and Propagation, vol. 63, no. 1, pp. 122–132, 2014.
- [15] S. Foo, *Liquid-crystal reconfigurable metasurface reflectors*, in Proc. 2017 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting. IEEE, 2017, pp. 2069–2070.
- [16] T. Ting, *Technology of liquid crystal based antenna*, Optics Express, pp. 17 138–17 153, 2019.
- [17] T. Singh and R. R. Mansour, *Characterization, optimization, and fabrication of phase change material germanium telluride based miniaturized DC–67 GHz RF switches*, IEEE Transactions on Microwave Theory and Techniques, vol. 67, no. 8, pp. 3237–3250, 2019.
- [18] N. El-Hinnawy, P. Borodulin, B. P. Wagner, M. R. King, E. B. Jones, R. S. Howell, M. J. Lee, and R. M. Young, *Low-loss latching microwave switch using thermally pulsed nonvolatile chalcogenide phase change materials*, Applied Physics Letters, vol. 105, no. 1, p. 013501, 2014.

- [19] T. Singh and R. R. Mansour, *Miniaturized reconfigurable 28 GHz PCM based 4-bit latching variable attenuator for 5G mmWave applications*, in Proc. IEEE MTT-S International Microwave Symposium (IMS). IEEE, 2020, pp. 53–56.
- [20] T. Singh and R. R. Mansour, *Loss compensated PCM GeTe-based latching wideband 3-bit switched true-time-delay phase shifters for mmWave phased arrays*, IEEE Transactions on Microwave Theory and Techniques, vol. 68, no. 9, pp. 3745–3755, 2020.
- [21] T. Asano and S. Noda, *Photonic crystal devices in silicon photonics*, Proceedings of the IEEE, vol. 106, no. 12, pp. 2183–2195, 2018.
- [22] T. Baba, *Photonic crystal devices for sensing*, in Proc. 2019 Conference on Lasers and Electro-Optics (CLEO), 2019, pp. 1–2.
- [23] W. Withayachumnankul, M. Fujita, and T. Nagatsuma, *Integrated silicon photonic crystals toward terahertz communications*, Advanced Optical Materials, vol. 6, no. 16, p. 1800401, 2018.
- [24] H. Wang, S. K. Gupta, B. Xie, and M. Lu, *Topological photonic crystals: a review*, Frontiers of Optoelectronics, pp. 1–23, 2020.
- [25] X. Zhou, Y. Wang, D. Leykam, and Y. D. Chong, *Optical isolation with nonlinear topological photonics*, New Journal of Physics, vol. 19, no. 9, p. 095002, 2017.
- [26] S. Yao and Z. Wang, *Edge states and topological invariants of non-hermitian systems*, Physical Review Letters, vol. 121, no. 8, p. 086803, 2018.
- [27] H. Hu, B. Huang, E. Zhao, and W. V. Liu, *Dynamical singularities of floquet higher-order topological insulators*, Physical Review Letters, vol. 124, no. 5, p. 057001, 2020.
- [28] E. O. Polat, G. Mercier, I. Nikitskiy, E. Puma, T. Galan, S. Gupta, M. Montagut, J. J. Piqueras, M. Bouwens, T. Durduran et al., *Flexible graphene photodetectors for wearable fitness monitoring*, Science Advances, vol. 5, no. 9, p. eaaw7846, 2019.
- [29] D. Suzuki, S. Oda, and Y. Kawano, *A flexible and wearable terahertz scanner*, Nature Photonics, vol. 10, no. 12, pp. 809–813, 2016.
- [30] M. Zhang and J. T. Yeow, *A flexible, scalable, and self-powered mid-infrared detector based on transparent PEDOT: PSS/graphene composite*, Carbon, vol. 156, pp. 339–345, 2020.
- [31] S. Ahn, D. Rourke, and W. Park, *Plasmonic nanostructures for organic photovoltaic devices*, Journal of Optics, vol. 18, no. 3, p. 033001, 2016.
- [32] C. C. Wang, W. C. Choy, C. Duan, D. D. Fung, E. Wei, F.-X. Xie, F. Huang, and Y. Cao, *Optical and electrical effects of gold nanoparticles in the active layer of polymer solar cells*, Journal of Materials Chemistry, vol. 22, no. 3, pp. 1206–1211, 2012.

- [33] X. Li, W. E. Sha, W. C. Choy, D. D. Fung, and F. Xie, *Efficient inverted polymer solar cells with directly patterned active layer and silver back grating*, Journal of Physical Chemistry C, vol. 116, no. 12, pp. 7200–7206, 2012.
- [34] O. Madelung, *Introduction to solid-state theory*, Vol. 2. Springer Science & Business Media, 2012.
- [35] K. Yao and Y. Liu, *Plasmonic metamaterials*, Nanotechnology Reviews, vol. 3, no. 2, pp. 177–210, 2014.
- [36] S. Zhang, W. Fan, N. Panoiu, K. Malloy, R. Osgood, and S. Brueck, *Experimental demonstration of near-infrared negative-index metamaterials*, Physical Review Letters, vol. 95, no. 13, p. 137404, 2005.
- [37] S. P. Burgos, R. De Waele, A. Polman, and H. A. Atwater, *A single-layer wide-angle negative-index metamaterial at visible frequencies*, Nature Materials, vol. 9, no. 5, pp. 407–412, 2010.
- [38] S. Sun, Q. He, S. Xiao, Q. Xu, X. Li, and L. Zhou, *Gradient-index metasurfaces as a bridge linking propagating waves and surface waves*, Nature Materials, vol. 11, no. 5, pp. 426–431, 2012.

Глава 15

Новые антенны

Обычные антенны рассчитаны на подключение к радиочастотным системам через коаксиальные кабели или микрополосковые линии. Антенны ТГц-диапазона радикально отличаются от них, потому что потери в линии передачи на терагерцовых частотах слишком велики. В более низком диапазоне частот (100–500 ГГц) были продемонстрированы интегральные схемы на основе кремния с интегрированной антенной на кристалле и антенной в корпусе [1, 2]. За пределами 500 ГГц можно использовать комбинированную технологию III–V/кремний, позволяющую использовать обычные антенны до 1 ТГц.

С другой стороны, поскольку ТГц-система может быть реализована на фотонной платформе, терагерцовые волны можно генерировать, пропуская фототок через фотопроводящую антенну (photoconductive antenna, PCA) или электрооптический кристалл. При освещении лазерными лучами на определенных частотах в некоторых полупроводниковых подложках, например InP и GaAs, может генерироваться ток фотопроводимости. Затем он может излучаться в пространство в виде ТГц-волны с помощью электродов антенны.

Наноподетекторы имеют схожий принцип работы, с той разницей, что они измеряют генерируемый фототок с помощью схемы, предназначенной для визуализации/считывания. Благодаря небольшому размеру и низкому энергопотреблению наноподетекторы демонстрируют огромный потенциал применения в мобильных и носимых приложениях.

Благодаря рассеивателям с тщательно выверенными формами и расстояниями между ними интеллектуальные поверхности, такие как отражающие и передающие решетки, могут служить плоскими «отражателями» или «линзами». В такие поверхности обычно встраивают реконфигурируемые компоненты, чтобы обеспечить гибкую подстройку к различным уровням освещенности радиоволнами. Метаповерхности поддерживают «умную» функциональность, обеспечивая заданную пользователем реакцию на определенный уровень освещенности. В будущем для программирования этой реакции можно будет использовать

цифровой контроллер, и такие интеллектуальные поверхности гарантированно обеспечат нам компактные, высокоэффективные и недорогие системы связи в ТГц-диапазоне.

В системе массивного MIMO 5G с момента ее внедрения в 2010 году был достигнут значительный прогресс [3]. Но поскольку следом неизбежно на сцену выйдет 6G, уже сейчас необходимо искать новые исследовательские перспективы для повышения производительности MIMO. В этой связи для достижения дополнительной глубины резкости предлагается использовать орбитальный угловой момент (orbital angular momentum, OAM), конечной целью которого является улучшение характеристик канала. Также сообщалось, что использование нескольких совместных структур беспроводной связи OAM и массивных MIMO позволяет получить мультипликативное усиление эффективности спектра [4, 5].

15.1. Антенна с фотопроводящей линзой

Как упоминалось ранее, PCA может преобразовывать энергию фотонов в ТГц-излучение. Однако эффективность фотопреобразования обычно невелика, что, в свою очередь, приводит к низкой излучаемой мощности (мкВт). Тем не менее мы можем использовать диэлектрическую линзу, чтобы коллимировать исходящие лучи и в конечном итоге увеличить усиление антенны.

В типичной антенне линза питается от тонкой подложки с высокой диэлектрической проницаемостью, на которой напечатаны плоские электроды антенны. К традиционным типам антенн относятся диполи, «бабочки», спиральные и логопериодические. К антенне прикладывают внешнее напряжение смещения, чтобы ускорить переходные фототоки, и в конечном итоге терагерцовая волна излучается в пространство. Чтобы уменьшить рассогласование и максимизировать характеристики антенны, нам необходимо разработать линзы оптимального размера и формы.

Исследователи активно разрабатывают новые методы повышения эффективности фотопреобразования. Например, в фотопроводящий зазор вносили наноструктуры [6] для увеличения рассеяния волн или возбуждения плазмонных волн [7], тем самым увеличивая эффективность фотопары. В качестве покрытий для увеличения поглощения света также использовались квантовые точки [8].

Для увеличения скорости визуализации среды были предложены структуры с решетками в фокальной плоскости. Один из вариантов включает в себя подключенные антенные решетки, которые напечатаны на питающей подложке и питают линзы с большой апертурой. Решетки можно сконфигурировать для управления лучом или форми-

рования луча, и эта конфигурация предназначена для сканирования окружающей среды, а также для телекоммуникационных систем. В другом варианте применяется близко расположенная линза, размещенная поверх антенных элементов, чтобы сформировать плотно упакованную решетку. Эта структура больше подходит для космических или астрономических приложений, где предпочтительно наличие функции многопиксельного сканирования. Эти два варианта [9] показаны на рис. 15.1.

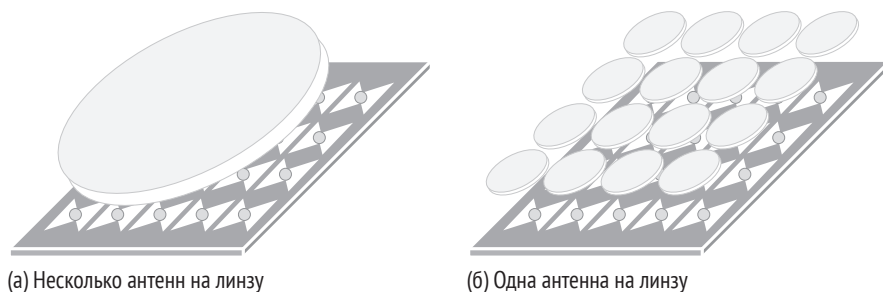


Рис. 15.1. Антенны с фотопроводящими линзами

15.2. Отражающие и передающие решетки

И отражающие, и передающие решетки находят множество применений, включая расширение диапазона связи, беспроводную передачу энергии, пространственную модуляцию и антенны с высоким коэффициентом усиления. Обе решетки имеют низкий профиль, высокую эффективность и поддерживают реконфигурацию. Это означает, что их можно легко интегрировать в самые разные системы, и поэтому они являются многообещающими кандидатами на использование в коммуникациях следующего поколения.

В отражающих решетках используются различные рассеиватели электромагнитного излучения, напечатанные на поверхности. При освещении фидерной антенной отражательные решетки ведут себя как традиционные рефлекторные антенны. Рассеиватели могут принимать форму микрополосковых пятен (или других типов резонансных структур), как показано на рис. 15.2. Каждый рассеиватель тщательно проектируют для создания фазового сдвига, имитирующего изогнутую поверхность отражателя и, следовательно, отражающего луч с заданными свойствами.

Рассеиватели могут быть сополяризованы без изменения поляризации падающего поля. Они также могут иметь двойную поляризацию или кросс-поляризацию, чтобы по-разному отражать падающие лучи с разной поляризацией или менять ее. Мы можем сделать рефлектор реконфигурируемым, чтобы в любой момент менять направление

отражения луча. Для создания реконфигурируемого фазового профиля могут применяться и независимо управляться такие современные перестраиваемые материалы, как жидкие кристаллы и графен.

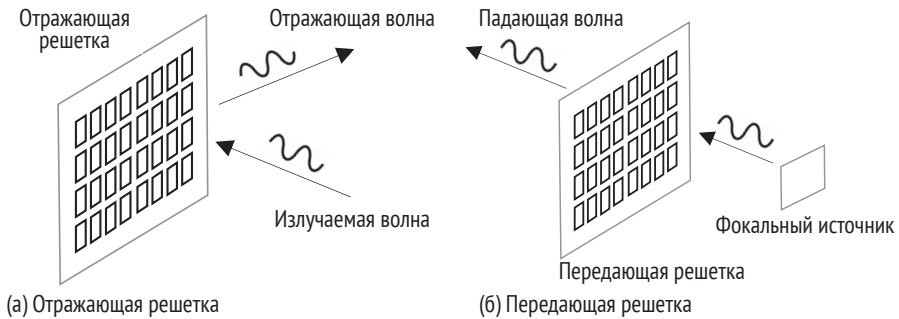


Рис. 15.2. Отражающие и передающие решетки

Точно так же передающая решетка состоит из плоской подложки с несколькими печатными резонаторами, обладающими заданным фазовым распределением. Когда решетка расположенного в фокусе источника освещает передающую решетку, падающая волна проходит через плоскость передающей решетки и затем преобразуется в луч с желаемой диаграммой направленности. В большинстве случаев передающая решетка ведет себя как плоская диэлектрическая линза. Стоит отметить, что функция управления лучом может быть реализована как в плоскости источника, так и в плоскости решетки. Например, сначала можно настроить расположенный в фокусе источник. Затем в плоскости передающей решетки манипулируют резонаторами для придания желаемых свойств исходящему лучу.

15.3. Метаповерхности

Метаповерхности предоставляют привлекательные решения по организации излучения для многих новых применений. Благодаря гибкости управления метаповерхности «программируются» с помощью цифровой платформы [10]. Вдобавок к этому они имеют небольшие размеры, что обеспечивает недорогую интеграцию с различными платформами. Напечатанные на гибкой подложке метаповерхности могут входить в состав носимых устройств, потенциально пригодных для связи, получения изображений и более сложных вариантов применения. Например, в [11] было продемонстрировано программируемое интеллектуальное метаповерхностное стекло, которое обеспечивает полное проникновение, частичное и полное отражение падающих радиоволн. Это стекло может найти применение для управления характеристиками канала в системах беспроводной связи.

В микроскопическом представлении каждый элемент элементарной ячейки может иметь частотную характеристику, выраженную в переменных значениях амплитуды и фазы. Для достижения независимого управления в каждой элементарной ячейке могут быть применены настраиваемые элементы. Макроскопически элементарные ячейки могут быть спроектированы так, чтобы формировать электромагнитно взаимосвязанную сеть, которая конструктивно выполняет заданные функции, например поглощения волн, подавления поверхностных волн, развязки антенны, формирования луча и т. д.

Для метаповерхностных антенн элементарные ячейки тщательно проектируют таким образом, чтобы согласовать амплитуды и фазу для получения желаемой диаграммы направленности в свободном пространстве. Каждая элементарная ячейка служит небольшим излучателем с определенными параметрами, что, в свою очередь, позволяет объединенному лучу от всех элементов принять необходимую форму. Антенны передатчика и приемника могут быть интегрированы на одной подложке, и мы можем реализовать отдельные волноводные структуры для подключения излучателей к входному или выходному сигнальному порту.

Одним из самых интересных применений, вероятно, является *метаповерхностная голография*. В голографии информация об амплитуде и фазе световых волн, рассеянных объектом, записывается и кодируется на фотопленке. Затем мы применяем алгоритм реконструкции изображения для создания трехмерной голограммы, которая выглядит как реальный объект. Метаповерхности могут использоваться для записи информации голограммы такого типа, поскольку элементарные ячейки могут действовать как «пиксели». Другими словами, каждый элемент может хранить определенные амплитудную и фазовую характеристики. При последующем освещении волнами определенной конфигурации отражения от этих элементов могут быть наложены друг на друга для создания трехмерной голограммы исходного объекта. Принцип работы показан на рис. 15.3.

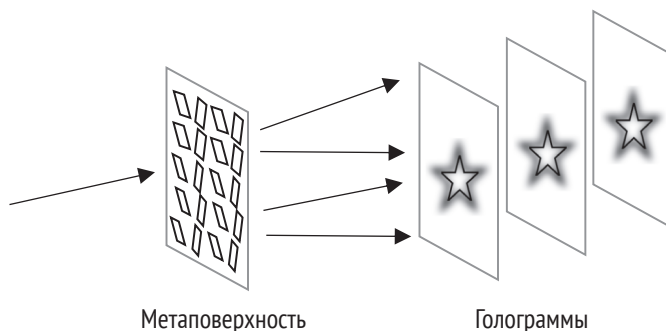


Рис. 15.3. Голограммы, созданные метаповерхностью

Еще одно интересное применение метаповерхности – маскировка. Она заключается в размещении тонкого метаповерхностного слоя на определенном расстоянии от маскируемого объекта, что, в свою очередь, позволяет создать «противофазную» деструктивную интерференцию, которая нейтрализует рассеяние. Конечным результатом является объект, «невидимый» на определенной частоте. Добавляя активные элементы метаповерхности, мы можем расширить рабочую полосу частот.

15.4. Нанопотодетекторы

Как упоминалось ранее, платформы формирования изображений с рабочей частотой выше 1 ТГц остаются проблемными из-за ограничений быстрогодействия кремниевых компонентов. Современные системы построения изображений «за пределами 1 ТГц» преимущественно основаны на больших оптических системах и слишком громоздки для интеграции. Поэтому необходима портативная и удобная измерительная система. Благодаря развитию нанотехнологий эта задача может быть решена с помощью нанопотодетекторов.

Последние разработки в области нанотехнологий открыли новые возможности для получения изображений в ТГц-диапазоне. Нанопотодетекторы небольшие, масштабируемые, экономичные и энергоэффективные, их можно интегрировать в различные системы построения изображений при достаточном уровне быстрогодействия.

Потенциальным кандидатом на использование в этой области является чувствительная матрица из углеродных нанотрубок (carbon nanotube, CNT). Используя фототермоэлектрический эффект, CNT могут преобразовывать освещающее ТГц-излучение в измеряемый фототок в цепи считывания. В статьях [12, 13] сообщается, что сканирующий массив CNT для получения изображений может достигать пространственного разрешения в несколько сотен микрон на частоте 1,4 ТГц. Кроме того, предлагаемый массив CNT может быть изготовлен на гибкой подложке для использования в носимых устройствах.

Графеновый фотодетектор работает аналогично сенсору CNT. Он может преобразовывать освещающее ТГц-излучение в измеримый фототок. В этом контексте в статье [14] сообщают, что гибкий браслет с графеновым фотодетектором можно использовать для наблюдения за состоянием здоровья. Кроме того, прозрачность графена означает, что его можно использовать для экранных антенн, интеллектуальных очков, носимых устройств и т. д.

15.5. Антенна на кристалле и антенна в корпусе

Достижения полупроводниковых технологий, таких как CMOS и SiGe BiCMOS, позволили нам создать интегральную схему ТГц-диапазона. В статьях [15–17] показано, что с помощью технологии SiGe HBT²⁰ вполне возможно изготавливать интегральные схемы с рабочей частотой до 700 ГГц. Оценки показывают, что предел производительности SiGe HBT может вскоре достичь или даже превысить 1 ТГц. Кремниевая терагерцовая интегральная схема обладает рядом преимуществ, таких как низкая стоимость, компактный размер, высокий выход и простота интеграции.

Удобный способ реализации ТГц-антенны подразумевает ее прямую интеграцию с входной схемой на кремниевой подложке. Однако размещение антенны на кристалле является сложной задачей из-за поверхностных волн, генерируемых в подложке. Поверхностные волны будут мешать излучению антенны и приведут к плохой работе устройства. Для борьбы с этим явлением была предложена линзовая антенна с обратным излучением. Применение такой антенны для получения изображений и зондирования в ТГц-диапазоне продемонстрировано в [18, 19].

Новое применение ТГц-излучения – это терагерцовое *сканирование в ближнем поле* (near-field imaging, NFI). В случае NFI пространственное разрешение изображения не ограничено дифракционным пределом и может применяться для получения изображений в микроскопическом или наноразмерном масштабе. NFI требует, чтобы антенна имела сильную связь в ближнем поле. Например, в [20] было доказано, что резонансные структуры, такие как *резонаторы с разъемным кольцом* (split-ring-resonators, SRR), способны создавать биометрические изображения отпечатков пальцев человека.

Подход «антенна в корпусе» дает нам еще один способ интеграции. Однако на терагерцовых частотах соединения между антенной и монолитной микроволновой интегральной схемой (monolithic microwave integrated circuit, MMIC) слишком велики. К счастью, в настоящее время разрабатываются эффективные методы построения корпусов для решения этой проблемы и минимизации потерь. Интегрированная волноводная рупорная антенна на частоте 300 ГГц была выполнена в многослойном пакете LTCC [21] и продемонстрирована в действии в сервисе загрузки данных KIOSK²¹.

²⁰ Silicon-germanium heterojunction bipolar transistor, кремний-германиевый биполярный транзистор с гетеропереходом. – *Прим. перев.*

²¹ <https://ieeexplore.ieee.org/document/7928447>.

15.6. Орбитальный угловой момент

Традиционно пропускную способность радиолинии увеличивали за счет мультиплексирования пространства, частоты, времени, кода и поляризации. Затем для улучшения характеристик передачи данных разработали ММО-антенну с различной поляризацией. В последнее время большое внимание привлек новый аспект мультиплексирования – орбитальный угловой момент (orbital angular momentum, OAM). Согласно новому подходу, при использовании OAM антенна может генерировать ортогональные моды, каждая из которых связана с различным орбитальным моментом. Например, сигнал может иметь фазовый коэффициент $e^{-j\varphi}$, $e^{-j2\varphi}$ и т. д. Каждый режим может нести различную информацию; следовательно, несколько режимов OAM могут сосуществовать и передавать данные одновременно по одному каналу связи. Правильно выделив требуемый режим OAM на стороне приемника, можно оптимизировать эффективность использования спектра.

Системам связи OAM было посвящено значительное число исследований. Например, в [22] было предложено генерировать волны OAM путем управления фазовыми модами с использованием круглой решетки с несколькими фазовращателями и сумматорами. В статье [23] продемонстрирована система мультиплексирования OAM-ММО, которая использует несколько однородных кольцевых антенн для генерации пяти режимов OAM. В статьях [24, 25] показано, что системы связи OAM позволяют достичь высокой скорости передачи данных.

Технология OAM имеет большой потенциал в сценариях беспроводной связи LOS, таких как транзитная сеть сотовой связи и межсоединения в центре обработки данных. Кроме того, комбинированная связь OAM-ММО значительно увеличит скорость передачи данных и обеспечит гораздо более высокую эффективность использования спектра. Например, в [5] сообщается о системе связи с мультиплексированием OAM-ММО, которая обеспечивает скорость передачи данных 100 Гбит/с на расстоянии 10 м. Предложенный метод мультиплексирования использует 11 режимов OAM на частоте 28 ГГц.

OAM – многообещающая технология передачи данных, поэтому сейчас активно рассматриваются способы ее применения в сетях 6G. Из-за особенностей передающей антенны массива OAM на практике диапазоны частот, которые позволяют 6G использовать системы OAM, будут превышать 20 ГГц, и, как таковые, диапазоны миллиметровых волн являются хорошими кандидатами. Было предпринято несколько попыток создания компактной антенной архитектуры и конструкции антенной решетки миллиметрового диапазона [26, 27]. Основная проблема, связанная с применением OAM, относится к реализации сценариев использования мобильности. В этом контексте необходимы технологические

прорывы, которые позволят нам использовать простые антенны на стороне пользовательского оборудования для демодуляции сигналов ОАМ, особенно когда речь идет о поддержке мобильных применений [28].

15.7. Источники

- [1] P. Hillger, J. Grzyb, R. Jain, and U. R. Pfeiffer, *Terahertz imaging and sensing applications with silicon-based technologies*, IEEE Transactions on Terahertz Science and Technology, vol. 9, no. 1, pp. 1–19, 2018.
- [2] Y. Zhang and J. Mao, *An overview of the development of antenna-in-package technology for highly integrated wireless devices*, Proceedings of the IEEE, vol. 107, no. 11, pp. 2265–2280, 2019.
- [3] T. L. Marzetta, *Noncooperative cellular wireless with unlimited numbers of base station antennas*, IEEE Transactions on Wireless Communications, vol. 9, no. 11, pp. 3590–3600, 2010.
- [4] W. Cheng, H. Zhang, L. Liang, H. Jing, and Z. Li, *Orbital-angular-momentum embedded massive MIMO: Achieving multiplicative spectrum-efficiency for mmWave communications*, IEEE Access, vol. 6, pp. 2732–2745, 2017.
- [5] D. Lee, H. Sasaki, H. Fukumoto, Y. Yagi, and T. Shimizu, *An evaluation of orbital angular momentum multiplexing technology*, Applied Sciences, vol. 9, no. 9, p. 1729, 2019.
- [6] S.-G. Park, Y. Choi, Y.-J. Oh, and K.-H. Jeong, *Terahertz photoconductive antenna with metal nanoislands*, Optics Express, vol. 20, no. 23, pp. 25 530–25 535, 2012.
- [7] H. Tanoto, J. Teng, Q. Wu, M. Sun, Z. Chen, S. Maier, B. Wang, C. Chum, G. Si, A. Danner et al., *Nano-antenna in a photoconductive photomixer for highly efficient continuous wave terahertz emission*, Scientific Reports, vol. 3, p. 2824, 2013.
- [8] E. O. Polat, G. Mercier, I. Nikitskiy, E. Puma, T. Galan, S. Gupta, M. Montagut, J. J. Piqueras, M. Bouwens, T. Durduran et al., *Flexible graphene photodetectors for wearable fitness monitoring*, Science Advances, vol. 5, no. 9, p. eaaw7846, 2019.
- [9] O. Yurduseven, *Wideband integrated lens antennas for terahertz deep space investigation*, Ph.D. dissertation, Delft University of Technology, 2016.
- [10] T. J. Cui, M. Q. Qi, X. Wan, J. Zhao, and Q. Cheng, *Coding metamaterials, digital metamaterials and programmable metamaterials*, Light: Science & Applications, vol. 3, no. 10, p. e218, 2014.
- [11] NTT DoCoMo, *DOCOMO conducts worlds first successful trial of transparent dynamic metasurface*, Tokyo, 2020.

- [12] D. Suzuki, S. Oda, and Y. Kawano, *A flexible and wearable terahertz scanner*, Nature Photonics, vol. 10, no. 12, pp. 809–813, 2016.
- [13] D. Suzuki, Y. Ochiai, and Y. Kawano, *Thermal device design for a carbon nanotube terahertz camera*, ACS Omega, vol. 3, no. 3, pp. 3540–3547, 2018.
- [14] M. Zhang and J. T. Yeow, *A flexible, scalable, and self-powered mid-infrared detector based on transparent PEDOT: PSS/graphene composite*, Carbon, vol. 156, pp. 339–345, 2020.
- [15] J. Grzyb, B. Heinemann, and U. R. Pfeiffer, *Solid-state terahertz superresolution imaging device in 130-nm SiGe BiCMOS technology*, IEEE Transactions on Microwave Theory and Techniques, vol. 65, no. 11, pp. 4357–4372, 2017.
- [16] J. Grzyb, B. Heinemann, and U. R. Pfeiffer, *A 0.55 THz near-field sensor with a μ m range lateral resolution fully integrated in 130 nm SiGe BiCMOS*, IEEE Journal of Solid-State Circuits, vol. 51, no. 12, pp. 3063–3077, 2016.
- [17] P. Hillger, R. Jain, J. Grzyb, W. Förster, B. Heinemann, G. MacGrogan, P. Mounaix, T. Zimmer, and U. R. Pfeiffer, *A 128-pixel system-on-a-chip for real-time super-resolution terahertz near-field imaging*, IEEE Journal of Solid-State Circuits, vol. 53, no. 12, pp. 3599–3612, 2018.
- [18] J. Grzyb and U. Pfeiffer, *Thz direct detector and heterodyne receiver arrays in silicon nanoscale technologies*, Journal of Infrared, Millimeter, and Terahertz Waves, vol. 36, no. 10, pp. 998–1032, 2015.
- [19] D. F. Filipovic, S. S. Gearhart, and G. M. Rebeiz, *Double-slot antennas on extended hemispherical and elliptical silicon dielectric lenses*, IEEE Transactions on Microwave Theory and Techniques, vol. 41, no. 10, pp. 1738–1749, 1993.
- [20] P. Hillger, R. Jain, J. Grzyb, W. Förster, B. Heinemann, G. MacGrogan, P. Mounaix, T. Zimmer, and U. R. Pfeiffer, *A 128-pixel system-on-a-chip for real-time superresolution terahertz near-field imaging*, IEEE Journal of Solid-State Circuits, vol. 53, no. 12, pp. 3599–3612, 2018.
- [21] T. Tajima, T. Kosugi, H.-J. Song, H. Hamada, A. El Moutaouakil, H. Sugiyama, H. Matsuzaki, M. Yaita, and O. Kagami, *Terahertz MMICs and antenna-in-package technology at 300 GHz for KIOSK download system*, Journal of Infrared, Millimeter, and Terahertz Waves, vol. 37, no. 12, pp. 1213–1224, 2016.
- [22] M. Klemes, H. Boutayeb, and F. Hyjazie, *Orbital angular momentum (OAM) modes for 2-D beam-steering of circular arrays*, in Proc. 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). IEEE, 2016, pp. 1–5.
- [23] H. Sasaki, D. Lee, H. Fukumoto, Y. Yagi, T. Kaho, H. Shiba, and T. Shimizu, *Experiment on over-100-Gbps wireless transmission with*

- OAM-MIMO multiplexing system in 28-GHz band*, in Proc. 2018 IEEE Global Communications Conference (GLOBECOM). IEEE, 2018, pp. 1–6.
- [24] Y. Ren, L. Li, G. Xie, Y. Yan, Y. Cao, H. Huang, N. Ahmed, Z. Zhao, P. Liao, C. Zhang et al., *Line-of-sight millimeter-wave communications using orbital angular momentum multiplexing combined with conventional spatial multiplexing*, IEEE Transactions on Wireless Communications, vol. 16, no. 5, pp. 3151–3161, 2017.
- [25] A. M. Yao and M. J. Padgett, *Orbital angular momentum: Origins, behavior and applications*, Advances in Optics and Photonics, vol. 3, no. 2, pp. 161–204, 2011.
- [26] M. Klemes, H. Boutayeb, and F. Hyjazie, *Minimal-hardware 2-D steering of arbitrarily large circular arrays (combining axial patterns of phase-modes)*, in Proc. 2016 IEEE International Symposium on Phased Array Systems and Technology (PAST). IEEE, 2016, pp. 1–8.
- [27] Z. Zhao, G. Xie, L. Li, H. Song, C. Liu, K. Pang, R. Zhang, C. Bao, Z. Wang, S. Sajuyigbe et al., *Performance of using antenna arrays to generate and receive mm-wave orbital-angular-momentum beams*, in Proc. IEEE Global Communications Conference (GLOBECOM). IEEE, 2017, pp. 1–6.
- [28] M. Klemes, *Reception of OAM radio waves using pseudo-doppler interpolation techniques: A frequency-domain approach*, Applied Sciences, vol. 9, no. 6, p. 1082, 2019.

Глава 16

Современные технологии ТГц-диапазона

Недавние разработки в области полупроводниковой технологии позволили преодолеть «запрещенную зону» в ТГц-диапазоне (вызванную отсутствием аппаратных средств для работы в ТГц-диапазоне) и стимулировали развитие различных применений терагерцевых частот. Что касается спектра, его терагерцевая область находится между миллиметровым диапазоном радиоволн и инфракрасным светом. Сигналы ТГц-диапазона могут проникать через диэлектрические материалы на разную глубину, открывая новые способы получения изображений. Благодаря эффектам теории дифракционного предела разрешение пространственного изображения в ТГц-спектре будет намного выше по сравнению с его аналогом в миллиметровом диапазоне. ТГц-излучение также неионизирующее, поскольку его частота ниже, чем в УФ-диапазоне, а это означает, что у него есть широкие перспективы применения в биомедицинских приложениях. Кроме того, различные материалы демонстрируют уникальные отклики в диапазоне 0,5–3 ТГц, что позволяет использовать такие отклики в качестве спектральных «отпечатков» для обнаружения и определения характеристик материала.

Растущий спрос на более высокие скорости передачи данных и меньшую задержку означает, что более высокие частоты и более широкая полоса пропускания становятся все более важными при разработке систем связи. В последнее время были исследованы различные архитектуры для систем терагерцевой связи и продемонстрированы испытательные стенды, основанные на двух различных подходах: электронный, где радиочастоты умножаются до ТГц; и фотонный, где оптические частоты делятся до ТГц. Следует отметить тот факт, что большинство этих систем разрабатываются в основном для связи на малых расстояниях внутри помещений, отчасти из-за высокого атмосферного затухания в ТГц-диапазоне. Однако этого можно до некоторой степени избежать, выбрав «окно частот», в котором потери в атмосфере низкие, например 140, 220 и 300 ГГц, как описано в главе 12.

В этой главе мы кратко описываем современные технологии ТГц-диапазона. Сначала обсудим компоненты систем ТГц-диапазона и сравним их характеристики в различных технологиях. Затем проанализируем производительность систем связи и формирования изображений в ТГц-диапазоне и в заключение обсудим проблемы, связанные с ТГц-диапазоном, и текущее состояние соответствующих исследований.

16.1. Компоненты ТГц-систем

Как упоминалось ранее, ТГц-система может быть реализована с использованием электронного или фотонного подхода. В то время как фотонный подход в основном нацелен на верхние частоты в ТГц-диапазоне, электронный подход может обеспечить решения в нижних частотах диапазона. Максимальная частота, на которой может работать терагерцовая электронная схема, зависит от максимальной частоты $f_{\max}$ транзистора, изготовленного по определенному твердотельному техпроцессу.

В традиционных технологиях CMOS и BiCMOS $f_{\max}$ транзистора находится в диапазоне от 200 ГГц до 350 ГГц [1]. Благодаря технологии SiGe BiCMOS транзистор может достигать $f_{\max}$ 0,5 ТГц или даже 0,75 ТГц [2]. Полупроводники типа III–V, такие как транзисторы с высокой подвижностью электронов (high electron mobility transistors, HEMT) на основе InP или биполярные транзисторы с гетеропереходом (HBT), могут поднять $f_{\max}$ более чем до 1 ТГц [3]. Соответствующие технологические дорожные карты были разработаны как для InP-, так и для SiGe-технологий [1]. Ожидается, что в ближайшем будущем терагерцовая электроника превысит 1 ТГц и достигнет 2 ТГц с использованием технологий SiGe CMOS и III–V.

16.1.1. Электронный подход

Из-за фундаментального ограничения частоты транзистора $f_{\max}$ электронный подход в основном нацелен на системы, работающие на более низких частотах ТГц-диапазона. Выход за пределы этих частот требует тщательного проектирования системы, чтобы уменьшить нелинейность устройства и гармонические эффекты, которые начинают оказывать существенное влияние.

В [4] было проведено сравнение современных источников ТГц-излучения как в CMOS, так и в SiGe-технологиях. В этом разделе мы приводим неполный перечень источников ТГц-излучения, выполненных по технологии InP HEMT/HBT. В табл. 16.1 приведены характеристики источников InP ТГц-диапазона за пределами 300 ГГц, а на рис. 16.1 и 16.2 даны сравнения источников ТГц-диапазона, изготовленных по

различным технологиям. Как показано на рисунках, большинство источников ТГц-диапазона работают на частотах от 0,2 ТГц до 0,5 ТГц с уровнем выходной мощности от -10 дБм до $+10$ дБм. При аналогичном уровне выходной мощности источники InP демонстрируют лучшую эффективность использования постоянного тока по отношению к радиочастоте (отношение выходной мощности радиосигнала к потребляемой мощности постоянного тока). Для диапазона частот выше 0,5 ТГц были представлены данные для всех трех источников ТГц-излучения, основанных на соответствующих технологиях. Источники CMOS и SiGe демонстрируют более низкий уровень выходной мощности, поскольку рабочая частота превышает $f_{\max}$, а эффективность снижается из-за гармонического эффекта. Более высокая выходная мощность достигается с помощью источников InP [5, 6], что важно для формирования ТГц-луча и управления им.

Таблица 16.1. Сравнение современных источников ТГц-излучения на основе InP

Технология	Частота (ТГц)	Процесс (нм)	P_{OUT} (дБм)	P_{DC} (мВт)	Источник
InP HBT	300	250	+5,3	87,4	[6]
InP DHBT	300	250	+4,8	88	[7]
InP DHBT	300	250	-5	46,2	[8]
InP DHBT	300	250	+1,5	148	[9]
InP DHBT	300	130	+4,7	75,6	[10]
InP DHBT	303	800	$-6,2$	37,6	[11]
InP DHBT	306	800	$-1,6$	36	[12]
InP DHBT	325	800	-7	40	[13]
InP DHBT	330	250	$-6,5$	13,5	[14]
InP HBT	413	250	$-5,6$	<115	[15]
InP DHBT	480	300	-11	15	[5]
InP HBT	487	250	$-8,9$	<115	[15]
InP HBT	573	250	$-19,2$	<115	[15]
InP HBT	591	250	$-17,4$	49,3	[6]
InP HBT	645	250	$-17,4$	49,3	[16]
InP HEMT	670	25	+2,55	1,7	[17]

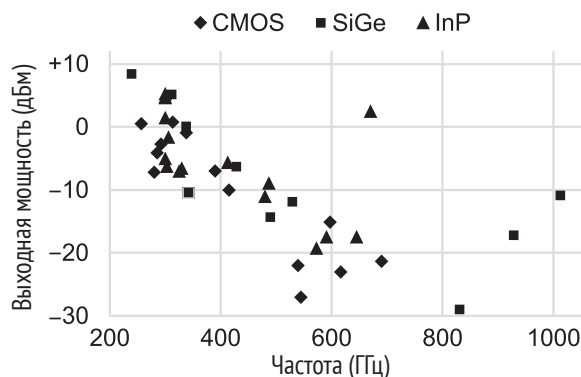


Рис. 16.1. Сравнение выходной мощности современных источников ТГц-диапазона, изготовленных по технологиям CMOS, SiGe и InP. Данные для CMOS и SiGe взяты из [4], а данные для InP – из табл. 16.1

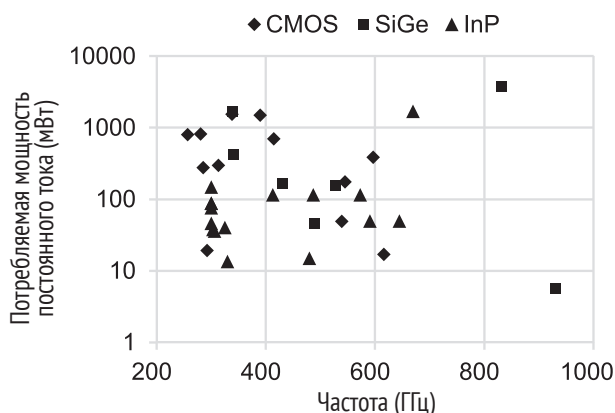


Рис. 16.2. Сравнение мощности постоянного тока, потребляемой современными источниками ТГц-диапазона в технологиях CMOS, SiGe и InP. Данные для CMOS и SiGe взяты из [4], а данные для InP – из табл. 16.1

В предыдущих работах сравнивались различные усилители мощности ТГц-диапазона, использующие технологии CMOS, SiGe, GaAs и InP [18–20]. Подробное сравнение различных технологий усилителей приведено в [20], хотя частота в основном не превышает 100 ГГц. В статье [18] сравниваются характеристики современных усилителей мощности на частотах 110–180 ГГц с использованием как CMOS-, так и SiGe-технологий и показано, что при выходной мощности до 20 дБм КПД преобразования постоянного тока в излучение, как правило, составляет менее 10 %. Усилители мощности на основе InP на частоте около 200 ГГц сравнивались в [19], где показано, что выходная мощность 20–30 дБм может быть сгенерирована с КПД также менее 10 %. В этом разделе был проведен неполный обзор литературы по усилителям мощности на час-

тотах выше 300 ГГц. В табл. 16.2 и на рис. 16.3 сравниваются характеристики усилителей мощности, выполненных по различным технологиям. Были продемонстрированы усилители мощности на основе GaAs и InP, работающие на частоте около 300 ГГц, хотя усилители на основе InP показали более высокий уровень выходной мощности. Дальнейшие работы [21, 22] расширили рабочие частоты усилителей на основе InP до 850 ГГц.

Таблица 16.2. Сравнение современных усилителей мощности ТГц-диапазона, изготовленных по различным технологиям

Технология	Процесс (нм)	Частота (ГГц)	Усиление (дБ)	P_{SAT} (дБм)	P_{DC} (мВт)	PAE	Ист.
CMOS SOI	32	210	15	4,6	40	6,00 %	[23]
SiGe BiCMOS	130	230	12,5	12	740	1,00 %	[24]
SiGe BiCMOS	130	215	25	9,6	–	0,50 %	[25]
GaN	50	190	12	14,1	–	1,20 %	[26]
InGaAs mHEMT	35	320	13,5	8,6	–	–	[27]
InGaAs mHEMT	35	320	12	7	–	–	[27]
InGaAs mHEMT	35	310	7	8,5	521	–	[28]
InGaAs mHEMT	35	294	15	4,8	–	–	[29]
InP HEMT	50	340	15	10	–	–	[30]
InP HBT	250	300	12	9,2	848	1,10 %,	[31]
InPHBT	250	300	13,4	13,5	–	–	[32]
InP HBT	130	325	10	9,4	243	2,20 %,	[33]
InP HBT	130	325	9,4	11,4	243	1,09 %	[33]
InPDHBT	250	325	11	1,13	–	0,60 %	[34]
InP HEMT	80	300	20	12	–	–	[35]
InP HEMT	80	300	14	9,5	–	–	[36]
InP DHBT	130	670	24	-4	–	–	[37]
InP DHBT	130	655	20	-0,7	–	–	[37]
InP HBT	130	585	20	2,8	455	–	[22]
InP HEMT	25	850	17	-0,3	60	–	[21]

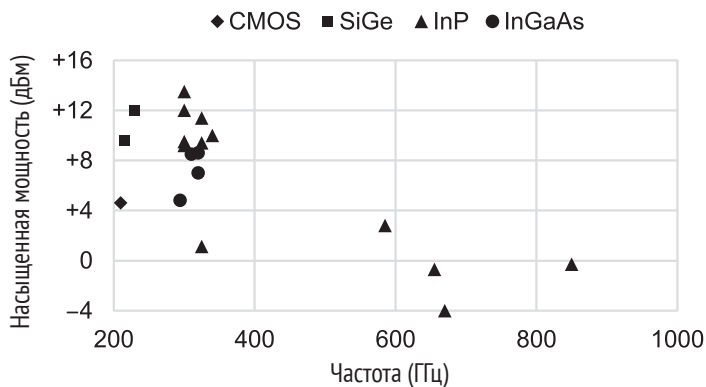
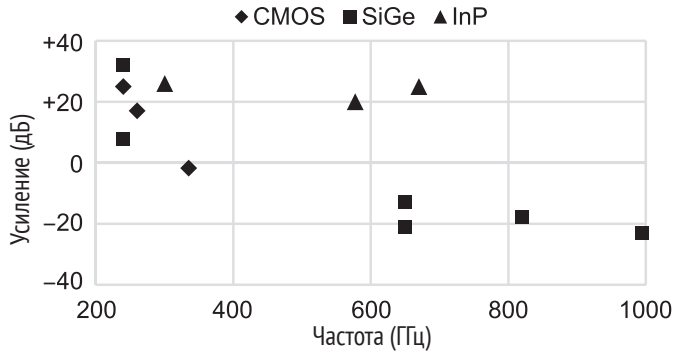


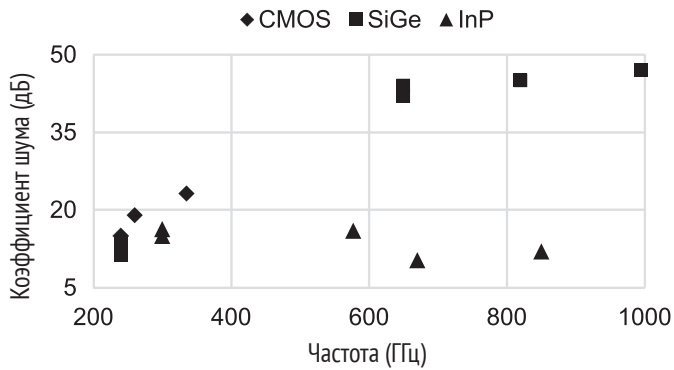
Рис. 16.3. Сравнение современных усилителей мощности, выполненных по различным технологиям. Данные для CMOS и SiGe взяты из [4], а данные для InP – из табл. 16.2

Приемники ТГц-диапазона можно разделить на гомодинные/гетеродинные и прямые детекторные. В статье [4] сравниваются различные приемники ТГц-диапазона, основанные на технологиях CMOS и SiGe, работающие в диапазоне от 200 ГГц до почти 1 ТГц. Здесь мы приводим неполный обзор приемников ТГц-диапазона, выполненных по технологиям InP и GaAs. В табл. 16.3, на рис. 16.4 и рис. 16.5 сравниваются характеристики приемников ТГц-диапазона в различных технологиях. Как показано на рис. 16.4, приемники InP и SiGe работали на частотах выше 500 ГГц. Сообщалось об усилении до 25 дБ на частоте 670 ГГц с использованием технологии InP HEMT [38]. Также в публикациях говорилось о нескольких трансиверах InP HEMT для использования в ТГц-связи [39–41]. Кремний-германиевые приемники в основном применяются для получения изображений и сканирования в ТГц-диапазоне [42].

Большинство прямых детекторов интегрированы с антеннами и поэтому известны как *прямые детекторы со связанной антенной*. Двумя наиболее распространенными формами интеграции являются антенны на кристалле и внешние гиперполусферические кремниевые линзовые антенны. Для оценки характеристик усиления и чувствительности этих детекторов используются такие критерии, как *чувствительность* и *эквивалентная мощность шума* (noise-equivalent power, NEP). По сравнению с гомодинными/гетеродинными приемниками, прямые детекторы вносят гораздо более высокий шум в рабочих диапазонах. В результате они в основном используются в сценариях, связанных со сканированием и визуализацией, где требования к шуму могут быть несколько ослаблены.



(а) Сравнение производительности



(б) Сравнение коэффициента шума

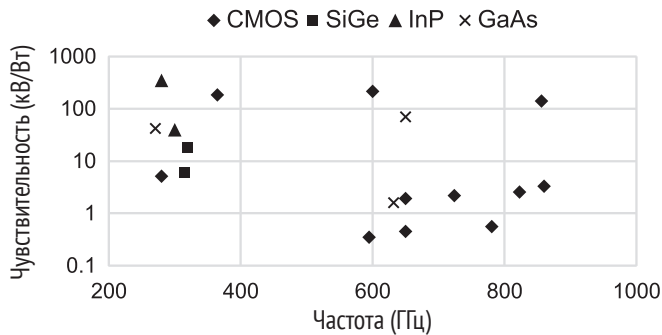
Рис. 16.4. Сравнение современных гомодинных/гетеродинных приемников, выполненных по различным технологиям.

Данные для CMOS и SiGe взяты из [4], а данные для InP – из табл. 16.3

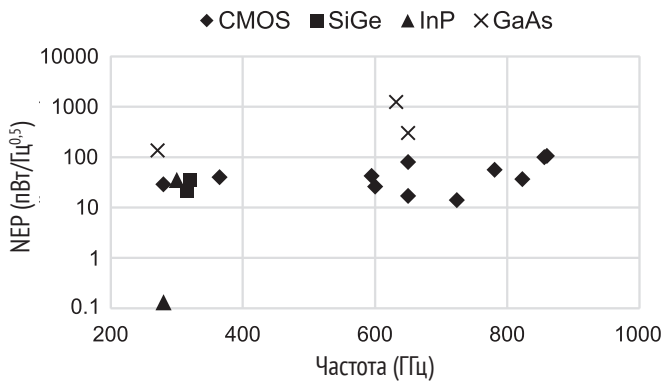
Таблица 16.3. Сравнение современных приемников ТГц-диапазона, выполненных по различным технологиям

Технология	Процесс (нм)	Частота (ГГц)	Усиление (дБ)	Коэф. шума (дБ)	Рассеиваемая мощность (мВт)	Ист.
InP HBT	130	577	20	16	–	[43]
InP HEMT	25	850	–	12	1160	[40]
InP HEMT	25	670	25	10,3	1800	[38]
InP HEMT	80	300	–	15	–	[39]
InP DHBT	250	300	26	16,3	482	[41]

Технология	Процесс (нм)	Частота (ГГц)	Чувствит. (кВ/Вт)	NEP (пВт/Гц ^{0,5})		Ист.
InP DHBT	250	280	350	0,13		[44]
InP HBT	250	300	40	35		[45]
GaAs HEMT	–	271	42	135		[46]
GaAs HEMT	–	632	1,6	1250		[46]
GaAs HEMT	–	650	70	300		[47]



(а) Сравнение показателей чувствительности



(б) Сравнение показателей NEP

Рис. 16.5. Сравнение современных приемников прямого детектирования терагерцового диапазона, выполненных по различным технологиям. Данные для CMOS и SiGe взяты из [4], а данные для InP – из табл. 16.3

16.1.2. Гибридный и фотонный подходы

Как упоминалось в предыдущей главе, непрямая запрещенная зона кремния препятствует интеграции в кристалл активных оптических

источников, таких как лазеры. Чтобы устранить это ограничение, требуется гетерогенная платформа III–V/кремний, которая обеспечит полную интеграцию различных оптических компонентов, таких как лазеры, оптические усилители и модуляторы, на одной кремниевой пластине. Это многообещающее направление разработок для следующего поколения фотонных интегральных схем.

При интеграции полупроводниковых компонентов III–V на кремний применяются два подхода: прямое выращивание материалов III–V на кремниевой подложке и гетерогенная интеграция посредством соединения пластин. Оба подхода продемонстрировали свою готовность к выпуску коммерческих продуктов.

Первая демонстрация гетерогенных III–V/кремниевых лазеров с распределенной обратной связью была проведена в 2008 г. [48]. С тех пор был достигнут значительный прогресс [49, 50]. Управляя локализацией электронов в квантовых ямах между слоями гетероструктуры (например, InGaAs и GaAs), лазеры на основе материалов III–V/кремния могут обеспечить улучшенные характеристики, такие как более быстрая модуляция и более высокая эффективность, по сравнению с обычными лазерами. В лазерах с *квантовыми точками* (quantum dot, QD) используются квантовые точки в активной области, и их можно выращивать непосредственно на кремниевой подложке. Использование квантовых точек в активной области позволяет удерживать носители заряда в трехмерном пространстве, что приводит к возникновению энергетических состояний, подобных тем, что свойственны атомам. Это свойство дает возможность обеспечить лучшую эффективность при высоких температурах, более низкий пороговый ток и более длительный срок службы по сравнению с лазерами на квантовых ямах и обычными лазерами.

Оптические усилители, как следует из названия, усиливают оптические сигналы в полупроводниковой усиливающей среде. Материалы типа III–V допускают более быстрое переключение, поскольку время жизни носителей короче из-за их прямой запрещенной зоны. Первая демонстрация оптического усилителя на основе III–V/кремния была проведена в 2007 г. [51], после чего был достигнут значительный прогресс. Обзор различных недавних достижений в области гетерогенных оптических усилителей из III–V/кремния представлен в [52]. Например, в [53] сообщили об оптическом усилителе из III–V/кремния с усилением до 28 дБ, а в [52] продемонстрировали оптический усилитель из III–V/кремния с выходной мощностью в режиме насыщения до 16,8 дБм. Гетерогенный оптический усилитель позволил найти новые применения фотонных интегральных схем, например в устройствах управления лучом, таких как LiDAR.

Оптические модуляторы переносят электрические сигналы в световую область. Обычные кремниевые оптические модуляторы ис-

пользуют эффект плазменной дисперсии в кремниевых устройствах, где показатель преломления материала изменяется в зависимости от изменения плотности свободного заряда. Это изменение приводит к изменениям фазы, которые впоследствии преобразуются в амплитудную модуляцию с помощью интерферометра или резонатора. Использование материалов III–V в гетерогенной платформе может помочь улучшить характеристики модулятора, предлагая большое изменение показателя преломления, высокую подвижность электронов и многое другое [49]. Кроме того, использование квантовых ям позволяет получить квантово-индуцированный эффект Старка (коэффициент поглощения материала меняется с изменением электрического поля) и может предложить решения быстрой коммутации для оптических модуляторов [54].

Фотодетектор на фотонной платформе функционально эквивалентен антенне. Он должен быть достаточно чувствительным, чтобы использовать мощность, которая соответствует требованию к коэффициенту битовых ошибок (bit error rate, BER) при заданной скорости передачи данных. Популярным кандидатом на материал для фотодетекторов является германий (Ge) из-за высокой пропускной способности, хорошей чувствительности и способности легко интегрироваться в платформы на основе CMOS. Фотоприемник на основе материалов III–V, как правило, может предложить сопоставимые или лучшие характеристики. Фотоприемник типа III–V с полосой пропускания более 67 ГГц и чувствительностью до 0,7 А/Вт был продемонстрирован в [55].

16.2. Системы ТГц-диапазона

Разработка различных компонентов ТГц-диапазона, пригодных для размещения на высокоинтегрированной платформе, позволила освоить множество применений ТГц-диапазона. Например, платформы на основе SiGe использовались в приложениях для беспроводной связи, сканирования и визуализации в ТГц-диапазоне, в основном на частотах ниже 300 ГГц [4]. Помимо этого, развитие технологий материалов III–V – например, с использованием InP – показало, что на их основе можно создавать быстродействующие ТГц-системы, как обсуждалось в предыдущих разделах. Гетерогенная интегрированная платформа III–V/кремний является многообещающим вариантом для создания как коммерческих, так и промышленных портативных ТГц-систем.

16.2.1. Терагерцовые системы связи

Беспроводная связь в терагерцовом диапазоне продемонстрировала большой потенциал во многих приложениях связи малого радиу-

са действия в определенных частотных «окнах» (например, 140, 220 и 300 ГГц). Целевая группа IEEE 802.15.3d исследовала спектр 252–325 ГГц и определила варианты использования, такие как киоски для загрузки данных, внутричиповая/внутриплата радиосвязь, беспроводная связь в центрах обработки данных, а также мобильные оконечные и транзитные каналы [56]. В этом разделе мы суммируем характеристики современной ТГц-системы в диапазоне 275–450 ГГц.

Как упоминалось в предыдущих разделах, спектр в терагерцовой области может испытывать большие потери на трассе из-за сильного атмосферного затухания. В табл. 16.4 оценивается бюджет линии связи на частотах 275–450 ГГц с усилением антенны при передаче (Tx) и приеме (Rx) 30 дБ, а также мощностью Tx 0 дБм, расстоянием между линиями 10 м и коэффициентом шума 15 дБ. Поскольку потери на трассе велики (~100 дБ), в качестве компенсации необходим узконаправленный стержневой луч.

Таблица 16.4. Оценка бюджета линии связи на частотах 275–450 ГГц

Мощность Tx (дБм)	Частота (ГГц)	Расстояние (м)	PL (дБ)	Атм. затухание (дБ/км)	Коеф. шума (дБ)	Усиление антенны Tx/Rx (дБ)	Мощность Rx (dBm)
0	275–296	10	–101,5	10	15	30	–56,6
0	306–313	10	–102,3	16	15	30	–57,4
0	318–333	10	–102,7	20	15	30	–57,9
0	356–450	10	–104,5	10	15	30	–59,6

В табл. 16.5 приведен неисчерпывающий перечень параметров последних современных систем беспроводной связи ТГц-диапазона и их характеристики в диапазоне от 275 ГГц до 450 ГГц. И электронный, и оптоэлектронный подходы продемонстрировали скорость передачи данных до 100 Гбит/с, хотя расстояние связи в большинстве случаев составляет менее двух метров. В статье [35] продемонстрирована система, использующая технологию InP – HEMT, которая обеспечивает скорость передачи данных до 120 Гбит/с на расстоянии 9,8 м. Также была создана оптоэлектронная система, состоящая из передатчика на фотонном диоде с односторонним движением носителя (uni-traveling-carrier photodiode, UTC-PD) и активного электронного приемника на основе технологии InGaAs HEMT, которая, как сообщается, способна обеспечивать скорость передачи данных до 100 Гбит/с на расстоянии 15 м.

Таблица 16.5. Сводка характеристик современного канала ТГц-диапазона

Технология	Частота (ГГц)	Скорость (Гбит/с)	Расстояние (м)	Мощность Тх (дБ)	Модуляция	Ист.
GaAs mHEMT	300	64	1	–4	QPSK	[57]
INP HEMT	300	20	0,8	+3	ASK	[58]
CMOS	300	56	0,05	–5,5	16QAM	[59]
CMOS	300	105	–	–5,5	32QAM	[60]
CMOS	300	20	0,1	–	16QAM	[61]
InP-HEMT	300	100	2,22	–	16QAM	[62]
InP-HEMT	300	120	9,8	–	16QAM	[35]
InGaAs	300	60	0,5	–7	16Q AM	[63]
Optoelectronic	280	100	0,5	–10	16QAM	[64]
Optoelectronic	300	10	0,3	–20	OOK	[65]
Optoelectronic	300	100	0,5	–16,1	16QAM	[66]
Optoelectronic	300	100	15	–8	32QAM	[67]
Optoelectronic	330	50	1	–10,5	ASK	[68]
Optoelectronic	350	100	2	–12	16QAM	[69]
Optoelectronic	350	100	2	+14	16QAM	[69]
Optoelectronic	385	32	0,5	–12	QPSK	[70]
Optoelectronic	400	60	0,5	–17	QPSK	[71]
Optoelectronic	400	60	0,5	–21	QPSK	[72]
Optoelectronic	450	132	1,8	+16	64QAM	[73]
Optoelectronic	350–475	120	0,5	–15	QPSK	[74]
Optoelectronic	400	106	0,5	–	QPSK	[75]
Optoelectronic	400	160	0,5	–17,5	QPSK	[76]

16.2.2. ТГц-визуализация и сенсорные системы

Терагерцовая часть спектра может найти широкое применение в области визуализации и сканирования, включая определение характеристик материалов, биомедицинскую визуализацию и биохимическое сканирование. В ТГц-спектроскопии применяется характерная реакция

образца на освещение ТГц-излучением, что позволяет использовать его для определения характеристик материала и сканирования предметов в целях безопасности. Благодаря более короткой длине волны отраженные сигналы можно использовать для распознавания жестов с повышенной точностью. Технология построения изображений в ближнем поле в ТГц-диапазоне может преодолеть фундаментальный дифракционный предел и предоставить изображения сверхвысокого разрешения, что открывает большие перспективы для медицинских применений, таких как визуализация биомолекул.

Спектроскопическую визуализацию в ТГц-диапазоне можно разделить на *спектроскопию во временной области* (time-domain spectroscopy, TDS) и *спектроскопию непрерывной волны* (continuous wave spectroscopy, CWS). В TDS генерируется импульс ТГц-диапазона, который направляется на образец материала. Переданные или отраженные сигналы собираются и затем преобразуются в частотную область. Этот записанный спектр содержит уникальный «отпечаток» образца, который можно использовать для определения свойств материала, из которого состоит образец. Для генерации широкополосного ТГц-сигнала можно использовать линзовую фотопроводящую антенну (которую мы опишем позже). Система TDS обычно громоздка, так как содержит вспомогательные оптические компоненты.

CWS работает по аналогичному принципу, но вместо генерации широкополосного импульса установка CWS генерирует узкополосный спектр для достижения высокочастотного разрешения. Электронная платформа на основе кремния может предложить компактное решение для реализации спектроскопии в более низком ТГц-диапазоне. В статье [77] продемонстрирован многодиапазонный (шесть гармоник 160 ГГц) набор микросхем визуализации, работающий в диапазоне от 160 до 1000 ГГц с использованием 250-нм процесса SiGe HBT. С развитием технологии III–V/кремния ожидаются дополнительные улучшения параметров оборудования спектроскопии.

Радиолокационная визуализация в ТГц-диапазоне использует принцип радарного дальномера и может обеспечить высокое разрешение/качество изображения. Спрос на радарные датчики в автомобильных системах, здравоохранении, мобильных устройствах и других приложениях в последнее время существенно увеличился – многообещающим кандидатом для удовлетворения таких требований является кремниевая платформа. В [78] сообщалось, что однокристалльный радиолокационный модуль с частотной модуляцией непрерывной волны (frequency modulated continuous wave, FMCW), работающий на частотах 210–270 ГГц с использованием 130-нм технологии SiGe HBT, смог достичь пространственного разрешения 2,57 мм.

Технология NFI (сканирование в ближнем поле) может создавать изображения со сверхвысоким разрешением в микрометровом или даже нанометровом диапазоне. Обычно формирование изображений в ближнем поле в ТГц-диапазоне выполняется с помощью сканирующей оптической микроскопии. Хотя разрешение может достигать нанометрового диапазона, система обычно громоздка и ее трудно интегрировать. Поэтому для получения изображений ближнего поля рассматривают встроенный на кристалл датчик SRR. Например, система формирования изображения ближнего поля с использованием 130-нм технологии SiGe была описана в [79], где датчик SRR применялся для считывания мощности ближнего поля (на расстоянии нескольких мкм), а затем сигнал подавался на измерительную схему. Сообщалось о достижении пространственного разрешения 10–12 мкм на частоте 550 ГГц.

16.3. Проблемы

Системы беспроводной связи как в миллиметровом, так и в ТГц-диапазонах испытывают большие потери в тракте передачи, хотя они более значительны в терагерцовом спектре. Например, потери на трассе длиной 10 м на 28 ГГц составляют 81 дБ, но увеличиваются до 101 дБ на 280 ГГц. Для компенсации высоких потерь на трассе обычно используются высоконаправленные антенные решетки. Чтобы генерировать высоконаправленный луч, равнофазная поверхность передаваемых лучей от всех антенных элементов должна быть перпендикулярна направлению распространения волны. В большинстве случаев для компенсации фазовых задержек, вызванных разнесением различных антенных элементов на определенные расстояния (обычно на половину длины волны), необходимы аналоговые фазовращатели. Фактический фазовый сдвиг в основном зависит от несущей частоты. Это становится проблемой в широкополосной системе, где используются несколько несущих, которые охватывают широкий частотный диапазон. Результирующий луч может рассеиваться с изменением частоты, что приводит к потере усиления решетки – эффекту, известному как перекося луча в миллиметровых волнах. В ТГц-системах этот эффект становится более серьезным, потому что полоса пропускания еще шире, а ширина луча чрезвычайно мала (такие лучи известны как «игловидные» или «карандашные» лучи). Сгенерированные лучи могут при изменении частоты отклоняться в разных направлениях, что приводит к еще более серьезным потерям усиления в решетке. Чтобы отличить эффект в ТГц-системах от эффекта в системах миллиметрового диапазона, этот эффект в ТГц-системах называется *расщеплением луча*. Недавно в [80, 81] для смягчения нежелательного эффекта предложили использовать метапо-

верхности (которые мы обсуждали в главе 15). Многообещающие результаты открыли новый путь к проектированию несложных систем связи с улучшенными характеристиками луча.

Хотя технология III–V/кремниевых полупроводников продемонстрировала пригодность к созданию масштабируемой и недорогой платформы для систем формирования изображений и связи в ТГц-диапазоне, несколько ключевых проблем остаются нерешенными. Первая проблема связана с широкополосными спектральными характеристиками материала и ограниченным диапазоном перестройки полупроводниковых источников, в результате чего рабочий диапазон источника ТГц-сигнала может оказаться недостаточным для покрытия всего спектрального диапазона сканируемого материала. Поэтому для спектроскопической визуализации необходимы источники ТГц-излучения, которые можно настраивать в широком диапазоне. Последние работы показали заметный прогресс; например, в [82] сообщается об источнике ТГц-диапазона с широким диапазоном перестройки (0,04–0,99 ТГц). Кроме того, для получения изображений в ТГц-диапазоне нужны настраиваемые источники с подходящими возможностями формирования луча, чтобы избежать громоздкости, присущей системам механического сканирования. Требуется компактная интеграция реконфигурируемых источников с большими антенными решетками ТГц-диапазона, и ожидается, что эти источники упростят реализацию передовых алгоритмов построения изображения, таких как сжатые выборки.

Достижения технологии полупроводников типа III-V/кремний позволили выполнять гетерогенную интеграцию различных высокопроизводительных компонентов, таких как электронные, фотонные, магнитные и графеновые компоненты. Поэтому для обеспечения оптимальных характеристик необходима комплексная платформа, которая была бы одновременно компактной и весьма разнородной, а также вбирала бы в себя все упомянутые компоненты. Более того, последние достижения в области плазмонных технологий открыли многообещающие перспективы интеграции плазмонных компонентов на кристалл кремния [83, 84]. Ожидается, что в будущем работа будет продолжена, несмотря на сложный характер этой области, с целью дальнейшего развития технологии переноса плазмонных компонентов на кремний до коммерческого уровня. Следовательно, оптимизированная интеграция этих гетерогенных компонентов на одной кремниевой пластине может привести нас к созданию электронно-фотонной/плазмонной системы, способной обеспечить высокую производительность наряду с оптимальным балансом между стоимостью, эффективностью и программируемостью, – системы для следующего поколения коммерческих и промышленных ТГц-устройств.

16.4. Источники

- [1] P. Garcia, A. Chantre, S. Pruvost, P. Chevalier, S. T. Nicolson, D. Roy, S. P. Voinigescu, and C. Garnier, *Will BiCMOS stay competitive for mmW applications?*, in Proc. 2008 IEEE Custom Integrated Circuits Conference. IEEE, 2008, pp. 387–394.
- [2] B. Heinemann, H. Rücker, R. Barth, F. Bärwolf, J. Drews, G. Fischer, A. Fox, O. Fursenko, T. Grabolla, F. Herzel et al., *SiGe HBT with f_x/f_{max} of 505 GHz/720 GHz*, in Proc. 2016 IEEE International Electron Devices Meeting (IEDM). IEEE, 2016, pp. 3.1.1–3.1.4.
- [3] M. Urteaga, Z. Griffith, M. Seo, J. Hacker, and M. J. Rodwell, *InP HBT technologies for THz integrated circuits*, Proceedings of the IEEE, vol. 105, no. 6, pp. 1051–1067, 2017.
- [4] U. R. Pfeiffer, R. Jain, J. Grzyb, S. Malz, P. Hillger, and P. Rodríguez-Vázquez, *Current status of terahertz integrated circuits – from components to systems*, in Proc. 2018 IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium (BCICTS). IEEE, 2018, pp. 1–7.
- [5] M. Hossain, N. Weimann, M. Brahém, O. Ostinelli, C. R. Bolognesi, W. Heinrich, and V. Krozer, *A 0.5 THz signal source with 11 dBm peak output power based on InP DHBT*, in Proc. 2019 49th European Microwave Conference (EuMC). IEEE, 2019, pp. 856–859.
- [6] J. S. Rieh, J. Yun, D. Yoon, J. Kim, and H. Son, *Terahertz InP HBT oscillators*, in Proc. 2018 IEEE International Symposium on Radio-Frequency Integration Technology (RFIT). IEEE, 2018, pp. 1–3.
- [7] J. Yun, D. Yoon, H. Kim, and J.-S. Rieh, *300-GHz InP HBT oscillators based on common-base cross-coupled topology*, IEEE Transactions on Microwave Theory and Techniques, vol. 62, no. 12, pp. 3053–3064, 2014.
- [8] J.-Y. Kim, H.-J. Song, K. Ajito, M. Yaita, and N. Kukutsu, *InP HBT voltage controlled oscillator for 300-GHz-band wireless communications*, in Proc. 2012 International SoC Design Conference (ISOC). IEEE, 2012, pp. 262–265.
- [9] D. Kim and S. Jeon, *A WR-3 band fundamental voltage-controlled oscillator with a wide frequency tuning range and high output power*, IEEE Transactions on Microwave Theory and Techniques, vol. 67, no. 7, pp. 2759–2768, 2019.
- [10] D. Kim and S. Jeon, *A 300-GHz high-power high-efficiency voltage-controlled oscillator with low power variation*, IEEE Microwave and Wireless Components Letters, vol. 30, no. 5, pp. 496–499, 2020.
- [11] T. K. Johansen, M. Hossain, S. Boppel, R. Doerner, V. Krozer, and W. Heinrich, *A 300 GHz active frequency tripler in transferred-substrate*

- InP DHBT technology*, in Proc. 2019 14th European Microwave Integrated Circuits Conference (EuMIC). IEEE, 2019, pp. 180–183.
- [12] M. Hossain, S. Boppel, W. Heinrich, and V. Krozer, *Efficient active multiplier-based signal source for 300 GHz system applications*, Electronics Letters, vol. 55, no. 23, pp. 1220–1221, 2019.
 - [13] M. Hossain, K. Nosaeva, N. Weimann, V. Krozer, and W. Heinrich, *A 330 GHz active frequency quadrupler in InP DHBT transferred-substrate technology*, in Proc. 2016 IEEE MTT-S International Microwave Symposium (IMS). IEEE, 2016, pp. 1–4.
 - [14] D. Yoon, J. Yun, and J.-S. Rieh, *A 310–340-GHz coupled-line voltage-controlled oscillator based on 0.25- μ m InP HBT technology*, IEEE Transactions on Terahertz Science and Technology, vol. 5, no. 4, pp. 652–654, 2015.
 - [15] M. Seo, M. Urteaga, J. Hacker, A. Young, Z. Griffith, V. Jain, R. Pierson, P. Rowell, A. Skalare, A. Peralta et al., *InP HBT IC technology for terahertz frequencies: Fundamental oscillators up to 0.57 THz*, IEEE Journal of Solid-State Circuits, vol. 46, no. 10, pp. 2203–2214, 2011.
 - [16] J. Yun, J. Kim, D. Yoon, and J.-S. Rieh, *645-GHz InP heterojunction bipolar transistor harmonic oscillator*, Electronics Letters, vol. 53, no. 22, pp. 1475–1477, 2017.
 - [17] A. Zamora, K. M. Leong, G. Mei, M. Lange, W. Yoshida, K. T. Nguyen, B. S. Gorospe, and W. R. Deal, *A high efficiency 670 GHz x36 InP HEMT multiplier chain*, in Proc. 2017 IEEE MTT-S International Microwave Symposium (IMS). IEEE, 2017, pp. 977–979.
 - [18] S. Daneshgar and J. F. Buckwalter, *Compact series power combining using subquarte wavelength baluns in silicon germanium at 120 GHz*, IEEE Transactions on Microwave Theory and Techniques, vol. 66, no. 11, pp. 4844–4859, 2018.
 - [19] M. Urteaga, Z. Griffith, M. Seo, J. Hacker, and M. J. Rodwell, *InP HBT technologies for THz integrated circuits*, Proceedings of the IEEE, vol. 105, no. 6, pp. 1051–1067, 2017.
 - [20] H. Wang, F. Wang, H. Nguyen, S. Li, T. Huang, A. Ahmed, M. Smith, N. Mannem, and J. Lee, *Power amplifiers performance survey 2000–present*, vol. 10, 2018.
 - [21] K. M. Leong, X. Mei, W. Yoshida, P.-H. Liu, Z. Zhou, M. Lange, L.-S. Lee, J. G. Padilla, A. Zamora, B. S. Gorospe et al., *A 0.85 THz low noise amplifier using InP HEMT transistors*, IEEE Microwave and Wireless Components Letters, vol. 25, no. 6, pp. 397–399, 2015.
 - [22] M. Seo, M. Urteaga, J. Hacker, A. Young, A. Skalare, R. Lin, and M. Rodwell, *A 600 GHz InP HBT amplifier using cross-coupled feedback stabilization and dual-differential power combining*, in 2013 IEEE MTT-S International

- Microwave Symposium Digest (MTT). IEEE, 2013, pp. 1–3.
- [23] Z. Wang, P.-Y. Chiang, P. Nazari, C.-C. Wang, Z. Chen, and P. Heydari, *A CMOS 210-GHz fundamental transceiver with OOK modulation*, IEEE Journal of Solid-State Circuits, vol. 49, no. 3, pp. 564–580, 2014.
 - [24] M. H. Eissa and D. Kissinger, *4.5 A 13.5 dBm fully integrated 200-to-255 GHz power amplifier with a 4-way power combiner in SiGe: C BiCMOS*, in Proc. 2019 IEEE International Solid-State Circuits Conference (ISSCC). IEEE, 2019, pp. 82–84.
 - [25] N. Sarmah, K. Aufinger, R. Lachner, and U. R. Pfeiffer, *A 200–225 GHz SiGe power amplifier with peak P_{sat} of 9.6 dBm using wideband power combination*, in Proc. ESSCIRC Conference 2016: 42nd European Solid-State Circuits Conference. IEEE, 2016, pp. 193–196.
 - [26] M. Ćwikliński, P. Brückner, S. Leone, C. Friesicke, R. Lozar, H. Maßler, R. Quay, and O. Ambacher, *190-GHz G-band GaN amplifier MMICs with 40 GHz of bandwidth*, in Proc. 2019 IEEE MTT-S International Microwave Symposium (IMS). IEEE, 2019, pp. 1257–1260.
 - [27] L. John, A. Tessmann, A. Leuther, P. Neiningner, and T. Zwick, *Investigation of compact power amplifier cells at THz frequencies using InGaAs mHEMT technology*, in Proc. 2019 IEEE MTT-S International Microwave Symposium (IMS). IEEE, 2019, pp. 1261–1264.
 - [28] L. John, P. Neiningner, C. Friesicke, A. Tessmann, A. Leuther, M. Schlechtweg, and T. Zwick, *A 280–310 GHz InAlAs/InGaAs mHEMT power amplifier MMIC with 6.7–8.3 dBm output power*, IEEE Microwave and Wireless Components Letters, vol. 29, no. 2, pp. 143–145, 2018.
 - [29] A. Tessmann, A. Leuther, V. Hurm, H. Massler, S. Wagner, M. Kuri, M. Zink, M. Riessle, H.-P. Stulz, M. Schlechtweg et al., *A broadband 220–320 GHz medium power amplifier module*, in Proc. 2014 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS). IEEE, 2014, pp. 1–4.
 - [30] V. Radisic, W. R. Deal, K. M. Leong, X. Mei, W. Yoshida, P.-H. Liu, J. Uyeda, A. Fung, L. Samoska, T. Gaier et al., *A 10-mW submillimeter-wave solid-state power-amplifier module*, IEEE Transactions on Microwave Theory and Techniques, vol. 58, no. 7, pp. 1903–1909, 2010.
 - [31] Z. Griffith, M. Urteaga, P. Rowell, and R. Pierson, *A 6–10 mW power amplifier at 290–307.5 GHz in 250 nm InP HBT*, IEEE Microwave and Wireless Components Letters, vol. 25, no. 9, pp. 597–599, 2015.
 - [32] J. Kim, S. Jeon, M. Kim, M. Urteaga, and J. Jeong, *H-band power amplifier integrated circuits using 250-nm InP HBT technology*, IEEE Transactions on Terahertz Science and Technology, vol. 5, no. 2, pp. 215–222, 2015.
 - [33] A. S. Ahmed, A. Simsek, M. Urteaga, and M. J. Rodwell, *8.6–13.6 mW series-connected power amplifiers designed at 325 GHz using 130 nm InP HBT*

- technology, in Proc. 2018 IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium (BCICTS). IEEE, 2018, pp. 164–167.
- [34] J. Hacker, M. Urteaga, D. Mensa, R. Pierson, M. Jones, Z. Griffith, and M. Rodwell, *250 nm InP DHBT monolithic amplifiers with 4.8 dB gain at 324 GHz*, in 2008 IEEE MTT-S International Microwave Symposium Digest. IEEE, 2008, pp. 403–406.
- [35] H. Hamada, T. Tsutsumi, G. Itami, H. Sugiyama, H. Matsuzaki, K. Okada, and H. Nosaka, *300-GHz 120-Gb/s wireless transceiver with high-output-power and high-gain power amplifier based on 80-nm InP-HEMT technology*, in Proc. 2019 IEEE BiCMOS and Compound semiconductor Integrated Circuits and Technology Symposium (BCICTS). IEEE, 2019, pp. 1–4.
- [36] H. Hamada, T. Kosugi, H.-J. Song, M. Yaita, A. El Moutaouakil, H. Matsuzaki, and A. Hirata, *300-GHz band 20-Gbps ASK transmitter module based on InP-HEMT MMICs*, in Proc. 2015 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS). IEEE, 2015, pp. 1–4.
- [37] J. Hacker, M. Urteaga, M. Seo, A. Skalare, and R. Lin, *InP HBT amplifier MMICs operating to 0.67 THz*, in 2013 IEEE MTT-S International Microwave Symposium Digest (MTT). IEEE, 2013, pp. 1–3.
- [38] W. Deal, K. Leong, A. Zamora, W. Yoshida, M. Lange, B. Gorospe, K. Nguyen, and G. X. Mei, *A low-power 670-GHz InP HEMT receiver*, IEEE Transactions on Terahertz Science and Technology, vol. 6, no. 6, pp. 862–864, 2016.
- [39] H. Hamada, T. Tsutsumi, H. Matsuzaki, T. Fujimura, I. Abdo, A. Shirane, K. Okada, G. Itami, H.-J. Song, H. Sugiyama et al., *300-GHz-band 120-Gb/s wireless front-end based on InP-HEMT PAs and mixers*, IEEE Journal of Solid-State Circuits, vol. 55, no. 9, pp. 2316–2335, 2020.
- [40] K. M. Leong, X. Mei, W. H. Yoshida, A. Zamora, J. G. Padilla, B. S. Gorospe, K. Nguyen, and W. R. Deal, *850 GHz receiver and transmitter front-ends using InP HEMT*, IEEE Transactions on Terahertz Science and Technology, vol. 7, no. 4, pp. 466–475, 2017.
- [41] S. Kim, J. Yun, D. Yoon, M. Kim, J.-S. Rieh, M. Urteaga, and S. Jeon, *300 GHz integrated heterodyne receiver and transmitter with on-chip fundamental local oscillator and mixers*, IEEE Transactions on Terahertz Science and Technology, vol. 5, no. 1, pp. 92–101, 2014.
- [42] K. Statnikov, J. Grzyb, B. Heinemann, and U. R. Pfeiffer, *160-GHz to 1-THz multi-color active imaging with a lens-coupled SiGe HBT chip-set*, IEEE Transactions on Microwave Theory and Techniques, vol. 63, no. 2, pp. 520–532, 2015.

- [43] M. Urteaga, Z. Griffith, M. Seo, J. Hacker, and M. J. Rodwell, *InP HBT technologies for THz integrated circuits*, Proceedings of the IEEE, vol. 105, no. 6, pp. 1051–1067, 2017.
- [44] C. Yi, M. Urteaga, S. H. Choi, and M. Kim, *A 280-GHz InP DHBT receiver detector containing a differential preamplifier*, IEEE Transactions on Terahertz Science and Technology, vol. 7, no. 2, pp. 209–217, 2017.
- [45] J. Yun, S. J. Oh, K. Song, D. Yoon, H. Y. Son, Y. Choi, Y.-M. Huh, and J.-S. Rieh, *Terahertz reflection-mode biological imaging based on InP HBT source and detector*, IEEE Transactions on Terahertz Science and Technology, vol. 7, no. 3, pp. 274–283, 2017.
- [46] E. Javadi, A. Laisauskas, M. Shahabadi, N. Masoumi, J. Zhang, J. Matukas, and H. G. Roskos, *Terahertz detection with a low-cost packaged GaAs high-electron-mobility transistor*, IEEE Transactions on Terahertz Science and Technology, vol. 9, no. 1, pp. 27–37, 2018.
- [47] D. M. Yermolaev, I. Khmyrova, E. Polushkin, A. Kovalchuk, V. Gavrilenko, K. Maremyanin, N. Maleev, V. Ustinov, V. Zemlyakov, V. A. Bespalov et al., *Detector for terahertz applications based on a serpentine array of integrated GaAs/InGaAs/AlGaAs field-effect transistors*, in Proc. 2017 International Conference on Applied Electronics (AE). IEEE, 2017, pp. 1–4.
- [48] A. W. Fang, E. Lively, Y.-H. Kuo, D. Liang, and J. E. Bowers, *A distributed feedback silicon evanescent laser*, Optics Express, vol. 16, no. 7, pp. 4413–4419, 2008.
- [49] T. Komljenovic, D. Huang, P. Pintus, M. A. Tran, M. L. Davenport, and J. E. Bowers, *Photonic integrated circuits using heterogeneous integration on silicon*, Proceedings of the IEEE, vol. 106, no. 12, pp. 2246–2257, 2018.
- [50] M. Tang, J.-S. Park, Z. Wang, S. Chen, P. Jurczak, A. Seeds, and H. Liu, *Integration of III-V lasers on Si for Si photonics*, Progress in Quantum Electronics, vol. 66, pp. 1–18, 2019.
- [51] H. Park, A. W. Fang, O. Cohen, R. Jones, M. J. Paniccia, and J. E. Bowers, *A hybrid AlGaInAs–silicon evanescent amplifier*, IEEE Photonics Technology Letters, vol. 19, no. 4, pp. 230–232, 2007.
- [52] M. L. Davenport, S. Skendžić, N. Volet, J. C. Hulme, M. J. Heck, and J. E. Bowers, *Heterogeneous silicon/III–V semiconductor optical amplifiers*, IEEE Journal of Selected Topics in Quantum Electronics, vol. 22, no. 6, pp. 78–88, 2016.
- [53] P. Kaspar, G. de Valicourt, R. Brenot, M. A. Mestre, P. Jennevé, A. Accard, D. Make, F. Lelarge, G.-H. Duan, N. Pavarelli et al., *Hybrid III-V/silicon SOA in optical network based on advanced modulation formats*, IEEE Photonics Technology Letters, vol. 27, no. 22, pp. 2383–2386, 2015.
- [54] D. A. Miller, *Device requirements for optical interconnects to silicon chips*,

- Proceedings of the IEEE, vol. 97, no. 7, pp. 1166–1185, 2009.
- [55] L. Shen, Y. Jiao, W. Yao, Z. Cao, J. van Engelen, G. Roelkens, M. Smit, and J. van der Tol, *High-bandwidth uni-traveling carrier waveguide photodetector on an InP-membrane-on-silicon platform*, Optics Express, vol. 24, no. 8, pp. 8290–8301, 2016.
 - [56] K. Sengupta, T. Nagatsuma, and D. M. Mittleman, *Terahertz integrated electronic and hybrid electronic–photonic systems*, Nature Electronics, vol. 1, no. 12, pp. 622–635, 2018.
 - [57] I. Kallfass, P. Harati, I. Dan, J. Antes, F. Boes, S. Rey, T. Merkle, S. Wagner, H. Massler, A. Tessmann et al., *MMIC chipset for 300 GHz indoor wireless communication*, in Proc. 2015 IEEE International Conference on Microwaves, Communications, Antennas and Electronic Systems (COMCAS). IEEE, 2015, pp. 1–4.
 - [58] H.-J. Song, T. Kosugi, H. Hamada, T. Tajima, A. El Moutaouakil, H. Matsuzaki, Y. Kawano, T. Takahashi, Y. Nakasha, N. Hara et al., *Demonstration of 20-Gbps wireless data transmission at 300 GHz for KIOSK instant data downloading applications with InP MMICs*, in Proc. 2016 IEEE MTT-S International Microwave Symposium (IMS). IEEE, 2016, pp. 1–4.
 - [59] K. Takano, K. Katayama, S. Amakawa, T. Yoshida, and M. Fujishima, *56-Gbit/s 16-QAM wireless link with 300-GHz-band CMOS transmitter*, in Proc. 2017 IEEE MTT-S International Microwave Symposium (IMS). IEEE, 2017, pp. 793–796.
 - [60] K. Takano, S. Amakawa, K. Katayama, S. Hara, R. Dong, A. Kasamatsu, I. Hosako, K. Mizuno, K. Takahashi, T. Yoshida et al., *17.9 A 105 Gb/s 300 GHz CMOS transmitter*, in Proc. 2017 IEEE International Solid-State Circuits Conference (ISSCC). IEEE, 2017, pp. 308–309.
 - [61] S. Hara, K. Takano, K. Katayama, R. Dong, S. Lee, I. Watanabe, N. Sekine, A. Kasamatsu, T. Yoshida, S. Amakawa et al., *300-GHz CMOS transceiver for terahertz wireless communication*, in Proc. 2018 Asia-Pacific Microwave Conference (APMC). IEEE, 2018, pp. 429–431.
 - [62] H. Hamada, T. Fujimura, I. Abdo, K. Okada, H.-J. Song, H. Sugiyama, H. Matsuzaki, and H. Nosaka, *300-GHz. 100-Gb/s InP-HEMT wireless transceiver using a 300-GHz fundamental mixer*, in Proc. 2018 IEEE/MTT-S International Microwave Symposium- IMS. IEEE, 2018, pp. 1480–1483.
 - [63] I. Dan, G. Ducournau, S. Hisatake, P. Szriftgiser, R.-P. Braun, and I. Kallfass, *A terahertz wireless communication link using a superheterodyne approach*, IEEE Transactions on Terahertz Science and Technology, vol. 10, no. 1, pp. 32–43, 2019.
 - [64] V. Chinni, P. Latzel, M. Zégaoui, C. Coinon, X. Wallart, E. Peytavit, J. Lampin, K. Engenhardt, P. Szriftgiser, M. Zaknounge et al., *Single-*

- channel 100 Gbit/s transmission using III–V UTC-PDs for future IEEE 802.15. 3D wireless links in the 300 GHz band*, Electronics Letters, vol. 54, no. 10, pp. 638–640, 2018.
- [65] E. Lacombe, C. Belem-Goncalves, C. Luxey, F. Ganesello, C. Durand, D. Gloria, and G. Ducournau, *10-Gb/s indoor THz communications using industrial si photonics technology*, IEEE Microwave and Wireless Components Letters, vol. 28, no. 4, pp. 362–364, 2018.
- [66] C. Castro et al., *32 GBd 16QAM wireless transmission in the 300 GHz band using a PINdiode for THz upconversion*, in Proc. Optical Fiber Communications Conference and Exhibition (OFC). Optical Society of America, 2019.
- [67] I. Dan, P. Szriftgiser, E. Peytavit, J.-F. Lampin, M. Zegaoui, M. Zaknoune, G. Ducournau, and I. Kallfass, *A 300-GHz wireless link employing a photonic transmitter and an active electronic receiver with a transmission bandwidth of 54 GHz*, IEEE Transactions on Terahertz Science and Technology, vol. 10, no. 3, pp. 271–281, 2020.
- [68] T. Nagatsuma and G. Carpintero, *Recent progress and future prospect of photonics-enabled terahertz communications research*, IEICE Transactions on Electronics, vol. 98, no. 12, pp. 1060–1070, 2015.
- [69] K. Liu, S. Jia, S. Wang, X. Pang, W. Li, S. Zheng, H. Chi, X. Jin, X. Zhang, and X. Yu, *100 Gbit/s THz photonic wireless transmission in the 350-GHz band with extended reach*, IEEE Photonics Technology Letters, vol. 30, no. 11, pp. 1064–1067, 2018.
- [70] G. Ducournau, K. Engenhardt, P. Szriftgiser, D. Bacquet, M. Zaknoune, R. Kassi, E. Lecomte, and J.-F. Lampin, *32 Gbit/s QPSK transmission at 385 GHz using coherent fibre-optic technologies and THz double heterodyne detection*, Electronics Letters, vol. 51, no. 12, pp. 915–917, 2015.
- [71] X. Yu, R. Asif, M. Piels, D. Zibar, M. Galili, T. Morioka, P. U. Jepsen, and L. K. Oxenløwe, *400-GHz wireless transmission of 60-Gb/s Nyquist-QPSK signals using UTC-PD and heterodyne mixer*, IEEE Transactions on Terahertz Science and Technology, vol. 6, no. 6, pp. 765–770, 2016.
- [72] X. Yu, R. Asif, M. Piels, D. Zibar, M. Galili, T. Morioka, P. U. Jepsen, and L. K. Oxenløwe, *60 Gbit/s 400 GHz wireless transmission*, in Proc. 2015 International Conference on Photonics in Switching (PS). IEEE, 2015, pp. 4–6.
- [73] X. Li, J. Yu, L. Zhao, W. Zhou, K. Wang, M. Kong, G.-K. Chang, Y. Zhang, X. Pan, and X. Xin, *132-Gb/s photonics-aided single-carrier wireless terahertz-wave signal transmission at 450 GHz enabled by 64QAM modulation and probabilistic shaping*, in Proc. Optical Fiber Communication Conference. Optical Society of America, 2019, pp. M4F–4.

- [74] S. Jia, X. Yu, H. Hu, J. Yu, T. Morioka, P. U. Jepsen, and L. K. Oxenløwe, *120 Gb/s multi-channel THz wireless transmission and THz receiver performance analysis*, IEEE Photonics Technology Letters, vol. 29, no. 3, pp. 310–313, 2017.
- [75] S. Jia, X. Pang, O. Ozolins, X. Yu, H. Hu, J. Yu, P. Guan, F. Da Ros, S. Popov, G. Jacobsen et al., *0.4 THz photonic-wireless link with 106 Gb/s single channel bitrate*, Journal of Lightwave Technology, vol. 36, no. 2, pp. 610–616, 2018.
- [76] X. Yu, S. Jia, H. Hu, M. Galili, T. Morioka, P. U. Jepsen, and L. K. Oxenløwe, *Ultrabroadband THz photonic wireless transmission*, in Proc. 2018 23rd Opto-Electronics and Communications Conference (OECC). IEEE, 2018, pp. 1–2.
- [77] K. Statnikov, J. Grzyb, B. Heinemann, and U. R. Pfeiffer, *160-GHz to 1-THz multi-color active imaging with a lens-coupled SiGe HBT chip-set*, IEEE Transactions on Microwave Theory and Techniques, vol. 63, no. 2, pp. 520–532, 2015.
- [78] J. Grzyb, K. Statnikov, N. Sarmah, B. Heinemann, and U. R. Pfeiffer, *A 210–270-GHz circularly polarized FMCW radar with a single-lens-coupled SiGe HBT chip*, IEEE Transactions on Terahertz Science and Technology, vol. 6, no. 6, pp. 771–783, 2016.
- [79] P. Hillger, R. Jain, J. Grzyb, W. Förster, B. Heinemann, G. MacGrogan, P. Mounaix, T. Zimmer, and U. R. Pfeiffer, *A 128-pixel system-on-a-chip for real-time superresolution terahertz near-field imaging*, IEEE Journal of Solid-State Circuits, vol. 53, no. 12, pp. 3599–3612, 2018.
- [80] A. Mehdipour, J. W. Wong, and G. V. Eleftheriades, *Beam-squinting reduction of leaky-wave antennas using Huygens metasurfaces*, IEEE Transactions on Antennas and Propagation, vol. 63, no. 3, pp. 978–992, 2015.
- [81] M. Faenzi, G. Minatti, D. Gonzalez-Ovejero, F. Caminita, E. Martini, C. Della Giovampaola, and S. Maci, *Metasurface antennas: New models, applications and realizations*, Scientific Reports, vol. 9, no. 1, pp. 1–14, 2019.
- [82] X. Wu and K. Sengupta, *Single-chip source-free terahertz spectroscope across 0.04–0.99 THz: Combining sub-wavelength near-field sensing and regression analysis*, Optics Express, vol. 26, no. 6, pp. 7163–7175, 2018.
- [83] T. Harter, S. Muehlbrandt, S. Ummethala, A. Schmid, S. Nellen, L. Hahn, W. Freude, and C. Koos, *Silicon-plasmonic integrated circuits for terahertz signal generation and coherent detection*, Nature Photonics, vol. 12, no. 10, pp. 625–633, 2018.
- [84] S. Moazeni, *CMOS and plasmonics get close*, Nature Electronics, vol. 3, no. 6, pp. 302–303, 2020.

Глава 17

Вычислительная техника после закона Мура

Закон Мура в течение последних нескольких десятилетий успешно предсказывал тенденции развития технологической индустрии. Но теперь, с развитием кремниевой литографии и миниатюризации электроники, ожидается, что к 2025 году кривая роста станет более плоской [1]. В ближайшей перспективе платформы на основе кремния будут продолжать способствовать повышению производительности интегральных схем и расширению функциональности за счет гетерогенной интеграции полупроводников типа III–V, фотонных/плазмонных и других передовых материалов, как упоминалось в главе 14. Новые вычислительные архитектуры, такие как специализированные вычислительные системы, оборудование и оптимизированные алгоритмы, помогут повысить производительность вычислений со стороны программного обеспечения. В долгосрочной перспективе две вычислительные технологии обещают совершить фундаментальный скачок вперед с точки зрения вычислительной производительности. Одна из них – это технологии, основанные на подражании человеческому мозгу, такие как нейроморфные вычисления и глубокое обучение, которые позволяют нам обрабатывать информацию более эффективно, как биологический мозг. Другая – квантовые вычисления, использующие преимущества полиномиальной суперпозиционной природы кубитов; это может значительно снизить вычислительную нагрузку. Мы обсудим упомянутые технологии в данной главе.

17.1. Постмуровская эра

В 1965 году Гордон Мур утверждал, что количество транзисторов в той или иной области вычислений будет удваиваться каждые два года [2]. В 1975 году его предсказание подтвердилось, и эта тенденция сохранялась до середины 2000-х годов. С тех пор кривая роста становилась все

более плоской, и все больше крепло мнение, что электронная отрасль вошла в эру, когда закон Мура утратил свою силу.

За последние 50 лет был достигнут значительный прогресс в уменьшении размеров компонентов. Например, длина затвора полевых транзисторов металл-оксид-полупроводник (metal-oxide-semiconductor field-effect transistors, MOSFET) уменьшилась с 10 мкм в начале 1970-х годов до 5 нм сегодня [2]. После достижения длины затвора 50 нм полевые транзисторы разделились на два разных семейства: плоская полностью обедненная структура «кремний на изоляторе» и трехмерная структура многозатворного полевого транзистора (fin field-effect transistor, FinFET). В настоящее время достижения в области литографии в дальнейшем (коротковолновом) ультрафиолете и множество усовершенствований в различных процессах привели к появлению транзисторов атомарного уровня. Полагаться исключительно на закон Мура больше нельзя; потребуются новые концепции устройств, архитектур и вычислительных парадигм.

Устройства на основе кремния продолжают развиваться в течение следующих нескольких лет. Например, для реализации вертикального наращивания полупроводниковых структур были разработаны многослойные интегральные схемы, а гетерогенная интеграция кремния с полупроводниками типа III–V и другими функциональными материалами продолжит повышать быстродействие и эффективность электронных компонентов. Кроме того, кремниевая фотоника – интеграция фотонных компонентов на кремниевом чипе – расширит функциональность устройств в фотонной области. Усилия в этих направлениях еще долго будут приносить свои плоды по мере развития нанотехнологий, эффективного управления энергопотреблением и усовершенствованного проектирования схем.

Массовое производство 7-нм FinFET-транзисторов началось в 2018 году, а предел 5 нм был достигнут только двумя годами позже, в 2020 году. Хотя уменьшение размера на первый взгляд не кажется значительным – с 7 нм до 5 нм, на самом деле это сокращение на 30 %, что весьма существенно. Следуя этой тенденции, мы можем рассчитывать на создание транзисторов атомарного уровня к 2022–2024 гг. В атомарном масштабе начнут преобладать квантовые эффекты, что позволит нам реализовать новые методы, которые невозможно использовать сегодня, и сделает традиционную схемную логику устаревшей. Ожидается, что к 2030 году технология 2 нм будет готова к выпуску первой партии устройств 6G. Долгосрочная дорожная карта развития технологии полупроводниковых чипов имеет решающее значение для успешной коммерциализации 6G.

В долгосрочной перспективе необходимы фундаментальные успехи в достижении высокой энергоэффективности манипулирования

информацией во всех контекстах. Чтобы добиться смены парадигмы в мире вычислений, необходимы технологии, которые могут преодолеть ограничения CMOS. Исторический опыт развития транзисторов показывает, что требуется около 10 лет, чтобы новации на уровне фундаментальной физики достигли массового применения. Ожидается, что в течение следующих двух десятилетий продолжающиеся исследования принесут свои плоды, кульминацией которых станут радикальные усовершенствования в процессах создания микросхем, новая вычислительная парадигма и связанные с ними практические применения.

Заслуживают внимания и экономические соображения. Ожидается, что инвестиции в серийное производство новых кремниевых чипов для технологии 5 нм составят 12 миллиардов долларов США. Количество поставщиков, которые могут построить такие объекты и имеют опыт их эксплуатации, чрезвычайно мало. Единовременные затраты на разработку специализированных интегральных схем (ASIC) в этих процессах составляют десятки миллионов долларов, а это означает, что новые производственные процессы имеют экономический смысл только для тех, у кого очень большие объемы.

17.2. Нейроморфные вычисления

С появлением ИИ вычислительные технологии, имитирующие функции мозга (например, нейронные сети и машинное обучение), произвели революцию в компьютерной индустрии. ИИ манипулирует данными так же, как биологический мозг, и может выполнять самообучение, чтобы распознавать скрытые закономерности в больших объемах данных. Первоначально ИИ был разработан для выполнения некоторых базовых функций, таких как распознавание речи и обработка изображений, но с тех пор он достиг значительного прогресса и способен выполнять намного более сложные функции, такие как игра в шахматы, сочинение музыки и даже написание статей.

Хотя недавний прогресс в области ИИ позволил нам автоматизировать обработку данных, нам все еще нужна вычислительная платформа, которая может обрабатывать информацию как человеческий мозг, чтобы реализовать настоящий машинный интеллект. Человеческий мозг состоит из миллиардов нейронов, связанных триллионами синапсов в трехмерном пространстве. Передача информации между нейронами сопровождается очень низкими энергозатратами. В случае нейронной сети ее взаимосвязанный характер позволяет ей обрабатывать несколько информационных потоков параллельно. Это позволяет выполнять вычисления намного быстрее, потребляя при этом очень мало энергии. Поэтому для создания интеллектуально мощных и в то же время энер-

гоэффективных машинных ресурсов желательно иметь вычислительную модель, имитирующую мозг.

С вычислительной точки зрения алгоритм глубокого обучения имитирует обработку информации человеческим мозгом. Используя огромный объем доступных данных, такой алгоритм обучается, извлекая высокоуровневые признаки из исходных данных с помощью многоуровневой сети, которая имитирует передачу сигналов нейронного возбуждения. Каждый слой учится извлекать все более абстрактные и репрезентативные признаки из предыдущего слоя и отправляет эти признаки в следующий слой для более точного извлечения. В конце концов, исходные данные распознаются на уровне выходного слоя.

Ранние аппаратные реализации *нейроморфных вычислений* использовали схемы на полевых транзисторах для имитации активности нейронов и синапсов [3]. Это привело к появлению «нейроморфных» кремниевых нейронов (silicon neurons, SiN), основанных на технологии сверхкрупной интеграции (very large scale integration, VLSI). Впоследствии было предложено множество типов SiN для имитации реальных нейронов с различной степенью сложности [4]. В этой статье представлен широкий спектр схем, обычно используемых для проектирования SiN. Например, синаптическая передача может быть смоделирована математически с помощью простого дифференциального уравнения первого порядка. Для реализации такой функции может использоваться схема интегрирования во времени. Однако реализовать систему, которая обеспечивает обширные соединения на основе этой схемы, сложно, потому что она громоздкая.

Для повышения степени соответствия нейроморфных алгоритмов природе мозга и достижения более высокой вычислительной эффективности были предложены цифровые реализации модели мозга на программируемых вентильных матрицах (programmable gate arrays, FPGA) или ASIC. Для улучшения обработки данных в нейроморфных вычислительных системах применяют оптимизированные интерфейсы памяти, такие как реконфигурируемые интерфейсы памяти и память с высокой пропускной способностью.

В последнее время большое внимание исследователей привлекает энергонезависимая память, поскольку она обеспечивает альтернативные способы хранения данных [5]. Одним из примеров такого устройства является *мемристор*, который записывает исторический профиль возбуждений – это уникальная способность. Мемристор, электронный эквивалент биологической памяти, попросту говоря, может открыть новое измерение в обработке данных в применении к нейроморфным вычислениям. Например, он позволяет записывать поиски исторических данных и более эффективно применять их в последующих поис-

ках, помогая нейронным сетям глубокого обучения выполнять задачи более «осмысленно».

Нейроморфные вычисления имитируют обработку информации на биологическом уровне и функционируют на компактной и энергоэффективной платформе. Эта технология, вдохновленная мозгом, является многообещающим претендентом на достижение уровня настоящего машинного интеллекта, и она будет продолжать развиваться по мере роста рынка искусственного интеллекта.

17.3. Квантовые вычисления

Квантовые вычисления – это новая вычислительная парадигма. В отличие от привычных бинарных операций в обычном компьютере, квантовый компьютер обрабатывает информацию на основе явлений квантовой механики. Для представления информации в квантовой механике используются *квантовые биты*, называемые *кубитами*, которые являются фундаментальной вычислительной единицей в квантовом компьютере. По сравнению с дискретной природой двоичных битов, кубиты могут принимать когерентные состояния от 0 до 1 на непрерывной сфере – так называемой *сфере Блоха*. Используя преимущества концепции суперпозиции, кубиты могут одновременно представлять информацию о 2^n (где n – количество кубитов) состояний. Поскольку кубит по сути представляет собой полиномиальную суперпозицию нескольких состояний одновременно, квантовый компьютер может обрабатывать все 2^n состояний параллельно, что приводит к значительному увеличению скорости вычислений.

За последнее десятилетие было разработано множество квантовых алгоритмов для демонстрации возможностей как квантовых, так и традиционных вычислительных платформ различных уровней сложности [6]. Например, квантовое преобразование Фурье занимает $O(n \log n)$ шагов, а по классическому алгоритму – $O(n2^n)$. Из этого можно сделать вывод, что вычислительная нагрузка снижается полиномиально.

Было проведено множество исследований по аппаратной реализации кубитов. Первоначально в качестве возможных способов реализации кубитов рассматривали ядерный магнитный резонанс (ЯМР) в жидкой среде и ионные ловушки [7, 8]. Однако в последнее время акцент сместился на кубиты на основе полупроводников. Одним из примеров является кремниевый кубит, в котором используется одна или несколько квантовых точек, емкостно связанных между двумя электродами кремниевого транзистора с двойным затвором. На основе разных уровней емкостной связи между этими квантовыми точками можно моделировать различные квантовые состояния. На рис. 17.1 показан пример системы с двойными квантовыми точками. Особо следует отметить тот

факт, что кремниевые кубиты имеют более длительное время спиновой когерентности, а также унаследовали все преимущества кремниевой электроники. С момента появления первого кремниевого кубита [9, 10] многие исследовательские лаборатории и предприятия сотрудничают в создании масштабируемых квантовых устройств. Например, в Канаде недавно объявили, что работают над запуском программы квантовых вычислений с IBM Q-HUB [11].

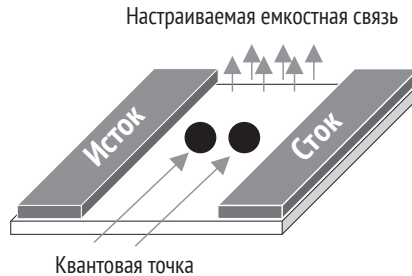


Рис. 17.1. Система с двойными квантовыми точками

Для управления кубитами и считывания их состояния необходима периферийная система. Функции контроллера могут включать квантовое усиление сигнала, преобразование с понижением частоты, оцифровку и многое другое. На рис. 17.2 показана блок-схема такой системы [12]. Для считывания кубитов применяются оптические каналы, а для облегчения многопоточковой передачи квантового сигнала между кубитами и контроллером предусмотрены мультиплексирование и демultipлексирование.

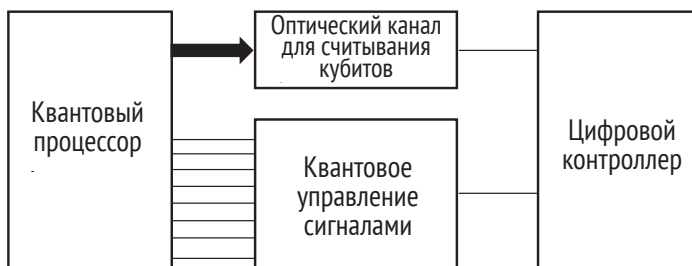


Рис. 17.2. Блок-схема периферийной системы управления/считывания кубитов [12]

Схема контроллера обычно работает при комнатной температуре, в то время как система квантовых вычислений нуждается в охлаждении до криотемпературы. Большой температурный градиент между ними может вызвать тепловой шум, который спровоцирует квантовые ошибки. Поэтому для исправления ошибок используются избыточные

кубиты, но это увеличивает задержку и снижает эффективность квантовых вычислений. Чтобы решить эту проблему, многие исследователи предложили системы контроллеров на основе технологии CMOS, работающих при глубоких криогенных температурах (менее 10 кельвинов), то есть схемы криоCMOS. Тепловой шум можно значительно снизить, если и квантовый процессор, и схема контроллера охлаждены до криотемператур. Это также может снизить сложность соединений между ними. В статьях [12, 13] было представлено несколько криоCMOS-компонентов, включая считывание кубитов, пассивные циркуляторы и малошумящие усилители (МШУ). Затем из этих компонентов, в качестве примера описанной выше системы контроллера, по 160-нм технологии CMOS была изготовлена микросхема.

КриоCMOS позволяет квантовому контроллеру и процессору работать при одинаковой температуре и в конечном итоге потенциально полностью интегрировать их в один и тот же чип. Хотя мы пока не располагаем глубоким пониманием того, как CMOS-схемы ведут себя при криогенных температурах, в области криофизики, новых моделей и моделирования на уровне системы уже проведены значительные исследовательские работы. Мы ожидаем, что благодаря прогрессу в области моделей криоCMOS-схем и функционирования квантовых систем кремниевые квантовые компьютеры с высокой производительностью, низким энергопотреблением и масштабируемой производительностью станут реальностью в ближайшем будущем.

17.4. Новые вычислительные архитектуры

Хотя конец закона Мура может стимулировать рост новых вычислительных парадигм, по-прежнему необходимо улучшать существующие платформы цифровых вычислений. В то время как обычная параллелизация может увеличить скорость вычислений примерно пропорционально количеству потоков или ядер, она не способна оптимизировать распределение задач, а это означает, что отдельным процессорам могут не достаться задачи, для которых они были разработаны. Появление графических процессоров сделало возможным обработку специализированных задач целевым оборудованием, а для выполнения определенных научных задач были разработаны другие специализированные процессоры/ускорители. Например, в [1] сообщается о специализированном суперкомпьютере для моделирования молекулярной динамики, который в 180 раз превышает производительность обычного высокопроизводительного компьютера. Специализация обеспечивает значительное повышение производительности вычислений за счет более эффективной оптимизации использования оборудования.

В краткосрочной перспективе специализированное оборудование может продолжать обеспечивать желаемое повышение производительности вычислений.

На одной гетерогенной вычислительной платформе можно разместить несколько специализированных типов процессоров/ускорителей. Например, центральный и графический процессоры могут быть интегрированы в одну и ту же схему, чтобы объединить преимущества того и другого. Центральный процессор фокусируется на задачах, связанных с операционной системой, тогда как графический процессор обрабатывает задачи, связанные с трехмерной графикой, и выполняет интенсивные научные вычисления. В последнее время встречается множество сообщений о разработке специализированных процессоров/ускорителей, включая криптографические сопроцессоры, блоки обработки тензорных данных и другие ускорители глубокого обучения. Интеграция центрального процессора и специализированного оборудования позволяет гораздо более эффективно решать специфические задачи.

С экономической точки зрения повышенные затраты и длительное время выполнения заказов связаны со спецификой внедрения передовых технологий. В то время как у крупных компаний есть финансовые ресурсы для приобретения специализированного оборудования высокого класса, у малых и средних компаний их нет, и поэтому они не могут в полной мере воспользоваться плодами прогресса. В стремлении полностью реализовать концепцию гетерогенных вычислений была предложена идея гибкого использования аппаратного обеспечения, в частности *чиплеты* [1]. В отличие от одиночного кремниевого чипа, объединяющего все специализированные процессоры/ускорители и центральный процессор для удовлетворения конкретных потребностей (стоимость обычно высока из-за конструктивной сложности), в концепции чиплета система разбита на функциональные блоки, каждый из которых выполняет минимально возможную задачу. В результате это позволяет производителям собирать вместе только те блоки, которые необходимы для конкретного применения, с гораздо меньшими затратами и более быстрым вводом в производство.

Гетерогенные вычислительные платформы нуждаются в новом программном обеспечении. Современные компьютерные архитектуры разделяют блоки хранения и обработки данных, что приводит к значительному снижению скорости вычислений при перемещении данных между ними. Эта проблема известна как «узкое место фон Неймана» или «стена памяти». Чтобы решить данную проблему, новое программное обеспечение должно минимизировать операции перемещения данных в дополнение к выполнению арифметических операций более высокого порядка.

При разработке нового программного обеспечения ключевым вопросом является перепроектирование существующих алгоритмов и программных сред для практической реализации преимуществ, предлагаемых специализированным оборудованием. Помимо оптимизации алгоритмов, еще одной проблемой являются затраты, связанные с перемещением данных. В публикации [1] указано, что общую сложность вычислений не следует рассчитывать только на основе количества операций с плавающей запятой; скорее, следует также учитывать сложность принудительного перемещения данных, в которых нуждается алгоритм. Следовательно, в новом программном обеспечении желательно использовать оптимизированную топологию доступа к памяти/данным, например неоднородный доступ к памяти.

Что касается оборудования, необходима полная переработка аппаратного обеспечения для решения проблемы узких мест в памяти. Например, желательно сокращение количества металлических межсоединений или их замена, поскольку они ограничивают энергоэффективность. Многообещающим вариантом решения этой проблемы является кремниевая фотоника, где могут применяться фотонные межсоединения и волноводы с низкими потерями для уменьшения расхода энергии на перемещение данных. Одним из таких примеров использования интегрированных фотонных технологий является проект ARPAe ENLITENED [14], в рамках которого были разработаны новые сетевые топологии на основе фотоники для передачи информации в среде центра обработки данных.

17.5. Источники

- [1] J. Shalf, *The future of computing beyond Moore's law*, Philosophical Transactions of the Royal Society A, vol. 378, no. 2166, p. 20190061, 2020.
- [2] Wikipedia, *Moore's law*, 2020, accessed Sept. 29, 2020. https://en.wikipedia.org/wiki/Moore's_law.
- [3] C. Mead, *Analog VLSI and neural systems*. Addison-Wesley Longman Publishing Co., 1989.
- [4] G. Indiveri, B. Linares-Barranco, T. J. Hamilton, A. Van Schaik, R. Etienne-Cummings, T. Delbruck, S.-C. Liu, P. Dudek, P. Häfliger, S. Renaud et al., *Neuromorphic silicon neuron circuits*, Frontiers in Neuroscience, vol. 5, p. 73, 2011.
- [5] Y. Chen, H. H. Li, C. Wu, C. Song, S. Li, C. Min, H.-P. Cheng, W. Wen, and X. Liu, *Neuromorphic computing's yesterday, today, and tomorrow – an evolutionary view*, Integration, vol. 61, pp. 49–61, 2018.

- [6] A. Steane, *Quantum computing*, Reports on Progress in Physics, vol. 61, no. 2, p. 117, 1998.
- [7] L. M. Vandersypen and I. L. Chuang, *NMR techniques for quantum control and computation*, Reviews of Modern Physics, vol. 76, no. 4, p. 1037, 2005.
- [8] P. Schindler, D. Nigg, T. Monz, J. T. Barreiro, E. Martinez, S. X. Wang, S. Quint, M. F. Brandl, V. Nebendahl, C. F. Roos et al., *A quantum information processor with trapped ions*, New Journal of Physics, vol. 15, no. 12, p. 123012, 2013.
- [9] R. Maurand, X. Jehl, D. Kotekar-Patil, A. Corna, H. Bohuslavskiy, R. Lavit'eville, L. Hutin, S. Barraud, M. Vinet, M. Sanquer et al., *A CMOS silicon spin qubit*, Nature Communications, vol. 7, p. 13575, 2016.
- [10] L. Hutin, R. Maurand, D. Kotekar-Patil, A. Corna, H. Bohuslavskiy, X. Jehl, S. Barraud, S. De Franceschi, M. Sanquer, and M. Vinet, *Si CMOS platform for quantum information processing*, in Proc. 2016 IEEE Symposium on VLSI Technology. IEEE, 2016, pp. 1–2.
- [11] CMC, *CMC becomes member of IBM Q-HUB at Université de Sherbrooke*, 2020, accessed Sept. 29, 2020. <https://www.cmc.ca/cmc-becomesmember-of-ibm-q-hub-at-universite-de-sherbrooke/>.
- [12] E. Charbon, *Cryo-CMOS electronics for quantum computing applications*, in Proc. ESSDERC 2019–49th European Solid-State Device Research Conference (ESSDERC). IEEE, 2019, pp. 1–6.
- [13] S. Schaal, A. Rossi, V. N. Ciriano-Tejel, T.-Y. Yang, S. Barraud, J. J. Morton, and M. F. Gonzalez-Zalba, *A CMOS dynamic random access architecture for radio-frequency readout of quantum devices*, Nature Electronics, vol. 2, no. 6, pp. 236–242, 2019.
- [14] ARPA-E, *ENLITENED*, 2020, accessed Sept. 29, 2020. <https://arpa-e.energy.gov/?q=arpa-e-programs/enlitened>.

Глава 18

Новые устройства

Беспроводные технологии оказали огромное влияние на общество. Сегодня большинство людей считают мобильные устройства неотъемлемой частью повседневной жизни. Смартфоны, которые полностью заменили своего предшественника, скромный функциональный телефон, больше не просто вещь, которую приятно иметь; скорее, это жизненная необходимость. Революция в области смартфонов началась с концепции использования беспроводных компьютеров для доступа в интернет через услуги мобильного широкополосного доступа. На сегодняшний день, когда мобильные приложения и сервисная модель ОТТ распространились почти повсеместно, мы стали свидетелями взрывного роста использования мобильных приложений. Сейчас нам доступен обширный набор приложений практически для всего: календарь, камера, калькулятор, навигатор, карта, кошелек, новости и общественный транспорт – это лишь некоторые из них. Другие приложения могут рекомендовать рестораны, торговые центры, парковки и онлайн-классы. В этой главе мы обсудим тенденции развития мобильных устройств и интерфейсов между людьми и устройствами.

18.1. Мобильные устройства будущего

Предполагается, что следующая революция мобильных устройств – точно так же, как смартфоны заменили обычные телефоны, – произойдет в течение срока службы 6G. Ожидается, что к 2030 году будущие устройства получат новые возможности, обеспечиваемые системой связи 6G, – сканирование и построение образов окружающей обстановки, тактильную связь, голографический дисплей, искусственный интеллект и другие функции.

Эти инновационные возможности радикально меняют не только конструкцию, но и роль мобильных устройств в нашей жизни. По мере дальнейшего развития мобильные устройства получают следующие способности:

- восприятие на уровне человека (например, неограниченная полоса пропускания для визуального/звукового восприятия на уровне человека и реалистичного общения между людьми [1]);
- сканирование окружающей среды (например, возможность получения мультиспектральных изображений с близкого расстояния [2] и высокоточного позиционирования);
- взаимодействие человека и киберпространства (например, голографические носимые дисплеи для взаимодействия человека с кибермиром [3]);
- сбор энергии (например, беспроводная зарядка и одновременная беспроводная передача информации и энергии).

Новые возможности превратят нынешний мир «интеллектуальных помощников» в мир «киберфизического слияния», как показано на рис. 18.1.

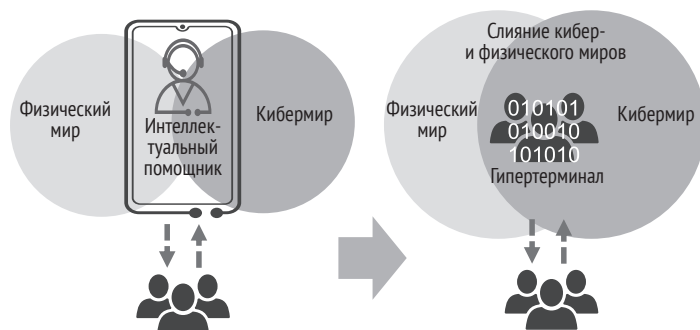


Рис. 18.1. Эволюция от интеллектуального помощника к «цифровому Я», которое является виртуальным представлением конечного пользователя или устройства

Современные смартфоны действуют как шлюзы, соединяющие физический мир с кибермиром, в основном обеспечивая доступ в интернет. Мы ожидаем, что в будущем мобильные беспроводные устройства превратятся в носимые гипертерминалы в мире киберфизического слияния – так называемое «цифровое Я». Эти устройства будут превосходить человеческие возможности по ряду направлений, таких как интеллектуальное распознавание и восприятие окружающей среды.

Учитывая вышеперечисленные возможности, можно выделить четыре основные тенденции разработки мобильных устройств, как показано на рис. 18.2.



Рис. 18.2. Четыре новые возможности позволяют реализовать четыре направления развития устройств будущего

Итак, благодаря технологическим достижениям новые устройства будут:

умнее – смартфоны станут не только мощнее, но и получают опцию дополненной реальности для автоматизации окружающей жизни.

Современные смартфоны не просто в миллионы раз мощнее компьютеров Национального управления по аэронавтике и исследованию космического пространства (НАСА), которые использовались во время миссии «Аполлон-11» [4], – они становятся все умнее по мере увеличения вычислительных возможностей. Вычислительная мощность до недавнего времени росла в соответствии с законом Мура, что способствовало развитию искусственного интеллекта и машинного обучения в последние годы. Кроме того, успехи, достигнутые в производстве полупроводников, открывают путь к более высокой вычислительной производительности, лучшей энергоэффективности, меньшему размеру микросхем и более высокой плотности транзисторов.

Чем мощнее становятся смартфоны, тем лучше они могут реализовать возможности искусственного интеллекта. Сегодня все большее количество смартфонов оснащается специализированными встроенными модулями нейронной обработки для вычислений с искусственным интеллектом. Алгоритмы машинного обучения могут использоваться для выполнения множества вычислительно-ресурсоемких задач, таких как дополненная реальность, распознавание лиц и распознавание голо-

са и речи. Для разработки инновационных приложений созданы многочисленные API-интерфейсы ИИ, такие как HiAI от Huawei, Caffe2Go от Facebook, TensorFlow Lite от Google и CoreML от Apple.

В то же время ресурсоемкие задачи могут быть перенесены с мобильных устройств в граничные облака; этот подход известен как парадигма *граничных вычислений* (edge computing). Граничные вычисления нуждаются в сверхвысокой скорости передачи данных, сверхнизкой задержке и сверхвысокой надежности, которые обеспечат беспроводные сети следующего поколения. В будущем устройства станут более интеллектуальными, а также обеспечат защиту конфиденциальности за счет использования распределенных вычислений и обучения благодаря граничным вычислениям, облачным вычислениям, а также мощным локальным процессорам, графическим процессорам и специализированному оборудованию для ускорения ИИ.

Достижения технологий связи малого радиуса действия и алгоритмов искусственного интеллекта позволили группам дронов, транспортных средств и роботов локально взаимодействовать друг с другом и с окружающей средой. Подобные группы устройств могут достигать своих целей, используя децентрализованное и самоорганизованное управление (например, интеллект роя и групповую робототехнику). По мере развития технологий ближней связи 6G и алгоритмов искусственного интеллекта / машинного обучения будет расти интеллектуальная составляющая устройств, а качество обслуживания и производительность будут улучшаться за счет более обширной автоматизации всех аспектов нашей жизни;

универсальнее – не только подключенность, но и новые возможности обнаружения, открывающие перспективы для новых применений.

В будущем поддержка беспроводной связи станет фундаментальной функцией любого устройства, поскольку она будет необходима для каждого человека, дома, организации и отрасли. Как рассказывалось в главе 2 о вариантах чрезвычайно захватывающего взаимодействия с человеком, сеть 6G может достигать пропускной способности в диапазоне Тбит/с, обеспечивая при этом задержку передачи менее 1 мс. Это поддерживает плавный обмен мультисенсорной информацией VR, включая видео с высокой реалистичностью, аудио и даже тактильные ощущения. Благодаря мультисенсорным возможностям устройства в будущем могут быть интегрированы с человеческим телом, например для устранения последствий инвалидности. Взаимное объединение с человеческим телом может способствовать развитию человечества, формируя кибернетические организмы.

Помимо связи, новые возможности сканирования будут поддерживаться объединенными технологиями связи и сканирования (рис. 18.3). Благодаря этому мобильные устройства будущего смогут поддерживать множество новых функций. Например, радиосвязь в ТГц-диапазоне может использоваться для визуализации и спектроскопии, предоставляя нам персональные «рентгеновские снимки» на молекулярном уровне, как упоминалось в главе 3. Кроме того, мониторинг состояния здоровья (сердцебиение, падения и т. д.), нарушение границ перемещения и распознавание жестов посредством бесконтактного взаимодействия с устройствами также могут быть реализованы с помощью беспроводного сканирования и машинного обучения.

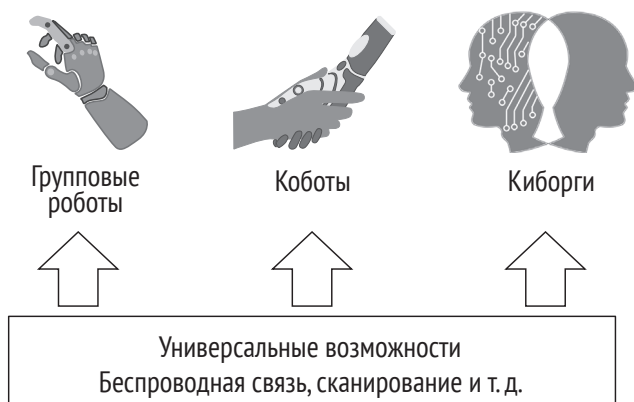


Рис. 18.3. Разнообразные возможности будущих устройств эпохи 6G

Более того, взаимодействие между людьми и машинами, а также между физическим миром и кибермиром станет еще богаче. Например, системы тактильных датчиков и исполнительных механизмов будут включать в себя виртуальную реальность, чтобы предоставить пользователям больше подсказок для ментального построения виртуального мира за пределами пространственных и временных ограничений, а оборудование, управляемое удаленно с помощью мозговых волн, значительно расширит сферу личной деятельности. Поддержка естественных человеческих связей (human bond communication, HBC) направлена на улучшение коммуникативного опыта людей за счет включения пяти сенсорных функций. Это позволит обмениваться более выразительной и целостной сенсорной информацией за счет использования высоконадежных, оперативных и интеллектуальных средств связи.

Для воплощения в жизнь этих возможностей необходимо значительно продлить срок службы батареи. Благодаря массовому использованию новых материалов и архитектуры батарей (например, графен, литий-фосфат железа и полимерные батареи) людям со встроенными

устройствами не нужно будет ежедневно заряжать батареи с помощью проводных зарядных устройств. Вместо этого будут доступны различные технологии сбора энергии, включая беспроводную передачу энергии, которая может эффективно и безопасно заряжать мобильные устройства на больших расстояниях (например, через комнату);

разнообразнее – будут подключены не только смартфоны, но и различные устройства, выполняющие функции датчиков и исполнительных механизмов.

Сегодня смартфоны являются доминирующим мобильным устройством среди индивидуальных потребителей услуг мобильной связи. В глобальном масштабе доля людей, использующих смартфоны, по прогнозам, превысит 80 % к 2025 году [5]. В то же время технологии беспроводных устройств продолжают развиваться, поскольку беспроводные сети стремятся предложить более высокую скорость, меньшую задержку и более высокую надежность. Кроме того, происходит смещение парадигмы от преимущественного обслуживания индивидуальных клиентов к обслуживанию как индивидуальных, так и разнообразных вертикальных/отраслевых клиентов.

Умные устройства, такие как смартфоны, планшеты, ноутбуки и даже умные автомобили, становятся неотъемлемой частью нашей жизни благодаря достижениям в области полупроводниковых технологий и программных платформ. В будущем появится широкий спектр ориентированных на человека и промышленных устройств, объединяющих передовые датчики, новые технологии визуализации и искусственный интеллект (рис. 18.4).

- **Устройства, ориентированные на человека:** мы прогнозируем развитие носимых устройств (гарнитуры VR, очки AR, умные часы и роботы-экзоскелеты), имплантируемых медицинских устройств и т. д.
- **Промышленные устройства:** к сетям 6G будут подключены автомобили, роботы/коботы и интеллектуальное производственное оборудование; повсюду будет разбросано большое количество недорогих, маломощных (или даже безбатарейных, называемых «пассивными») устройств IoT, обслуживающих умный город и дом, умное здравоохранение, мониторинг загрязнений, отслеживание перемещений и другие приложения.

Эти устройства будут служить датчиками и исполнительными механизмами киберфизической системы, объединяя вычисления, сетевые и физические процессы.

В связи со стремительным ростом числа разнообразных устройств требования к взаимосвязанности будут значительно выше. Например,

смартфоны будут автоматически подключаться к телевизорам для улучшенного просмотра видео, а носимые устройства будут подключаться к смартфонам для передачи собранной информации. Но какое устройство будет служить в будущем в качестве *якорного устройства*? Такое якорное устройство должно обладать надежной беспроводной связью, длительным временем автономной работы и вычислительными ресурсами; хорошие кандидаты – смартфоны, гарнитуры VR, роботы-экзоскелеты и умные автомобили. Благодаря сети 6G якорные устройства и подключенные к ним устройства обеспечат бесперебойную и согласованную работу в интересах потребителей;

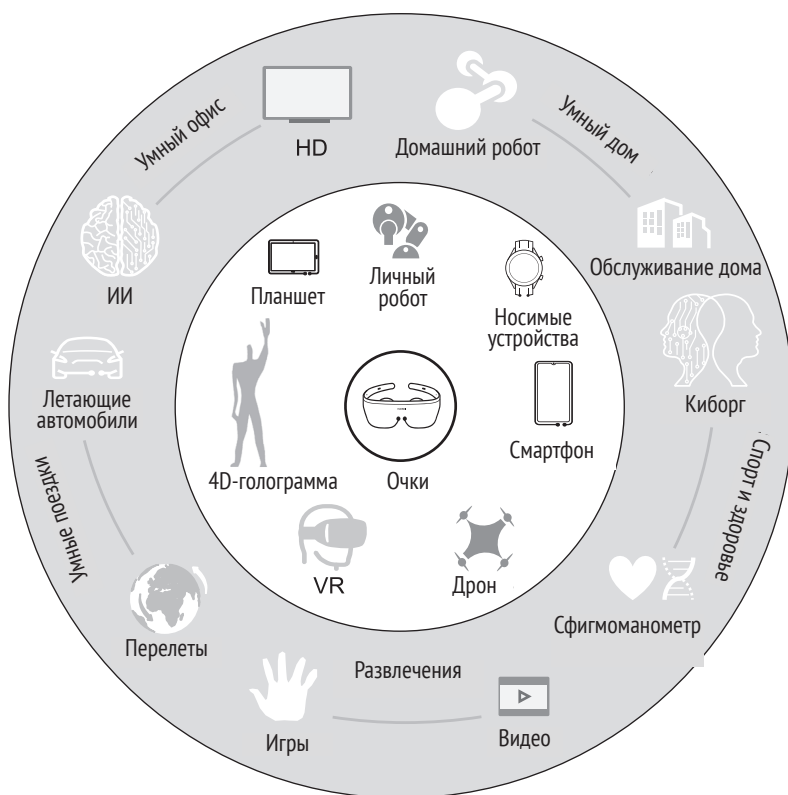


Рис. 18.4. Пример ориентированных на человека приложений и разнообразных устройств

распределенное – потребителям будут служить не только физические, но и виртуальные устройства, обеспечивающие защиту конфиденциальности и новые бизнес-модели.

По мере развития технологии облачных вычислений и IoE и благодаря возможности подключения 6G все больше физических устройств 6G будут иметь виртуального двойника в облаке (рис. 18.5). Благодаря децент-

рации персональные данные пользователей могут храниться на виртуальных устройствах под контролем пользователей, а не на мобильных устройствах или в принадлежащих третьей стороне централизованных хранилищах, что избавляет пользователей от опасений по поводу потенциальных проблем с конфиденциальностью. На основе концепции виртуализированных устройств появятся общие устройства 6G, которые обычно находятся в общественных местах и используются по запросу. Благодаря интеграции биометрии, искусственного интеллекта, персонализированной автоматической настройки и технологий защиты конфиденциальности устройства 6G смогут предоставлять простые в использовании функции, такие как биометрическая аутентификация и гиперперсонализированные конфигурации. Это позволит конечным пользователям получать доступ к желаемым услугам в любое время и в любом месте через общие устройства. Перечень типов общих устройств 6G будет весьма обширным, включая арендуемые автомобили, конференц-залы, облачные устройства и любые другие устройства с возможностями ввода/вывода и вычислительными возможностями, которые можно использовать совместно или арендовать в общественных местах.



Рис. 18.5. Концепция виртуального устройства

Кроме того, некоторые устройства 6G будут поддерживать облачные вычисления. Клиенты смогут приобретать устройства 6G с поддержкой частных облачных вычислений – такие устройства помогут клиентам удовлетворить более высокие требования к вычислительным хранилищам для более интеллектуальной голосовой поддержки, хранения личных изображений и видео и т. д. Эти устройства могут взаимодействовать с общедоступными облачными сервисами через соединения 6G для поддержки расширенной функциональности. Кроме того, благодаря наличию виртуальных двойников физические устройства смогут переносить ресурсоемкие вычислительные задачи в облако, тем самым продлевая срок службы батареи и избегая проблем с перегревом.

18.2. Интерфейс мозга и устройств будущего

Концепция интерфейса мозг–компьютер (brain–computer interface, BCI) восходит к 1970-м годам, когда Жак Видаль попытался создать систе-

матический подход к коммуникации между человеком и компьютером [6, 7]. Первоначально исследователи стремились использовать ВСІ для диагностики неврологических расстройств и заболеваний головного мозга (например, для контроля эпилептических припадков). Система записывала и обрабатывала сигналы электроэнцефалографии (ЭЭГ) кожи головы и извлекала из данных интересные особенности для диагностики эпилепсии. В 2004 году в статье [8] была продемонстрирована система ВСІ, которая может контролировать многомерные движения с использованием неинвазивного подхода. Этот прорыв вызвал огромный интерес к исследованиям ВСІ в научных и промышленных кругах. Как показано на рис. 18.6, по крайней мере типичная система ВСІ состоит из следующих компонентов: обработка сигналов, выделение признаков, классификация или распознавание образов для преобразования данных в машинные команды и интерфейсы приложений (например, компьютеры или роботизированные манипуляторы). Интерфейсы приложений, наконец, будут передавать сенсорную информацию обратно пользователю, образуя двустороннюю связь.



Рис. 18.6. Схема системы ВСІ

В последнее десятилетие технология ВСІ эволюционировала настолько, чтобы воплотить в жизнь достижения, которые когда-то считались научной фантастикой. Например, в [9] было показано, что нейропротез может способствовать восстановлению двигательной активности после травмы позвоночника. После вживления электродов в мозг парализованный пациент может управлять роботизированным протезом с помощью своего разума. Более того, телекинез (перемещение предметов силой мысли) когда-то считался сверхъестественным явлением, но теперь в некотором смысле он возможен. Например, в статье [10] показано, что мы можем управлять полетом дрона с помощью мозговых волн. Совсем недавно в [11] было продемонстрировано, что с помощью

нейрочипов, имплантированных в мозг животного, можно точно предсказать движения его суставов. Это открывает целый ряд возможностей для лечения слепоты, паралича, психических заболеваний и т. д.

Для регистрации активности мозга использовались как инвазивные, так и неинвазивные методы. Во втором случае используются электроды, помещаемые на кожу головы, в то время как в первом используются интракортикальные и стереоэлектродэнцефалографические (stereoelectroencephalography, SEEG) электроды, расположенные над корой головного мозга. Инвазивный метод, по сравнению с неинвазивным, обеспечивает сигналы с более высоким пространственно-временным разрешением и более низким уровнем шума. В этом контексте в статье [11] показано, что нейрочип размером с монету, содержащий тысячи электродов, может быть имплантирован в мозг животного для сбора внутрикортикальных сигналов ЭЭГ. Как бы многообещающе это ни звучало, пока существуют достаточно серьезные препятствия для широкого использования этой технологии среди людей.

Кроме того, развитию технологии BCI также способствовали достижения в области глубокого обучения. Обычно локализация источника ЭЭГ (обратная задача) решалась путем итеративного вычисления прямой задачи ЭЭГ до тех пор, пока моделированные сигналы ЭЭГ не совпадали с измерениями в пределах желаемых уровней допуска. Глубокое обучение более эффективно решает обратную задачу за счет использования большого количества доступных данных [12, 13].

Помимо пассивного считывания сигналов мозга, двунаправленная связь BCI может обеспечивать сенсорную обратную связь от контролируемого объекта (например, протеза конечности). Исследования в этом направлении показали, что соматосенсорная обратная связь может быть обеспечена посредством электрической стимуляции периферических и/или корковых нервов пациентов [14, 15]. Кортикальная стимуляция включает стимуляцию определенных соматосенсорных областей коры головного мозга, которые связаны с определенными сенсорными функциями и частями тела. Эти области обычно локализованы, и для точной стимуляции требуются электроды микроскопических размеров. Кроме того, обратная связь в реальном времени необходима для моделирования естественной биомеханики, опосредованной центральной нервной системой. В данном контексте критически значимой становится скорость соединения (т. е. пропускная способность канала для обмена данными между мозгом и машиной). Авторы [16] сделали еще один шаг вперед и показали, что сверхплотный чип с микроэлектродной матрицей может записывать корковые сигналы в головном мозге грызунов. Размер упакованного нейрочипа составляет менее $23 \times 18,5 \times 2$ мм, и сообщается, что он способен поддерживать многоканальную стиму-

ляцию с целью нейронной модуляции, хотя это еще предстоит продемонстрировать. Ожидается, что имплантируемые нейрочипы с новыми нейронными сенсорами будут способствовать дальнейшему развитию этого направления исследований [17].

Периферийная стимуляция связана с концепцией киборга, которая в основном известна из научно-фантастических фильмов, где организм человека улучшают, добавляя в него роботизированные компоненты. Последние достижения в бионике показали, что эра киборгов имеет все шансы стать реальностью. В публикациях [18, 19] в качестве перспективы будущего продемонстрировали мионевральный интерфейс агонист–антагонист (agonist–antagonist myoneural interface, AMI), который позволяет передавать команды мозга конечной мышце для динамического контроля протеза и обеспечивает проприоцептивную сенсорную обратную связь с мозгом. Согласно этой методике, хирурги сначала анализируют прямые и обратные сигналы, связанные с группой мышц в остаточной конечности пациента во время ампутации. Когда пациент совершает движение, электромиографические сигналы целевой группы мышц записываются и впоследствии преобразуются в машинные команды для управления сервоприводами протезируемой конечности. И соответственно, наоборот, когда протезный сустав движется, генерируются соответствующие сенсорные сигналы, которые затем отправляются в мозг через нервы, позволяя пациенту почувствовать движение. Интерфейс AMI был продемонстрирован в выступлении на TED [20], где протезированная конечность помогала пациенту бегать, лазать и танцевать. AMI представляет собой многообещающую основу, в рамках которой мы можем объединить робототехнику и биологических людей. Похоже, что эпоха людей-киборгов не за горами.

Мы ожидаем, что в ближайшие несколько десятилетий нейронные протезы, управляемые мозгом, станут реальностью и улучшат качество жизни парализованных пациентов и пациентов с ограниченными возможностями. Управляя роботизированными протезами с помощью человеческого разума, пациенты смогут ходить, перемещать предметы и выполнять другие действия, связанные с конечностями.

18.3. Новые носимые устройства

Носимые устройства за последние несколько лет испытали взрывной рост популярности благодаря их новым функциональным возможностям. Носимые электронные датчики могут снимать различные физиологические параметры организма, например частоту сердечных сокращений, электрокардиограмму, температуру кожи, движения тела и артериальное давление. Эти устройства имеют множество применений, в том числе в спорте, военном оборудовании, здравоохранении и по-

вседневной жизни. С появлением 6G на сцене появятся новые носимые устройства, реализующие новые варианты применения.

В статье [21] показано, что растяжимая тактильная перчатка с плетеным массивом датчиков способна определять некоторые формы объектов, оценивать их вес и формировать тактильные паттерны при захвате предметов. В ходе исследования взаимодействие руки с различными объектами регистрировалось с помощью 548 датчиков, установленных на тактильной перчатке. Затем глубокую сверточную нейросеть обучили извлекать тактильные сигнатуры человеческого восприятия. Эти сигнатуры могут быть весьма полезны при разработке роботов и протезов.

В [22] было продемонстрировано широкополосное и гибкое носимое устройство ТГц-визуализации на основе пленки из карбоновых нанотрубок при комнатной температуре. Терагерцовый сканер на основе нанотрубок может выполнять пассивную визуализацию как плоских, так и изогнутых образцов. Эта технология облегчит использование удобных и портативных устройств ТГц-диапазона во многих сценариях, включая проверки безопасности и мониторинг состояния здоровья.

Для применений в мониторинге состояния здоровья были продемонстрированы носимые устройства на основе графена. В частности, в статье [23] показано, что графеновый фотодетектор, сенсibilизированный полупроводниковыми квантовыми точками (GQD-PD), может применяться при создании гибких и прозрачных носимых устройств. Уже сегодня на основе предложенного фотодетектора GQD-PD можно разработать ряд действующих прототипов. Например, был показан браслет с GQD-PD и встроенным источником света, измеряющий частоту сердечных сокращений как в режимах отражения, так и в режимах передачи. Авторы [24] продемонстрировали гибкий и прозрачный браслет, интегрированный с графеном и поли(3,4-этилендиокситиофен)-поли(4-стиролсульфонат) (PEDOT:PSS)/графеновым фотодетектором среднего инфракрасного диапазона. Браслет можно использовать для наблюдения за состоянием здоровья, отслеживая изменения молекулярного состава кожных выделений тела.

18.4. Источники

- [1] W. Wang, X. Deng, L. Ding, and L. Zhang, *Brain-inspired intelligence and visual perception*. Springer, 2019.
- [2] V. C. Coffey, *Multispectral imaging moves into the mainstream*, Optics and Photonics News, vol. 23, no. 4, pp. 18–24, 2012.
- [3] A. Maimone, A. Georgiou, and J. S. Kollin, *Holographic near-eye displays for virtual and augmented reality*, ACM Transactions on Graphics, vol. 36, no. 4, pp. 1–16, 2017.

- [4] *Your smartphone is millions of times more powerful than the Apollo 11 guidance computers*. <https://www.zmescience.com/science/news-science/smartphone-power-compared-to-apollo-432/>.
- [5] GSM. Association et al., *Global mobile trends 2020 new decade, new industry?* 2019.
- [6] J. J. Vidal, *Toward direct brain-computer communication*, Annual Review of Biophysics and Bioengineering, vol. 2, no. 1, pp. 157–180, 1973.
- [7] Wikipedia, *Brain computer interface*. [https://en.wikipedia.org/wiki/Brain %E2 %80 %93computer_interface](https://en.wikipedia.org/wiki/Brain%E2%80%93computer_interface).
- [8] J. R. Wolpaw and D. J. McFarland, *Control of a two-dimensional movement signal by a noninvasive brain–computer interface in humans*, Proceedings of the National Academy of Sciences, vol. 101, no. 51, pp. 17 849–17 854, 2004.
- [9] D. Borton, M. Bonizzato, J. Beauparlant, J. DiGiovanna, E. M. Moraud, N. Wenger, P. Musienko, I. R. Minev, S. P. Lacour, J. D. R. Millán et al., *Corticospinal neuroprostheses to restore locomotion after spinal cord injury*, Neuroscience Research, vol. 78, pp. 21–29, 2014.
- [10] CTV News, «“Pure thought”: Edmonton graduates using brain waves to fly drones». <https://www.ctvnews.ca/sci-tech/pure-thought-edmonton-graduates-using-brain-waves-to-fly-drones-1.4343715>.
- [11] A. Regalado, «Elon Musk’s Neuralink is neuroscience theater». <https://www.technologyreview.com/2020/08/30/1007786/elon-musks-neuralink-demo-update-neuroscience-theater>.
- [12] S. Cui, L. Duan, B. Gong, Y. Qiao, F. Xu, J. Chen, and C. Wang, «EEG source localization using spatio-temporal neural network», China Communications, vol. 16, no. 7, pp. 131–143, 2019.
- [13] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, «Deep convolutional neural network for inverse problems in imaging», IEEE Transactions on Image Processing, vol. 26, no. 9, pp. 4509–4522, 2017.
- [14] D. W. Tan, M. A. Schiefer, M. W. Keith, J. R. Anderson, and D. J. Tyler, «Stability and selectivity of a chronic, multi-contact cuff electrode for sensory stimulation in human amputees», Journal of Neural Engineering, vol. 12, no. 2, p. 026002, 2015.
- [15] G. A. Tabot, J. F. Dammann, J. A. Berg, F. V. Tenore, J. L. Boback, R. J. Vogelstein, and S. J. Bensmaia, «Restoring the sense of touch with a prosthetic hand through a brain interface», Proceedings of the National Academy of Sciences, vol. 110, no. 45, pp. 18 279–18 284, 2013.
- [16] E. Musk et al., «An integrated brain–machine interface platform with thousands of channels», Journal of Medical Internet Research, vol. 21, no. 10, p. e16194, 2019.

- [17] C. Li and W. Zhao, «Progress in the brain–computer interface: An interview with Bin He», *National Science Review*, vol. 7, no. 2, pp. 480–483, 2020.
- [18] T. R. Clites, M. J. Carty, J. B. Ullauri, M. E. Carney, L. M. Mooney, J.-F. Duval, S. S. Srinivasan, and H. M. Herr, «Proprioception from a neurally controlled lower-extremity prosthesis», *Science Translational Medicine*, vol. 10, no. 443, p. eaap8373, 2018.
- [19] S. Srinivasan, M. Carty, P. Calvaresi, T. Clites, B. Maimon, C. Taylor, A. Zorzos, and H. Herr, «On prosthetic control: A regenerative agonist–antagonist myoneural interface», *Science Robotics*, vol. 2, no. 6, 2017.
- [20] H. Herr, «New bionics let us run, climb and dance». <https://ed.ted.com/lessons/g8KC49mB>.
- [21] S. Sundaram, P. Kellnhofer, Y. Li, J.-Y. Zhu, A. Torralba, and W. Matusik, «Learning the signatures of the human grasp using a scalable tactile glove», *Nature*, vol. 569, no. 7758, pp. 698–702, 2019.
- [22] D. Suzuki, S. Oda, and Y. Kawano, «A flexible and wearable terahertz scanner», *Nature Photonics*, vol. 10, no. 12, pp. 809–813, 2016.
- [23] E. O. Polat, G. Mercier, I. Nikitskiy, E. Puma, T. Galan, S. Gupta, M. Montagut, J. J. Piqueras, M. Bouwens, T. Durduran et al., «Flexible graphene photodetectors for wearable fitness monitoring», *Science Advances*, vol. 5, no. 9, p. eaaw7846, 2019.
- [24] M. Zhang and J. T. Yeow, «A flexible, scalable, and self-powered mid-infrared detector based on transparent PEDOT : PSS/graphene composite», *Carbon*, vol. 156, pp. 339–345, 2020.

Резюме части IV

Приближающаяся эра 6G вызвала большой интерес к разработке аппаратных технологий, необходимых для подготовки инфраструктуры к предстоящему сдвигу парадигмы. Уже накоплен опыт практического применения миллиметрового диапазона, теперь пора приступить к освоению ТГц-диапазона – пока хотя бы нескольких частотных полос. Ожидается, что комбинация миллиметрового и терагерцового диапазонов обеспечит скорость передачи данных в диапазоне Тбит/с и доставку изображений с высоким разрешением. С появлением новых спектров, вариантов использования и аппаратных технологий (таких как интеллектуальная поверхность) возникла острая необходимость в разработке обновленной модели канала, которая может точно отражать распространение радиоволн. С точки зрения аппаратного обеспечения, гетерогенная платформа на материалах типа III–V может еще больше повысить производительность кремниевой системы за счет оптимизации каждого компонента на одном кристалле. Для достижения более высоких характеристик на кремниевый чип могут быть добавлены новые материалы, такие как фотонные кристаллы, фотоэлектрические элементы и плазмонные поверхности. Новые конструкции антенны на кристалле и в корпусе, наряду с технологией компактных линз, такой как RIS, могут обеспечить более точную реализацию желаемых характеристик антенны, а также уменьшить размер системы. ТГц-технология позволит использовать новые методы связи и визуализации, но реализация терагерцовых систем, основанных на электронике, оптоэлектронике и фотонике, будет зависеть от сценария использования и рабочих частот. Характеристики этих систем обсуждались в части IV. Было показано, что структура, использующая одновременно как совместный орбитальный угловой момент, так и массивную архитектуру MIMO (многообещающая технология передачи для 6G), улучшает производительность системы и обеспечивает мультипликативный выигрыш в эффективности использования спектра. Наконец, поскольку закон Мура скоро перестанет действовать, а кремниевая технология приближается к физическому пределу, мы рассчитываем на фундаментальный скачок вычислительной мощности, который станет следствием развития двух новых концепций – нейроморфных и квантовых вычислений.

Часть **V**

**Технологии-кандидаты
для применения
в радиоинтерфейсе 6G**

Введение в часть V

При проектировании сети радиодоступа 5G разработчики ориентировались на концепцию IoE, которая нашла воплощение в различных сценариях использования, таких как eMBB, URLLC и mMTC. Три типичных требования для этих сценариев: скорость передачи данных 20 Гбит/с при полосе пропускания 100 МГц, надежность 99,999 % при задержке радиоинтерфейса 1 мс и 1 млн устройств на квадратный километр с более широким покрытием соответственно.

Для реализации оптимизированного метода поддержки универсальных сценариев применения и широкого спектрального диапазона в 5G применили новый унифицированный радиоинтерфейс, обладающий как гибкостью, так и адаптируемостью. Этот радиоинтерфейс изначально предусматривает сегментирование сети радиодоступа (radio access network, RAN), что облегчает мультиплексирование услуг, повышение эффективности использования спектра, увеличение возможности подключения и уменьшение задержки. Чтобы смягчить влияние географических «сот», которые создают для устройств нежелательный эффект «перехода через границу», радиодоступ 5G рассматривает по отдельности специфические для пользовательского оборудования физические сигналы и идентификаторы (ID) «соты», предоставляя сетям больше возможностей для выбора оптимального «луча» или «лучей» для обслуживания пользователей, реализуя так называемую архитектуру без сот, ориентированную на пользователя (user-centric no-cell, UCNC), чтобы улучшить взаимодействие с пользователем на границе соты с помощью нативно поддерживаемых схем совместной передачи.

Что касается 6G, то, как подробно описано в главе 1, мы ожидаем наступления новой эры, в которой к одной сети будут подключены люди, вещи и интеллект с новыми функциями, такими как обнаружение и сетевой ИИ, в дополнение к существующим сценариям использования 5G, которые получают дальнейшее развитие. В этом контексте радиоинтерфейс 6G должен соответствовать как принципиально новым KPI, так и гораздо более высоким или более строгим версиям прежних KPI, по сравнению с 5G. Сетям 6G необходимо поддерживать еще больший диапазон спектра и более широкую полосу пропускания, чтобы обеспечивать чрезвычайно высокую скорость передачи данных и сканирование с высоким разрешением. Для достижения этих новых сложных целей нам не обойтись без значительных изменений в устройстве радиоинтерфейса 6G. Мы должны полностью переосмыслить идеологию радиоинтерфейса

начиная с самых первых шагов разработки 6G. В связи с этим мы предлагаем несколько революционных изменений парадигмы.

От «мягкого» радиоинтерфейса к интеллектуальному

В радиоинтерфейс 5G не заложено универсальное решение всех задач. Зато его гибкость и настраиваемость (т. е. «мягкость») позволяют нам оптимизировать радиоинтерфейс для различных сценариев использования, таких как eMBB, URLLC и mMTC в рамках единой структуры.

Конструкция радиоинтерфейса 6G основана на сочетании искусственного интеллекта, управляемого моделями и данными, и, как ожидается, должна обеспечить индивидуальную оптимизацию радиоинтерфейса от предварительной конфигурации до самообучения. Персонализированный радиоинтерфейс может настраивать схему сети и параметры передачи на уровне пользовательского оборудования и служб, чтобы максимизировать индивидуальный пользовательский опыт без ущерба для пропускной способности системы в целом. Его можно легко масштабировать для поддержки URLLC с почти нулевой задержкой. Кроме того, новый простой и гибкий механизм обработки сигнала минимизирует накладные расходы и задержку прохождения.

От традиционного внешнего ИИ к нативному

В настоящее время технологии искусственного интеллекта и машинного обучения применяются для постоянного улучшения характеристик 5G. В эпоху 6G искусственный интеллект будет встроенной функцией радиоинтерфейса, образующей интеллектуальный физический уровень и отвечающей за управление доступом к среде передачи данных (media access control, MAC). Она не будет ограничиваться оптимизацией сетевого управления (например, балансировкой нагрузки и энергосбережением), заменой некоторых нелинейных или невыпуклых алгоритмов в модулях приемопередатчиков или компенсацией недостатков в нелинейных моделях.

Встроенный интеллект сделает физический уровень 6G более продвинутым и эффективным. Он облегчит оптимизацию структурных элементов физического уровня и процедурное проектирование, включая возможное изменение архитектуры приемопередатчика. Вдобавок ко всему он поможет реализовать новые возможности обнаружения и позиционирования, что, в свою очередь, значительно повлияет на конструкцию компонентов радиоинтерфейса. Обнаружение и позиционирование с помощью искусственного интеллекта также сделает возможным недорогое и высокоточное формирование луча. MAC образует интеллектуальный контроллер на основе одно- или многоагентного обучения с подкреплением, включая совместное машинное обучение для узлов сети и пользовательского оборудования. Например, с помо-

стью многопараметрической совместной оптимизации и индивидуального или совместного обучения может быть получен огромный прирост производительности с точки зрения пропускной способности системы, пользовательского опыта и энергопотребления.

От мер энергосбережения к конструктивно заданной экономичности

Минимизация энергопотребления как сетевыми узлами, так и оконечными устройствами должна быть первостепенной задачей при проектировании радиоинтерфейса 6G. В отличие от механизма энергосбережения 5G, где низкое энергопотребление является дополнительной функцией или дополнительным режимом, энергосбережение в 6G будет встроенной функцией и режимом работы по умолчанию. Мы ожидаем, что благодаря интеллектуальному управлению использованием мощности, стратегии энергопотребления по запросу и применения других новых технологий (например, схемы тестирования каналов с помощью сканирования/позиционирования) в сети 6G и оконечных устройствах будет значительно повышена энергоэффективность.

От простого подключения к интеграции функций подключения и сканирования

Сканирование не только предоставляет новые функции и, следовательно, новые возможности для бизнеса, но и способствует развитию связи. Например, с одной стороны, сеть связи может служить средством сканирования окружающей среды с высоким разрешением и широким охватом. Но, с другой стороны, сканирование может генерировать полезную информацию (такую как местоположение, доплеровский сдвиг, направление луча и визуализация), которая полезна для оптимизации связи. Кроме того, возникает возможность предложить новую услугу визуализации окружающей среды с помощью клиентских оконечных устройств.

Современные требования к 6G включают создание единой сети с функциями сканирования и связи, которые должны быть интегрированы в единую структуру на этапе проектирования радиоинтерфейса. Мы надеемся, что тщательно спроектированная сеть связи и сканирования будет предлагать полные возможности сканирования, а также более полно соответствовать всем KPI в области связи.

От пассивного управления лучом по запросу к упреждающему управлению лучом, ориентированному на оборудование пользователя

Передача данных посредством направленного луча очень важна для высоких частот, таких как миллиметровые волны. При использовании

высоконаправленных антенн требуются большие усилия для создания и поддержания точного совмещения лучей передатчика и приемника. Управление лучом в системе 5G направлено на адаптацию направления луча к изменению среды передачи и перемещениям пользовательского оборудования. Однако подобный подход сталкивается с проблемами, связанными с дополнительными затратами на обучение, задержкой доступа и высоким энергопотреблением, особенно для более широкого спектра.

В системе 6G ожидается более сложное управление лучом из-за использования диапазонов более высоких частот. К счастью, с помощью новых технологий, таких как сканирование, расширенное позиционирование и искусственный интеллект, мы можем заменить традиционные механизмы трассировки луча, обнаружения потери луча и восстановления луча упреждающими схемами формирования, трассировки и подстройки луча с приоритетом пользовательского оборудования. Кроме того, может быть реализована мобильность без передачи обслуживания, по крайней мере на физическом уровне. Эти новые интеллектуальные технологии формирования диаграммы направленности и управления лучом, ориентированные на пользователя, позволят максимально повысить качество обслуживания пользовательского оборудования и общую производительность системы.

Более того, ожидаемые RIS и новые типы мобильных антенн, например антенны, размещенные на беспилотных летательных аппаратах (БПЛА), позволяют нам перейти от пассивной работы с каналом, который есть в нашем распоряжении, к активному формированию каналов и управлению ими. Благодаря использованию массива антенных решеток с учетом потребностей канала и движущихся распределенных антенн трасса канала может быть построена или перестроена для получения канала передачи с оптимальным набором характеристик.

От отслеживания смены канала к ее прогнозированию

Для обеспечения высоконадежной беспроводной связи необходима точная информация о канале. В настоящее время выбор канала основан на результатах зондирования (измерения) канала с помощью опорного сигнала (reference signal, RS). На практике сложно получать информацию о канале в реальном времени из-за задержки измерения и предоставления результата, а также из-за желания избежать непроизводительных затрат на измерение канала. Также стоит отметить, что так называемое «старение канала» ухудшает производительность, особенно при работе с высокоскоростным пользовательским оборудованием.

Развитие технологии тестирования каналов с помощью ИИ на основе данных сканирования и позиционирования может привести к переходу от выбора каналов на основе RS к выбору с учетом текущего состоя-

ния окружающей среды в реальном времени. Обладая информацией, полученной в результате сканирования/локализации, мы можем значительно упростить процесс поиска луча. Упреждающее отслеживание и прогнозирование канала могут предоставить информацию о канале в реальном времени и решить проблему старения канала. Кроме того, новая технология выбора каналов сводит к минимуму накладные расходы на выбор канала и энергопотребление для сетевых и оконечных устройств.

От наземных и спутниковых систем к интеграции наземных и неназемных систем

Спутниковые системы были включены в последние версии 5G как расширения систем наземной связи. Ожидается, что в системе 6G интеграция наземных и неназемных систем обеспечит полное покрытие земной поверхности и пропускную способность по запросу. В 6G наземные и неназемные системы, спутниковые группировки, БПЛА, NAPS, дроны и т. д. будут рассматриваться как новые типы движущихся сетевых узлов, которые требуют новых подходов к архитектуре. Объединение наземных и неназемных систем позволит повысить эффективность совместного обслуживания нескольких соединений, более гибкое совместное использование функциональных возможностей и более быструю коммутацию кросс-соединений. Эти новые механизмы будут иметь большое значение для достижения глобального покрытия и подлинно глобальной мобильности с низким энергопотреблением.

От работы с несколькими несущими к сверхгибкому использованию спектра

5G поддерживает агрегацию несущих частот суб6G и миллиметрового диапазона, а также допускает перекрестную работу несущих TDD и FDD. Интеллектуальное использование спектра и управление канальными ресурсами являются важными аспектами проектирования 6G. Для поддержки беспрецедентных скоростей передачи данных, необходимых для 6G, будут исследованы более высокочастотные спектры с более широкой полосой пропускания (например, верхний предел диапазонов частот от миллиметрового до ТГц-диапазона). Однако более высокие частоты страдают от серьезных потерь на трассе и атмосферного поглощения. В свете этого при разработке радиоинтерфейса 6G мы должны подумать о том, как эффективно использовать эти новые спектры совместно с другими низкочастотными диапазонами. Более того, в эпоху 6G потребуются более продвинутый полнодуплексный режим, хотя он уже применяется в 5G. Нам следует рассмотреть возможность разработки упрощенного механизма, обеспечивающего быстрое

переключение между несущими и гибкое двунаправленное назначение ресурсов спектра в 6G. Также ожидается, что определение унифицированной структуры кадра и сигналов для FDD, TDD и полнодуплексного режима упростят системные операции и поддержат сосуществование пользовательского оборудования с различными возможностями дуплексной связи.

От РЧ-прозрачных систем к РЧ-оптимальным

Процессы и алгоритмы обработки сигналов основной полосы частот обычно разрабатываются без тщательного учета характеристик аналоговых и радиочастотных компонентов из-за сложности моделирования вносимых ими искажений и нелинейности. Это приемлемо для низких частот, особенно с эффектами линеаризации, такими как цифровое предварительное искажение характеристик усилителя мощности. В эпоху 6G ожидается, что конструкция системы на физическом уровне будет учитывать зависящие от частоты ухудшения или ограничения канала, особенно в терагерцовой части спектра. Благодаря встроенным функциям искусственного интеллекта также возможно совместное проектирование и оптимизация использования различных полос частот.

Глава 19

Структура интеллектуального радиointерфейса

19.1. Почему нужен новый радиointерфейс

Создатели *нового радио* (new radio, NR) добились значительного прогресса в разработке новых технологий радиодоступа, востребованных в различных сценариях использования. Радиointерфейсы NR показали значительное улучшение с точки зрения производительности, гибкости, масштабируемости и эффективности по сравнению с постепенно эволюционирующей сетью 4G (LTE).

Новая система программного радиointерфейса обеспечивает единую структуру для поддержки частотных диапазонов ниже 6 ГГц и выше 6 ГГц, таких как миллиметровые волны, как для лицензированного, так и для нелицензионного доступа. Настраиваемый унифицированный радиointерфейс также поддерживает как канал Uu между RAN и пользователями, так и прямые каналы между устройствами. Гибкая масштабируемая нумерология²² позволяет оптимизировать параметры передачи для различных диапазонов спектра и для различных услуг.

Унифицированный радиointерфейс в NR также является *конструктивно полным* (самодостаточным) в частотной области. Самодостаточность в частотной области позволяет поддерживать более гибкое разделение RAN за счет совместного использования ресурсов канала между различными службами как по частоте, так и по времени. Принцип самодостаточности сигналов в этой области также обеспечивает полную совместимость с будущими выпусками NR. Более того, полнота во временной области обеспечивает быструю обратную связь для услуг с малой задержкой.

Исходя из обсуждений технологических тенденций в главе 1 и сценариев использования и ключевых показателей эффективности в части II книги, можно предположить, что радиосеть 6G будет намного сложнее,

²² Переменный коэффициент разнесения поднесущих частот. – Прим. перев.

чем сеть 5G с точки зрения функциональности. Структура будущего радиоинтерфейса может оказаться чрезвычайно сложной, поскольку новые применения, требования и ключевые показатели эффективности сопровождаются попытками минимизировать затраты. Этим и обусловлена потребность в революционном радиоинтерфейсе 6G. Структура нового радиоинтерфейса 6G должна быть более интеллектуальной и экологичной по сравнению с NR. Она должна отвечать всем требованиям 6G наиболее продуктивным образом с точки зрения эффективности развертывания, стоимости, энергопотребления и сложности. Чтобы создать такую конструкцию, необходимо тщательно изучить и оптимально использовать все соответствующие компоненты технологии радиоинтерфейса, включая ИИ, новый спектр, наземную связь и сканирование.

Цель этой главы – предоставить обзор интеллектуальных составляющих радиоинтерфейса. Раздел 19.2 содержит всесторонний обзор текущего состояния технологий искусственного интеллекта и машинного обучения в беспроводной связи. Затем в разделе 19.2.1 мы обсудим ожидаемые новые разработки и потенциальные направления исследований радиоинтерфейсов.

19.2. Обзор существующих технологий

На сегодняшний день назрела насущная необходимость в оптимизации OPEX и CAPEX как для операторов сетей, так и для поставщиков устройств. С точки зрения сетевых операторов, наиболее важными факторами оптимизации OPEX и CAPEX являются эффективность использования фрагментированного спектра и оптимизация энергоэффективности. Последнее всегда было решающим фактором успешной работы сети после развертывания. Кроме того, важным аспектом взаимодействия с пользователем, влияющим на потребление различных услуг, является время автономной работы конечного устройства. Следовательно, для совершенствования пользовательского опыта в сети 6G крайне важно минимизировать энергопотребление конечного оборудования. С этой целью в разделе 19.2.1 мы предоставляем краткое введение в принципы использования спектра и механизмы энергосбережения при проектировании радиоинтерфейса NR.

Кроме того, как мы уже неоднократно упоминали, многообещающими технологиями-кандидатами для внедрения в телекоммуникации являются технологии искусственного интеллекта, особенно машинное обучение. В разделах 19.2.2 и 19.2.3 мы представим обзор трех ключевых аспектов применения методов ИИ и машинного обучения в беспроводной связи: методы на физическом уровне и на уровне доступа к среде, а также алгоритмы ИИ и соответствующие архитектуры.

Что касается физического уровня, продолжают исследования того, как использовать ИИ и машинное обучение для оптимизации конструкции компонентов и повышения производительности алгоритмов, таких как кодирование каналов и обслуживание ММО. Что касается МАС (управления доступом к среде), исследования показали, что для решения сложных задач оптимизации с применением улучшенных стратегий необходимо использовать технологии ИИ и машинного обучения. В разделе 19.2.3 обсуждается несколько примеров, таких как использование глубокого обучения с подкреплением для оптимизации функциональности МАС в части планирования каналов и управления мощностью.

19.2.1. Использование спектра и энергоэффективность в NR

В настоящее время существуют гибкие механизмы для совместной работы со спектром в NR, которые позволяют оператору эффективно задействовать несколько доступных ресурсов спектра. Чтобы эффективно применять эти ресурсы спектра, в NR используются агрегирование несущих (carrier aggregation, CA) и двойное подключение (dual connectivity, DC), что позволяет увеличить полосу пропускания пользовательского оборудования.

В случае CA одному пользовательскому оборудованию назначаются несколько компонентных несущих (component carriers, CC). В случае DC пользовательское оборудование может одновременно передавать и принимать данные по множеству CC из двух групп сот через главную и вторичную базовые станции.

В технологии NR CA поддерживается гибкое агрегирование спектра, включая агрегирование частот ниже 6 ГГц и выше 6 ГГц, а также агрегирование FDD и TDD. В технологии NR DC для обеспечения более тесного взаимодействия между главной и вторичной базовыми станциями поддерживается несколько сетевых архитектур, включая LTE-NR DC и NR-NR DC. В LTE-NR DC пользовательское устройство сначала подключается к радиосети и базовой сети LTE, а затем подключается к радиосети NR посредством реконфигурации управления радиоресурсами (radio resource control, RRC).

Энергоэффективность – это ключевой показатель эффективности беспроводных сетей как на стороне конечных устройств, так и на стороне сети. Гибкая и масштабируемая структура системы в 5G NR позволяет использовать различные стандартизированные методы энергосбережения для адаптации к различным нагрузкам и типам трафика в сетях. В частности, стандарт NR поддерживает адаптивный опорный сигнал (избегая использования постоянно присутствующего опорного

сигнала) и гибкое отключение ресурсов, а также обеспечивает прямую совместимость для реализации энергоэффективной сети, механизма прерывистого приема (discontinuous reception, DRX) и неактивного состояния RRC.

Что касается неактивного состояния RRC, которое было введено в NR версии 15 в дополнение к состояниям RRC-соединения и RRC-ожидания, его назначение состоит в том, чтобы реализовать более быстрый и более эффективный возврат к состоянию RRC-соединения. Это позволит осуществлять передачу данных с меньшими издержками на служебные сигналы, меньшей задержкой и меньшим энергопотреблением. Переход в состояние RRC-соединения из неактивного состояния в NR становится более энергоэффективным.

19.2.2. ИИ и машинное обучение на физическом уровне

Недавно было проведено исследование того, как применять алгоритмы ИИ и машинного обучения к компонентам физического уровня [1–3]. Ниже кратко представлены некоторые примеры, включая использование машинного обучения для моделирования и оценки каналов, кодирования каналов, модуляции, MIMO и разработки формы сигналов.

Метод ИИ / машинного обучения может использоваться для извлечения характерных признаков беспроводного канала во временной, частотной и пространственной областях. Например, в [4] представлена модель нейронной сети на основе RNN, таких как LSTM и стробируемый рекуррентный блок, предназначенная для изучения свойства временной корреляции беспроводного канала. Обученная модель может прогнозировать изменение канала, что обеспечивает даже более точную информацию о канале, чем оценка канала на основе пилот-сигнала в сценариях глубокого замирания. Чтобы извлечь больше признаков из данных беспроводного канала, в [5] был введен метод встраиваемых доменов, в котором встроенные данные подаются в сеть преобразователя для предварительного обучения модели канала. Предварительно обученная модель канала может использоваться во многих задачах нисходящей связи, таких как прогнозирование канала, картирование каналов и позиционирование.

Хотя кодирование каналов обычно разрабатывается в соответствии с разумными теоретическими рекомендациями, они не всегда реализуемы на практике. Типичным примером является тот факт, что производительность декодирования полярных кодов по списку последовательной отмены (successive cancellation list, SCL) плохо поддается анализу. В работе [6] структура на основе обучения с подкреплением использовалась для поиска полярных кодов с заданным размером списка в SCL-декодере.

Нейронные сети также использовались непосредственно в качестве декодеров. Преимущества нейронной сети в качестве канального декодера включают снижение сложности, как в [7], и лучшую компенсацию нелинейности, как в [8]. Еще одна область применения методов машинного обучения – совместное кодирование источника и канала (joint source and channel coding, JSCC). В работе [9] описан фреймворк JSCC на основе автоэнкодера, который продемонстрировал лучшее качество восстановления информации по сравнению с традиционными методами.

В области модуляции популярной задачей машинного обучения является распознавание типов модуляции неизвестных радиосигналов. В [10] сверточные нейросети используются для распознавания множества типов цифровой и аналоговой модуляций. Существуют и демодуляторы на основе нейронных сетей. Например, в [11] полносвязные сети используются для аппроксимации апостериорного логарифмического максимума (log-MAP), применяемого при мягкой демодуляции. Автоэнкодер используется для разработки схем модуляции, в которых выполняется как геометрическое, так и вероятностное формирование сигнала [12].

Методы машинного обучения широко применяются в системах MIMO. Прежде всего они несколькими способами упрощают подбор каналов. Например, в [13] каналы в одном наборе антенн и одном частотном диапазоне отображаются на каналы в другом наборе антенн в другом частотном диапазоне через полностью подключенные сети, что демонстрирует возможность прямого получения канала нисходящей линии связи по результатам сканирования канала восходящей линии связи даже в системах FDD. В [14] и [15] нейросети на основе автоэнкодера были обучены сжимать информацию о состоянии канала (channel state information, CSI), которая затем передается с меньшими затратами. Автоэнкодер также используется для разработки предварительного кодирования, особенно когда речь идет о неполной CSI [16] и есть шанс добиться большей эффективности. Методы машинного обучения можно использовать при разработке приемников MIMO. Например, в [17] для оценки канала MIMO использовалась специально разработанная глубокая нейросеть на основе глубокой априорной сети.

В области обработки сигналов [19] в каналах с зашумлением и замираниями Рэлея предлагается использовать глубокую комплексную сверточную сеть для декодирования сигнала с ортогональным частотным разделением (orthogonal frequency-division multiplexing, OFDM) с квадратурной амплитудной модуляцией напрямую в биты данных. В то же время цель генерации сигналов на основе машинного обучения в основном заключается в уменьшении отношения пиковой мощ-

ности к средней мощности (peak-to-average power ratio, PAPR) сигналов OFDM. В [20] была предложена основанная на автоэнкодере система, которая будет отображать модулирующие символы на поднесущие OFDM в передатчике и восстанавливать их отображение в приемнике. В результате значение PAPR при передаче символов OFDM снижается сильнее, чем любым другим способом. В [21] для генерации сигнала OFDM с помощью глубокой нейросети был выбран упорядоченный набор поднесущих, отображающих специально разработанные модулирующие символы, что в перспективе способствует значительному уменьшению PAPR сигнала.

19.2.3. Машинное обучение на уровне доступа к среде (MAC)

В области организации доступа к среде некоторые работы, такие как [22], полагаются либо на обучение с учителем, либо на обучение без учителя. При обучении с учителем сначала формулируются задачи оптимизации, чтобы выполнить разметку данных. При обучении без учителя задачи оптимизации формулируются для того, чтобы установить цели оптимизации. Однако задачи принятия решений на уровне MAC весьма чувствительны к изменениям среды канала. Модели на основе глубокого обучения с подкреплением (deep reinforcement learning, DRL) не только принимают решения, но и динамически корректируют политику принятия решений, что делает их полезными на уровне MAC. Сжатый обзор применения глубокого обучения с подкреплением в коммуникациях и сетях представлен в [23]. Ниже мы резюмируем некоторые типичные случаи.

Традиционные схемы адаптивной модуляции и кодирования в основном являются *реактивными*, т. е. регулируют схему модуляции и кодирования на основе обратной связи от приемников. В [24] агент DRL применялся для определения настроек модуляции и кодирования путем изучения опыта и взаимодействия с другими агентами, что привело к более эффективному решению, принимаемому агентом *проактивно*, т. е. на опережение.

Распределение ресурсов – еще одна важная функция системы на уровне MAC. Распределяемым ресурсом может быть, среди прочего, возможность доступа, возможность передачи, мощность или спектр. В [25] агент DRL по распределению возможности доступа помогал устройствам выбирать свои собственные шаблоны доступа, то есть порядок использования тайм-слотов, а также определять вероятность каждого выбранного шаблона. Моделирование показало, что метод на основе DRL превосходит фиксированные протоколы переключения режимов Tx/Rx.

В исследованиях [26] и [27] агент DRL применялся в качестве планировщика МАС на базовой станции для поиска распределения возможностей передачи; была достигнута оптимальная производительность в случае трафика с полным буфером, и получено увеличение до 30 % по схеме пропорциональной справедливости в случае трафика с неполным буфером.

В [28] был предложен метод глубокой Q-сети для решения проблемы разделения спектра при организации связи между транспортными средствами (vehicle-to-vehicle, V2V) и между транспортными средствами и инфраструктурой (vehicle-to-infrastructure, V2I). Агент DRL обучался централизованно и распределялся по каждому каналу V2V для принятия решений на местном уровне. Удалось повысить пропускную способность системы без сложного централизованного управления. Аналогичные проблемы совместного использования спектра рассматривались в сверхплотных сетях миллиметрового диапазона [29] и гетерогенных сетях на основе LTE [30], где предложены распределенный алгоритм Q-обучения и гетерогенные многоцелевые распределенные стратегии на основе обучения с подкреплением соответственно.

В [31] рассмотрена проблема управления мощностью и устранения взаимных помех для сценария сверхплотных малых сот. Были предложены методы на основе как RL, так и DRL; их цель состояла в том, чтобы максимизировать пропускную способность целевой соты при минимизации мощности передачи этой соты, чтобы ограничить взаимные помехи с соседними сотами. Управление мощностью в многопользовательских сотовых сетях, где учитывались как межсотовые, так и внутрисотовые помехи, рассматривалось в [32]. Для решения проблемы использовались методы DRL, такие как глубокая Q-сеть и глубокий детерминированный градиентный спуск по стратегиям.

19.3. Ожидания от новых разработок и потенциальные направления исследований

Устройство радиоинтерфейса 6G должно обеспечить совершенно новую основу для эффективной поддержки всех известных и будущих технологий радиодоступа в мобильной сети. Ниже перечислены основные предпочтительные характеристики такого радиоинтерфейса:

- более интеллектуальный и экологичный, с нативным ИИ и возможностью энергосбережения;
- обеспечивает более гибкое использование спектра вплоть до ТГц;
- поддерживает эффективную интеграцию связи и сканирования;
- содействует более тесной интеграции наземной и неназемной связей;

- обеспечивает более простой протокол и механизм управления с низкими накладными расходами и сложностью.

В этом разделе мы обсудим наши ожидания от инфраструктуры интеллектуального радиointерфейса 6G, которая будет предлагать индивидуальную конфигурацию для каждого отдельного пользовательского применения и гибкие услуги с меньшими накладными расходами. Эта персонализированная инфраструктура состоит из интеллектуального физического уровня с поддержкой ИИ, интеллектуального контроллера MAC и интеллектуального протокола, за которым следует потенциально новое устройство сквозного канала (E2E) на основе ИИ.

19.3.1. Персонализированный радиointерфейс с поддержкой ИИ

Интеллектуальный протокол и механизмы передачи сигнала являются неотъемлемой частью персонализированного радиointерфейса с поддержкой ИИ, который изначально будет поддерживать интеллектуальный физический уровень и интеллектуальный доступ к среде. В этом и будет заключаться отличие гибкого радиointерфейса NR от интеллектуального радиointерфейса 6G.

- **Интеллектуальный физический уровень:** технологии ИИ и машинного обучения способны обрабатывать массивные выборки данных, решать проблемы нелинейного отображения и создавать саморазвивающиеся механизмы передачи. Возможно, на основе ИИ удастся разработать универсальный модуль оптимизации различных функций физического уровня, чтобы сделать его более адаптивным и гибким. В данном контексте нам необходим интеллектуальный механизм реконфигурирования физических модулей, ориентированный на поддержку сценариев быстрой обработки массивных данных, таких как сканирование, локализация и глубокое погружение в виртуальную реальность.
- **Интеллектуальный контроллер MAC:** методы машинного обучения способны предоставить прогнозы и стратегию. Из собранных беспроводных данных можно многому научиться. Исходя из этого, мы можем представить себе ситуацию, в которой в состав системы включены интеллектуальные контроллеры MAC, благодаря чему упрощается оперативная настройка алгоритмов и параметров радиointерфейса. Благодаря кумулятивному методу обучения этот интеллектуальный контроллер автономно обучится настолько, чтобы запоминать опыт и принимать правильные решения. Благодаря распространяемым контроллером MAC согласованным настройкам различных модулей приемопередатчи-

ка и взаимодействию между многими сетевыми объектами можно ожидать огромного прироста производительности.

- **Интеллектуальный протокол и передача сигналов:** для реализации интеллектуального физического уровня и интеллектуального MAC, а также интеграции этих уровней необходимы новый протокол и механизмы передачи сигналов. Ожидается, что 6G предложит соответствующий механизм для обеспечения эффективной работы интеллектуального физического уровня и MAC.

В следующих разделах мы предоставим некоторые варианты использования ИИ на физическом уровне. Мы также предоставим более подробное обсуждение некоторых проблем на уровне MAC, решению которых могут способствовать технологии ИИ и машинного обучения.

Интеллектуальный физический уровень с поддержкой ИИ

Недавние достижения в области искусственного интеллекта и машинного обучения побудили исследователей в области беспроводной связи использовать соответствующие методы для разработки модулей физического уровня для систем беспроводной связи следующего поколения. Некоторые основанные на машинном обучении проекты предусматривают адаптацию к нелинейным факторам в модулях физического уровня; это имеет место, например, при канальном кодировании, модуляции и проектировании формы сигнала, которые традиционно основаны на математических моделях. Более того, ожидается, что многочисленные данные о каналах и окружающей среде, собранные с помощью сканирования в эпоху 6G, могут быть использованы для облегчения проектирования физического уровня с помощью машинного обучения, например данные о состоянии канала в системе MIMO. Ниже показано, в чем может заключаться выигрыш от применения машинного обучения и ИИ для основных модулей физического уровня:

- **кодирование и декодирование канала:** кодирование канала необходимо для надежной передачи данных по зашумленным каналам. Хороший код канала может приблизительно равняться пределу Шеннона. Хотя теория информации Шеннона обеспечивает цель или критерий оценки для кода канала, она не дает прямых указаний по его разработке. Реализация канального кодирования в рамках теории информации в основном основана на предположении о канале с аддитивным белым гауссовым шумом. Однако на практике большинство каналов являются каналами с замиранием. В этом случае при разработке кодирования каналов отсутствуют даже теоретические критерии, и здесь нам может помочь машинное обучение. Декодирование также вызывает затруднения, поскольку обычно оно связано с высокой вычис-

лительной сложностью. Иногда приходится делать упрощающие предположения, чтобы декодировать коды с доступной сложностью, но за это приходится платить снижением производительности. Следовательно, обучаемые модели также применимы в канальных декодерах, где процесс декодирования может быть смоделирован как задача классификации;

- **модуляция и демодуляция:** основная задача модуля модуляции – отображение нескольких битов в передаваемый символ с учетом достижения высокой спектральной эффективности при ограниченной полосе пропускания. В системах беспроводной связи широко используются классические схемы модуляции, такие как M-QAM. Подобные методы квадратурной модуляции гарантируют низкую сложность демодуляции в приемнике. Было обнаружено, что существуют и другие совокупности схем модуляции, использующие дополнительные характеристики сигнала, такие как неевклидово расстояние и вероятностные распределения. Следовательно, методы машинного обучения можно применить для формирования и разработки подходящих комбинаций для конкретных сценариев применения;
- **MIMO и приемник:** MIMO – очень привлекательная технология, которая повышает надежность и пропускную способность беспроводной связи. С ростом числа антенных элементов системы MIMO увеличиваются не только преимущества, но и сложность проектирования и управления. Поэтому было бы очень полезно использовать методы машинного обучения для правильного проектирования модулей, связанных с MIMO, например для таких задач, как обратная связь на основе CSI, выбор антенны, предварительное кодирование, а также оценка и обнаружение канала. Большинство алгоритмов машинного обучения могут быть развернуты в режиме автономного обучения / онлайн-вывода, что может решить проблему больших накладных расходов на обучение;
- **форма волны и множественный доступ:** генерация волны заданной формы отвечает за отображение информационных символов в сигналы, подходящие для распространения в виде электромагнитных волн. Обычный модуль для генерации сигналов можно заменить модулем глубокого обучения. Например, методы на основе глубокого обучения позволяют проектировать сигналы сложной формы без использования модуля DFT. В качестве следующего шага вперед можно было бы напрямую проектировать новую форму волны, чтобы полностью заменить стандарт OFDM, установив некоторые конкретные требования, например ограни-

чение PAPR или низкий уровень внеполосного излучения. Важно, что этот подход будет поддерживать асинхронную передачу и, таким образом, позволит избежать больших накладных расходов на сигналы синхронизации, в которых нуждается массив оконечных устройств. Отсюда также следует возможность хорошей локализации во временной области для предоставления услуг с малой задержкой и эффективной передачи малых пакетов.

От системы с искусственным интеллектом мы можем ожидать революционных новаций. Среди возможных направлений будущих исследований можно выделить следующие:

- **оптимизация и обновление параметров элементов физического уровня:** оптимизированные параметры каждого модуля, такие как кодирование, модуляция, параметры ММО, оказывают большое влияние на производительность систем связи. Оптимизированные параметры могут динамически изменяться из-за быстро изменяющихся во времени характеристик канала физического уровня в реальной среде. Применение нейросетей и методов машинного обучения значительно упрощает получение оптимизированных параметров по сравнению с традиционными схемами. Кроме того, традиционная оптимизация параметров выполняется для каждого составляющего блока по отдельности, например в модели кодовой модуляции с перемежением битов [33], в то время как согласованная оптимизация нескольких блоков может обеспечить дополнительный прирост показателей. К тому же, чтобы адаптироваться к быстро меняющемуся во времени состоянию канала, можно использовать самообучение оптимизированных параметров с помощью ИИ для дальнейшего повышения показателей;
- **выбор канала:** получение информации о беспроводном канале и среде передачи всегда было фундаментальным аспектом проектирования системы беспроводной связи. Исторические данные канала и данные сканирования хранятся в виде наборов данных, на основе которых с помощью нейросети можно построить радиокарту среды. Обладая радиокартой, информацию о канале можно получить не только путем обычных измерений, но и путем сопоставления ее с другой информацией, такой как местоположение;
- **формирование и сопровождение луча:** поскольку несущая частота достигает миллиметрового или даже ТГц-диапазона, в беспроводной связи широко применяются такие операции, как формирование, наведение и сопровождение луча. В этом контексте важно располагать эффективными алгоритмами формирова-

ния и сопровождения луча. Методы машинного обучения могут применяться для комплексной оптимизации выбора антенны, динамического формирования диаграммы направленности и предварительного кодирования;

- **сканирование и позиционирование:** данные высокого качества имеют основополагающее значение для технологий машинного обучения и ИИ. Ожидается, что в системах 6G благодаря большой пропускной способности, новому спектру, плотной сети и большому количеству каналов LOS могут быть получены и измеренные параметры канала, и данные сканирования/позиционирования. На основе этих данных при помощи обученной нейросети можно составить радиокарту окружающей среды, где информация о канале связана с соответствующим местоположением или информацией об окружающей среде; следовательно, можно улучшить структуру физического уровня.

Интеллектуальный MAC-контроллер с поддержкой ИИ

Контроллер на уровне MAC играет важную роль в обеспечении бесперебойной работы сети радиодоступа. Он принимает множество ключевых решений в течение жизненного цикла системы связи, таких как формирование луча, управление лучом, использование спектра, выделение ресурсов канала, адаптация MCS, управление гибридным автоматическим повторным запросом, адаптация режима Tx/Rx, управление мощностью и снижение взаимных помех. Среда беспроводной связи очень динамична из-за меняющихся условий канала, трафика, нагрузки, помех и т. д.

В общем случае производительность системы улучшается, если параметры передачи могут адаптироваться к быстро меняющейся среде. Однако традиционные методы решения вышеупомянутых проблем на основе классической теории оптимизации обычно представляют собой NP-трудную задачу и слишком сложны для реализации. В этом случае машинное обучение может стать мощным инструментом создания интеллектуального контроллера для оптимизации беспроводной передачи на уровне MAC.

Когда мы переходим к интеллектуальной системе контроллера MAC, изначально использующей машинное обучение, нам необходимо аккуратно решить две проблемы:

- **одинокый агент или мультиагент:** в отличие от моделей DRL, используемых в других областях, где одного агента может быть достаточно для большинства приложений, для систем беспроводной связи требуется многоагентная система DRL. Несмотря на то что каждый контроллер базовой станции может принимать ин-

дивидуальные решения, система всегда лучше работает в случае совместно принимаемых решений нескольких базовых станций. К счастью, почти все алгоритмы DRL имеют многоагентные аналоги. Однако процедура обучения многоагентной модели DRL намного сложнее, чем одноагентной. Выбор между одним и несколькими агентами – это, по сути, компромисс между качеством и сложностью обучения;

- **совместная или индивидуальная оптимизация:** будучи ограниченными вычислительной мощностью, традиционные алгоритмы обычно применяются в небольшом объеме относительно всей системы. Например, многие решения в сотовых сетях принимаются внутри соты, что приводит только к локальным оптимумам. С помощью нескольких агентов совместная оптимизация может быть достигнута в большем объеме. Взаимодействие между агентами – эффективный способ построения интеллектуального контроллера MAC.

После обсуждения интеллектуального контроллера в начале этой главы мы пришли к выводу, что интеллектуальные функции на уровне MAC в будущем могут стать ключевыми компонентами интеллектуального контроллера. Более того, ожидается, что совместная оптимизация этих компонентов обеспечит еще лучшие показатели. Речь идет о следующих ключевых компонентах:

- **интеллектуальное управление точкой передачи/приема (transmit/receive point, TRP):** при разработке 5G были исследованы единая TRP и совместная передача с несколькими TRP, например макросоты, малые соты, пикосоты, фемтосоты, выносные радиомодули, ретрансляционные узлы и т. д. Всегда было сложно разработать эффективную схему управления TRP, соблюдая правильный компромисс между производительностью и сложностью. Для таких типичных задач, как выбор TRP, включение/выключение TRP, управление мощностью и распределение ресурсов, как правило, трудно найти оптимальное решение. Особенно это касается крупномасштабной сети. Вместо использования сложного математического метода оптимизации можно использовать машинное обучение, которое имеет меньшую сложность и может адаптироваться к условиям сети. Например, для поддержки интеллектуального управления TRP с целью интеграции наземных и неназемных сетей можно спроектировать сеть политик в одно- или многоагентной модели DRL;
- **интеллектуальное управление лучом:** несколько антенн (или антенная решетка с фазовым сдвигом) передатчика могут дина-

мически формировать один или несколько лучей на основе условий канала для направленной передачи сигнала одному или нескольким пользователям. Приемник также должен точно настроить свою антенную решетку по направлению приходящего луча. В этом случае обучаемую модель можно использовать для изучения изменений окружающей среды и более точного управления лучом в течение очень короткого периода времени. При помощи методов машинного обучения можно сгенерировать правила для управления фазовыми сдвигами антенных элементов, после чего антенные решетки будут работать более «разумно» в различных условиях;

- **интеллектуальная система модуляции и кодирования:** адаптивная модуляция и кодирование (adaptive modulation and coding, AMC) – важный механизм, помогающий системе следовать за динамикой беспроводного канала. Традиционные алгоритмы AMC полагаются на обратную связь от получателя сигнала для принятия ответного решения. Быстро меняющиеся каналы вкуче с задержками в планировании часто делают обратную связь устаревшей. Чтобы решить эту проблему для определения настроек модуляции и кодирования, можно использовать машинное обучение. Благодаря обучению на собственном опыте и взаимодействию с другими агентами интеллектуальный агент с большей вероятностью примет лучшее решение, причем сделает это проактивно;
- **интеллектуальная стратегия HARQ:** помимо комбинирования алгоритмов для достижения избыточности на физическом уровне, работа процедуры HARQ (hybrid automatic repeat request, гибридный автоматический повторный запрос) также влияет на показатели, например на конечные возможности передачи и на распределение ресурсов между новыми и повторными передачами. Для достижения глобальной оптимизации необходимо рассматривать проблему с межуровневой точки зрения, где машинное обучение подходит благодаря большому количеству информации, доступной из различных источников;
- **интеллектуальная координация режимов Tx/Rx:** в сети со множеством взаимодействующих участников координация между ними является ключом к эффективности. В реальных системах как внутренние условия, такие как состояние беспроводного канала и буфера, так и поведение других игроков очень динамичны. По этой причине их чрезвычайно сложно, а то и невозможно предсказать традиционными методами. Несмотря на предпринимаемые усилия, существующих в настоящее время решений

этой проблемы может быть недостаточно из-за все возрастающих масштабов сети, например из-за большего числа игроков в будущем. В этой ситуации машинное обучение может обеспечить сокращение накладных расходов на координацию режима Tx/Rx и улучшить общую производительность системы;

- **интеллектуальное устранение взаимных помех:** борьба со взаимными помехами всегда была ключевой задачей для сотовых сетей. Помехи меняются динамически, и без обратной связи в реальном времени их трудно измерить точно. Ожидается, что машинно обучаемая модель изучит ситуацию с помехами на базовой станции и пользовательском оборудовании по отдельности или совместно. Затем обученная модель может автоматически настроить глобальную оптимальную стратегию, чтобы минимизировать взаимные помехи, тем самым достигая максимальной эффективности использования спектра и мощности;
- **интеллектуальное распределение ресурсов канала:** планировщик распределения ресурсов канала можно рассматривать как «мозг» сотовой сети, поскольку он определяет распределение возможностей передачи, определяя показатели системы. Помимо возможностей передачи, интеллектуальные агенты также могут управлять другими радиоресурсами, такими как спектр, порт антенны и коды расширения спектра, вместе с интеллектуальным управлением точками приема/передачи (см. выше). Координация распределения радиоресурсов между множеством базовых станций будет способствовать повышению общих показателей сети;
- **интеллектуальное управление мощностью:** ослабление радиосигналов и нестабильность характеристик беспроводных каналов означают, что необходимо контролировать мощность в беспроводной связи. С одной стороны, необходимо гарантировать покрытие, чтобы пользователи на границе ячеек могли получать свою информацию. В то же время нужно свести к минимуму помехи для других пользователей. Поэтому управление мощностью и устранение взаимных помех обычно оптимизируются совместно. Ожидается, что вместо решения сложной задачи оптимизации, которая должна повторяться при изменении среды, альтернативным подходом станет использование машинного обучения.

Интеллектуальный протокол и сигнальные механизмы с поддержкой ИИ

Как мы уже выяснили, для поддержки специализированных структур радиоинтерфейса и, таким образом, для поддержки различных услуг и устройств в сети желательно располагать интеллектуальным физическим

уровнем и контроллером MAC. Но для поддержки этих опций на фундаментальном уровне сети 6G необходимы новый протокол и механизм передачи сигнала. Это позволит персонализировать соответствующий радиоинтерфейс с настраиваемыми параметрами для удовлетворения конкретных требований, при этом минимизируя накладные расходы на работу с сигналом и максимизируя эффективность использования спектра всей системы с помощью персонализированных технологий искусственного интеллекта. Ниже приведены некоторые примеры интеллектуального протокола и механизма передачи сигнала:

- **сверхгибкая структура кадра и гибкая передача сигналов:** например, сверхгибкая структура кадра в персонализированной структуре радиоинтерфейса может быть спроектирована с более гибкими параметрами формы сигнала и продолжительности передачи. Эти параметры могут быть приспособлены для различных требований из широкого диапазона сценариев, например для чрезвычайно низкой задержки 0,1 мс. В результате для каждого параметра в системе доступно множество вариантов. Структура управляющих сигналов должна быть предельно простой и гибкой, требующей лишь нескольких форматов управляющих сигналов, в то время как управляющая информация может иметь переменный размер. Она также может иметь прямую совместимость, без необходимости вводить новый формат для будущих разработок 6G;
- **интеллектуальное использование спектра:** как было сказано в главе 12, потенциальный спектр для 6G может состоять из полос низко- и среднечастотного диапазона, миллиметровых волн, ТГц-диапазона и даже видимого света. Таким образом, совокупная ширина спектра 6G намного шире, чем у 5G, и разработка высокоэффективной системы для поддержки такого широкого спектрального диапазона является сложной задачей.

В сетях 4G и 5G для совместного использования нескольких частей широкого спектра применяются обе схемы – как CA (carrier aggregation, агрегация несущих частот), так и DC (dual connectivity, двойное подключение). В 5G для обеспечения гибкого использования спектра применяется несколько схем DC. В связи с большим количеством комбинаций несущих частот для 6G требуется новый простой и эффективный интеллектуальный радиоинтерфейс, способный поддерживать весь спектральный диапазон.

Текущее распределение спектра и структура кадра обычно связаны с дуплексным режимом, FDD или TDD, что может накладывать ограничения на эффективное использование спектра. Ожидается, что полный дуплекс достигнет зрелости в эпоху 6G и что создава-

емые беспроводные сети будут содержать все больше и больше узлов (конечных пользователей и точек доступа) с различными полосами частот.

Примерами структуры связи, которая не ограничивается направлениями восходящей и нисходящей линий, являются связь D2D, связь IAB, неназемная связь и т. д. Таким образом, интеллектуальный радиоинтерфейс должен предоставлять собой независимую от дуплекса технологию с адекватной конфигурацией, способной уместить в себя различные узлы и типы связи. В этом случае для поддержки всех дуплексных режимов и узлов можно спроектировать однокадровую структуру, а схемы распределения ресурсов в интеллектуальном радиоинтерфейсе смогут выполнять эффективные передачи по множественным радиоканалам;

- **нативное интеллектуальное энергосбережение:** при проектировании экологически безопасной сети на базе 6G энергосбережение является очевидным и важным требованием. Разработку новой системы следует с самого начала вести в контексте энергосбережения, чтобы минимизировать энергопотребление как конечных устройств, так и сетевых узлов.

Радиоинтерфейс 6G предусматривает интеллектуальное управление MIMO и лучом, интеллектуальное использование спектра и точное позиционирование. Это может значительно снизить энергопотребление как устройств, так и сетевых узлов по сравнению с традиционными технологиями, особенно при передаче массивных данных. Таким образом, радиоинтерфейс 6G представляет собой основу энергосбережения системы.

Ожидается, что с помощью интеллектуальных технологий в 6G может быть значительно сокращена продолжительность передачи данных. В результате оконечное устройство может дольше оставаться в «спящем» режиме, когда оно не имеет активного доступа к сети или не взаимодействует с ней. Это один из аспектов нативного энергосбережения, что особенно важно для энергоэффективных устройств и экологически безопасных сетей.

Эффективные каналы передачи могут быть спроектированы с учетом минимизации сигналов управления и количества переходов между состояниями или изменений режима мощности для достижения максимальной экономии энергии оконечными устройствами и сетевыми узлами. В то же время, поскольку сеть 6G должна поддерживать сценарии использования со сверхмалой задержкой, такие как улучшенный URLLC (или URLLC+), ожидается, что эти схемы собственного энергосбережения будут обес-

печивать гибкие функциональные возможности. Они не будут мешать сверхбыстрому доступу к сетям и сверхвысокой передаче данных; примером является интеллектуальное управление режимами питания и работой.

Поскольку радиоинтерфейс будет персонализирован для каждого устройства, разные типы устройств будут иметь разные требования к энергопотреблению. Тогда можно будет без труда персонализировать решения по энергосбережению для различных типов устройств, при этом соблюдая требования к качеству связи.

19.3.2. Организация сквозного канала на основе ИИ и нерешенные проблемы

Большинство недавних попыток внедрения машинного обучения в радиосети сводились к замене одного или нескольких традиционных компонентов физического уровня их усовершенствованной версией. Предполагается, что в эпоху 6G базовая сетевая структура будет неразрывно связана с технологией искусственного интеллекта. Это будет отличать ее как от 5G, так и от предыдущих систем беспроводной связи, в которых методы машинного обучения изначально даже не рассматриваются. В системе 6G технологии радиосвязи и искусственного интеллекта с первого шага будут развиваться вместе в направлении реализации чрезвычайно эффективной обработки информации, сканирования среды передачи и интеллектуальных возможностей радиоинтерфейса.

На рис. 19.1 показан совершенно другой путь применения машинного обучения, то есть устройство канала E2E. В этом случае цепь передачи и модули могут быть реорганизованы. Некоторые ранее описанные функции могут отсутствовать. Во время передачи и приема информации передатчик будет интеллектуально согласовывать изменения источника информации в реальном времени через нейронные сети. Приемник также будет извлекать полезную информацию из полученных данных в соответствии с конкретными задачами. Это уже не универсальный подход к обработке данных, основанный на уникальном критерии (доставка и восстановление сообщений без ошибок), как описано в работе Шеннона [34].

Есть много теорий, используемых для понимания искусственного интеллекта. Среди них теория информационных узких мест [35] является наиболее подходящей для задач коммуникации на физическом уровне, поскольку она использует идею взаимной информации из теории информации в качестве показателя оптимизации. Основанная на теории информационных узких мест, архитектура автоэнкодера (АЕ) является одним из ключевых инструментов для проектирования сквозных кана-

лов связи на основе ИИ. Более подробную информацию о теории информационных узких мест и архитектурах автоэнкодеров также можно найти в части III этой книги.

Вышеупомянутый интеллектуальный канал связи может открыть новые возможности и потенциально стать базовой структурой связи в будущем. Однако методы машинного обучения требуют больших объемов данных. Чтобы задействовать ИИ в беспроводной связи, необходимо будет собирать, хранить и передавать постоянно растущие объемы данных. В отличие от данных с простыми характеристиками, например в области компьютерного зрения или обработки естественного языка, характеристики беспроводных данных охватывают обширные диапазоны в нескольких областях, например от миллиметровых до терагерцевых частот, распространение в космосе, в открытом наземном пространстве, в помещениях, передачу текста, голоса и видео. Эти данные настолько отличаются друг от друга, что будет трудно объединить их сбор, обработку и использование в рамках единой системы.

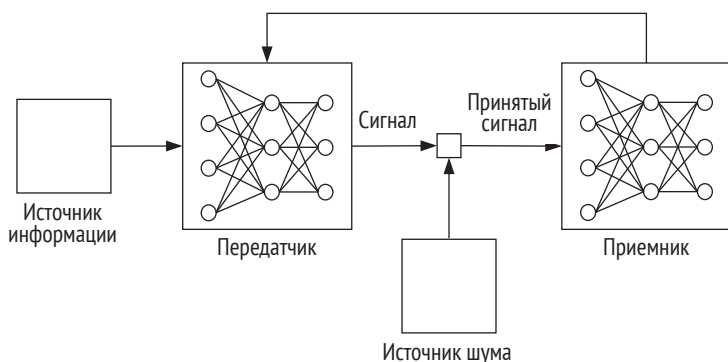


Рис. 19.1. Интеллектуальный сквозной канал связи (E2E)

19.4. Источники

- [1] Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao, *Application of machine learning in wireless networks: Key techniques and open issues*, IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3072–3108, 2019.
- [2] Q. Mao, F. Hu, and Q. Hao, *Deep learning for intelligent wireless networks: A comprehensive survey*, IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 2595–2621, 2018.
- [3] C. Zhang, P. Patras, and H. Haddadi, *Deep learning in mobile and wireless networking: A survey*, IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2224–2287, 2019.

- [4] Y. Huangfu, J. Wang, R. Li, C. Xu, X. Wang, H. Zhang, and J. Wang, *Predicting the mumble of wireless channels with sequence-to-sequence models*, in Proc. 2019 IEEE 30th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). IEEE, 2019, pp. 1–7.
- [5] Y. Huangfu, J. Wang, C. Xu, R. Li, Y. Ge, X. Wang, H. Zhang, and J. Wang, *Realistic channel models pre-training*, in Proc. 2019 IEEE Globecom Workshops. IEEE, 2019, pp. 1–6.
- [6] L. Huang, H. Zhang, R. Li, Y. Ge, and J. Wang, *AI coding: Learning to construct error correction codes*, IEEE Transactions on Communications, vol. 68, no. 1, pp. 26–39, 2019.
- [7] T. Gruber, S. Cammerer, J. Hoydis, and S. ten Brink, *On deep learning-based channel decoding*, in Proc. 2017 51st Annual Conference on Information Sciences and Systems (CISS). IEEE, 2017, pp. 1–6.
- [8] Y. He, J. Zhang, C.-K. Wen, and S. Jin, *TurboNet: A model-driven DNN decoder based on max-log-MAP algorithm for turbo code*, in Proc. 2019 IEEE VTS Asia-Pacific Wireless Communications Symposium (APWCS). IEEE, 2019, pp. 1–5.
- [9] N. Farsad, M. Rao, and A. Goldsmith, *Deep learning for joint source-channel coding of text*, in Proc. 2018 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). IEEE, 2018, pp. 2326–2330.
- [10] T. J. O'Shea, J. Corgan, and T. C. Clancy, *Convolutional radio modulation recognition networks*, in Proc. International Conference on Engineering Applications of Neural Networks. Springer, 2016, pp. 213–226.
- [11] O. Shental and J. Hoydis, *Machine learning: Learning to softly demodulate*, in Proc. 2019 IEEE Globecom Workshops. IEEE, 2019, pp. 1–7.
- [12] M. Stark, F. A. Aoudia, and J. Hoydis, *Joint learning of geometric and probabilistic constellation shaping*, in Proc. 2019 IEEE Globecom Workshops. IEEE, 2019, pp. 1–6.
- [13] M. Alrabeiah and A. Alkhateeb, *Deep learning for TDD and FDD massive MIMO: Mapping channels in space and frequency*, in Proc. 2019 53rd Asilomar Conference on Signals, Systems, and Computers. IEEE, 2019, pp. 1465–1470.
- [14] C.-K. Wen, W.-T. Shih, and S. Jin, *Deep learning for massive MIMO CSI feedback*, IEEE Wireless Communications Letters, vol. 7, no. 5, pp. 748–751, 2018.
- [15] T. Wang, C.-K. Wen, S. Jin, and G. Y. Li, *Deep learning-based CSI feedback approach for time-varying massive MIMO channels*, IEEE Wireless Communications Letters, vol. 8, no. 2, pp. 416–419, 2018.

- [16] F. Sotrohi, H. V. Cheng, and W. Yu, *Robust symbol-level precoding via autoencoder-based deep learning*, in Proc. 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020, pp. 8951–8955.
- [17] E. Balevi, A. Doshi, and J. G. Andrews, *Massive MIMO channel estimation with an untrained deep neural network*, IEEE Transactions on Wireless Communications, vol. 19, no. 3, pp. 2079–2090, 2020.
- [18] H. He, C.-K. Wen, S. Jin, and G. Y. Li, *A model-driven deep learning network for MIMO detection*, in Proc. 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP). IEEE, 2018, pp. 584–588.
- [19] Z. Zhao, M. C. Vuran, F. Guo, and S. Scott, *Deep-waveform: A learned OFDM receiver based on deep complex convolutional networks*, arXiv preprint arXiv:1810.07181, 2018.
- [20] M. Kim, W. Lee, and D.-H. Cho, *A novel PAPR reduction scheme for OFDM system based on deep learning*, IEEE Communications Letters, vol. 22, no. 3, pp. 510–513, 2017.
- [21] B. Wang, Q. Si, and M. Jin, *A novel tone reservation scheme based on deep learning for PAPR reduction in OFDM systems*, IEEE Communications Letters, vol. 24, no. 6, June 2020.
- [22] W. Cui, K. Shen, and W. Yu, *Spatial deep learning for wireless scheduling*, IEEE Journal on Selected Areas in Communications, vol. 37, no. 6, pp. 1248–1261, 2019.
- [23] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, *Applications of deep reinforcement learning in communications and networking: A survey*, IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3133–3174, 2019.
- [24] M. P. Mota, D. C. Araujo, F. H. C. Neto, A. L. de Almeida, and F. R. Cavalcanti, *Adaptive modulation and coding based on reinforcement learning for 5G networks*, in Proc. 2019 IEEE Globecom Workshops. IEEE, 2019, pp. 1–6.
- [25] A. Destounis, D. Tsilimantos, M. Debbah, and G. S. Paschos, *Learn2MAC: Online learning multiple access for URLLC applications*, arXiv preprint arXiv:1904.00665, 2019.
- [26] J. Wang, C. Xu, Y. Huangfu, R. Li, Y. Ge, and J. Wang, *Deep reinforcement learning for scheduling in cellular networks*, in Proc. 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2019, pp. 1–6.
- [27] C. Xu, J. Wang, T. Yu, C. Kong, Y. Huangfu, R. Li, Y. Ge, and J. Wang, *Buffer-aware wireless scheduling based on deep reinforcement learning*, in Proc. 2020 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2020, pp. 1–6.

- [28] L. Liang, H. Ye, and G. Y. Li, *Spectrum sharing in vehicular networks based on multiagent reinforcement learning*, IEEE Journal on Selected Areas in Communications, vol. 37, no. 10, pp. 2282–2292, 2019.
- [29] C. Fan, B. Li, C. Zhao, W. Guo, and Y.-C. Liang, *Learning-based spectrum sharing and spatial reuse in mm-wave ultradense networks*, IEEE Transactions on Vehicular Technology, vol. 67, no. 6, pp. 4954–4968, 2017.
- [30] G. Alnwaimi, S. Vahid, and K. Moessner, *Dynamic heterogeneous learning games for opportunistic access in LTE-based macro/femtocell deployments*, IEEE Transactions on Wireless Communications, vol. 14, no. 4, pp. 2294–2308, 2014.
- [31] L. Xiao, H. Zhang, Y. Xiao, X. Wan, S. Liu, L.-C. Wang, and H. V. Poor, *Reinforcement learning-based downlink interference control for ultra-dense small cells*, IEEE Transactions on Wireless Communications, vol. 19, no. 1, pp. 423–434, 2019.
- [32] F. Meng, P. Chen, L. Wu, and J. Cheng, *Power allocation in multi-user cellular networks: Deep reinforcement learning approaches*, IEEE Transactions on Wireless Communications, 2020.
- [33] G. Caire, G. Taricco, and E. Biglieri, *Bit-interleaved coded modulation*, IEEE Transactions on Information Theory, vol. 44, no. 3, pp. 927–946, 1998.
- [34] C. E. Shannon and W. Weaver, *The mathematical theory of communication*. University of Illinois Press, 1964.
- [35] N. Tishby, F. C. Pereira, and W. Bialek, *The information bottleneck method*, arXiv preprint physics/0004057, 2000.

Глава 20

Интеграция наземной и неназемной связи

20.1. Почему нужна интеграция наземной и неназемной связи

Интеграция неназемной сети (non-terrestrial network, NTN) в наземную сеть беспроводного доступа будет играть важную роль в достижении глобального покрытия, обеспечивая высокую степень доступности и надежности для сетей 6G даже в случае стихийных бедствий. Такая интеграция имеет важное значение для преодоления пробелов в покрытии недостаточно обслуживаемых территорий путем расширения покрытия сотовых сетей за счет неназемных узлов, что будет иметь ключевое значение для обеспечения доступа «в любом месте и в любое время» и предоставления услуг мобильной широкополосной связи для необслуживаемых или недостаточно обслуживаемых регионов в таких средах, как, например, океаны, горы, леса или другие отдаленные районы, где сложно развернуть наземные точки доступа или базовые станции. Благодаря широкому охвату неназемных узлов доставка услуг и приложений может быть расширена на большое количество новых пользователей, расположенных либо в городских, либо в удаленных районах [1].

В дополнение к очевидному преимуществу от расширения охвата потребителей интеграция наземных и неназемных сетей может предоставить новый набор услуг и приложений, таких как безразрывное соединение, дистанционное сканирование, пассивное сканирование и определение местоположения, навигация, отслеживание и автономная доставка. Это требует наличия единой структуры сети, в которой узлы неназемной сети (например, спутниковые группировки с межспутниковыми линиями связи), сетевые узлы на борту БПЛА и HAPS, а также наземные сетевые узлы могут в равной степени рассматриваться как базовые станции с точки зрения их функциональных возможностей.

Пользовательские устройства могут беспрепятственно получать доступ к наземным и неназемным базовым станциям. В качестве альтернативы неназемные сетевые узлы, такие как спутниковые группировки с межспутниковыми линиями связи, могут рассматриваться как узлы интегрированного доступа и обратной связи (integrated access and backhaul, IAB) с системной глубокой интеграцией, централизованной координацией, возможностями связи, кешированием или мобильными граничными вычислениями (mobile edge computing, MEC).

Для спутников или сетевых узлов HAPS инфраструктура узлов передачи всегда доступна, поэтому услуги по запросу могут быть доставлены потребителю путем освещения заданного региона точечными лучами со спутников, HAPS или БПЛА, что повышает надежность и эластичность всей сети. Такое развертывание основано на механизме обслуживания и требует эффективной координации узлов передачи как в наземных, так и в неназемных сетях.

Для HAPS по требованию и для временных HAPS улучшенная IAB может предложить решение по уплотнению сети по запросу. БПЛА и низковысотные HAPS, которые иногда могут работать как ретрансляторы с использованием технологии IAB, способны обеспечивать краткосрочное транзитное соединение без необходимости в плотной наземной инфраструктуре. Технология IAB также имеет решающее значение для обеспечения возможности подключения летательных аппаратов, таких как авиалайнеры и грузовые беспилотные летательные аппараты, как в качестве воздушных пользователей, так и в качестве воздушных базовых станций, тем самым обеспечивая доступ в сеть удаленным пользователям или пользователям и базовым станциям в перегруженных частях наземной сети. Очевидно, что предотвращение перегрузки как в наземных, так и в неназемных сетях при одновременном подавлении взаимных помех при наличии узлов IAB требует тесной интеграции между различными подсистемами.

Точное позиционирование является одним из ключевых требований для сетей 6G, и его соблюдение может быть облегчено в рамках интегрированной наземной и неназемной сети. Например, широкое покрытие и особенности LOS неназемной сети могут способствовать лучшему приему опорных сигналов определения местоположения. В такой среде пользователи могут принимать сигналы LOS от многих базовых станций для лучшего определения местоположения. Кроме того, можно измерять отражения сигналов спутникового вещания для пассивного обнаружения и определения местоположения близлежащих объектов.

Чтобы еще больше увеличить бюджет канала для доступа к неназемной сети, такие средства, как мощные ресурсы бортовой обработки, большие антенные решетки для генерации узких лучей, передовые

решения радиointерфейса и развертывание по требованию плотных неназемных TRP, должны быть объединены для борьбы с большими потерями на трассе и мобильностью в контексте спутниковой связи.

20.2. Обзор существующих решений

Консорциум 3GPP изучил различные сценарии и технологии, позволяющие интегрировать неназемные компоненты в новую радиосвязь 5G [1]. Такие исследования предлагают разные варианты для каждого из сценариев использования 5G, то есть eMBB, mMTC и URLLC. В зависимости от базового сценария преимущества каждого предполагаемого варианта использования могут быть наилучшим образом достигнуты путем интеграции различного набора неназемных компонентов, таких как БПЛА, HAPS и спутники на разных орбитах.

В табл. 20.1 приведены основные характеристики (высота и задержка распространения) и возможные варианты использования различных компонентов неназемной сети [1].

Таблица 20.1. Характеристики и возможные варианты использования для различных неназемных компонентов

Тип NTN	Высота	Задержка распространения	Возможные применения
Спутник GEO	35 786 км	120 мс	Широкое покрытие (eMBB, mMTC) Средний охват (eMBB) Вещание публичного контента (eMBB) Обратный канал фиксированной и мобильной связи Связь между отдаленными пользователями
Спутник LEO	400–1600 км	1,3–5 мс	Обратный канал фиксированной и мобильной связи Связь между отдаленными пользователями Вещание медиаконтента Услуги IoT с широким охватом
Спутник VLEO	100–400 км	0,33–1,3 мс	Обратный канал фиксированной и мобильной связи Связь между отдаленными пользователями Вещание медиаконтента Услуги IoT с широким охватом Широкополосный интернет

Тип NTN	Высота	Задержка распространения	Возможные применения
HAPS	15–25 км	50–83 мкс	Обратный канал воздушных и наземных станций Связь между отдаленными пользователями Вещание медиаконтента IoT с локальным подключением
UAV	0,1–10 км	0,33–33 мкс	Обратный канал наземных станций Воздушный ретранслятор / TRP Точка доступа по запросу Региональные аварийно-спасательные службы

Хотя обычные спутники на геостационарной околоземной орбите (GEO) хорошо подходят для трансляции общедоступного и популярного контента (например, мультимедийного контента, сообщений по безопасности движения и/или обновлений для подключенных автомобилей) на локальный сервер, они могут быть не очень полезны для сценариев применения, чувствительных к задержкам. Трех спутников GEO, развернутых на высоте 35 786 км над экватором и относительно неподвижных по отношению к поверхности Земли, достаточно для обеспечения глобального покрытия (за исключением полярных регионов).

Спутники на низкой околоземной орбите (LEO) могут обеспечить лучший баланс между широким покрытием и задержкой распространения/потерями на трассе. Спутники GEO и LEO, покрывающие большую территорию, могут использоваться для обеспечения непрерывного обслуживания подвижных сот на наземных, морских или авиационных платформах без необходимости переключения на другую соту или луч в пределах сотен километров. Они также могут служить транзитным каналом для фиксированных сот, особенно в удаленных районах. Хотя спутники LEO могут быть недоступны для наземных пользователей из-за чрезмерных потерь на трассе, предполагается, что по мере развития технологии спутниковых антенн LEO пользовательское оборудование в конечном итоге получит прямой доступ к неназемной сети 6G. В качестве альтернативы оборудование на борту летательных аппаратов, таких как БПЛА или HAPS, может служить в качестве концентратора для обеспечения неназемного доступа удаленных пользователей, у которых отсутствует спутниковое соединение с использованием технологии IAB.

Благодаря последним достижениям в области космической техники стало возможным стабилизировать спутники на орбитах VLEO (между

100–400 км), что намного ниже, чем у обычных спутников LEO. В сочетании с другими технологическими возможностями, которые существенно снижают затраты на запуск спутников на более низкие орбиты, такие разработки позволяют нам создавать большие группировки спутников VLEO для приложений мобильного широкополосного доступа. меньшая высота этих спутников обеспечивает лучший бюджет линии связи, а также меньшую задержку связи для наземных пользователей. Кроме того, меньшая высота над уровнем моря облегчает уменьшение зоны охвата луча, тем самым повышая возможность повторного использования частоты в фиксированном покрытии. Например, SpaceX планирует запустить группировку из примерно 12 000 спутников LEO и VLEO, чтобы к 2027 году обеспечить глобальный доступ в интернет.

Летательные аппараты с промежуточной высотой размещения, такие как HAPS, имеют преимущество в том, что они обеспечивают радиointерфейс с меньшими потерями на трассе для пользователей с ограниченным бюджетом мощности, в то же время используя функцию HAPS LOS и другие преимущества (например, большие антенные решетки) для эффективной связи со спутниками или наземными станциями в транспортной сети. HAPS могут обеспечить превосходный компромисс между задержкой распространения / потерями на тракте и покрытием для базовых станций в удаленных районах, а также могут служить в качестве магистрали (обеспечивая возможности связи и вычислений) для удаленных базовых станций или станций в городских районах, где наземная сеть перегружена в часы пик [2, 3]. В [4] авторы предложили использовать HAPS в качестве супермакробазовых станций для дополнения покрытия наземной сети в городских регионах, а также для других целей. К потенциальным применениям относятся интернет вещей, интеллектуальные транспортные системы, беспилотные летательные аппараты высокой грузоподъемности, предоставление услуг по запросу во время чрезвычайных ситуаций и других непредсказуемых событий и частичное уменьшение вычислительной нагрузки, что обеспечивается за счет предоставления HAPS возможностей хранения и вычислений в дополнение к инфраструктуре связи. В частности, некоторые исследователи предлагают архитектуру, в которой HAPS работает как промежуточный узел между воздушными и наземными узлами и узлами спутникового мегасозвездия в интегрированной вертикальной гетерогенной сети. Они также предлагают интеграцию реконфигурируемых интеллектуальных поверхностей в полезную нагрузку HAPS, тем самым обеспечивая энергоэффективную обратную связь для воздушных и наземных базовых станций.

Платформы с меньшей высотой размещения, такие как БПЛА, по-прежнему могут использоваться для обеспечения плотного развер-

тывания, поскольку их зона действия может быть ограничена локальной областью. Гибкость настройки траектории БПЛА также обеспечивает дополнительную степень свободы для развертывания [5]. Простота развертывания делает их эффективным решением для создания точек доступа по запросу (например, для освещения определенного события или спортивной игры) или для формирования специальной сети для служб экстренного реагирования.

Благодаря прогрессу в области бортовых антенн, усилителей мощности, больших антенных решеток и возможностей обработки сигналов наземные платформы воздушного и космического базирования значительно улучшились, особенно с точки зрения их способности предоставлять услуги с высокой пропускной способностью.

Хотя пропускная способность неназемных узлов передачи может быть сопоставима, а в некоторых случаях даже превосходить проектную цель базовых станций 5G, их емкость зоны все еще весьма ограничена по сравнению с достижимой емкостью зоны наземной сети. Эффективность использования спектра на канальном уровне в неназемной сети остается низкой из-за ограниченного бюджета канала.

Следующие технические ограничения были определены как критически важные направления для улучшения качества обслуживания в неназемной сети:

- **ограниченная эффективность использования спектра:** благодаря развитию возможностей бортовой обработки данных технология уменьшения помех в совмещенном канале может использоваться и в спутниковой связи. Применяемые в настоящее время схемы повторного использования полос частот, обычно обеспечивающие коэффициенты повторного использования частот от 3 до 4, не подходят для достижения высокой общей эффективности использования спектра. Усовершенствованные схемы передачи, такие как совместная передача с участием нескольких спутников, могут повысить пропускную способность пользователей; однако такие передовые схемы передачи требуют дальнейшего изучения и рассмотрения и еще не готовы к практическому применению;
- **отсутствие адаптивного покрытия по запросу:** из-за отсутствия координации планирования между спутниками и ограничений стоимости бортовых антенн радиопокрытие неназемных платформ обычно либо фиксировано, либо планируется заранее, независимо от текущих требований к услуге. Поэтому пока не удастся в полной мере воспользоваться преимуществами от предоставления адаптивных услуг по запросу через спутниковую связь;

- **накладные расходы на мобильность и управление лучом:** обработка сигналов на высоком уровне и потребление энергии во время процедур поддержания мобильности из-за быстрого движения спутников приведут к чрезмерной нагрузке на систему доставки сигнала, особенно с учетом потенциально огромного количества мобильных пользователей в сетях 6G. В остронаправленных лучевых каналах связи накладные расходы на управление лучом могут стать недопустимо высокими для спутников VLEO из-за их орбитального движения, что делает процедуры перемещения луча и восстановления после потери луча, а также отслеживания движения пользовательского оборудования еще более трудными для реализации. Кроме того, такие процедуры далеки от оптимальных с точки зрения задержки. Следовательно, оптимизация процедур поддержания мобильности и управления лучом является важной областью исследований для сетей 6G;
- **отсутствие поддержки сканирования:** поддержка сканирования в настоящее время недоступна в сетях NR; однако ожидается, что сканирование будет играть более заметную роль в сетях 6G, где сканирование на основе наземных и неназемных сетей может работать вместе для улучшения пользовательского опыта;
- **собственная задержка:** большинство существующих неназемных платформ не оборудованы бортовыми базовыми станциями. Развертывание низковысотных платформ, таких как спутники VLEO, HAPS и беспилотные летательные аппараты, с полным набором возможностей обработки сигналов на борту и стабильными линиями межспутниковой связи должно обеспечить основу для обеспечения передачи с малой задержкой;
- **отсутствие тесной интеграции между наземными и неназемными сетями:** интеграция неназемных сетей в NR R-16 является дополнением к NR R-15. Философия проектирования наземно ориентированных систем NR по своей сути не оптимизирована для бортового радиочастотного оборудования и обеспечения мобильности платформы неназемной сети. Широко признано, что для интеграции неназемных компонентов в наземную сеть все еще необходимы улучшения в неназемной сети, открывающие доступ к более широкому спектру услуг и применений, основанных на полностью интегрированной наземной и неназемной сети. Однако очевидно, что эти услуги и применения не могут быть эффективно реализованы той или иной сетью по отдельности. Следовательно, требуется тесная интеграция совместно работающих наземных и неназемных сетей. Обратите внимание, что этот тип интеграции отличается от подхода к интегрированной

неназемной сети, рассмотренного в 3GPP в контексте 5G NR R-16 и более поздних версий [1], где интеграция направлена только на улучшение радиоинтерфейса NR для использования неназемных сетей в качестве расширения наземных сетей, а не на достижение тесной интеграции между наземными и неназемными сетями.

20.3. Ожидания от новой системы и потенциальные направления исследований

20.3.1. Интегрированная многоуровневая сеть

Следствием интеграции неназемных компонентов в систему наземной связи является гетерогенная сеть, состоящая из нескольких уровней, как показано на рис. 20.1. Основная цель проектирования интеграции наземных и неназемных сетей состоит в том, чтобы достичь лучших общих характеристик за счет эффективной совместной работы нескольких каналов, более гибкого совместного использования функциональных возможностей и более быстрого переключения каналов физического уровня между наземными и неназемными сетями.

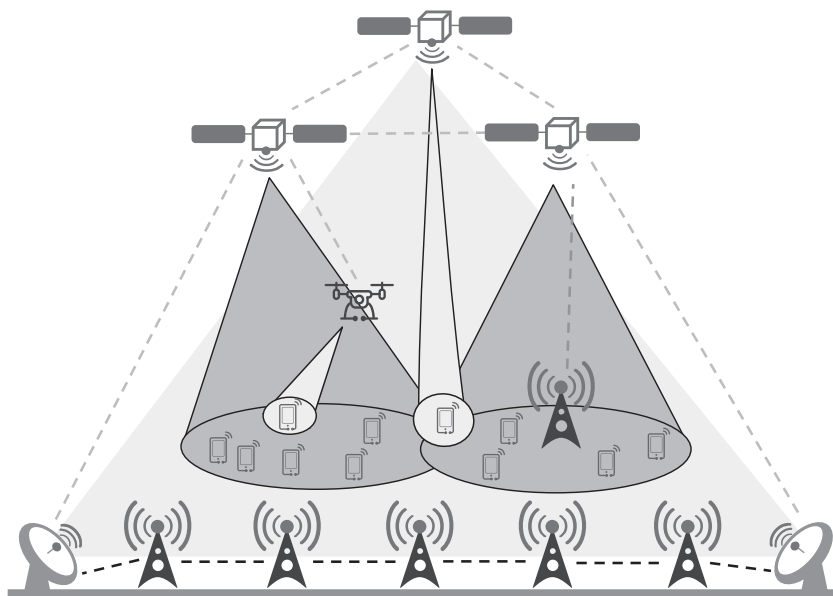


Рис. 20.1. Интегрированная наземная и неназемная сеть
(серые пунктирные линии обозначают беспроводные транзитные соединения, а черные пунктирные линии – оптоволоконные соединения)

Неоднородность и разнообразие в такой среде могут потенциально повысить надежность и покрытие сети, предоставляя пользовате-

лям несколько соединений на разных уровнях интегрированной сети (включая спутники на разных орбитах, HAPS и другие воздушные или наземные точки доступа). Однако стоит отметить, что каждый слой по отдельности не может достичь желаемой производительности. Например, хотя конкретное пользовательское оборудование или IoT-устройство может осуществлять доступ к нескольким наземным базовым станциям, спутникам VLEO или даже HAPS, каждый канал доступа не обязательно будет надежным на протяжении всего времени работы. В результате критически важна возможность быстрого выбора и переключения на лучший из имеющихся каналов доступа, а для повышения надежности необходимо гибко использовать разные пути, что требует тесной интеграции различных уровней. В 5G NR наземные и неназемные системы предназначены для работы по отдельности. Однако в сетях 6G будущего ожидается полная интеграция функций и операций, а также управления ресурсами и мобильностью обеих систем. В следующих разделах мы обсудим различные подходы к интеграции и ожидаемые улучшения в каждом случае.

Различные подходы к интеграции

Интеграция на уровне приложений – это простейшая форма нестрогой интеграции. Чтобы различные подсистемы могли сосуществовать друг с другом, каждая подсистема реализует все функциональные возможности своего уровня и интерфейс прикладного уровня, чтобы позволить пользователю обмениваться информацией с сервером приложений и от него через несколько соединений между разными подсистемами. Этот подход легко реализовать, но он может оказаться неэффективным из-за отсутствия координации между различными подсистемами [6].

Интеграция на уровне базовых сетей – это еще один нестрогий подход, который обычно применяется для взаимодействия между сотовыми сетями 3GPP (такими как LTE) и сетями, не относящимися к 3GPP (такими как WLAN). В этом подходе радиодоступ обеих сетей работает раздельно, но между базовыми сетями существует определенное взаимодействие. Пользователям предоставляется некоторая информация о сетях и конфигурации, относящаяся к политикам операторов [6]. Затем пользователи выбирают одну из RAN для пересылки пакетов на основе конфигурации сети, качества радиолинии, предпочтений пользователя и т. д. Таким образом, пользователи устанавливают соединения через лучшую RAN. Однако решение пользователя основано на локальных сигнальных механизмах или измерениях, что приводит к неоптимальному выбору. Кроме того, ресурсы радиодоступа могут использоваться неэффективно, поскольку каждая часть ресурсов отдельно управляется другой RAN.

Другое альтернативное решение – *тесная интеграция на основе RAN*, при которой разные подсистемы сходятся на уровне RAN, возможно, совместно используя одни и те же ресурсы [6]. Примером интеграции на основе RAN является решение 3GPP для установления двойного/множественного подключения через разные технологии радиодоступа (например, LTE и NR), которые могут совместно использовать полосу пропускания от одного и того же оператора. В зависимости от транзитных соединений конвергенция здесь может варьироваться от частичной координации, такой как решения со множественным подключением LTE/NR, до полноценной интеграции, которая соответствует единому решению технологии радиодоступа (RAT) с использованием единого радиointерфейса между наземными и неназемными сетями. При частичной координации каждая подсистема обеспечивает свой вклад в устранение помех между различными TRP или подсистемами, тогда как в решении с одной RAT все TRP наземных и неназемных сетей координируются одним и тем же централизованным узлом управления (control unit, CU) в каждом регионе, а региональные узлы связаны между собой через интерфейс транзитного соединения с высокой пропускной способностью. Используя унифицированный уровень управления, сеть совместно управляет радиоресурсами наземных и неназемных сетей, адаптируя параметры физического уровня унифицированного радиointерфейса на основе мгновенных условий канала, тем самым эффективно используя ресурсы, одновременно повышая надежность и качество обслуживания для различных конечных устройств.

Наземные и неназемные сети 5G по существу сводятся в одну структуру: из-за прямой совместимости 5G NR интеграция неназемных сетей выполняется после завершения спецификаций наземных сетей 5G.

Напротив, мы предполагаем, что в сети 6G будет изначально единая конструкция E2E, поскольку наземные и неназемные сети будут разрабатываться как единое целое с первого дня. Наземные узлы и наземные сети, а также неназемные узлы и неназемные сети – это просто разные реализации одной системы. Следовательно, после развертывания созвездий LEO сеть 6G сможет полностью покрыть землю. Наземные сети будут обеспечивать более высокие скорости передачи данных для конечных пользователей, а неназемные сети будут обеспечивать базовые скорости передачи данных в местах, где нет наземных сетей. В этой архитектуре 6G можно использовать несколько авиационных или космических уровней для дополнения покрытия наземной сети, обеспечивая дополнительное покрытие по запросу в городских регионах. В частности, может оказаться нерентабельным покрывать спорадические (как по времени, так и по местоположению) пиковые потребности пользователей за счет дорогостоящего развертывания уплотненной

наземной инфраструктуры. Используя широкое покрытие неназемных сетей, можно эпизодически предоставлять неназемные ресурсы тем частям сетей, которые отчаянно нуждаются в дополнительном покрытии. Использование неназемных ресурсов должно совместно координироваться посредством тесной интеграции различных уровней в многоуровневой сети 6G, в которой ресурсы разных TRP (на разных уровнях) управляются одним и тем же объектом уровня управления.

Проблемы и потенциальные направления исследований

Последние несколько десятилетий беспроводные сети в основном состояли из статических наземных точек доступа. Однако, учитывая распространенность БПЛА, HAPS и спутников VLEO и желание интегрировать спутниковую связь в сотовые сети, можно обоснованно предположить, что перспективные системы больше не будут горизонтальными и двумерными. Возникающие трехмерные вертикальные сети состоят из множества подвижных и высотных точек доступа (кроме геостационарных спутников), таких как БПЛА, HAPS и спутники VLEO, как показано на рис. 20.2.

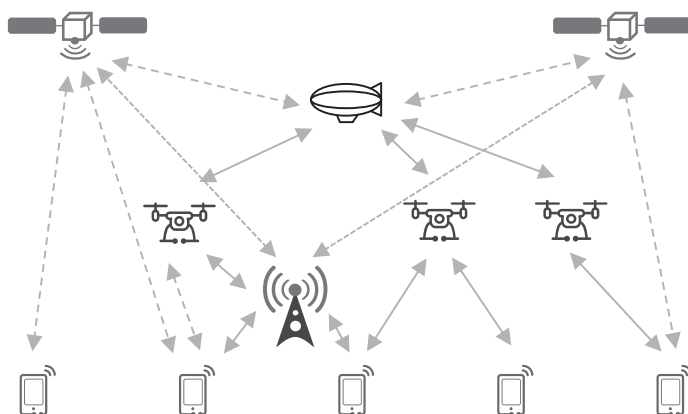


Рис. 20.2. Дроны, HAPS и спутники VLEO – новые компоненты интегрированной сети 6G

Новой задачей для 6G является поддержка разнообразных и разнородных точек доступа, а отсюда вытекает потребность в самоорганизации для беспрепятственной интеграции в сеть нового БПЛА или оказавшегося в зоне видимости низкоорбитального спутника без необходимости перенастраивать пользователей. Из-за их относительной близости к земле БПЛА, HAPS и спутники VLEO могут выполнять функции, аналогичные наземным базовым станциям, и, таким образом, могут рассматриваться как базовые станции нового типа в эпоху 6G, хотя и создают новый набор проблем, которые нужно будет преодолеть. Они могут ис-

пользовать тот же радиointерфейс и полосы частот, что и в системах наземной связи, но при этом требуется новый подход для планирования соты, получения соты и передачи обслуживания между неназемными узлами доступа или между наземными и неназемными узлами доступа. Более того, как и их наземные аналоги, неназемные узлы и их клиенты нуждаются в адаптивной и динамической беспроводной транзитной сети для поддержания связи. Поддержка таких разнообразных и разнородных точек доступа требует от системы способности к самоорганизации для беспрепятственного включения нового БПЛА в сеть без необходимости перенастройки с высокими накладными расходами. Такие решения (основанные на виртуализированном радиointерфейсе) должны упростить получение сот и TRP, а также маршрутизацию данных и управления, чтобы эффективно и беспрепятственно интегрировать эти воздушные узлы с базовой наземной сетью. Следовательно, добавление и удаление воздушных точек доступа должно быть в значительной степени прозрачным для конечных пользователей, не говоря уже об операциях физического уровня, таких как синхронизация UL/DL, формирование диаграммы направленности, измерение канала и обратная связь, связанных с точками доступа.

Наземные и неназемные сети 6G нацелены на совместное использование единого устройства физического уровня и MAC, чтобы один и тот же модемный чип, оснащенный интегрированным стеком протоколов, мог поддерживать как наземную, так и неназемную связь. Несомненно, общий набор микросхем имеет смысл с точки зрения затрат, но его довольно сложно изготовить на практике из-за различных требований к конструкции со стороны наземных и неназемных сетей, на которые могут влиять такие факторы, как параметры сигнала физического уровня, форма волны и модуляция. Например, системы спутниковой связи могут иметь строгие требования PAPR. Хотя нумерология NR оптимизирована для связи с малой задержкой, спутниковая связь должна быть способна выдерживать длительную задержку передачи. В следующем разделе будет предоставлено более подробное обсуждение потенциальных улучшений для неназемной связи. Ожидается, что достижения всего нескольких параметров будет достаточно для гибкого приведения структуры физического уровня/MAC в соответствие различным сценариям развертывания с нативной поддержкой наземной связи с воздуха или из космоса.

20.3.2. Улучшенная неназемная связь

Хотя неназемные компоненты, такие как беспилотные летательные аппараты, HAPS и спутники VLEO, будут частью сетевой инфраструктуры 6G, в основном обеспечивая функциональные возможности, ана-

логичные наземным базовым станциям, все еще есть возможности для улучшений с точки зрения конструкции неназемных узлов, чтобы преодолеть строгие требования к бюджету канала, особенно для высотных платформ. Действительно, с учетом ожидаемого прогресса в разработке встроенных радиочастотных модулей и возможностей обработки у исследователей будет достаточный простор для реализации потенциальных прорывов в разработке радиоинтерфейса 6G. При этом следует ориентироваться на два общих принципа проектирования:

- преодоление проблем, связанных с неназемной связью;
- использование свойств, характерных для неназемных узлов.

Возвращаясь к ограничениям в существующей неназемной связи, о которых говорилось в разделе 20.2, ниже мы рассмотрим соответствующие решения и потенциальные направления исследований для повышения эффективности неназемных сетей.

Технология передачи с высокой эффективностью спектра

Общая эффективность использования спектра существующих спутниковых систем намного ниже, чем у сотовых сетей, отчасти из-за неэффективного бюджета линии связи, вызванного большими расстояниями передачи, а отчасти из-за сильных помех в спутниковой связи. Последний фактор сохранит свое влияние, несмотря на развитие возможностей обработки на борту и/или улучшение качества приемника пользователя. В настоящее время применяется многократное использование частоты для уменьшения помех в совмещенном канале от соседних лучей за счет низкой эффективности использования спектра. Многолучевое предварительное кодирование – это очень зрелый и эффективный метод, используемый в сотовых сетях для уменьшения помех в совмещенном канале и достижения полного повторного использования частот в контексте спутниковой связи, улучшающий общий пользовательский опыт в интегрированной наземной и неназемной сети. Для реализации эффективного сбора информации о канале требуются схемы обратной связи по каналу с минимальными издержками. Стратегия предварительного кодирования, которая учитывает ограничения RF и другие ограничения развертывания, требует дальнейшего изучения.

Поляризационное мультиплексирование – еще одна многообещающая спутниковая технология, которая может повысить эффективность использования спектра, учитывая, что спутники обычно оснащены антеннами с круговой поляризацией. Благодаря свойствам канала LOS сигналы с правой круговой поляризацией (right-hand circular polarization, RHCP) и левой круговой поляризацией (left-hand circular polarization,

ЛНСП) могут быть невосприимчивыми к кросс-поляризации, несмотря на большое расстояние передачи. Поляризация обеспечивает новое ортогональное измерение в дополнение к традиционным измерениям, таким как время и частота, и требует дальнейшего изучения при проектировании неназемных сетевых систем. Используя преимущество хорошей развязки разных поляризаций в контексте спутниковой связи, соседние точечные лучи могут использовать одну и ту же полосу частот с разными поляризациями, чтобы избежать межканальных помех. Другие усовершенствованные схемы передачи, такие как ММО [7] и пространственная модуляция [8], также могут быть применены к двум изолированным каналам поляризации для дальнейшего повышения эффективности передачи.

Совместная передача через несколько спутников – еще одно многообещающее, хотя и сложное решение для повышения эффективности передачи. Ожидается, что фактическая скорость передачи увеличится, если пользователь сможет принимать сигналы от нескольких спутников одновременно или если несколько спутников могут совместно обнаруживать сигналы восходящей линии связи от одного пользователя, используя методы распределенной системы ММО. Последний вариант будет чрезвычайно полезен при устранении узких мест в бюджете канала связи из-за ограниченной мощности передачи пользователя в восходящем канале. Пропускная способность канала такой распределенной схемы ММО должна быть оптимизирована с точки зрения топологии и бюджета канала согласованных спутников. Для группировок LEO требуются дополнительные усилия для решения проблем синхронизации и приема сигналов с/от различных спутников.

Умное покрытие по запросу

Сеть на основе созвездий должна быть интеллектуальной, чтобы обеспечивать гибкое покрытие по запросу (по требованию), как показано на рис. 20.3. Предполагая, что у нас есть идеальные транзитные соединения между компонентами неназемной сети, инфраструктуру, развернутую в небе, можно рассматривать как пул ресурсов, и все ресурсы, содержащиеся в ней, можно планировать одновременно, чтобы повысить общую эффективность использования ресурсов. Сеть должна знать потребности пользователей и динамически координировать доступные ресурсы для удовлетворения этих потребностей, управляя распределением временных/частотных ресурсов и развертывая узконаправленные точечные лучи. В случае внезапного спроса, возникающего в определенной зоне «горячей точки», все ресурсы видимого небосвода должны быть направлены на обслуживание этой зоны, чтобы повысить плотность подключения региона. Точно так же, когда один из спутниковых узлов выходит из строя, другие спутники должны иметь

возможность немедленно заполнить пробел в зоне покрытия, вызванный отказавшим узлом, чтобы уменьшить внезапную потерю пропускной способности.

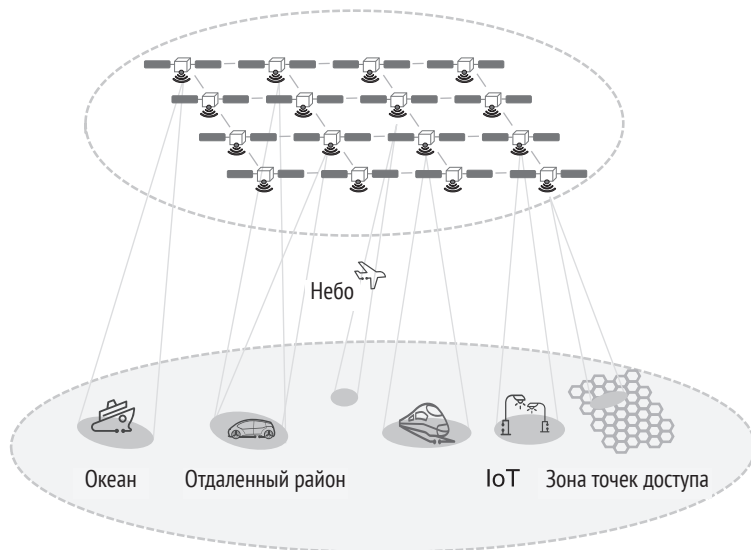


Рис. 20.3. Спутниковое покрытие по запросу

Картина спутниковых лучей должна динамически корректироваться в соответствии с изменяющимися условиями помех по мере того, как спутники движутся по своим орбитам. Поскольку требования к качеству предоставляемой спутником услуги в данный момент времени могут значительно отличаться от области к области, а воспринимаемая спутником структура требований будет время от времени меняться из-за движения спутника, планировщик ресурсов должен соответствующим образом адаптироваться, чтобы перебросить достаточное количество частотных, временных и энергетических ресурсов из слабо нагруженных областей на обслуживание востребованных точек доступа. При этом для пользователей в слабо загруженных областях по-прежнему сохраняется отклик базовой станции, и схема распределения ресурсов может обновляться в реальном времени, когда зона обслуживания спутника перемещается в новую область. Алгоритм планирования должен учитывать любые ограничения, касающиеся возможностей бортового оборудования и требований к услугам, связанным с геолокацией.

Эффективное отслеживание перемещений

Спутниковая сеть должна обеспечивать эффективное отслеживание перемещений, чтобы справляться с частым переключением точечных лучей, вызванным движением спутников, не относящихся к GEO, тем

самым уменьшая объем служебных данных, сокращая время прерывания и снижая энергопотребление.

Широко применяемая в наземных сетях стратегия срабатывания механизмов повторного выбора и передачи обслуживания (хэндовера) на основе мощности сигнала не будет работать в контексте спутниковой связи, так как здесь не работает «эффект ближнего-дальнего». В результате требуются новые показатели, относящиеся к относительному положению пользователей и спутниковых точечных лучей. Процедуры хэндовера и повторного выбора также должны быть оптимизированы с учетом особенностей конкретных сценариев перехода, включая переключение внутри спутника, между спутниками и между спутником и сотой. Поскольку движение спутника преобладает над относительными изменениями положения между спутником и пользователями, мы можем уменьшить накладные расходы на обслуживание сигнального механизма, вызванные переключением, прогнозируя это движение.

Переключение спутникового точечного луча может привести к сбою соединения в случаях, когда ресурсы нового точечного луча недоступны. Чтобы гарантировать непрерывность обслуживания, следует использовать механизм резервирования ресурсов, способный гарантировать пользователям доступность требуемых ресурсов после переключения. Таким образом, процедурам мобильности в неназемных сетях могут помочь наземные сети, и наоборот.

Высокая точность и быстрое позиционирование

Мегасозвездия на низкой околоземной орбите могут обеспечить встроенную возможность определения местоположения, чтобы уменьшить зависимость от внешней GNSS на стороне пользовательского оборудования, а также обеспечить еще более удобное взаимодействие с пользователем. Значения геометрического снижения точности (geometric dilution of precision, GDOP) мегасозвездия будут значительно улучшены по сравнению с существующими GNSS [9]. Поскольку данный подход изначально обеспечивает аналогичную точность, можно предположить, что позиционирование на основе мегасозвездий позволит снизить требования к точности бортовых генераторов опорной частоты и допустить увеличение орбитальных ошибок из-за возмущений. За счет использования усовершенствованных схем позиционирования в сочетании с системой связи может быть достигнуто быстрое и точное измерение времени прибытия (TOA) и/или частоты прибытия (FOA). Например, планирование лучей позиционирования в выделенных слотах позволяет избежать помех во время передачи данных. Однако планирование лучей позиционирования и опорных сигналов для достижения этой цели при одновременном снижении накладных расходов является проблемой, которую еще предстоит решить.

20.4. Источники

- [1] 3GPP, *Study on new radio (NR) to support non-terrestrial networks*, 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.811, 10 2019, version 15.2.0. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3234>.
- [2] O. Kodheli, E. Lagunas, N. Maturo, S. K. Sharma, B. Shankar, J. Montoya, J. Duncan, D. Spano, S. Chatzinotas, S. Kisseleff et al., *Satellite communications in the new space era: A survey and future challenges*, arXiv preprint arXiv:2002.08811, 2020.
- [3] P. Wang, J. Zhang, X. Zhang, Z. Yan, B. G. Evans, and W. Wang, *Convergence of satellite and terrestrial networks: A comprehensive survey*, IEEE Access, vol. 8, pp. 5550–5588, 2019.
- [4] G. Kurt, M. G. Khoshkholgh, S. Alfattani, A. Ibrahim, T. S. Darwish, M. S. Alam, H. Yanikomeroglu, and A. Yongacoglu, *A vision and framework for the high altitude platform station (HAPS) networks of the future*, arXiv preprint arXiv:2007.15088, 2020.
- [5] H. Wang, H. Zhao, W. Wu, J. Xiong, D. Ma, and J. Wei, *Deployment algorithms of flying base stations: 5G and beyond with UAVs*, IEEE Internet of Things Journal, vol. 6, no. 6, pp. 10 009–10 027, 2019.
- [6] S. Andreev, M. Gerasimenko, O. Galinina, Y. Koucheryavy, N. Himayat, S.-P. Yeh, and S. Talwar, *Intelligent access network selection in converged multi-radio heterogeneous networks*, IEEE Wireless Communications, vol. 21, no. 6, pp. 86–96, 2014.
- [7] A. Byman, A. Hultkonen, P.-D. Arapoglou, M. Bertinelli, and R. De Gaudenzi, *MIMO for mobile satellite digital broadcasting: From theory to practice*, IEEE Transactions on Vehicular Technology, vol. 65, no. 7, pp. 4839–4853, 2015.
- [8] P. Henarejos and A. I. Pérez-Neira, *Dual polarized modulation and reception for next generation mobile satellite communications*, IEEE Transactions on Communications, vol. 63, no. 10, pp. 3803–3812, 2015.
- [9] T. G. Reid, B. Chan, A. Goel, K. Gunning, B. Manning, J. Martin, A. Neish, A. Perkins, and P. Tarantino, *Satellite navigation for the age of autonomy*, in Proc. 2020 IEEE/ION Position, Location and Navigation Symposium (PLANS). IEEE, 2020, pp. 342–352.

Глава 21

Интеграция сканирования и связи

21.1. Почему нужна интеграция сканирования и связи

Сотовые сети изначально были разработаны для беспроводной связи, и быстро растущий спрос на применения, связанные с определением местоположения, привлек значительное внимание к исследованиям позиционирования в сотовых сетях. Как описано в главе 3, некоторые из наиболее интригующих сценариев использования 6G включают построение картины окружающей среды посредством высокоточного позиционирования, отображения и реконструкции, а также распознавания жестов/действий. *Сканирование* будет новой услугой 6G, которую можно охарактеризовать как процесс получения информации об окружающей среде. Этот процесс состоит из операций, которые можно условно отнести к одной из двух следующих категорий:

- **радиочастотное сканирование:** фактически это отправка радиосигнала и изучение окружающей среды путем приема и обработки отраженных сигналов;
- **визуальное сканирование (визуализация):** использование изображений и видео, полученных из окружающей среды (например, с помощью камеры) для построения пространственной картины.

Посылая электромагнитную волну и получая эхо, радиочастотное сканирование может извлекать информацию об объектах в окружающей среде, например о существовании, текстуре, расстоянии, скорости, форме и ориентации. В современных системах радиочастотное сканирование сводится к классическому радару, который используется для обнаружения, локализации и отслеживания пассивных объектов, то есть объектов, которые не зарегистрированы в сети. Существо-

ющие системы радиочастотного сканирования имеют два основных ограничения:

- они автономны и ориентированы на конкретное применение, т. е. они не взаимодействуют с другими радиочастотными системами;
- они нацелены только на пассивные объекты и не могут использовать отличительные признаки активных объектов, то есть объектов, зарегистрированных в сети.

Сегодня основная цель проектирования беспроводной сети – оптимизация производительности связи, что включает в себя повышение спектральной эффективности и надежности при минимизации задержки и энергопотребления. Таким образом, сейчас как никогда необходимы более информативные, эффективные и гибкие системы связи, которые улучшают качество обслуживания и охват за счет разнообразных сценариев использования, равно как и сенсорные системы, которые предоставляют необходимые знания об окружающей среде. Традиционная практика включает в себя две разные подсистемы, которые обмениваются ограниченным объемом информации для достижения определенного уровня познания, но этот подход имеет много недостатков, таких как большие накладные расходы, большой объем данных, неэффективность и громоздкость подсистем. Альтернативный подход основан на том факте, что в будущих беспроводных системах ожидается значительный сдвиг в сторону более высоких частот (таких как миллиметровые волны и даже ТГц-диапазон) с более широким доступным спектром. Это позволит системам связи обладать функциональными характеристиками радиолокационных систем. Следовательно, чтобы уменьшить как мощность, так и форм-фактор, некоторые аппаратные компоненты могут использоваться совместно двумя типами систем, например антенны, усилители мощности и генераторы. Кроме того, для повышения эффективности эти системы могут использовать такие совместные ресурсы, как время и спектр.

По мере появления новых технологий будут продолжаться усовершенствования в отношении высоких скоростей передачи данных, низкой задержки и широких возможностей соединения в системах беспроводной связи. Это откроет пользователям доступ к другим функциям системы, таким как сканирование/визуализация и локализация, тем самым представив множество инновационных применений и повысив производительность будущей беспроводной системы. В интегрированной системе сканирования и связи (integrated sensing and communication, ISAC) сканирование и связь будут двумя взаимодополняющими функциями. Следуя этой тенденции, позиционирование в 5G (которое ограничено активными устройствами) будет расширено за

счет дополнительных функций обнаружения. Для этого придется ввести новые метрики (такие как точность и разрешающая способность), чтобы они служили новыми ключевыми показателями эффективности в 6G, заменяя KPI точности позиционирования в 5G. Как описано в главе 3, в настоящее время предлагаются ключевые показатели эффективности, основанные на различных сценариях применения, точность которых может составлять от 1 до 10 см, а разрешение – до 1 мм. Эти цели зависят от сопутствующих технологий, которые мы обсудим далее.

21.2. Обзор существующих решений

Проект партнерства 3-го поколения (3rd Generation Partnership Project, 3GPP) работает над интеграцией связи и позиционирования со времен GSM [1], и недавно было запланировано «усовершенствование позиционирования для удовлетворения требований к высокой точности и задержке в промышленных условиях внутри помещений» в качестве функции в 5G NR, документ R17 [2]. Однако, как и в глобальной навигационной спутниковой системе, функция определения местоположения в 5G NR предназначена только для активных объектов (на основе устройств), т. е. устройства обмениваются сигналами с несколькими сетевыми узлами, а затем оценивают информацию о местоположении локально на устройстве или удаленно через сетевой объект. Однако существует множество сценариев, в которых требуется обнаружение и локализация пассивных объектов, например сканирование окружающей среды, распознавание жестов и наблюдение за зонами ограниченного доступа.

Обнаружение пассивных объектов относится к сфере традиционных исследований в области радиолокации. На протяжении многих лет радар использовался во многих областях, включая управление движением в аэропортах и морских портах, дистанционное сканирование земной поверхности, высокоточное обнаружение небольших деформаций поверхности, измерения обезлесения, а также мониторинг вулканов и землетрясений. Совсем недавно тенденция изменилась, и теперь сюда добавились круиз-контроль и предотвращение столкновений, а также мониторинг состояния здоровья по сердцебиению, дыханию и движению голосовых связок. Поскольку подобные сенсоры используются повсеместно, сценарии, в которых применяется радар, превзошли все ожидания с того момента, когда он был впервые разработан и спроектирован, например Google Soli [3]. Хотя 5G NR изначально не поддерживает радиолокационное сканирование с точки зрения конструкции системы, проводится значительное количество исследований по использованию существующих сотовых сигналов для моностатического или бистатического обнаружения целей [4, 5]. Например, [6] использует

формы сигналов нисходящей линии связи в LTE и NR в качестве сигналов освещения радара для целей сканирования и исследует проблему полнодуплексного режима в моностатической конфигурации.

Как правило, существуют ограничения, связанные с прямым применением физического сигнала в традиционной сотовой системе (включая 5G NR) для сканирования. Во-первых, хотя опорные сигналы, используемые для оценки канала и отслеживания фазы в 5G NR, потенциально могут быть повторно применены для сканирования, система не располагает достаточными ресурсами по части спектра, времени или пространства для выполнения этой операции с высокой точностью и разрешением [7]. Это связано с тем, что спектр является очень ограниченным ресурсом с точки зрения связи, а опорные сигналы тщательно разрабатываются, чтобы минимизировать накладные расходы. Кроме того, общение не требует очень точной оценки окружающей среды. Во-вторых, пассивные бистатические или мультистатические радиолокационные системы, использующие сотовые сигналы, имеют ограничения с точки зрения проектирования системы. Причина этого в основном заключается в том, что пассивному когерентному радиолокационному приемнику требуется отдельная направленная антенна (прямой канал) для приема сигналов прямой видимости в качестве эталона синхронизации для когерентного детектирования [8]. Тем не менее в действительности во многих случаях такой опорный сигнал LOS не существует из-за шума, помех и эффектов замирания. Эта проблема еще больше усугубляется тем фактом, что различные сигналы излучаются со стороны базовой станции посредством формирования луча (цифрового и/или аналогового), направляемого пользователям в разных направлениях, и поэтому принятый отраженный сигнал может и не быть задержанной и ослабленной версией опорного сигнала LOS. В результате когерентное детектирование нецелесообразно, а пассивный радар работает непредсказуемо. Кроме того, хотя сигнал прямой видимости обычно полезен для связи, крайне важно подавить сильный сигнал прямой видимости, просачивающийся в канал наблюдения в пассивной радиолокационной системе, чтобы извлечь слабые эхо-сигналы от целей. Это делает требования к динамическому диапазону пассивных радиолокационных приемников намного более жесткими, чем для приемников связи. И последнее, но не менее важное: в пассивной мультистатической конфигурации нет взаимодействия между узлами Tx и Rx, что значительно затрудняет уменьшение помех и проведение когерентной обработки.

Достигаемое при помощи сканирования формирование изображений с высоким разрешением с использованием радиочастотных сигналов (6G-визуализация) позволяет получить больше информации об объекте

и предоставить базовые данные для классификации и распознавания объектов [9]. Визуализацию можно по существу смоделировать как обратную задачу электромагнитного рассеяния [10]. В основе визуализации лежит использование электромагнитных волн для облучения цели и восстановление информации о цели путем сбора рассеянных эхо-сигналов. Исследования в области визуализации можно разделить на две категории: 2D-визуализация и 3D-томография. Двумерное изображение используется для хорошо проводящих целей, а трехмерная дифракционная томография применяется для диэлектрических тел, как показано на рис. 21.1. В традиционных системах визуализации расстояние между передатчиком и целью обычно велико; поэтому модель 2D-изображения может быть упрощена как линейная обратная задача в дальней зоне [11]. В трехмерной томографии необходимо решать нелинейные обратные задачи для получения количественной информации, такой как диэлектрическая проницаемость и проводимость, которые, в свою очередь, необходимо преобразовать в результат визуализации. По сравнению с 2D-визуализацией, томография более сложна, поскольку эффект многократного рассеяния приводит к нелинейной проблеме визуализации [12]. В последние годы исследователи предложили метод визуализации с высоким разрешением на основе множественных наблюдений с использованием пространственно-временного случайного поля излучения [13]. В 2013 году была предложена когерентная модель визуализации, основанная на теории восприятия компрессии при туберкулезе кости и вычислении корреляции функции Грина [14], в которой позиционирование цели достигается с помощью специально разработанной антенны из метаматериала. С развитием метаматериалов и настраиваемых материалов точность этой системы визуализации может быть улучшена для достижения более высокого разрешения.

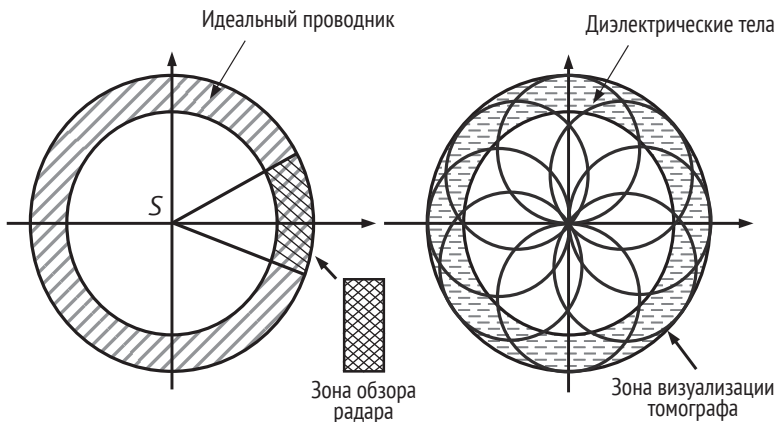


Рис. 21.1. Наблюдение с помощью радарной визуализации и томографии

В сетях 6G сканирование можно рассматривать как услугу, параллельную связи, поскольку она может работать и приносить доход независимо от основного назначения сети. Это означает, что система должна быть спроектирована так, чтобы легко переходить от простого обмена данными к совмещению сканирования и обмена данными, таким образом удовлетворяя KPI для обеих услуг.

Уровень взаимодействия между системами сканирования и связи можно разделить на следующие категории [15]:

- **существование:** две системы рассматривают друг друга как источники помех и поэтому не обмениваются никакой информацией между собой;
- **сотрудничество:** две системы разрабатываются отдельно, однако они обмениваются информацией между собой с целью уменьшения межсистемных помех;
- **совместное/интегрированное проектирование:** две системы спроектированы так, чтобы вести себя как единая унифицированная система.

Большинство исследований, касающихся интегрированного проектирования, в основном сосредоточены на совместной форме волны, о чем пойдет речь в разделе 22.2.4. Основная проблема этого подхода возникает из-за того, что связь и сканирование имеют противоречивые KPI. В частности, разработки в области связи в основном направлены на максимальное повышение спектральной эффективности, тогда как оптимальная форма сигнала для сканирования нацелена на точные оценки и высокое разрешение. Иными словами, когда требуется только оценка дальности, оптимальной формой сигнала считывания является дельта-подобная функция автокорреляции во временной области наряду с высокоэффективной обработкой, позволяющей оценить параметр при наличии сильного шума или интерференции.

Хотя характеристика формы сигнала является важным аспектом ISAC, она не охватывает всей картины. Некоторые исследовательские работы предлагают объединение схем сканирования и связи, выходящее за рамки формы сигнала. В качестве примеров можно отметить, что в [16] была предложена суперпозиция сигналов радара и связи в области мощности, а в [17] был предложен совместный метод в области мощности и пространственной области. Однако почти вся существующая литература сосредоточена на аспектах интегрированного проектирования на уровне каналов и не включает рассмотрение проблемы на уровне системы. На наш взгляд, проектирование системной архитектуры должно гарантировать, что полная интеграция систем связи и сканирования полностью реализует ожидаемые преимущества той и другой услуги.

21.3. Ожидания от новой сети и потенциальные направления исследований

Заглядывая в будущее, мы ожидаем, что 6G будет работать в более высоких частотных диапазонах с более широкой полосой пропускания (например, на терагерцовых частотах), а сверхмассивные антенные решетки станут более доступными. Это предоставит нам уникальную возможность расширить область применения сотовых сетей от чистой связи до комбинированной услуги связи и сканирования. На наш взгляд, разработка сети 6G должна естественным образом прийти к интегрированному решению для связи и сканирования, в отличие от существующего решения, в котором основные компоненты (включая 5G NR, GNSS и радарные системы) работают независимо, как показано на рис. 21.2. Кроме того, по мере развития новых технологий и достижений 6G (включая внедрение метаматериалов, реконфигурируемых интеллектуальных антенн и искусственного интеллекта) это интегрированное решение будет еще более оправданным, особенно с учетом того, что эти технологии являются ключевыми движущими силами для обеих систем.

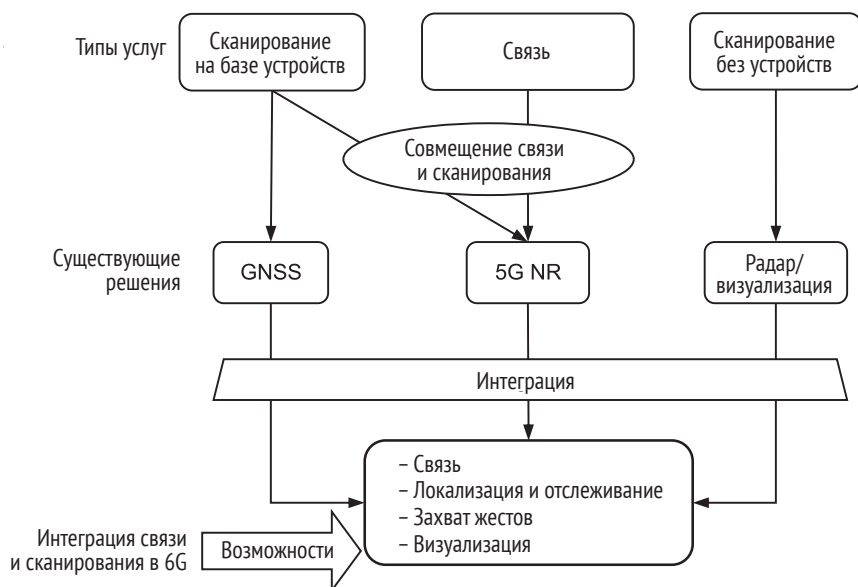


Рис. 21.2. Обзор предлагаемых решений ISAC

Как и любая другая новая технология, ISAC имеет как возможности, так и проблемы. Чтобы раскрыть огромный потенциал этой технологии, надо сначала найти ответы на следующие вопросы:

- как могут сосуществовать модули связи и сканирования в интегрированном решении на системном уровне и как конструкция на системном уровне может обеспечить более эффективное сосуществование?
- какую выгоду интегрированное решение принесет услугам связи и сканирования? В частности, какие проблемы в существующих системах связи и сканирования решит интеграция и какие новые вызовы это повлечет?

Ответы на эти вопросы помогают нам определить новые потенциальные направления исследований, обсуждаемые в следующих разделах.

21.3.1. Аспекты проектирования интегрированной системы сканирования/связи

С помощью рис. 21.3 можно дать простое объяснение интеграции связи и сканирования.

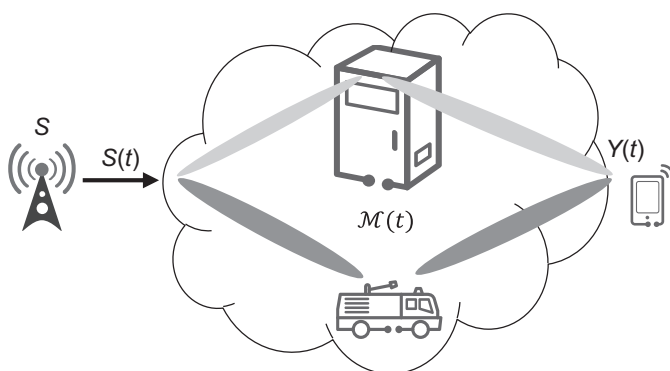


Рис. 21.3. Связь и сканирование на одной платформе

На схеме показан радиочастотный источник S , передающий сигнал $S(t)$, который распространяется через среду связи $\mathcal{M}$ и принимается или обрабатывается одним или несколькими приемниками. При обмене данными радиочастотный сигнал $S(t)$ сначала подвергается обработке в источнике, который встраивает в него информацию (d_s). Теоретические границы скорости передачи хорошо понятны и могут быть выражены как предел Шеннона $I(S(t); Y(t) | \mathcal{M}(t))$, где $Y(t)$ обозначает принятый сигнал на предполагаемом приемнике, а $I(x; y|z)$ – взаимную информацию между переменными x и y данной переменной z . В предыдущей формулировке предполагается, что все переменные являются функциями времени (t). Что касается операции сканирования, в радиочастотный сигнал вносит изменения среда связи, а не источник сигнала. Другими словами, мы можем рассматривать сканирование как особый метод

связи, в котором информация (d_m), встроенная в среду связи, передается приемнику (приемникам) посредством радиочастотного сигнала, который может излучать третья сторона. Как и в общении, мы можем определить ограничение скорости сканирования в виде $I(Y(t); \mathcal{M}(t)|S(t))$ [18]. Иными словами, ключевое отличие сканирования от традиционной связи заключается в том, совмещены ли источник информации и источник радиочастотного сигнала.

Этот подход предполагает, что сканирование и связь могут быть представлены как общая задача связи, в которой излучаемым сигналом можно управлять с помощью как исходных данных (d_s), так и информации о среде (d_m). Это отправная точка для определения структуры будущей сети ISAC, в которой узлы могут выполнять обе функции со следующими преимуществами:

- **более эффективное использование ресурсов:** интегрированная конструкция обеспечивает эффективное совместное использование ресурсов, а не консервативное их разделение (например, FDMA и TDMA);
- **уменьшение помех между сигналами сканирования и связи:** само собой разумеется, что основным преимуществом интегрированной конструкции будет предотвращение или уменьшение помех между двумя системами, когда они совместно используют ресурсы во временной, частотной или пространственной области;
- **лучшее качество связи:** интеграция позволяет улучшить качество связи по двум направлениям. Первое предполагает подстройку связи под среду с помощью более эффективных методов формирования диаграммы направленности и уменьшения помех благодаря полученной информации. Второе подразумевает более гибкую и предсказуемую связь, которая работает в оптимальных условиях с использованием информации о текущем и будущем статусах канала. Этот подход основан на знании характеристик окружающей среды и прогнозировании изменений в ней. Фактически, зная среду связи, мы можем узнать основную причину изменений в информации о состоянии канала, не нуждаясь в наблюдении за ней;
- **более эффективное сканирование:** сеть ISAC обеспечивает более эффективное сканирование, выполняемое по запросу. Другими словами, сканирование больше не является только услугой, управляемой приложением, но теперь может запускаться и по запросу от другого сетевого узла;
- **меньшая потребляемая мощность:** если учесть, что большая часть электроэнергии в высокочастотных системах рассеивает-

ся на аналоговом интерфейсе (например, усилители мощности и АЦП/ЦАП), мощность может использоваться гораздо более эффективно, если системы хотя бы частично интегрированы.

Основная проблема при анализе системы ISAC – нахождение пределов быстродействия. Одной из первых попыток в этом направлении исследований была работа [19], в которой была предложена структура ISAC в одноцелевой системе. Одним из важных выводов в этой работе является то, что, поскольку полная энергия передаваемого источником сигнала фиксирована, необходимо найти компромисс между отраженными и поглощенными волнами. Предлагаемая структура ISAC является очень ограниченным примером в том смысле, что учитывается только моностатическое сканирование, то есть передатчик и приемник сигнала сканирования расположены вместе. Другим хорошо известным примером ISAC в парадигме когерентной связи является процесс «оценки канала», в котором некоторые «известные» сигналы, называемые опорными сигналами или пилот-сигналами, передаются вместе с сигналом, несущим информацию, и приемники сигнала связи и сигнала сканирования представляют собой одно и то же устройство. В этом случае на передаваемый радиочастотный сигнал влияют как исходная информация, так и канал распространения. Однако это пример ISAC, в котором информация сканирования (о канале распространения) не имеет ценности сама по себе и используется только для того, чтобы помочь приемнику декодировать исходные данные. В целом же в рамках ISAC сканирование рассматривается как отдельная услуга, которая не обязательно используется для улучшения связи. Даже если сканирование должно улучшить качество связи, полезность сканирования может выходить за рамки узкого назначения.

В будущей системе 6G будут по-прежнему использоваться передовые технологии для дальнейшего повышения производительности сети мобильной связи. Некоторые из этих технологий, имеющих ключевое значение для системы 6G и охватывающих оконечные устройства и сетевую инфраструктуру, открывают перед нами следующие перспективы:

- более широкий и высокочастотный спектр с большей полосой пропускания;
- усовершенствованная конструкция антенны с очень большими решетками и метаповерхностью;
- более согласованная работа базовых станций и пользовательского оборудования;
- передовые методы подавления помех;
- интегрированная усовершенствованная обработка сигналов и ИИ.

Существует пять потенциальных уровней интеграции ISAC. Первые два уровня касаются совместного использования доступных физических ресурсов (включая спектр), оборудования и радиочастотных цепей. На третьем уровне рассматривается совместное использование ресурсов и возможностей цифровой обработки сигналов, таких как алгоритмы и модули обработки на физическом уровне. Переход на следующий уровень интеграции может потребовать разработки интерфейса протокола для обеспечения межуровневого, межмодульного и межузлового обмена информацией. Наконец, наиболее эффективный и идеальный уровень предполагает совместное использование всех доступных ресурсов и информации для повышения производительности обеих служб. Таким образом, в сети 6G могут сосуществовать разные уровни интеграции в соответствии с конкретными сценариями использования, ключевыми показателями эффективности и расходами на реализацию. Например, для транспортных средств с экстремальными требованиями к задержке связи и вероятности распознавания или обнаружения объекта могут потребоваться все пять уровней интеграции.

Хотя эта многоуровневая структура может показаться простой в реализации из-за очевидной схожести между услугами связи и сканирования, для реализации такой интегрированной системы необходимо решить следующие основные проблемы:

- **разные возможности узлов:** по сравнению с 5G в будущих беспроводных сетях 6G будет еще больше беспроводных узлов с разнообразным набором возможностей в плане обработки сигналов, памяти, полосы пропускания и радиочастотного тракта. Это усложнит интеграцию по сравнению с обычной связью, поскольку параметры узла имеют гораздо большее влияние на производительность сканирования;
- **полудуплексные узлы:** хотя полнодуплексные приемопередатчики уже прототипированы в существующих системах с приемлемой производительностью и, как ожидается, будут коммерциализированы в ближайшем будущем, нам еще предстоит пройти долгий путь, прежде чем в сетях будут присутствовать только полнодуплексные узлы. Поскольку в будущих сетях останется большое количество полудуплексных устройств, полудуплексные узлы затруднят сканирование, особенно в моностатическом режиме, когда передатчики и приемники расположены рядом;
- **ограниченное покрытие сканирования:** одна из основных проблем сканирования в сотовых сетях заключается в том, что дальность сканирования увеличивается как корень четвертой степени из мощности передачи, а не второй, как в случае связи.

Следующие направления исследований приблизят нас к интеграции связи и сканирования:

- **анализ производительности обобщенных сетей ISAC:** имеющиеся выводы о фундаментальных границах производительности, полученные для простой структуры, представленной в [19], было бы желательно распространить на более общие варианты многоузловых многоцелевых сетей с мультистатистическими конфигурациями сканирования и найти ответы на следующие вопросы: (1) каков наилучший компромисс между характеристиками сканирования и пропускной способностью канала связи; (2) как разработать практические схемы, приближающие к такому компромиссу, например формы сигналов, таблицы кодирования, структуры и протоколы. Одна попытка исследований в этом направлении представлена в [20], где объекты сканирования рассматриваются как виртуальные приемники энергии, а затем ISAC моделируется как задача распределения ресурсов информации и энергии по беспроводному каналу. С точки зрения теории информации, сканируемые объекты – это радиорелейные станции, которые принимают сканирующий сигнал определенной формы и пересылают его обратно передатчику со своей собственной информацией о параметрах среды, встроенной в эховолну. Необходимы дополнительные подходы и идеи, чтобы получить хорошее представление о границах возможностей сетей ISAC в различных сценариях;
- **конструкция сети радиодоступа для ISAC:** наиболее важная проблема, которую необходимо решить, заключается в том, как эта интеграция влияет на конструкцию сети радиодоступа на различных уровнях. На физическом уровне нам понадобится следующее:
 - ❖ архитектура сети, обеспечивающая гибкое и здоровое сосуществование между сигналами связи и сканирования, а также соответствующими конфигурациями, тем самым гарантируя, что рабочие характеристики систем связи и сканирования не будут ухудшены;
 - ❖ общесистемные решения для совместного использования сканирующих возможностей различных узлов, включая сетевые узлы и пользовательские устройства;
 - ❖ механизмы передачи сигналов, которые подразумевают взаимодействие между сетевыми объектами и обеспечивают возможность проектирования и настройки связанных параметров;

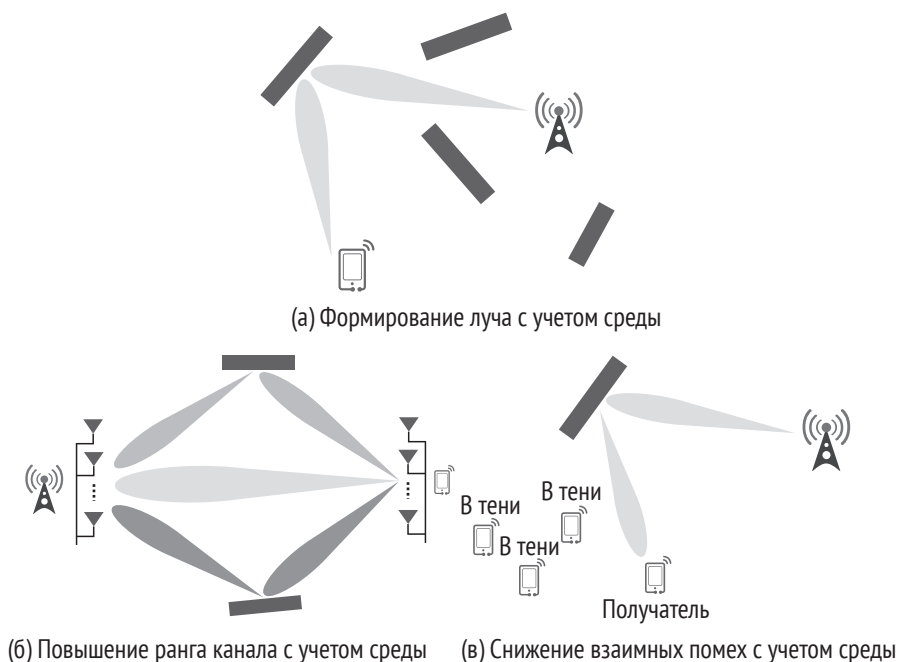


Рис. 21.4. Примеры содействия связи с помощью сканирования

- **связь при содействии сканирования:** хотя в будущем сканирование будет представлено как отдельная услуга, все же полезно исследовать, как информацию, полученную с помощью сканирования, можно использовать в связи. Самым тривиальным преимуществом сканирования будет определение характеристик окружающей среды, что позволяет организовать связь по более детерминированным и предсказуемым каналам благодаря знаниям о среде распространения. Некоторые примеры содействия услуге связи за счет сканирования показаны на рис. 21.4, который демонстрирует, как знания об окружающей среде, полученные с помощью сканирования, могут улучшить связь. Первый пример иллюстрирует, как знания об окружающей среде используются для оптимизации формирования луча в направлении оконечного устройства (формирование луча с учетом среды), а второй пример показывает, как эти знания могут быть использованы для применения всех потенциальных степеней свободы в канале распространения (повышение ранга канала с учетом среды). Последний пример показывает, как осведомленность о среде распространения помогает уменьшить или смягчить взаимные помехи между оконечными устройствами. Степень, в которой сканирование приносит пользу связи, не

должна ограничиваться улучшением пропускной способности и уменьшением помех. Отдельное интересное направление исследований можно было бы посвятить расширению функционального охвата связи с помощью данных сканирования и изучению того, как функции, которые обычно реализуются системой связи, могут выполняться подмодулем сканирования. Это приведет к огромной экономии как с точки зрения накладных расходов, так и с точки зрения задержки;

- **связь, основанная на сканировании:** другой взгляд на парадигму конструкции ISAC подразумевает использование альтернативного подхода к связи посредством сканирования. Как говорилось ранее, связь и сканирование можно рассматривать как общую платформу, в которой на радиосигнал от источника в другом месте накладываются данные о среде, а затем он принимается третьей стороной. При таком подходе нет необходимости ограничиваться совместным размещением источников данных и радиосигнала, и это будет особенно полезно в сценариях, где для сбора данных разворачиваются устройства с ограниченными возможностями обработки (большинство устройств IoT в будущих системах). Вместо передачи собранных данных устройства могут манипулировать радиочастотным сигналом, передаваемым другим источником в радиочастотной области (в отличие от основной полосы частот), для экономии значительного количества энергии. Этот тип связи, совмещенный с функцией сканирования, называется *связью с обратным рассеянием* [21]. Другой пример – *связь под влиянием среды* [22], когда характеристики среды прохождения сигнала намеренно изменяются для передачи информации;
- **сканирование при поддержке связи:** платформа связи позволяет нам достигать более эффективного и интеллектуального сканирования, соединяя сканирующие узлы между собой. В сети подключенных пользователей может быть реализовано сканирование по запросу, то есть сканирование может быть выполнено на основе запроса другого узла или делегировано другому узлу. Кроме того, опираясь на поддержку связи, можно организовать совместное сканирование, при котором несколько узлов сканирования получают информацию об окружающей среде. Все эти расширенные функции требуют тщательно спроектированной RAN, чтобы обеспечить связь между сканирующими узлами по каналам DL, UL и SL с прицелом на минимальные издержки и максимальную эффективность сканирования, что является интересной темой исследования;

- **позиционирование с помощью сканирования:** *активная локализация*, также называемая позиционированием, включает в себя локализацию пользовательского оборудования посредством передачи или приема сигналов к нему или от него. Главное преимущество этого метода – простота в эксплуатации. Хотя точное знание местоположения пользовательского оборудования чрезвычайно высоко ценится, его трудно получить из-за многих факторов, включая многолучевое распространение, несовершенную синхронизацию времени/частоты, ограниченные возможности выборки/обработки на стороне пользователя и ограниченный динамический диапазон оконечного устройства. С другой стороны, *пассивная локализация* включает в себя получение информации о местоположении активных или пассивных объектов путем обработки эхо-сигналов переданного сигнала в одном или нескольких местоположениях. По сравнению с активной локализацией пассивная локализация через сканирование имеет следующие явные преимущества:

- ❖ она помогает в выявлении каналов LOS и уменьшении остаточного смещения NLOS;
- ❖ на нее гораздо меньше влияют ошибки синхронизации между устройством пользователя и сетью;
- ❖ она может улучшить разрешение и точность позиционирования в случаях, когда полоса пропускания локализации ограничена целевым устройством пользователя.

Принимая во внимание эти моменты, потенциальные исследования можно бы сосредоточить на том, как пассивная локализация путем сканирования устраняет недостатки активной локализации. Иными словами, вопрос в том, как мы можем наслаждаться лучшим из обоих миров. Что касается пассивной локализации, стоит упомянуть *проблему сопоставления* (matching problem). Она связана с тем, что полученные эхо-сигналы не имеют уникальной сигнатуры, позволяющей однозначно сопоставить их с объектами (и их скрытыми переменными местоположения), от которых они отражаются. Это полностью отличается от локализации с использованием активных сигналов или маяков, когда сигнатура, соответствующая маяку или ориентиру, однозначно идентифицирует связанные с ними объекты. Поэтому нам нужны продвинутые решения, чтобы связать наблюдения с местоположением активных устройств, что существенно повысит точность и разрешение активной локализации.

21.3.2. Конструкция и алгоритмы оборудования радиочастотного сканирования

В этом разделе мы сначала углубимся в технические проблемы и потенциальные направления исследований в области радиочастотного сканирования, а затем рассмотрим конкретные технические проблемы, связанные со сканированием на основе визуализации.

По сравнению с 5G использование гораздо более широкой полосы пропускания и ресурсов антенной решетки в системе 6G ISAC сулит нам как новые возможности, так и новые проблемы. Наиболее очевидным преимуществом с точки зрения сканирования является улучшенное разрешение и точность, а основные проблемы можно резюмировать следующим образом:

- **фрагментация ресурсов:** являясь подсистемой 6G, радиочастотное сканирование не может одновременно использовать всю полосу пропускания, временной интервал и антенные элементы, поэтому ресурсы, выделенные для сканирования, будут фрагментарно разбросаны по разным областям. Следовательно, одна из основных проблем связана с эффективным формированием сигнала и координацией распределения ресурсов во временной, частотной и пространственной областях без ухудшения спектральной эффективности и производительности сканирования;
- **высокая сложность реализации:** что касается передачи данных в сканировании 6G, большая полоса пропускания сопровождается очень высокими частотами дискретизации, широкополосными сигналами и нагруженными каналами обработки. Это порождает серьезные проблемы в виде сложности вычислений и высокого энергопотребления;
- **сниженное покрытие системы:** поскольку в 6G рабочая частота стремится к более высоким пределам, покрытие системы уменьшается при сохранении количества базовых станций и общей потребляемой мощности. Кроме того, NLOS-характер распространения сигнала усугубляет проблему покрытия из-за суровых условий окружающей среды (например, внутри помещений и в городских условиях). Следовательно, даже несмотря на то, что высокая частота может потенциально обеспечить высокое разрешение и высокую точность сканирования, системе может не хватать узлов для постоянного предоставления высококачественных услуг;
- **трудность реализации высокого разрешения:** в 6G сканирование можно использовать в сценариях, где требуется высокое раз-

решение. В то же время реализация высокого разрешения ставит перед разработчиками такие задачи, как сканирование в ближнем поле с помощью больших антенных решеток, разработка специальных радиосигналов для сокращения времени сканирования, сканирование объектов со сложной внутренней структурой и т. д.

Для решения этих проблем мы можем предложить несколько перспективных направлений исследований:

- **совместная разработка сигналов для радиочастотного сканирования:** подсистемы связи и сканирования в системе 6G будут совместно использовать ресурсы во временной, частотной и пространственной областях. Следовательно, важно координировать эти ресурсы, чтобы удовлетворить технические требования обеих подсистем. Фактически сканирование может не занимать всю полосу пропускания, временные интервалы и элементы антенной решетки, что означает, что мы можем уменьшить выделение ресурсов для сканирования за счет разреженного распределения. Учитывая, что объект сканирования или параметры окружающей среды обычно не изменяются резко из-за физических ограничений реального мира (т. е. канал не изменяется внезапно за период когерентности), для разделения частотно-временного подпространства можно использовать предыдущие показатели. В результате мгновенная полоса пропускания каждого временного интервала будет только дискретным сегментом полной пропускной способности, а разрешение составной постобработки будет таким же, как и производительность при полной пропускной способности. В системе 6G потребность в ценных ресурсах передачи может быть уменьшена за счет разреженной структуры сигнала сканирования – при условии сохранения сверхвысокого разрешения. Мы можем использовать сэкономленные ресурсы, чтобы гарантировать качество услуг связи, и, следовательно, стоит рассмотреть вопрос о разрежении сигналов сканирования в объединенном пространстве сигналов (частотно-временном пространстве) без ухудшения характеристик сканирования;
- **алгоритм сжатого радиочастотного сканирования:** сжатая выборка (compressed sensing, CS) и теория дискретизации представляют нам математическую основу для приема и обработки широкого класса аналоговых сигналов с частотой дискретизации ниже, чем критерий Найквиста. По своей сути CS включает в себя восстановление разреженных векторов большой размерности из довольно небольшого количества измеренных данных, и его можно легко расширить до разреженных сигналов на любом подходящем базисе.

Вдобавок ко всему CS может эффективно снизить вычислительную нагрузку. Хотя традиционные методы радиочастотного сканирования надежно оценивают целевые параметры, они требуют, чтобы сигнал дискретизировался с частотой не ниже критерия Найквиста. Например, они используют согласованную фильтрацию (matched filtering, MF) или сжатие импульсов, чтобы максимизировать отношение сигнал/шум (signal-to-noise ratio, SNR) в присутствии аддитивного белого гауссова шума. В некоторых случаях также могут использоваться другие фильтры для оптимизации различных показателей, таких как отношение пиковых и боковых лепестков (peak to sidelobe ratio, PSLR) и интегрированное отношение боковых лепестков (integrated sidelobe ratio, ISLR) [23–25]. Разрешение сканирования обратно пропорционально носителю функции неоднозначности, тем самым ограничивая возможность выполнения более точного разрешения для близко расположенных целей. В результате для достижения высокого разрешения по диапазону многие современные системы радиочастотного сканирования используют широкую полосу пропускания, обычно от нескольких сотен МГц и вплоть до десятков ГГц. Это требует высокоскоростных АЦП и приводит к значительным накладным расходам вычислительной мощности. Чтобы повысить возможности алгоритмов высокого разрешения с широкой полосой пропускания, было выдвинуто предложение о дискретизации сигнала и оценке параметров цели с частотой ниже Найквиста [26]. В дополнение к частотной области была также предложена аналогичная субдискретизация в доплеровской области [27]. Радиочастотные сканирующие устройства с антенными решетками сталкиваются с сопоставимыми проблемами дискретизации в пространственной области, и поэтому стоит изучить, как обработка по Найквисту с пространственной субдискретизацией может использоваться в устройствах с массивом ММО с меньшим количеством антенных элементов без снижения углового разрешения;

- **сканирование с высоким разрешением в ближнем поле:** при определенных характеристиках канала большое количество антенн может полностью задействовать путь распространения для передачи, тем самым увеличивая пропускную способность в канале связи. Кроме того, благодаря развитию высокочастотных технологий и микросхем антенны с большой решеткой сантиметрового уровня могут быть широко интегрированы и развернуты на высоких частотах в будущем, когда будет доступен спектр суб-ТГц- или ТГц-частот. После этого сканирование с высоким раз-

решением, основанное на антеннах с большой решеткой, будет рассматриваться как нативная функция системы ISAC. Кроме того, учитывая ограничения в отношении мощности передачи и потерь в тракте передачи на более высоких частотах (например, ТГц), было бы более естественно использовать функцию ISAC в пределах малого радиуса действия оконечного устройства. Например, оборудование пользователя или перчатки, упомянутые в главе 3, будут сканировать среду в пределах нескольких десятков метров или даже нескольких метров, что предположительно означает работу системы в области ближнего поля.

Обычные сканирующие системы в основном работают за счет излучения электромагнитного сигнала в свободном пространстве и приема эхо-сигналов от отражателей/рассеивателей. Сигнал, используемый для визуализации, может излучаться в ближнем или дальнем поле в зависимости от размера апертуры антенны D и рабочей длины волны передатчика. В табл. 21.1 показана зависимость расстояния ближней зоны от размера антенной решетки и частоты.

Таблица 21.1. Расстояние ближней зоны для различных антенн

Частота, ГГц	Расстояние ближнего поля (м), вычисленное как $2D^2$		
	256 элементов (16×16)	1024 элемента (32×32)	2048 элементов (45×45)
3,5	11	44	87
6	6	26	51
10	4	15	30
39	0,98	3,9	7,8
73	0,5	2,1	4,2
140	0,3	1,1	2,2

При визуализации в дальней зоне электромагнитные волны приходят как плоские волны. В то же время, если исходить из предположения, что мы работаем в дальнем поле, небольшая разница в расстоянии мало влияет на амплитуду эхо-сигналов, и, по сути, мы можем принять расстояние до каждой точки на цели примерно равным радарному расстоянию. Используя упрощенную формулу радиолокационного эхо-сигнала, мы можем восстановить изображение цели с помощью быстрого преобразования Фурье. Однако в случае ближнего поля разность фаз между центральной точкой и краем цели относительно велика; поэтому необходи-

мо использовать гипотезу сферической волны. Таким образом, в формуле визуализации появляется нелинейный фазовый член, и при уменьшении расстояния приближенная обработка не может использоваться для уменьшения вычислительной нагрузки;

- **одновременная активная и пассивная локализация** (concurrent active and passive localization, CAPL): проблема локализации может показаться устрашающей, если мы не знаем позиции пользователей, осуществляющих навигацию, и не располагаем сведениями о среде распространения сигнала. В то же время одно из фундаментальных открытий в области робототехники предлагает решения, которые одновременно локализуют сканирующего пользователя в неизвестном местоположении и итеративно создают карту окружающей среды. Эта парадигма получила название одновременного определения местоположения и отображения (simultaneous location and mapping, SLAM) и основана на предпосылке, что отображение и позиционирование не являются отдельными проблемами. Другими словами, если мы будем рассматривать их как независимые проблемы, то будут получены неоптимальные ответы.

Взаимосвязанный характер предыдущей проблемы был также математически доказан пионерами робототехники в 1990-х годах [28, 29]. Проблема связи процессов картирования/локализации также имеет логическое объяснение: неточная карта напрямую влияет на точность позиционирования сканирующего устройства, а неточное местоположение напрямую превращается в неточную карту. Следовательно, если бы сканирующее устройство было установлено в статической точке загроможденной сложной карты, а не на мобильном навигационном устройстве, измерения, вероятно, были бы такими же. Это также означает, что независимо от того, где установлен сканирующий модуль (на навигационном устройстве или в фиксированной точке), единственная проблема заключается в нахождении объектов, с которыми взаимодействует сенсорный сигнал. Это взаимодействие может быть отражением, приемом или передачей. Короче говоря, между пассивной и активной локализациями не только отсутствует концептуальная разница, но и эти две проблемы действительно связаны.

Все это говорит о том, что одновременные пассивная и активная локализации (CAPL) с использованием только портативных устройств в качестве сканеров сталкиваются с серьезными проблемами. К счастью, появление множества вариантов использования 6G, таких как беспилотные летательные аппараты, транспортные средства и устройства IoT, может облегчить реализацию

CARL. С помощью БПЛА, транспортных средств, устройств IoT и пользовательских устройств, помогающих сетям локализовать объекты и строить карту окружающей среды, сотовые системы могут создавать виртуальную среду в киберпространстве;

- **совместное многоузловое радиочастотное сканирование:** радиочастотное сканирование посредством совместной работы относится к узлам сканирования, которые делятся своими наблюдениями друг с другом и пытаются достичь общего консенсуса в визуализации окружающей среды, и было показано, что это значительно улучшает характеристики локализации [30]. Конкретный процесс включает в себя узлы-участники, формирующие динамическую опорную сетку посредством распределенной передачи и обработки. Такое взаимодействие снижает неопределенность измерений и обеспечивает больший охват, а также более высокую точность и разрешение измерения за счет объединения данных.

21.4. Источники

- [1] R. S. Campos, *Evolution of positioning techniques in cellular networks, from 2G to 4G*, Wireless Communications and Mobile Computing, vol. 2017, Article ID 2315036, 17 pages, 2017. [Online]. Available: <https://doi.org/10.1155/2017/2315036>.
- [2] B. Bertenyi, *5G in Release 17 – strong radio evolution*, Technical Report, 2019. <https://www.3gpp.org/news-events/2019-5g-in-release-17-%E2%80%93%E2%80%93strong-radio-evolution>.
- [3] J. Lien, *Soli radar-based perception and interaction in pixel 4*, Technical Report. <https://ai.googleblog.com/2020/03/soli-radar-based-perception-and.html>.
- [4] M. Bica, K.-W. Huang, V. Koivunen, and U. Mitra, *Mutual information based radar waveform design for joint radar and cellular communication systems*, in Proc. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2016, pp. 3671–3675.
- [5] M. Schmidhammer, S. Sand, M. Soliman, and F. de Ponte Muller, *5G signal design for road surveillance*, in Proc. 2017 14th Workshop on Positioning, Navigation and Communications (WPNC). IEEE, 2017, pp. 1–6.
- [6] C. B. Barneto, T. Riihonen, M. Turunen, L. Anttila, M. Fleischer, K. Stadius, J. Ryyänänen, and M. Valkama, *Full-duplex OFDM radar with LTE and 5G NR waveforms: Challenges, solutions, and measurements*, IEEE Transactions on Microwave Theory and Techniques, vol. 67, no. 10, pp. 4042–4054, 2019.

- [7] R. M. Rao, V. Marojevic, and J. H. Reed, *Probability of pilot interference in pulsed radar-cellular coexistence: Fundamental insights on demodulation and limited CSI feedback*, IEEE Communications Letters, vol. 24, no. 8, pp. 1678–1682, Aug. 2020.
- [8] D. E. Hack, L. K. Patton, B. Himed, and M. A. Saville, *Detection in passive MIMO radar networks*, IEEE Transactions on Signal Processing, vol. 62, no. 11, pp. 2999–3012, 2014.
- [9] J. Yan, X. Feng, and P. Huang, *High resolution range profile statistical property analysis of radar target*, in Proc. 6th International Conference on Signal Processing, 2002, vol. 2. IEEE, 2002, pp. 1469–1472.
- [10] M. Bertero and P. Boccacci, *Introduction to inverse problems in imaging*. CRC Press, 1998.
- [11] D. M. Sheen, D. L. McMakin, and T. E. Hall, *Three-dimensional millimeter-wave imaging for concealed weapon detection*, IEEE Transactions on Microwave Theory and Techniques, vol. 49, no. 9, pp. 1581–1592, 2001.
- [12] J. P. Guillet, B. Recur, L. Frederique, B. Bousquet, L. Canioni, I. Manek-Hönninger, P. Desbarats, and P. Mounaix, *Review of terahertz tomography techniques*, Journal of Infrared, Millimeter, and Terahertz Waves, vol. 35, no. 4, pp. 382–411, 2014.
- [13] Y. Guo, D. Wang, X. He, and B. Liu, *Super-resolution staring imaging radar based on stochastic radiation fields*, in Proc. 2012 IEEE MTT-S International Microwave Workshop Series on Millimeter Wave Wireless Technology and Applications. IEEE, 2012, pp. 1–4.
- [14] J. Hunt, T. Driscoll, A. Mrozack, G. Lipworth, M. Reynolds, D. Brady, and D. R. Smith, *Metamaterial apertures for computational imaging*, Science, vol. 339, no. 6117, pp. 310–313, 2013.
- [15] B. Paul, A. R. Chiriyath, and D. W. Bliss, *Survey of RF communications and sensing convergence research*, IEEE Access, vol. 5, pp. 252–270, 2016.
- [16] X. Zheng, T. Jiang, and W. Xue, *A composite method for improving the resolution of passive radar target recognition based on WiFi signals*, EURASIP Journal on Wireless Communications and Networking, vol. 2018, no. 1, p. 215, 2018.
- [17] J. A. Zhang, X. Huang, Y. J. Guo, J. Yuan, and R. W. Heath, *Multibeam for joint communication and radar sensing using steerable analog antenna arrays*, IEEE Transactions on Vehicular Technology, vol. 68, no. 1, pp. 671–685, 2018.
- [18] M. R. Bell, *Information theory and radar waveform design*, IEEE Transactions on Information Theory, vol. 39, no. 5, pp. 1578–1597, 1993.
- [19] M. Kobayashi, G. Caire, and G. Kramer, *Joint state sensing and communication: Optimal tradeoff for a memoryless case*, in Proc. 2018

- IEEE International Symposium on Information Theory (ISIT). IEEE, 2018, pp. 111–115.
- [20] F. Liu, C. Masouros, A. Petropulu, H. Griffiths, and L. Hanzo, *Joint radar and communication design: Applications, state-of-the-art, and the road ahead*, IEEE Transactions on Communications, vol. 68, no. 6, pp. 3834–3862, 2020.
 - [21] H. Stockman, *Communication by means of reflected power*, Proceedings of the IRE, vol. 36, no. 10, pp. 1196–1204, 1948.
 - [22] A. K. Khandani, *Media-based modulation: A new approach to wireless transmission*, in Proc. 2013 IEEE International Symposium on Information Theory. IEEE, 2013, pp. 3050–3054.
 - [23] N. Levanon, *Radar principles* (First Edition). John Wiley & Sons, 1988.
 - [24] J. E. Cilliers and J. C. Smit, *Pulse compression sidelobe reduction by minimization of $l_{\text{sub } p}$ -norms*, IEEE Transactions on Aerospace and Electronic Systems, vol. 43, no. 3, pp. 1238–1247, 2007.
 - [25] J. George, K. Mishra, C. Nguyen, and V. Chandrasekar, *Implementation of blind zone and range-velocity ambiguity mitigation for solid-state weather radar*, in Proc. 2010 IEEE Radar Conference. IEEE, 2010, pp. 1434–1438.
 - [26] Y. C. Eldar, *Sampling theory: Beyond bandlimited systems*. Cambridge University Press, 2015.
 - [27] J. Akhtar, B. Torvik, and K. E. Olsen, *Compressed sensing with interleaving slow-time pulses and hybrid sparse image reconstruction*, in Proc. 2017 IEEE Radar Conference (RadarConf). IEEE, 2017, pp. 0006–0010.
 - [28] S. Thrun, W. Burgard, and D. Fox, *A probabilistic approach to concurrent mapping and localization for mobile robots*, Autonomous Robots, vol. 5, no. 3–4, pp. 253–271, 1998.
 - [29] H. Durrant-Whyte and T. Bailey, *Simultaneous localization and mapping: Part I*, IEEE Robotics & Automation Magazine, vol. 13, no. 2, pp. 99–110, 2006.
 - [30] M. Z. Win, Y. Shen, and W. Dai, *A theoretical foundation of network localization and navigation*, Proceedings of the IEEE, vol. 106, no. 7, pp. 1136–1165, 2018.

Глава 22

Новые формы сигналов и схемы модуляции

22.1. Почему нужны новые формы сигнала и схемы модуляции

Современные системы связи 4G и 5G в полосах частот ниже 6 ГГц сталкиваются с обширными замираниями из-за многолучевого распространения. Как стандарты LTE, так и стандарты NR для борьбы с этим нежелательным явлением используют OFDM в качестве основной формы сигнала. Этот стандарт также совместим с MIMO, который является одним из наиболее важных методов достижения высокой эффективности использования спектра. Однако по сравнению с сигналами с одной несущей OFDM имеет более высокое отношение пиковой мощности к средней мощности (PAPR). В сценариях, где передача по восходящей линии связи ограничена зоной покрытия, также поддерживается сигнал OFDM с дискретным преобразованием Фурье и расширением спектра с низким PAPR (DFT-s-OFDM).

Что касается модуляции, как LTE, так и NR используют обычные семейства QAM с кодами Грея. QAM облегчает проектирование системы за счет простой демодуляции с разделением реального и мнимого компонентов, но за счет некоторой потери коэффициента усиления. Кроме того, для работы с сигналами DFT-s-OFDM для дальнейшего уменьшения PAPR для сценариев расширенного покрытия с очень низкой эффективностью использования спектра введена модуляция по схеме двоичной фазовой манипуляции $\pi/2$ (binary phase shift keying, BPSK).

И хотя формы сигналов с несколькими несущими, в частности OFDM и обычный QAM, вероятно, продолжат играть центральную роль в системах беспроводного доступа в будущем, могут быть введены новые формы сигналов и схемы модуляции, поскольку появляются новые варианты использования устройств и спектра, о чем будет подробно сказано ниже.

Во-первых, низкий PAPR имеет решающее значение для многих применений. Например, высокочастотная часть спектра (ТГц-диапазон) будет играть важную роль в удовлетворении постоянно растущего спроса на более высокие скорости передачи данных и новые типы трафика. Низкий PAPR будет жизненно важным для любой формы сигнала и схемы модуляции на этих высоких частотах из-за проблем, связанных с проектированием эффективных высокочастотных широкополосных усилителей мощности, преодолением возникающих потерь на трассе и обработкой относительно плоского канала с очень редким рассеянием в пространственно-временной области [1]. Чтобы решить такие проблемы, мы можем ввести новые формы сигналов и схемы модуляции. Сказанное верно и в других сценариях, где передатчик и приемник расположены близко друг к другу, а связь осуществляется по каналу с преобладанием прямой видимости, который остается плоским. В таких сценариях преимущества использования циклического префикса OFDM (CP-OFDM) для борьбы с замираниями из-за многолучевого распространения исчезают, но накладные расходы циклического префикса (CP) остаются. Поскольку высокая пропускная способность является ключевым требованием для связи на малых расстояниях, нежелательно достигать низкого PAPR за счет снижения пропускной способности. В спутниковой связи из-за ограничений мощности спутников и нелинейности усилителей мощности требуется новая форма сигнала с низким PAPR. Низкое значение PAPR также является ключевым требованием для достижения низкого энергопотребления для недорогих устройств с простыми усилителями мощности, низкими вычислительными возможностями и ограниченными возможностями источника питания.

Во-вторых, для недорогих устройств сложность является еще одной ключевой проблемой в дополнение к низкому PAPR. Форма сигнала и схема модуляции должны иметь низкую сложность обработки с хорошей устойчивостью к фазовому шуму, смещению несущей частоты, смещению синхронизации, нелинейности и т. д. Это связано с тем, что дешевое оборудование вносит радиочастотные искажения на разных уровнях.

В-третьих, в случае быстро движущихся объектов нельзя игнорировать эффект Доплера, приводящий к временной избирательности в беспроводных каналах. Если каналы являются LOS-доминирующими, эффект Доплера может быть скомпенсирован на стороне передатчика или приемника путем оценки с помощью пилот-сигнала или априорного знания. Однако если каналы обладают двойной избирательностью, особенно в системах MIMO, мы должны рассмотреть усовершенствованные схемы, которые имеют низкие накладные расходы и сложность, а также отвечают требованиям конкретного случая использования.

В-четвертых, ключевые требования для сценария использования URLLC – низкая задержка и высокая надежность. Это означает, что желаемая форма сигнала и схема модуляции должны иметь короткие временные интервалы и достаточную производительность декодирования без ущерба для других показателей производительности, таких как эффективность использования спектра.

И последнее, но не менее важное: ISAC еще больше повлияет на форму сигнала для беспроводных систем будущего. Для связи и сканирования желательно использовать одну и ту же форму сигнала, но это означает, что при разработке формы сигнала необходимо учитывать дополнительные характеристики и требования, уникальные для сканирования. В частности, форма сигнала, подходящая для связи, обычно имеет высокую спектральную эффективность и низкое внеполосное излучение, тогда как оптимальная форма сигнала для сканирования целей преследует другие целевые показатели, такие как точность и высокое разрешение. В частности, когда требуется только оценка дальности, решение сводится к дельта-подобной автокорреляции во временной области для лучшей оценки временной задержки. Кроме того, форма волны, используемая для сканирования, требует достаточного выигрыша при обработке, чтобы обеспечить возможность измерения параметров в сценариях с сильным шумом и помехами, замираниями из-за многолучевого распространения и недостатками оборудования (например, отсутствие синхронизации времени/частоты, фазовый шум и нелинейность усилителя). Хотя такие ошибки – в пределах разумного – допустимы в системах связи, они могут серьезно снизить точность сканирования.

22.2. Обзор существующих решений

Форма сигнала и схема модуляции для сотовой связи в основном учитывают следующие требования:

- широкий спектр вариантов использования, включая eMBB, mMTC и URLLC;
- высокую эффективность использования спектра для удовлетворения экспоненциально растущего трафика данных в сценариях использования eMBB;
- достаточно широкий охват;
- унифицированную структуру для нисходящего, восходящего и прямого каналов, хорошо совместимую с MIMO для высокой эффективности использования спектра;
- единый дизайн для низких и высоких частот;
- низкую сложность, простоту внедрения и хорошую энергоэффективность.

У каждой формы сигнала есть свои преимущества и недостатки, а это означает, что не существует такого варианта, который превзойдет все другие формы сигнала в соответствии с указанными выше требованиями. В следующих разделах обсуждаются различные формы сигналов.

22.2.1. Сигналы с несколькими несущими

Сигналы с несколькими несущими очень эффективно используют спектр, но делают это за счет довольно высокого PAPR из-за случайного сложения сигналов от разных поднесущих. Ниже мы обсудим несколько вариантов сигналов с несколькими несущими, подчеркнув их преимущества и недостатки. Подробные оценки характеристик и сравнение различных форм сигналов см. в [2].

- **CP-OFDM с ограничением спектра:** CP-OFDM, используемый в нисходящей линии связи LTE, допускает очень гибкую обработку в частотной области, включая мультиплексирование пользователей и каналов, выделение ресурсов и агрегацию несущих. Он имеет очень низкую сложность на стороне передатчика и приемника и обеспечивает хорошие показатели по коэффициенту блочных ошибок (block error rate, BLER). Кроме того, он совместим с MIMO, что является ключом к достижению высокой эффективности использования спектра и/или надежности. Тем не менее, несмотря на эти преимущества, он имеет очень высокое внеполосное излучение (out-of-band emission, OOBЕ). Для решения этой проблемы предлагаются методы ограничения спектра на основе фильтрации и оконного преобразования, которые показаны на рис. 22.1.

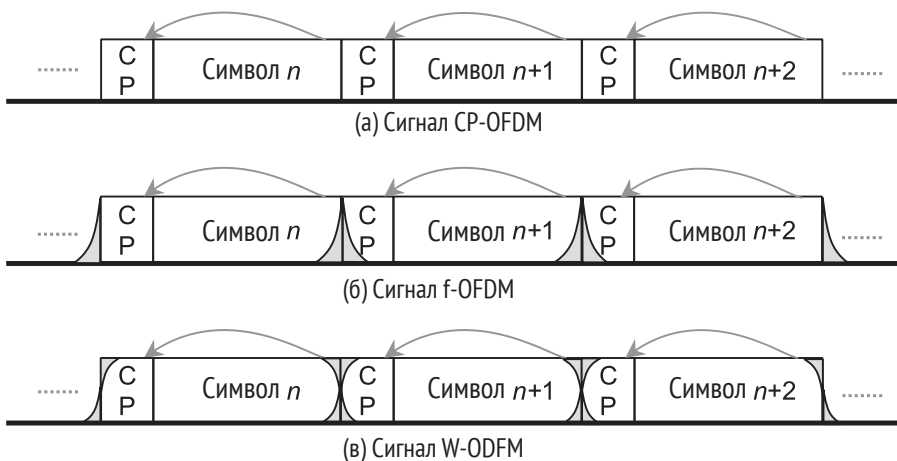


Рис. 22.1. Иллюстрации методов ограничения спектра на основе фильтрации и оконного преобразования, применяемых во временной области

- **Фильтрованный OFDM (f-OFDM):** фильтрация на основе поддиапазонов применяется к сигналам OFDM для подавления межполосных помех, если они используют разные нумерологии или асинхронную передачу между собой [3, 4]. Ортогональность во временной области между последовательными символами OFDM в каждом поддиапазоне нарушается намеренно, чтобы достичь более низкого OOBЕ с незначительным снижением других показателей. По сравнению с обычным OFDM, f-OFDM поддерживает смешанную нумерологию и асинхронную передачу по поддиапазонам, устраняя необходимость в глобальной синхронизации. Кроме того, f-OFDM потребляет значительно меньше ресурсов защитной полосы, используя спектр более эффективно (благодаря хорошим показателям OOBЕ и BLER), при приемлемой вычислительной сложности.
- **Оконный OFDM (W-OFDM):** для сглаживания перехода между последовательными символами OFDM применяется непрямоугольное окно свертки во временной области, что снижает сложность обработки [5]. Однако часть длины префикса используется окнами, что ограничивает показатели W-OFDM из-за уменьшения его эффективной длины.

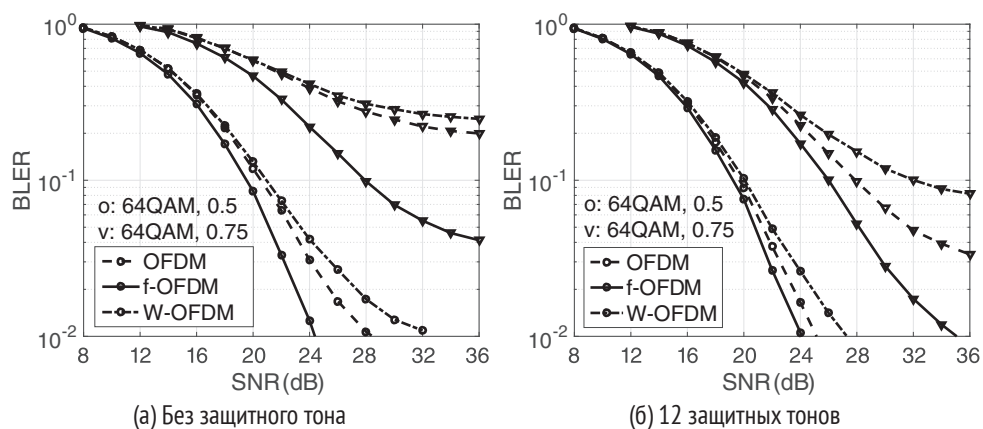


Рис. 22.2. Сравнение BLER для f-OFDM, W-OFDM и OFDM в случае смешанной нумерологии DL

На рис. 22.2 и 22.3 сравнивается уровень BLER для f-OFDM, W-OFDM и OFDM в случаях смешанной нумерологии DL [6] и UL [7], соответственно, с номерами защитного тона 0 и 12. Сравнения основаны на канале NLOS с многоотводной линией задержки (TDL-C) с разбросом задержки 1000 нс в полосе 4 ГГц и с идеальной оценкой канала. На рис. 22.2 пространство поднесущих для целевого поддиапазона и мешающего

поддиапазона составляет 15 кГц и 30 кГц соответственно. На рис. 22.3 промежутки поднесущих для целевого пользовательского устройства и двух создающих помехи устройств составляют 15 кГц и 30 кГц соответственно, а мешающее устройство имеет мощность на 5 дБ больше, чем целевое. Цифры показывают, что f-OFDM обеспечивает гораздо лучшие параметры. Однако до тех пор, пока передаваемый сигнал соответствует требованиям к радиочастотному каналу, для разных сценариев могут быть выбраны разные схемы, чтобы достичь баланса между сложностью и качеством. Кроме того, приемник может реализовать фильтрацию или оконный метод для подавления помех от соседних диапазонов. В 5G CP-OFDM используется без явных указаний на ограничение спектра на основе фильтрации или оконного управления – это зависит от практических реализаций.

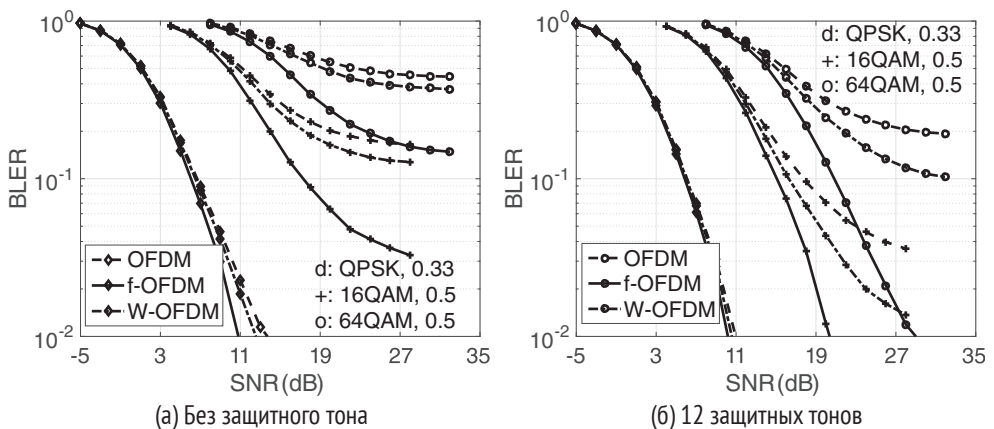


Рис. 22.3. Сравнение BLER для f-OFDM, W-OFDM и OFDM в случае смешанной нумерологии UL

- **Универсальная фильтрованная множественная несущая (UFMC):** UFMC применяет фильтрацию поддиапазонов к каждому символу OFDM, и из-за линейной свертки фильтрации циклический префикс заменяется расширениями символов для формирования защитных интервалов между символами [8]. Длина защитного интервала ограничивает длину фильтра UFMC и, следовательно, характеристики ООБЕ. Поскольку UFMC не имеет циклического префикса, он более сложен с точки зрения демодуляции и более чувствителен к синхронизации времени по сравнению с CP-OFDM.
- **Обобщенное мультиплексирование с частотным разделением (GFDM):** GFDM применяет фильтрацию поднесущих для достижения хороших характеристик ООБЕ. Для подавления по-

мех между несущими (ICI) необходимы поднесущие высокого порядка, круговая фильтрация и концевые биты, поскольку поднесущие расположены близко друг к другу и не взаимно ортогональны [9]. Однако для уменьшения ICI, остающегося после фильтрации, особенно для модуляции высокого порядка, по-прежнему необходимы продвинутые приемники, что усложняет структуру пилот-сигнала и передачу MIMO. Кроме того, GFDM обрабатывается поблочно; хотя издержки на циклический префикс могут быть уменьшены, это увеличивает задержку обработки и, следовательно, не подходит для передачи с малой задержкой.

- **OFDM со спектральным прекодированием (SP-OFDM):** SP-OFDM применяет прекодер к символам данных перед модуляцией OFDM, чтобы уменьшить OOBЕ [10]. Это сокращение ограничено, особенно при небольшой полосе пропускания. Хотя предварительное кодирование создает помехи между несущими, их можно уменьшить (например, путем умножения на инверсию прекодера) в приемнике, если приемник имеет информацию прекодера. Следовательно, сложность декодирования выше, и может потребоваться дополнительная обработка сигнала. Кроме того, предварительное кодирование может вызвать внутрисполосную пульсацию, ухудшающую характеристики детектора.
- **Гребенчато фильтруемый многочастотный сигнал со смещением QAM (FBMC-OQAM):** в FBMC-OQAM каждая поднесущая фильтруется индивидуально, обеспечивая очень хорошие характеристики OOBЕ [11]. По сравнению с другими системами на основе префиксов или дополнений (например, CP-OFDM), FBMC-OQAM не нуждается в защитном интервале с сохранением служебных данных синхронизации. Однако из-за фильтрации поднесущих длина фильтра обычно очень велика (например, более чем в три раза превышает длительность символа). В результате FBMC-OQAM не подходит для передачи коротких символов, как это необходимо для приложений с малой задержкой. Кроме того, он использует сложную схему оценки канала и не может быть легко расширен до MIMO из-за неортогональности в комплексных областях, что ограничивает область его применения.
- **Гребенчато фильтруемый многочастотный сигнал с QAM (FBMC-QAM):** он разработан для смягчения усложнения формы сигнала FBMC-OQAM, когда четные и нечетные поднесущие используют разные фильтры [12]. FBMC-QAM оптимизирует два типа фильтров для минимизации внутренних взаимных помех между несущими, вызванных неортогональностью в комплексных областях. Это обеспечивает низкий OOBЕ за счет использо-

вания длинного фильтра в силу природы фильтрации на основе поднесущих. Однако FBMC-QAM не может полностью устранить внутренние взаимные помехи, которые ведут к снижению характеристик канала даже при использовании усовершенствованного нелинейного приемника. Кроме того, неортогональность усложняет применение FBMC-QAM в сценариях MIMO и неортогонального множественного доступа (NOMA).

- **Взвешенная круговая свертка FBMC-OQAM (WCC-FBMC-OQAM):** для удаления «хвостов», вызванных линейной сверткой фильтра поднесущих в схеме FBMC-OQAM, используется взвешенная круговая свертка [13]. Это устраняет временные издержки, не вызывая появления взаимных помех между поднесущими или межсимвольных помех. Кроме того, взвешенное управление периодической сверткой во временной области используется для обеспечения плавных переходов по краям сигнала, тем самым решая проблему резких краев сигнала во временной области, вызванных круговой сверткой. WCC-FBMC-OQAM обеспечивает исключительный показатель ООВЕ с ограниченными накладными расходами на периодическую свертку во временной области. Хотя он удаляет хвосты сигнала посредством фильтрации, он обладает другими недостатками FBMC-OQAM (например, несовместим с технологией MIMO, которая является ключом к достижению высокой эффективности использования спектра).
- **Гибко настраиваемый OFDM (FC-OFDM):** FC-OFDM – это структура, обеспечивающая сосуществование FBMC-OQAM и OFDM в одном быстром преобразовании Фурье (FFT), где защитные поднесущие зарезервированы для разделения помех от FBMC-OQAM к OFDM [14]. Другими словами, это позволяет объединенной структуре передачи поддерживать эти две формы сигнала. Однако для подавления ООВЕ OFDM требуется дополнительная фильтрация или периодическая свертка. Кроме того, коэффициент перекрытия FBMC-OQAM (длина фильтра, деленная на размер FFT) ограничивается значением 1 для достижения низкой сложности обработки, что сводит на нет преимущество по ООВЕ общего FBMC-OQAM, использующего большой коэффициент перекрытия (например, 4).
- **Ортогональное частотно-временное пространство (OTFS):** OTFS использует двумерное преобразование Фурье для преобразования данных частотно-временной области в область доплеровской задержки, и эти преобразованные данные могут передаваться с помощью традиционного модулятора OFDM [15]. OTFS предполагает, что каналы являются разреженными и неизменными

ми в области доплеровской задержки, где данные свертываются с этими каналами. Это может обеспечить потенциальный выигрыш в высокоскоростных сценариях, но это происходит за счет большой длительности символа и очень высокой сложности выравнивающей фильтрации, не говоря уже о применении MIMO. Обратите внимание, что CP-OFDM с дополнительными опорными сигналами демодуляции (DMRS) для пользователей с высокой мобильностью будет иметь очень хорошие характеристики, особенно в каналах с преобладанием LOS.

- **Спектрально эффективное мультиплексирование с частотным разделением каналов (SEFDM) и мультиплексирование в частотной области с перекрытием (OVFDM):** по сравнению с OFDM, сигнал SEFDM/OVFDМ использует интервал между поднесущими, который меньше ширины поднесущей [16, 17]. Следовательно, сигналы сжимаются в частотной области за счет наличия взаимных помех. Более высокое сжатие означает, что сигнал занимает менее широкую полосу, но это увеличивает внутренние помехи. Для подавления этих помех можно использовать усовершенствованный приемник. Однако этот подход обычно сложен, особенно для высокой модуляции и большого количества поднесущих. Если в системе используется мощное кодирование (например, полярный код), дополнительный выигрыш в пропускной способности может оказаться невозможным. Поскольку высокая модуляция чувствительна к шуму и помехам, из-за внутренних взаимных помех может произойти потеря емкости.
- **Векторный OFDM (V-OFDM):** в векторе OFDM [18] данные N поднесущих каждого OFDM делятся на K субвекторов длины N/K . К каждому субвектору применяется обратное быстрое преобразование Фурье (IFFT), и результирующие K субвекторов перемежаются для генерации перемеженного вектора длины N , который передается вместе с циклическим префиксом. По сравнению с OFDM, в котором используется N поднесущих, V-OFDM уменьшает размер FFT, чтобы упростить обработку в передатчике и уменьшить PAPR. Но поскольку сигналы чередуются, спектр не локализован. Кроме того, среди K субвекторов существуют межсимвольные помехи, что приводит к более высокой сложности выравнивания в V-OFDM, чем в CP-OFDM.
- **Мультиплексирование с неортогональным частотным разделением каналов (NOFDM) / импульсное мультиплексирование OFDM (P-OFDM) / гребенчато фильтруемый OFDM (FB-OFDM):** вместо использования прямоугольной формы импульса (как используется в CP-OFDM) NOFDM/P-OFDM/FB-OFDM

рассматривает форму импульса как еще одну степень свободы для удовлетворения различных требований к сигналу [19–21]. Например, форма импульса может быть спроектирована так, чтобы иметь резкое затухание в частотной области и улучшить частотную локализацию по сравнению с OFDM. Различные формы сигналов с несколькими несущими, такие как CP-OFDM, W-OFDM, GFDM и FBMC-OQAM, могут совместно использовать аналогичную обработку для передачи, но они используют разные функции формы импульса и интервалы символов для разных форм сигналов. Это означает, что преимущества и недостатки различаются в зависимости от используемых импульсных функций и символьных интервалов.

- **Вейвлет-OFDM:** вместо использования быстрого преобразования Фурье (как это делается в OFDM) вейвлет-OFDM использует дискретное вейвлет-преобразование (discrete wavelet transform, DWT) [22]. В частности, обратное дискретное вейвлет-преобразование (IDWT) и DWT применяются на сторонах передатчика и приемника соответственно, с использованием дискретных гребенчатых фильтров. Подобно схемам фильтрации поднесущих, вейвлет-OFDM обеспечивает низкий уровень OOBЕ из-за перекрытия символов (хвосты во временной области длинных фильтров) и не требует защитного интервала во времени. Из-за длинных хвостов во временной области он не подходит для передачи с малой задержкой. Более того, его адаптация к MIMO требует дальнейшего углубленного исследования.
- **Мультиплексирование с разделением по Лагранжу и Вандермонду (LVDM):** LVDM обобщает OFDM с заполнением нулями (zero padding, ZP-OFDM), при котором передатчик модулирует вектор передаваемых символов с помощью матрицы Лагранжа, а приемник демодулирует принятый вектор символов с помощью матрицы Вандермонда [23]. Сигнатуры для построения матриц Лагранжа и Вандермонда оптимизируются в соответствии с каналами замирания, что требует знания о каналах. Если сигнатуры были распределены по единичному кругу с неизвестными каналами, LVDM сокращается до ZP-OFDM. Для определения прироста характеристик необходимы дополнительные исследования (например, сравнение BLER в кодированных системах).
- **Форма сигнала, основанная на преобразовании линейно-частотной модуляции:** в качестве развития преобразования Фурье, используемого в OFDM, в [24] была предложена форма сигнала на основе дробного преобразования Фурье, а в [25] была предложена форма сигнала на основе аффинного преобразования

Фурье. Их цель – обрабатывать изменяющиеся во времени каналы в условиях возникновения эффекта Доплера. Введение свободы в параметр линейной частотной модуляции делает систему более устойчивой к внутренним взаимным помехам по сравнению с OFDM в случаях, когда наблюдаются многолучевые каналы с линейной зависимостью между задержками и доплеровскими сдвигами. Впрочем, сравнение с OFDM нуждается в дальнейших исследованиях.

- **Неортогональная множественная несущая на основе базовой функции Слепиана (Slepian-basis-based non-orthogonal multi-carrier, SNMC):** в SNMC [26] одна поднесущая мультиплексируется множеством ортогональных базовых функций Слепиана, которые переносят информацию данных (например, символы QAM). Поскольку концентрация энергии в базовой функции Слепиана оптимальна как во временной, так и в частотной областях, SNMC обеспечивает очень хорошие характеристики ООВЕ. Хотя базовая функция Слепиана ортогональна в одной поднесущей, все же взаимные помехи существуют из-за неортогональности между поднесущими. Следовательно, для подавления взаимных помех необходим усовершенствованный приемник. Данная проблема также усложняет схему пилотного сигнала и оценки канала. Ожидается, что в сценариях MIMO использование SNMC будет намного сложнее по сравнению с OFDM.

22.2.2. Формы сигналов с одной несущей

Сигналы с одной несущей (например, DFT-s-OFDM, принятые в LTE и NR UL) более выгодны с точки зрения PAPR по сравнению с формами сигналов с несколькими несущими. Подобные сигналы также более устойчивы к фазовому шуму с относительно простой оценкой и компенсацией во временной области. Однако сигналы с одной несущей имеют некоторые общие недостатки, которые ограничивают сценарии их применения. В частности, поддержка операций мультиплексирования с частотным разделением каналов (frequency-division multiplexing, FDM), таких как разделение пользователей, разделение опорного сигнала и данных и агрегация несущих, ухудшит значение PAPR. Кроме того, выделение частотных ресурсов и отображение данных для пользователей, обслуживаемых сигналами с одной несущей, слишком ограничены, чтобы поддерживать низкий PAPR и разумную сложность обработки. В случае многолучевых частотно-избирательных беспроводных каналов применение простой коррекции в частотной области (применяемой в OFDM) приведет к потере показателей. Причем эта потеря намного выше в сценариях MIMO, где OFDM может реализовывать

предварительное кодирование в частотной области (даже для каждой поднесущей), не влияя на PAPR. И наоборот, для поддержания низкого PAPR сигналы с одной несущей могут использовать только полнополосное предварительное кодирование. Ниже перечислены некоторые типичные формы сигналов с одной несущей.

- **Одинокная несущая на основе расширения DFT:** большинство многочастотных сигналов, рассмотренных ранее, можно преобразовать в соответствующие версии с одной несущей. Типичным примером этого является DFT-s-OFDM. В данном случае предварительное кодирование DFT переносит входные сигналы в частотную область. Для дальнейшего уменьшения PAPR может применяться формирование спектра в частотной области (frequency-domain spectral shaping, FDSS), которое обычно подразумевает увеличение полосы пропускания (т. е. за счет снижения эффективности использования спектра). Фильтр FDSS не может быть прозрачным для модуляции высокого порядка, поскольку он значительно снижает качество декодирования, потенциально требуя дополнительных служебных сигналов. Сигналы частотной области отображаются на желаемую полосу пропускания, а затем переносятся обратно во временную область путем обратного быстрого преобразования Фурье. В сигнал обычно добавляется циклический префикс для обеспечения возможности однократной коррекции в частотной области на стороне приемника. Хотя выделение ресурсов совместимо с OFDM с точки зрения обработки в частотной области, оно ограничено несколькими вариантами, чтобы упростить операцию DFT.
- **SC-QAM/SC-FDE:** это традиционный сигнал с одной несущей, где вся обработка выполняется во временной области [27]. Подобно DFT-s-OFDM, SC-QAM/SC-FDE добавляет циклический префикс, нули или уникальные слова (UW), чтобы обеспечить однократное выравнивание частотной области на стороне приемника. По сравнению с DFT-s-OFDM, SC-QAM/SC-FDE имеет меньшую сложность (потому что ни DFT, ни IFFT не используются на стороне передатчика) и менее гибок с точки зрения операций в частотной области. В частности, SC-QAM/SC-FDE поддерживает только передачу/прием с полной полосой пропускания без учета частотного разделения пользователей, каналов или опорных сигналов.
- **ZT/UW-DFT-s-OFDM:** это вариант DFT-s-OFDM, в котором нули или уникальные слова добавляются перед дискретным преобразованием Фурье для замены циклического префикса [28]. В DFT-s-OFDM служебные данные циклического префикса явля-

ются фиксированными и общими для всех пользователей, тогда как длина нулей или уникального слова может зависеть от пользователя, чтобы динамически регулировать служебные данные на основе задержки расширения и распространения в ZT/UW-DFT-s-OFDM. Это потенциально может снизить некоторые накладные расходы для пользователей с малым разбросом задержки. Однако из-за отсутствия префикса продолжительность символа отличается от всех сигналов на основе циклического префикса, описанных ранее. Таким образом, ZT/UW-DFT-s-OFDM не может мультиплексироваться с этими пользователями в одной и той же полосе. Разброс задержки канала должен быть известен на сторонах базовой станции и пользовательского оборудования, для чего потребуются дополнительные служебные данные. Кроме того, параметры модуляции более высокого порядка будут ограничены из-за отсутствия характеристик кольцевой свертки каналов.

- **Быстрее, чем Найквист (FTN) / мультиплексирование с перекрытием во временной области (OVTDM):** в отличие от SEFDM/OVFDM, этот сигнал упаковывает символы во временной области для передачи со скоростью FTN [17, 29]. Недостатки FTN/OVTDM такие же, как и у SEFDM/OVFDM; а именно сложный приемник с потенциально ограниченным увеличением пропускной способности или вообще без него по сравнению с системами, которые используют мощную схему кодирования. Обратите внимание, что OVFDM и OVTDM совместно называются мультиплексированием X-области с перекрытием (overlapped X domain multiplexing, OVXDM) [17].

22.2.3. Схемы модуляции

По сравнению с обычной QAM другие схемы модуляции могут обеспечить лучший коэффициент усиления, более низкий PAPR и лучшую устойчивость к радиочастотным искажениям. Некоторые из таких схем модуляции обсуждаются ниже.

- **Повернутая QAM:** это обычная QAM после того, как к ней применили сдвиг (поворот) по фазе. Например, квадратурная фазовая манипуляция с поворотом $\pi/4$ (QPSK) применяется в узкополосной UL интернета вещей (NB-IoT), где поворот фазы равен $\pi/4 \times (n \bmod 2)$, а n – индекс символа. Дополнительное вращение фазы может уменьшить сдвиг фазы между соседними модулированными символами, что, в свою очередь, может уменьшить PAPR за счет использования сигналов с одной несущей. Однако

рост PAPR чрезвычайно ограничивает модуляцию более высокого порядка выше QPSK. Кроме того, повернутая QAM обычно используется вместе с непрозрачным FDSS для DFT-s-OFDM, чтобы обеспечить хорошие характеристики PAPR и BLER.

- **Нерегулярная QAM:** в 5G было предложено несколько схем нерегулярной QAM, включая 1D/2D-оптимизированную неоднородную QAM, вероятностную кодированную модуляцию [30] и амплитудно-фазовую манипуляцию (APSK). Эти схемы обладают потенциалом для обеспечения более высоких коэффициентов усиления формирования, улучшения PAPR и устойчивости к фазовому шуму. Однако такой выигрыш достигается за счет более высокой сложности демодуляции.
- **Интерполяция совокупности:** чтобы уменьшить фазовый сдвиг между соседними модулированными символами, в [31] была предложена интерполяция совокупности. В частности, модулированная совокупность сначала интерполируется по гладкой траектории с постоянной огибающей (для QPSK) или траектории с почти постоянной огибающей (для QAM) перед операцией дискретного преобразования Фурье в DFT-s-OFDM. Интерполяция достигается с помощью простого решетчатого кодера с Q -состоянием, где Q – исходный размер совокупности. Более высокий коэффициент интерполяции подразумевает более ограниченную траекторию совокупности, что приводит к меньшему PAPR за счет более высокой сложности интерполяции. FDSS может формировать компромисс между эффективностью использования спектра и PAPR, используя различные фильтры формирования спектра. Для выравнивания каналов может использоваться стандартный эквалайзер с одним отводом в частотной области. Выровненный символ затем подается в решетчатый декодер Q -состояний для получения логарифмического отношения правдоподобия (log likelihood ratio, LLR) для декодера с прямым исправлением ошибок (forward error correction, FEC). Интерполяция совокупности может обеспечить чрезвычайно хороший уровень PAPR, но достигает этого за счет сложности передачи и приема. В настоящее время необходимы дальнейшие исследования для изучения уровня BLER при различных компромиссах между эффективностью использования спектра и уровнем PAPR.
- **Многомерная модуляция:** многомерная модуляция использует больше степеней свободы для повышения качества канала за счет сложности. Она может улучшить коэффициент различения символов по сравнению с традиционными комбинациями QAM. Также она может предоставить новый способ организации массового

доступа пользователей, такой как схема множественного доступа с разреженным кодом (SCMA) [32, 33]. Кроме того, для некогерентного детектирования без пилот-сигнала могут применяться графмановы совокупности, что позволяет значительно снизить накладные расходы на пилот-сигнал в определенных случаях (например, mMTC и массивные MIMO) за счет увеличения сложности детектирования [34].

- **Индексная модуляция:** при индексной модуляции индексы составляющих блоков в системах связи используются для передачи дополнительной информации. Два применения индексной модуляции – это пространственная модуляция и OFDM с индексной модуляцией [35]. Пространственная модуляция передает информацию с использованием индексов передающих антенн в дополнение к традиционной схеме символьной модуляции (например, QAM). С другой стороны, OFDM с индексной модуляцией передает информацию с использованием индексов поднесущих передачи. Индексная модуляция может обеспечить хорошую энергоэффективность, но эффективность использования спектра будет ниже по сравнению с традиционными системами связи QAM.

22.2.4. Формы сканирующих сигналов

Сканирование и способы повторного использования сигналов связи вызвали большой интерес в академических кругах. Некоторые из типичных форм сканирующих сигналов обсуждаются ниже.

- **Сигналы с несколькими несущими:** формы сигналов с несколькими несущими и, в частности, CP-OFDM были доминирующими вариантами для связи из-за их высокой эффективности использования спектра, масштабируемости и гибкости благодаря введению циклического префикса. Поэтому многие исследователи рассматривали возможность использования подобных сигналов и для сканирования. Учитывая тот факт, что циклические префиксы могут ухудшить автокорреляцию во временной области [36], принят новый подход к обработке в частотной области, который может эффективно оценивать параметры CP-OFDM с максимальным выигрышем от обработки [37]. Кроме того, было доказано, что CP-OFDM свободен от проблемы доплеровского сдвига диапазона, таким образом, оценки дальности и доплеровского сдвига могут рассматриваться в CP-OFDM как независимые задачи [38]. Кроме того, параметры CP-OFDM, такие как расстояние до несущей, величина защитного интервала, длина кадра и

форма пилот-сигнала, могут быть адаптированы для оптимизации работы и устойчивости детектирования и передачи данных [39]. Однако подобные преимущества основываются на идеальной синхронизации (времени и частоты) между передатчиком и приемником – хотя идеальная синхронизация может оказаться невозможной, особенно для бистатического сканирования, когда передатчик и приемник сигнала сканирования не расположены рядом. В этом случае циклический префикс может не обеспечивать каких-либо преимуществ в качестве сканирования, и можно рассматривать формы сигналов с несколькими несущими без префикса. Основным недостатком отказа от префикса является сложность детектирования данных (из-за межсимвольных помех), которую необходимо устранить. Так или иначе, еще одной важной проблемой для применений сканирования, где чрезвычайно важна энергоэффективность, является большой PAPR сигналов с несколькими несущими (с префиксом или без него).

- **Непрерывная волна с частотной модуляцией (FMCW):** этот сигнал традиционно используется для радаров. Он имеет свойства, подходящие для сканирования, в том числе хорошую автокорреляцию, устойчивость к недостаткам оборудования и низкий PAPR. Однако он не пригоден для передачи данных, что очень важно для ISAC. Чтобы сделать его более пригодным для коммуникаций, некоторые исследователи предложили модификацию FMCW. Например, Саддик и др. в 2007 году предложили использовать повышающую линейную частотную модуляцию для связи и понижающую для радара [40], а Хан и Ву в 2010 году предложили непрерывную волну с трапецеидальной частотной модуляцией (TFMCW) [41], которая представляет собой новую схему с мультиплексированием циклов радара и связи во временной области. Хотя эти схемы обеспечивают эффективное мультиплексирование данных связи и сигналов сканирования, они по-прежнему страдают низкой эффективностью использования спектра из-за частотно-модулированных сигналов сканирования.
- **Сигналы с одной несущей:** эти сигналы основаны на расширении кодовой области совместных сигналов радара и связи [42–47], в котором на характеристики радара влияет автокорреляция последовательностей. Однако длинный расширяющий код, который приводит к хорошей автокорреляции, снижает эффективность использования спектра для связи. В этом случае оценка доплеровского эффекта не является тривиальной и требует более сложных алгоритмов.

22.3. Ожидания от новых разработок и потенциальные направления исследований

Ниже кратко изложены требования к сценариям использования, устройствам и спектру, которые имеют отношение к форме волны и схеме модуляции. Ожидается, что формы сигналов 6G и схемы модуляции должны соответствовать этим требованиям, что может потребовать наличия более чем одной конструкции системы.

- **Очень высокая частота:** ключевыми требованиями являются низкий PAPR для покрытия, высокая устойчивость к радиочастотным искажениям (например, фазовому шуму), низкая сложность для широкополосных операций и совместимость с MIMO. Особенности канала – большие потери на трассе с разреженным рассеянием.
- **Спутниковая связь:** основными требованиями являются низкий PAPR для покрытия и низкая сложность с целью энергосбережения. Особенности канала – большие потери на трассе из-за больших расстояний и высокая скорость движения спутника.
- **Связь малого радиуса действия:** основными требованиями являются высокая пропускная способность (например, совместимость с MIMO) и устойчивость к радиочастотным искажениям. В каналах преобладает прямая видимость (LOS) с плоской частотной характеристикой.
- **Недорогие устройства:** основными требованиями являются низкий PAPR, низкая сложность с целью энергосбережения и высокая надежность. Особенности канала зависят от сценариев применения. Например, в узкополосных сценариях с широким покрытием у каналов относительно плоская частотная характеристика.
- **Высокая мобильность:** основными требованиями являются высокая устойчивость к эффекту Доплера и совместимость с MIMO. Каналы избирательны по времени или дважды избирательны из-за эффекта Доплера.
- **URLLC:** ключевые требования – сверхнизкая задержка и высокая надежность (например, совместимость с MIMO). Особенности канала зависят от сценариев применения.
- **ISAC:** в дополнение к упомянутым ранее требованиям со стороны связи ключевыми требованиями со стороны сканирования являются точность оценки и высокое разрешение, особенно при радиочастотных искажениях (например, ошибка синхронизации времени/частоты, фазовый шум и нелинейность). Особенности

канала – высокие потери на трассе при двустороннем сканировании и возможность совместного использования многолучевых беспроводных каналов с коммуникациями.

Исходя из этих требований, при проектировании систем 6G следует учитывать следующие факторы, некоторые из которых уже были рассмотрены в техническом отчете [48] для NR за пределами 52,6 ГГц.

- **Эффективность усилителя мощности:** ожидается, что эффективность усилителей мощности на высоких частотах снизится, а некоторые варианты использования (например, спутниковая связь) чувствительны к потребляемой мощности. Поэтому необходимо использовать в первую очередь сигналы с низким PAPR, чтобы минимизировать ухудшение эффективности усилителей мощности. Кроме того, некоторые транзисторы с высокой подвижностью электронов в ТГц-диапазоне не могут обеспечивать непрерывную передачу энергии, а это означает, что в будущем может потребоваться форма сигнала с прерывистыми импульсами [49].
- **Динамические диапазоны аналого-цифрового преобразователя (АЦП) и цифроаналогового преобразователя (ЦАП):** более широкая полоса пропускания приводит к увеличению трудностей в обеспечении большого эффективного числа битов (ENOB) в АЦП/ЦАП при заданной мощности потребления. Но для того, чтобы приспособиться к сигналам основной полосы частот с более высоким PAPR, в передающем ЦАП требуется более высокий ENOB. Кроме того, недорогие устройства могут использовать ЦАП/АЦП с низким значением ENOB. Эти факторы необходимо учитывать, поскольку все они зависят от формы волны и схем модуляции.
- **Точность и внеполосное излучение модулированного сигнала:** весь радиотракт проектируется и настраивается в соответствии с требованиями к радиосигналу, такими как спектральная маска излучения, коэффициент утечки по соседнему каналу, внутриполосное и внеполосное излучение и величина вектора ошибок. Формы сигналов и схемы модуляции должны соответствовать этим требованиям, чтобы обеспечивать соответствующие характеристики качества внутриполосного сигнала и минимизировать помехи в соседнем канале и влияние на сигналы в соседних каналах. Ширина занимаемой полосы сигнала и защитная полоса для данной полосы пропускания канала определяют использование частоты, что критично для достижения высокой пропускной способности (например, для связи на малых расстояниях).

- **Сложность оборудования и качество связи:** учитывая высокую скорость передачи данных и высокую частоту дискретизации на фоне использования недорогих высокочастотных оконечных устройств или энергочувствительной спутниковой связи, при разработке методов генерации/модуляции и приема/демодуляции сигнала необходимо соблюдать компромисс между сложностью оборудования и характеристиками канала связи.
- **Гибкость спектра:** различия в вариантах использования и распределения частот в разных странах могут потребовать поддержки разных полос пропускания. Следовательно, при проектировании следует учитывать гибкость спектра.
- **Устойчивость к сдвигу по времени и частоте и к фазовому шуму:** смещение несущей частоты и фазовый шум намного выше с несовершенными усилителями мощности и кварцевыми генераторами и более критичны для высоких частот. Кроме того, с увеличением несущей частоты и относительной скорости движения увеличиваются доплеровский сдвиг и рассеивание. Для недорогих устройств, в дополнение к смещению частоты и фазовому шуму (в зависимости от рабочей частоты), также следует учитывать смещение по времени, поскольку для них может быть доступна только грубая синхронизация шкалы времени. Для реализации сканирования с высоким разрешением требуется накопление сигналов, которое зависит от когерентности эхо-сигналов. Таким образом, фазовый шум является основным фактором, влияющим на работу системы.
- **Совместимость с MIMO:** MIMO – это хороший способ повышения эффективности использования спектра. Связь в миллиметровом и ТГц-диапазонах может поддерживать как минимум MIMO 2×2 с двумя поляризациями, даже для каналов чистой прямой видимости. Следовательно, новые формы сигналов должны быть легко расширены до операций MIMO с разумной сложностью.
- **Сканирование:** здесь самая большая проблема – противоречие между KPI связи и сканирования. В области связи основными задачами являются максимальное повышение эффективности использования спектра и ООВЕ. Однако в случае измерения дальности и доплеровского сдвига оптимальная форма сигнала направлена на достижение максимального разрешения и точности оценки.

С учетом требований и соображений, изложенных выше, мы можем предложить некоторые потенциальные направления исследований.

- **Низкое значение PAPR:** многочастотные сигналы с передовыми методами уменьшения PAPR и сигналы с одной несущей с передо-

выми методами модуляции – это два многообещающих направления исследований, которые могут снизить PAPR без ущерба для эффективности использования спектра. Более того, оптимизированные для ИИ демодуляторы также могут содействовать уменьшению PAPR.

- **Низкая сложность:** ориентированные на недорогие устройства формы сигналов предназначены для узкополосных сценариев, в которых сложность обработки невысока для обеспечения энергосбережения. Также желательно обеспечить устойчивость к ограничениям, создаваемым недорогим оборудованием (например, смещения по времени и частоте). Примером этого является двухпозиционная модуляция/демодуляция, которую легко реализовать, если целевая скорость передачи данных и скорость доступа соответствуют требованиям конкретного сценария. С другой стороны, в миллиметровом/терагерцовом диапазоне сигналы рассчитаны на очень широкую полосу пропускания. В частности, желательна простая форма сигнала на основе выравнивающих кодов, чтобы избежать высокого энергопотребления.
- **Частотно-временная локализация:** сигнал с улучшенной частотно-временной локализацией повысит эффективность использования спектра для связи на малых расстояниях с высокими требованиями к скорости передачи данных. Он также снизит требования к временной синхронизации для доступа к недорогим устройствам по каналу UL. Помимо методов фильтрации и оконного преобразования Фурье, могут потребовать исследования и другие методы (например, FTN). Кроме того, URLLC требует хорошей локализации во времени для достижения относительно короткой продолжительности символа и хорошего качества декодирования.
- **Эффективность использования спектра:** этот показатель может быть улучшен за счет использования хорошей частотно-временной локализации, а дальнейшее повышение эффективности применения спектра может быть реализовано за счет использования оптимизированных для ИИ частотных комбинаций и демодуляторов высокого порядка, которые могут помочь улучшить форму сигнала при разумной сложности демодуляции. Для приложений, чувствительных к издержкам, вносимым пилот-сигналом, также могут быть рассмотрены некогерентная модуляция и демодуляция без пилот-сигнала.
- **Высокая мобильность:** для высокоскоростных сценариев использования с двойной селективностью каналов следует рассмотреть усовершенствованные схемы, которые соответствуют на-

кладным расходам, сложности и другим требованиям (например, низкая задержка для применений, чувствительных к задержке). За основу можно взять OFDM с дополнительными DMRS для отслеживания канала в системах NR.

- **Устойчивость к радиочастотным искажениям:** традиционные методы, основанные на обработке сигналов, такие как отслеживание фазового шума и цифровое предварительное искажение для компенсации нелинейности усилителя мощности, вызвали большой интерес в академических кругах. Однако существует значительный разрыв между реальными и идеальными показателями. Можно рассмотреть методы на основе искусственного интеллекта для дальнейшего сокращения разрыва и достижения почти идеальных характеристик.
- **ISAC:** несмотря на то что значительный объем исследований направлен на разработку сигнала для систем ISAC, все еще есть много возможностей для поиска формы сигнала, которая обеспечивает баланс между хорошей связью и параметрами сканирования из-за их противоречивых требований к сигналу. В частности, при разработке формы сигнала следует учитывать, нужно ли использовать сигналы ISAC только для целей сканирования, для передачи данных с возможностью сканирования или для того и другого в равной степени.

22.4. Источники

- [1] C. Lin and G. Y. L. Li, *Terahertz communications: An array-of-subarrays solution*, IEEE Communications Magazine, vol. 54, no. 12, pp. 124–131, 2016.
- [2] X. Zhang, L. Chen, J. Qiu, and J. Abdoli, *On the waveform for 5G*, IEEE Communications Magazine, vol. 54, no. 11, pp. 74–80, 2016.
- [3] J. Abdoli, M. Jia, and J. Ma, *Filtered ofdm: A new waveform for future wireless systems*, in Proc. 2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2015, pp. 66–70.
- [4] X. Zhang, M. Jia, L. Chen, J. Ma, and J. Qiu, *Filtered-ofdm-enabler for flexible waveform in the 5th generation cellular networks*, in Proc. 2015 IEEE Global Communications Conference (GLOBECOM). IEEE, 2015, pp. 1–6.
- [5] R. Zayani, Y. Medjahdi, H. Shaiek, and D. Roviras, *WOLA-OFDM: A potential candidate for asynchronous 5G*, in Proc. 2016 IEEE Globecom Workshops. IEEE, 2016, pp. 1–5.

- [6] Huawei and HiSilicon, *Waveform evaluation updates for case 2*, 3rd Generation Partnership Project (3GPP), RAN1 (R1) 166120, Aug. 2016. https://www.3gpp.org/ftp/tsg_an/wg1_r11/TSGR1_86/Docs/.
- [7] Huawei and HiSilicon, *Waveform evaluation updates for case 4*, 3rd Generation Partnership Project (3GPP), RAN1 (R1) 166091, Aug. 2016. https://www.3gpp.org/ftp/tsg_ran/wg1_r11/TSGR1_86/Docs/.
- [8] F. Schaich and T. Wild, *Waveform contenders for 5G OFDM vs. FBMC vs. UFMC*, in Proc. 2014 6th International Symposium on Communications, Control and Signal Processing (ISCCSP). IEEE, 2014, pp. 457–460.
- [9] N. Michailow, M. Matth  , I. S. Gaspar, A. N. Caldevilla, L. L. Mendes, A. Festag, and G. Fettweis, *Generalized frequency division multiplexing for 5th generation cellular networks*, IEEE Transactions on Communications, vol. 62, no. 9, pp. 3045–3061, 2014.
- [10] X. Huang, J. A. Zhang, and Y. J. Guo, *Out-of-band emission reduction and a unified framework for precoded OFDM*, IEEE Communications Magazine, vol. 53, no. 6, pp. 151–159, 2015.
- [11] F. Schaich, *Filterbank based multi carrier transmission (FBMC) evolving OFDM: FBMC in the context of WiMAX*, in Proc. 2010 European Wireless Conference (EW). IEEE, 2010, pp. 1051–1058.
- [12] C. Kim, K. Kim, Y. H. Yun, Z. Ho, B. Lee, and J.-Y. Seol, *QAM-FBMC: A new multicarrier system for post-OFDM wireless communications*, in Proc. 2015 IEEE Global Communications Conference (GLOBECOM). IEEE, 2015, pp. 1–6.
- [13] M. J. Abdoli, M. Jia, and J. Ma, *Weighted circularly convolved filtering in OFDM/OQAM*, in Proc. 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). IEEE, 2013, pp. 657–661.
- [14] H. Lin, *Flexible configured OFDM for 5G air interface*, IEEE Access, vol. 3, pp. 1861–1870, 2015.
- [15] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, *Orthogonal time frequency space modulation*, in Proc. 2017 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2017, pp. 1–6.
- [16] X. Liu, T. Xu, and I. Darwazeh, *Coexistence of orthogonal and non-orthogonal multicarrier signals in beyond 5G scenarios*, in Proc. 2020 2nd 6G Wireless Summit (6G SUMMIT). IEEE, 2020, pp. 1–5.
- [17] D. Li, *Overlapped multiplexing principle and an improved capacity on additive white gaussian noise channel*, IEEE Access, vol. 6, pp. 6840–6848, 2017.

- [18] X.-G. Xia, *Precoded and vector OFDM robust to channel spectral nulls and with reduced cyclic prefix length in single transmit antenna systems*, IEEE Transactions on Communications, vol. 49, no. 8, pp. 1363–1374, 2001.
- [19] W. Kozek and A. F. Molisch, *Nonorthogonal pulses shapes for multicarrier communications in doubly dispersive channels*, IEEE Journal on Selected Areas in Communications, vol. 16, no. 8, pp. 1579–1589, 1998.
- [20] Z. Zhao, M. Schellmann, Q. Wang, X. Gong, R. Boehnke, and W. Xu, *Pulse shaped OFDM for asynchronous uplink access*, in Proc. 2015 49th Asilomar Conference on Signals, Systems and Computers. IEEE, 2015, pp. 3–7.
- [21] X. Yu, Y. Guanghai, Y. Xiao, Y. Zhen, X. Jun, and G. Bo, *FB-OFDM: A novel multicarrier scheme for 5G*, in Proc. 2016 European Conference on Networks and Communications (EuCNC). IEEE, 2016, pp. 271–276.
- [22] S. Galli, H. Koga, and N. Kodama, *Advanced signal processing for PLCs: Wavelet- OFDM*, in Proc. 2008 IEEE International Symposium on Power Line Communications and Its Applications. IEEE, 2008, pp. 187–192.
- [23] K. Tourki, R. Zakaria, and M. Debbah, *Lagrange Vandermonde division multiplexing*, in Proc. 2020 IEEE International Conference on Communications (ICC). IEEE, 2020, pp. 1–6.
- [24] M. Martone, *A multicarrier system based on the fractional Fourier transform for timefrequency-selective channels*, IEEE Transactions on Communications, vol. 49, no. 6, pp. 1011–1020, 2001.
- [25] T. Erseghe, N. Laurenti, and V. Cellini, *A multicarrier architecture based upon the affine Fourier transform*, IEEE Transactions on Communications, vol. 53, no. 5, pp. 853–862, 2005.
- [26] X. Yang, X. Wang, and J. Zhang, *A new waveform based on Slepian basis for 5G system*, in Proc. 2016 Wireless Days Conference (WD). IEEE, 2016, pp. 1–4.
- [27] F. Pincaldi, G. M. Vitetta, R. Kalbasi, N. Al-Dhahir, M. Uysal, and H. Mheidat, *Singlecarrier frequency domain equalization*, IEEE Signal Processing Magazine, vol. 25, no. 5, pp. 37–56, 2008.
- [28] G. Berardinelli, F. M. Tavares, T. B. Sørensen, P. Mogensen, and K. Pajukoski, *Zero-tail DFT-spread-OFDM signals*, in Proc. 2013 IEEE Globecom Workshops. IEEE, 2013, pp. 229–234.
- [29] J. B. Anderson, F. Rusek, and V. Öwall, *Faster-than-Nyquist signaling*, Proceedings of the IEEE, vol. 101, no. 8, pp. 1817–1830, 2013.
- [30] O. İşcan, R. Böhnke, and W. Xu, *Probabilistic shaping using 5G new radio polar codes*, IEEE Access, vol. 7, pp. 22 579–22 587, 2019.
- [31] MediaTek Inc., *A new DFTS-OFDM compatible low PAPR technique for NR uplink waveforms*, 3rd Generation Partnership Project (3GPP),

- RAN1 (R1) 1609378, Oct. 2016. https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_86b/Docs/.
- [32] H. Nikopour and H. Baligh, *Sparse code multiple access*, in Proc. 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). IEEE, 2013, pp. 332–336.
 - [33] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, *Scma codebook design*, in Proc. 2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall). IEEE, 2014, pp. 1–5.
 - [34] R. H. Gohary and H. Yanikomeroglu, *Noncoherent MIMO signaling for block-fading channels: Approaches and challenges*, IEEE Vehicular Technology Magazine, vol. 14, no. 1, pp. 80–88, 2019.
 - [35] E. Basar, *Index modulation techniques for 5G wireless networks*, IEEE Communications Magazine, vol. 54, no. 7, pp. 168–175, 2016.
 - [36] B. Paul, A. R. Chiriyath, and D. W. Bliss, *Survey of RF communications and sensing convergence research*, IEEE Access, vol. 5, pp. 252–270, 2016.
 - [37] C. Sturm, E. Pancera, T. Zwick, and W. Wiesbeck, *A novel approach to OFDM radar processing*, in Proc. 2009 IEEE Radar Conference. IEEE, 2009, pp. 1–4.
 - [38] M. Braun, C. Sturm, and F. K. Jondral, «*Maximum likelihood speed and distance estimation for OFDM radar*», in Proc. 2010 IEEE Radar Conference. IEEE, 2010, pp. 256–261.
 - [39] M. Braun, C. Sturm, A. Niethammer, and F. K. Jondral, *Parametrization of joint OFDMbased radar and communication systems for vehicular applications*, in Proc. 2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications. IEEE, 2009, pp. 3020–3024.
 - [40] G. N. Saddik, R. S. Singh, and E. R. Brown, *Ultra-wideband multifunctional communications/radar system*, IEEE Transactions on Microwave Theory and Techniques, vol. 55, no. 7, pp. 1431–1437, 2007.
 - [41] L. Han and K. Wu, *Radar and radio data fusion platform for future intelligent transportation system*, in Proc. 7th European Radar Conference. IEEE, 2010, pp. 65–68.
 - [42] K. Mizui, M. Uchida, and M. Nakagawa, *Vehicle-to-vehicle 2-way communication and ranging system using spread spectrum technique: Proposal of double boomerang transmission system*, in Proc. Vehicle Navigation and Information Systems Conference. IEEE, 1994, pp. 153–158.
 - [43] S. Lindenmeier, K. Boehm, and J. F. Luy, *A wireless data link for mobile applications*, IEEE Microwave and Wireless Components Letters, vol. 13, no. 8, pp. 326–328, 2003.

- [44] S. Xu, Y. Chen, and P. Zhang, *Integrated radar and communication based on DS-UWB*, in Proc. 2006 3rd International Conference on Ultrawideband and Ultrashort Impulse Signals. IEEE, 2006, pp. 142–144.
- [45] Z. Lin and P. Wei, *Pulse amplitude modulation direct sequence ultra wideband sharing signal for communication and radar systems*, in Proc. 2006 7th International Symposium on Antennas, Propagation & EM Theory. IEEE, 2006, pp. 1–5.
- [46] Z. Lin and P. Wei, *Pulse position modulation time hopping ultra wideband sharing signal for radar and communication system*, in Proc. 2006 CIE International Conference on Radar. IEEE, 2006, pp. 1–4.
- [47] M. Bocquet, C. Loyez, C. Lethien, N. Deparis, M. Heddebaut, A. Rivenq, and N. Rolland, *A multifunctional 60-GHz system for automotive applications with communication and positioning abilities based on time reversal*, in Proc. 7th European Radar Conference. IEEE, 2010, pp. 61–64.
- [48] 3GPP, *Study on requirements for NR beyond 52.6 GHz*, 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.807, Jan. 2020, version 16.0.0. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3522>.
- [49] J. M. Jornet and I. F. Akyildiz, *Femtosecond-long pulse-based modulation for terahertz band communication in nanonetworks*, IEEE Transactions on Communications, vol. 62, no. 5, pp. 1742–1754, 2014.

Глава 23

Новое канальное кодирование

23.1. Почему нужно новое кодирование

Канальное кодирование является фундаментальным аспектом беспроводной связи. От 2G до 5G в беспроводных системах всегда использовались самые современные технологии канального кодирования. Инновации в канальном кодировании, например сверточные коды для 2G, турбокоды для 3G и 4G, а также полярные коды и коды контроля четности с низкой плотностью (low-density parity-check code, LDPC) для 5G, ускорили развитие технологий кодирования. В соответствии с законом Мура продолжают появляться кодеры и декодеры со все более высокой производительностью и более низким энергопотреблением при меньших затратах, что позволяет использовать широкий спектр передовых технологий канального кодирования. По состоянию на 2020 год технология канального кодирования почти достигла теоретического предела Шеннона для канала аддитивного белого гауссова шума (additive white Gaussian noise, AWGN) с разумной стоимостью реализации.

Канальное кодирование используется для устранения влияния неблагоприятных и динамически меняющихся сред каналов, обеспечивая надежную связь и высокое качество обслуживания. Поскольку 6G будет охватывать множество разнообразных вариантов использования, некоторые из которых имеют экстремальные требования, такие как сверхвысокая скорость передачи данных, сверхнизкие задержка и энергопотребление, необходимы инновации в кодировании каналов 6G, способные обеспечить построение оптимального кода для заданного состояния канала и сценария использования. Этот вопрос обсуждается в разделе 23.2.

Фундаментальное значение теоремы Шеннона заключается в том, что она позволяет рассматривать канальное кодирование и кодирование источника информации как два подхода к достижению максимальной пропускной способности канала. Это верно, если размер информационного блока очень велик, но это увеличивает сложность реализации

и время ожидания как для кодера, так и для декодера. Таким образом, использование текущих схем раздельного кодирования источника и канала для достижения сверхвысоких скоростей передачи данных со сверхмалой задержкой не является оптимальным. Вместо этого темой исследований в области 6G станет межуровневый подход к совместной разработке кодирования источника и канала. Об этом пойдет речь в разделе 23.3. Такие исследования будут важны для оптимизации характеристик системы в применениях 6G.

Помимо классической теоремы Шеннона, есть повод пересмотреть более широкое определение интеллектуальной связи. Уоррен Уивер в своей совместной с Шенноном книге, опубликованной в 1949 году [1], предложил трехуровневую иерархию интеллектуальной связи: первый уровень – технические проблемы (это теорема Шеннона), второй уровень – семантические проблемы, а третий – проблемы эффективности. Что касается 6G, семантическая связь имеет множество применений в большинстве случаев использования 6G, особенно взаимодействия машина–машина и человек–машина. В иерархии интеллектуальной связи канал беспроводной передачи использует основанное на теореме Шеннона внутреннее канальное кодирование и внешний семантический канал. В этом семантическом канале коды источника извлекают интересующую семантику, отличную от простого сжатия семантики, в сторону информационной энтропии. Интеграция канала беспроводной передачи и внешнего семантического канала – это новая фундаментальная область исследований в 6G, которая уже привлекает некоторых исследователей [2]. Семантическое извлечение обычно выполняется с помощью машинного обучения с использованием метода, определяемого данными. Мы вернемся к этому позже в разделе 23.3.

Канальное кодирование может применяться к P2P-коммуникациям по каналам AWGN, а также к связи мультиточка–точка (MP2P), точка–мультиточка (P2MP) и мультиточка–мультиточка (MP2MP) по не-AWGN или нестационарным каналам. В сети 6G с большим количеством коммуникаций машина–машина и человек–машина для повышения общей эффективности использования спектра в сценариях связи MP2P, P2MP и MP2MP может применяться сетевое кодирование. Основная идея сетевого кодирования – объединить несколько пакетов в один для лучшего кодирования. Другой важный аспект использования сетевого кодирования – это подавление помех в цифровой или аналоговой области. Выигрыш от использования сетевого кодирования могут получить все устройства в сети. Сетевое кодирование обычно выполняется на прикладном уровне, но оно может выполняться и на физическом уровне в 6G из-за требований к задержке и пропускной способности. Этот момент более подробно обсуждается в разделе 23.4.

23.2. Схемы канального кодирования

23.2.1. История вопроса

С момента публикации теоремы Шеннона в 1948 году в различных системах связи были предложены и реализованы разнообразные схемы канального кодирования. В ранних схемах использовались алгебраические коды, чьи информационные биты и закодированные биты удовлетворяли линейным алгебраическим соотношениям. Их устройство было направлено на максимальное увеличение кодового расстояния с помощью линейных алгебраических факторов, чтобы максимизировать возможности исправления ошибок. Среди первых схем канального кодирования были коды Хэмминга [3], коды Голея [4] и коды Рида–Мюллера (RM) [5]. Эти конструкции кода были ограничительными по своей сути, поэтому поддерживаемая ими длина и скорость кода были ограничены. Открытие циклических кодов [6] привело к созданию гораздо более гибких схем кодирования. Двумя яркими примерами этого являются коды Бозе–Чаудхури–Хоквенгема (BCH) [7, 8] и коды Рида–Соломона [9]. Эти алгебраические циклические коды демонстрируют отличные характеристики кодового расстояния, обеспечивая хорошую производительность в программных декодерах [10]. Мощные схемы мягкого декодирования, такие как списочное декодирование [11] и декодирование на основе порядковой статистики [12], обеспечивают достоверность, достигающую максимального правдоподобия для многих коротких кодов.

Канальное кодирование стало ключевым фактором, помогающим беспроводным системам приблизиться к пределу Шеннона. Хотя алгебраические коды позволили передавать цифровые сигналы на большие расстояния, их достижения далеки от предела Шеннона из-за их относительно короткой длины. В системах беспроводной связи используются различные схемы кодирования. Например, стандарт 2G применял сверточные коды [13], тогда как 3G и 4G использовали турбокоды [14]. Современные схемы кодирования, такие как турбокоды, приближаются к теоретической пропускной способности канала за счет большего размера блока, что значительно улучшает спектральную эффективность мобильной связи. Успех турбоитеративного декодера привел к повторному открытию кодов контроля четности с низкой плотностью (LDPC). Было показано, что нерегулярные конструкции на основе кодов LDPC [15] приближаются к пределу Шеннона с зазором всего в 0,0045 дБ. Первыми кодами, которые доказали, что они достигают теоретической пропускной способности канала и сокращают разрыв, были полярные коды [16].

Канальное кодирование, позволяющее достичь теоретической пропускной способности, помогает улучшить многие аспекты пользова-

тельского опыта. Например, высокий коэффициент кодирования расширяет покрытие сети и улучшает качество обслуживания, особенно на границах соты. Кроме того, высокораспараллеленный декодер увеличивает пиковые скорости передачи данных; а простая, удобная для аппаратной реализации схема существенно снижает энергопотребление, тем самым увеличивая срок службы батареи. Эти преимущества сыграли ключевую роль в успехе беспроводных устройств предыдущих поколений. Стандарты 5G NR предусматривают использование полярных кодов и кодов LDPC для каналов управления и каналов данных соответственно.

Для сценариев использования 6G потребуются характеристики, превосходящие возможности 5G: более высокие скорости передачи данных, более высокая надежность, меньшая сложность и меньшее энергопотребление. Они также требуют более разнообразного набора KPI, которого не было в предыдущих поколениях. В следующих разделах обсуждаются ключевые требования и ключевые показатели эффективности, знание которых необходимо для проектирования кодирования каналов 6G, а также некоторые принципы проектирования.

23.2.2. Целевые KPI канального кодирования 6G

Канальное кодирование 6G продолжит улучшать показатели в существующих сценариях использования 5G, а именно в eMBB, URLLC и mMTC. Эти улучшения будут включать повышение пиковой скорости передачи данных до уровня Тбит/с (сегодняшняя скорость декодирования данных eMBB составляет 10–20 Гбит/с), устранение ошибок декодирования блока для URLLC и приближение показателей декодирования коротких блоков для mMTC к пределу производительности конечной длины [17]. Для других вариантов использования, таких как автономное вождение, промышленное оборудование и спутниковая связь, требуются специфичные целевые KPI, включая субмиллисекундную задержку, надежность от семи до десяти девяток, высокую энергоэффективность и длительное время автономной работы. На рис. 23.1 показаны KPI, достигаемые полярными кодами 5G NR для каналов управления и кодами LDPC 5G NR для каналов данных, по отношению к KPI разнесенного канального кодирования 6G. Некоторые из типичных целевых KPI описаны ниже.

- **Пропускная способность Тбит/с:** поскольку связь 6G станет коммерчески доступной к 2030 году, приложения VR/AR станут массовыми, и многие из них будут использовать более широкую полосу пропускания 6G для беспроводной передачи данных со скоростью 1 Тбит/с на короткие расстояния. Таким образом, ключевым KPI для 6G будет пиковая скорость передачи данных 1 Тбит/с.



Рис. 23.1. Различные целевые KPI для 6G

- Показатели кодов небольшой длины:** в самых разнообразных сценариях межмашинной связи будут использоваться коды небольшой длины (т. е. короче 200 бит). Оптимизация качества кодирования улучшит охват и пойдет на пользу многим соответствующим применениям. Однако в настоящее время схемы канального кодирования, используемые в 5G, имеют отставание в 1–2 дБ по сравнению с показателями кода конечной длины в силу ограничений практического декодера с разумной сложностью реализации. 6G не только сократит этот пробел, но и упростит алгоритмы декодирования для повышения энергоэффективности декодера.
- Низкая задержка кодирования и декодирования и длительный срок службы батареи:** ключевые показатели эффективности для конкретных применений, такие как высокая энергоэффективность, длительное время автономной работы и низкая сложность декодирования, зависят от схемы кода. Их следует рассматривать в сочетании с задержкой радиоинтерфейса, надежностью передачи и покрытием. Кроме того, схемы кода должны быть достаточно гибкими, чтобы адаптироваться к различным комбинациям KPI для разных сценариев использования. Например, применения с высокой надежностью и малой задержкой мо-

гут смириться с ограничениями покрытия или мобильности, тогда как другие применения могут допускать умеренную задержку декодирования, чтобы получить более высокий выигрыш от кодирования (например, для компенсации замираний, вызванных высокой мобильностью).

По мере развития кремниевых технологий для применений 6G станут доступными больше схем канального кодирования. И поскольку сложные алгоритмы декодирования могут быть реализованы с меньшими затратами, при проектировании 6G будут дополнительно исследованы компромиссы между увеличением производительности и сложностью реализации.

23.2.3. Принципы проектирования канального кодирования 6G

Единое и гибкое канальное кодирование

В 5G было сложно выбрать схему канального кодирования, которая удовлетворяет множеству требований в нескольких случаях использования – не существует единой схемы кодирования канала, подходящей для всех. Эта проблема останется в 6G, где различные варианты использования создадут новые проблемы для кодирования каналов. Поскольку настраиваемые схемы канального кодирования для конкретных случаев использования увеличивают сложность реализации, существует сильная мотивация либо разработать семейство кодов, которые повторно используют оборудование и подходят к различным сценариям и спецификациям, либо разработать структуру, которая эффективно адаптирует несколько кодов к различным применениям.

Такая структура, объединяющая коды компонентов или базовые коды, должна быть унифицированной и гибкой. Она также должна быть адаптивной, чтобы соответствовать требованиям различных сценариев использования 6G. Два примера устройства такой структуры кратко представлены ниже.

Пример 1: единая и гибкая структура на основе полярных кодов

Полярные коды демонстрируют превосходные характеристики в широком диапазоне длин и скоростей кода. Благодаря оптимизированной схеме полярные коды достигают стабильных показателей при одноканальной гранулярности [18] с точки зрения длины блока кодирования информации. Они также поддерживают несколько декодеров [19], в диапазоне от декодирования низкой сложности с последовательной отменой (successive cancellation, SC) до более сложного списочного SC-декодирования (successive cancellation list, SCL) с более высокими пока-

зателями. Также могут найти применение декодеры с мягкой отменой (soft cancellation) [20] или декодеры с распространением уверенности (belief propagation) для поддержки мягкого вывода битов в усовершенствованных итеративных приемниках. А в сценариях, где желательна высокая пропускная способность, может быть реализован декодер с переворачиванием битов (bit-flipping) с параллельным планированием. Хотя декодеры на основе SC по своей сути являются последовательными, для сценариев URLLC можно ввести некоторый параллелизм [21]. В сценариях mMTC, где типичны короткие блоки, полярные коды могут быть оптимизированы с точки зрения кодового расстояния для достижения сопоставимых характеристик с кодами Рида–Маллера и расширенными кодами Боуза–Чоудхури–Хоквингема [22]. Для приложений со сверхвысокой пропускной способностью применимы новые схемы G_N -смежно-групповых кодов (G_N -coset code, описаны ниже), которые поддерживают полностью распараллеленное декодирование [20]. Схемы на основе полярных кодов либо адаптируют декодер, либо оптимизируют коды для определенных целей. Таким образом, ядро декодирования и элементы обработки могут быть повторно использованы в аппаратных средствах, а описание кода может быть объединено в одну структуру.

Пример 2: адаптация нескольких кодов

Введение нескольких кодов требует схемы адаптации кода. Коды LDPC – особенно квазициклические Raptor-подобные²³ LDPC [23] – поддерживают схемы с мелкой гранулярностью и HARQ. Эти коды, которые были предметом обширных исследований, хорошо работают в сценариях с высокой скоростью и большой длиной блока. Однако их показатели и эффективность реализации резко снижаются при небольшой скорости и длине. В сценариях с очень короткой длиной блока полярные коды и алгебраические коды (такие как опорный сигнал BCH) более конкурентоспособны, чем коды LDPC [24]. В сценариях с очень большой длиной блока пространственно связанные коды [25] с компонентными кодами, такими как коды LDPC [26], полярные коды и другие обобщенные связанные структуры имеют потенциал достижения оптимального баланса между сложностью и производительностью. Схемы согласования скорости [27] и HARQ пространственно связанных кодов требуют дальнейшего изучения. Чтобы эффективно сочетать сильные стороны этих кодов, подход к адаптации кода может связывать каждую схему канального кодирования с конкретным вариантом использования или типом пользовательского устройства.

²³ Raptor (или Rapid Tornado) – это первый известный класс кодов с линейным кодированием и декодированием по времени. Изобретен Амином Шокроллахи в 2001 г. – *Прим. перев.*

Кодирование для высокой пропускной способности

Один из важнейших целевых KPI сети 6G – высокая пропускная способность. Чтобы справиться с более высокой сложностью и большим потреблением энергии, требуются мощные канальные кодеры и декодеры. Например, для достижения пропускной способности 1 Тбит/с требуется примерно 50-кратное увеличение пропускной способности по сравнению с ограниченными ресурсами 5G. В настоящее время ни коды LDPC, ни полярные коды не могут достичь эффективности площади кристалла более 1 Тбит/с на мм² и эффективности мощности менее 1 пДж/бит [28]. Декодер LDPC ограничен производительностью структуры на основе кремния, тогда как полярный декодер ограничен из-за своей архитектуры последовательного декодирования. Таким образом, целью дальнейших исследований является новая архитектура декодирования, которая обеспечивает высокую эффективность и быструю реализацию.

При разработке схемы кодирования для скоростей порядка 1 Тбит/с следует отдавать приоритет низкой сложности и высокой степени параллелизма декодирования, уделяя особое внимание аппаратной реализации [29]. Например, коды, реализованные с помощью более простых схем маршрутизации, потребуют меньшей площади кристалла и будут демонстрировать более высокую степень регулярности как при кодировании, так и при декодировании. В качестве альтернативы некоторые существующие коды могут быть оптимизированы для более высокого параллелизма, более высокой регулярности и более низкой сложности либо на стороне декодирования [30], либо как при кодировании, так и при декодировании [31].

Параллелизованный полярный код (семейство G_N -смежно-групповых кодов) включает полярные коды и коды Рида–Маллера и может поддерживать высокопроизводительное декодирование с параллелизмом и регулярностью. G_N -смежные коды могут быть линейными блочными кодами с той же образующей матрицей, что и полярные коды, но с разными наборами информации. Алгоритм параллельного декодирования может быть использован на факторном графе кодов G_N -кода [31]. Этот алгоритм демонстрирует как высокий параллелизм, так и высокую регулярность. Он обрабатывает G_N -смежные коды как коды конкатенации и декодирует внутренние коды параллельно. Чтобы избежать последовательного декодирования внешних кодов, алгоритм либо строит эквивалентный граф декодирования, меняя местами внешний и внутренний коды в предыдущем графе факторов, как показано на рис. 23.2, либо декодирует внутренние коды на обоих графах декодирования.

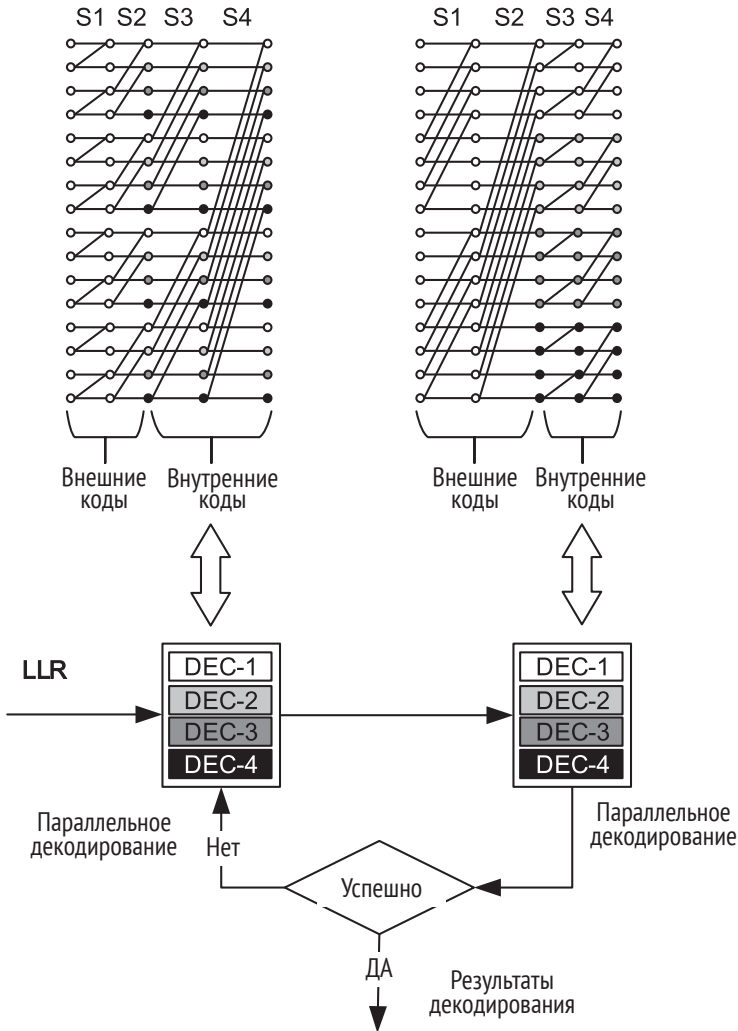


Рис. 23.2. Схема кодирования для скоростей порядка Тбит/с

Короткие коды

Чтобы сократить отставание в 1 дБ от теоретической границы исправления ошибок конечной длины, можно рассмотреть полярные коды и короткие алгебраические коды, такие как коды Рида–Маллера и Боуза–Чоудхури–Хоквингема / Рида–Соломона [24]. Полярные коды, обрабатываемые декодером списка последовательных отмен, обеспечивают оптимальный баланс между производительностью и сложностью в сценариях с малой длиной блока и поэтому подходят для каналов управления 5G. По той же причине в некоторых сценариях 6G, в которых будут

преобладать короткие блоки, также могут применяться полярные коды для каналов данных с дополнительным преимуществом повторного использования устаревшего оборудования 5G. С увеличением размера списка короткие полярные коды приближаются к границе случайного объединения кодов в области с низким коэффициентом ошибок [24]. Кроме того, для дальнейшей оптимизации конструкций кода для декодеров списков могут использоваться методы искусственного интеллекта [32, 33]. Были проведены некоторые исследования быстрых и упрощенных реализаций декодеров SCL [19, 34], и в настоящее время изучаются следующие области:

- показатели конечной длины можно улучшить, изменив поляризационные ядра, чтобы увеличить скорость поляризации [35]. К примерам этого подхода можно отнести сверточные полярные коды [36, 37], полярные коды с большими двоичными ядрами [38], полярные коды со смешанным ядром [39] и недвоичные полярные коды [40];
- для дальнейшего улучшения характеристик полярных кодов сверточные коды с коррективкой по поляризации (polarization-adjusted convolutional, PAC) [41] могут быть объединены с помощью полярных преобразований. Благодаря методам декодирования Фано эта схема способна обеспечить достижение теоретических границ;
- схемы кодирования, основанные на конкатенации внешнего кода, могут быть унифицированы с использованием предварительно преобразованных полярных кодов, где минимальное кодовое расстояние гарантированно улучшится [22].

Алгебраические коды в основном использовались в приложениях оптической связи и хранения данных, но для применения в беспроводной связи их схемы кодирования необходимо доработать, исходя из следующих принципов:

- кодовая длина имеет большое влияние на характеристики. Необходимо исследование того, как создавать эффективные коды короткого канала с хорошими характеристиками минимального расстояния;
- требуется адаптивность к различным размерам пакетов. Необходимо изучить практические методы адаптации скорости и длины для высокоэффективных коротких кодов;
- следует разработать новые алгоритмы декодирования, обеспечивающие баланс между вычислительной сложностью и сложностью реализации. В этом отношении декодирование на основе поряд-

ковой статистики [12] является жизнеспособным кандидатом, но требует дальнейшей оптимизации для минимизации сложности. Непрактичный декодер не может быть реализован, даже если он обеспечивает близкие к теоретическому пределу показатели.

Кодирование для критически важных приложений

В дополнение к обычным широкополосным приложениям 6G предусматривает множество критически важных применений, требующих гарантированно высочайших показателей. Существующие схемы кодирования далеко не всегда способны предоставить такие гарантии даже теоретически, поэтому в некоторых случаях необходимо разрабатывать коды для конкретных применений.

- **Чрезвычайно высокая надежность**, от 10^{-7} до 10^{-10} без нижней границы допустимых ошибок и с хорошим быстродействием. Хотя при использовании LDPC и других итеративно декодируемых кодов наличие минимального уровня ошибок неизбежно [42], его можно снизить ниже 10^{-10} . Теоретически было доказано, что полярные коды и связанные конструкции кодов не имеют минимального уровня ошибок и поэтому подходят для упомянутых применений. Для дальнейшего повышения надежности могут использоваться дополнительные функции, такие как недвоичные или сверточные конструкции.
- **Чрезвычайно низкая задержка**, от 1 мс до 0,1 мс, для поддержки автономного вождения и промышленной автоматизации. Эти приложения обычно требуют высокой надежности наряду с чрезвычайно низкой задержкой кодирования и декодирования. Здесь могут пригодиться алгебраические коды, но существующие схемы требуют декодирования на основе порядковой статистики, что слишком сложно для приложения с ограниченной задержкой. Для поддержки такого варианта использования необходимо разработать новые конструкции кода и схемы декодирования. Было продемонстрировано, что задержка декодирования для коротких полярных кодов очень мала [21] и может быть дополнительно снижена с помощью сегментированного декодирования [43] или параллельного декодирования [44].
- **Чрезвычайно низкая мощность и стоимость** для безбатарейных и энергособирающих конструкций со сроком службы от 10 лет до бесконечности. Интеллектуальные устройства будут модулировать и отражать принятые радиочастотные сигналы или собирать и накапливать радиочастотную энергию окружающей среды для сеансов радиочастотных передач. В таких устройствах

обычно используется недорогое оборудование, требующее чрезвычайно низкого энергопотребления. Это может быть достигнуто за счет совместной конструкции кодовой модуляции, в которой канальное кодирование разрабатывается совместно с формой сигнала и опорными сигналами [45].

- **Чрезвычайно высокая плотность подключения**, до 100 млн устройств на км². В плотно развернутых системах IoT следует ожидать возникновения многочисленных коллизий. Коды должны содержать как идентификаторы пользовательского оборудования, так и данные. Желательно иметь схему кодирования совместной последовательности с чрезвычайно большим пространством последовательности/кода, поддерживающую некогерентное детектирование/декодирование.

23.3. Совместное кодирование источника и канала

23.3.1. Предыстория исследований

Совместное кодирование источника и канала (joint source and channel coding JSCC) является предметом обширных исследований уже более 40 лет [46]. Верхние границы скорости кодирования канала для передач конечной длины были получены путем введения дисперсии канала [17]. Пару лет спустя были получены нижние границы для скорости кодирования источника с потерями при конечной длине блока с использованием исходной информации с d -наклоном [47]. Результаты этих работ были объединены в [48], послужив доказательством того, что JSCC превосходит раздельное кодирование источника и канала (SSCC) с точки зрения общей скорости в областях конечной длины. На рис. 23.3 показаны различия между JSCC и SSCC.

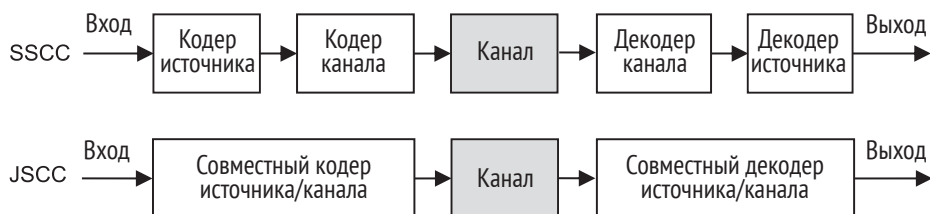


Рис. 23.3. Иллюстрация различий между SSCC и JSCC

Эти работы доказывают, что кодирование источника конечной длины часто бывает несовершенным, и последующий канальный декодер может использовать оставшуюся избыточность. Это вызвало серьезный

интерес к JSCC. Например, высокоуровневая или апостериорная информация кодов источника (например, Хаффмана [49], JPEG [50], арифметические [51] и Лемпеля–Зива [52]) может содействовать каналному декодированию на основе разреженных графов. В [53] была предложена совместная схема декодирования на основе полярных кодов и языкового декодера, в которой для исправления большинства ошибок декодирования на ранних стадиях используется словарь. Стремясь применять память источников, некоторые исследователи построили совместное декодирование марковского источника и канала на основе предположения, что некоторые источники могут быть смоделированы как марковский (или скрытый марковский) процесс [54, 55]. Для решения проблемы дистрибутивного сжатия многопользовательских источников были предложены различные модели источников и схемы совместного декодирования каналов, основанные на известной теореме Слепиана–Вольфа [56, 57]. Хотя эти реализации JSCC превосходят аналогичные SSCC, они сильно зависят от схем сжатия источника и страдают от плохой обобщенности. Другими словами, если схема сжатия источника изменяется, объединенный декодер должен быть перепроектирован заново.

Еще одна область JSCC, привлекающая большое внимание исследователей, – это беспроводная передача изображений/видео. Например, в [58] предложили схему аналоговой передачи под названием SoftCast. В этой схеме пиксели исходного изображения линейно кодируются в аналоговые коэффициенты и передаются через OFDM, а приемник декодирует пиксели с качеством непрерывного канала. Для решения проблемы низкой эффективности использования спектра и слабой степени сжатия, присущих аналоговой передаче, были предложены некоторые гибридные схемы аналогово-цифровой передачи. Например, в [59] предлагается передавать цифровой базовый уровень с помощью традиционных цифровых методов, а «уточняющий» уровень – с помощью псевдоаналогового и пространственного мультиплексирования. В [60] отображение Шеннона–Котельникова было адаптировано для дальнейшего сжатия аналоговых коэффициентов. Также недавно обсуждалась возможность использования JSCC для передачи изображений в дальний космос. Примером этого является работа [61], в которой предлагается использовать основанный на кодах Raptor способ линейного кодирования для замены традиционного энтропийного кодирования, тем самым повышая устойчивость к ошибкам декодирования и избегая чрезмерно длительного времени повторной передачи. Как показывают эти работы, JSCC может значительно уменьшить задержку передачи и обеспечить приемлемое качество изображения/видео даже при плохом состоянии канала. Однако все предлагаемые схемы нацелены на кон-

кретный сценарий: передачу изображения/видео. Совместная схема кодирования источника и кодирования каналов сильно интегрирована, что затрудняет широкое распространение такой конструкции.

Некоторые исследователи изучали возможность объединения методов машинного обучения с JSCC. Например, в [62] предлагается использовать для беспроводной передачи изображения JSCC на основе автоэнкодера, где зашумленный канал связи представлен как промежуточный уровень в паре кодер–декодер. В [63] и [64] была исследована совместная передача–распознавание на основе глубокой нейросети для беспроводной передачи данных на сервер с целью последующего распознавания. В дополнение к изображениям также может использоваться JSCC на основе машинного обучения для защиты передачи текста по зашумленному каналу, как описано в [65] и [66]. Подобно предыдущим работам, эти подходы JSCC с опорой на машинное обучение могут улучшить адаптируемость каналов и увеличить пространство оптимизации по сравнению с традиционными JSCC; тем не менее соответствующие схемы ориентированы на конкретную задачу, а не на обобщенное решение. В статьях [65] и [66] обсуждалось, как сохранить семантическую информацию предложений вместо минимизации количества битовых ошибок при беспроводной передаче. В следующем разделе мы исследуем ключевые моменты, связанные с извлечением и защитой семантики в машинных коммуникациях.

23.3.2. JSCC на основе машинного обучения

Развитие искусственного интеллекта и машинного обучения, несомненно, повлияет на JSCC. С точки зрения 6G семантическая связь Уивера имеет множество применений в большинстве сценариев использования, связанных с межмашинным обменом. Эта архитектура включает внутренний канал, основанный на классическом канальном кодировании, и внешний семантический канал, основанный на глубоком обучении. Интеграция этих двух каналов является предметом исследований, некоторые результаты уже опубликованы [2].

Одной из важных целей межмашинного взаимодействия 6G является создание новой парадигмы кодирования источника для связи между объектами искусственного интеллекта с помощью машинного обучения. В поисках продвинутого кодирования источника для агентов ИИ основной движущей силой является машинное обучение на основе глубоких нейронных сетей. В частности, обученная нейросеть разделяет многомерное пространство, при этом обучение рассматривается как «интерполятор», а его вывод рассматривается как «экстраполятор». И наоборот, глубокая нейросеть также является редуктором размерности, который сжимает или кодирует исходную информацию, удаляя не-

релевантные и менее важные компоненты из низкоразмерного представления. Кроме того, глубокая нейросеть может быть рассмотрена как расширитель измерений, декомпрессор или декодер источника из его низкоразмерного представления.

Как правило, нейросети кодера и декодера подключаются через архитектуру автоэнкодера, которая в целом отражает теорему скорости-искажения: увеличение размера узкого места уменьшает искажение, тогда как уменьшение размера увеличивает искажение. Теоретические основы семантической коммуникации обсуждались в главе 8. Возможные варианты реализации JSCC для связи между объектами искусственного интеллекта могут включать следующие варианты:

- **JSCC за пределами классической теории информации Шеннона:** целью классического кодирования источника является сжатие потока данных от источника информации. Это делается, чтобы минимизировать стоимость передачи и максимизировать пропускную способность сети. В большинстве случаев исходное кодирование оптимизировано в первую очередь для высокой достоверности человеческого восприятия. На передающей стороне данные максимально сжимаются с сохранением информации по Шеннону. На принимающей стороне получатель распаковывает информацию без влияния на качество восприятия человеком. В 6G, однако, целью кодирования источника является не сохранение точности человеческого восприятия, а, скорее, обеспечение межмашинной связи. Возможно дальнейшее сжатие потока данных, даже если кодирование источника применяется к большим данным в машинном обучении (т. е. при обучении глубокой нейросети);
- **новая метрика, выходящая за рамки теоремы скорости-искажения:** в практических применениях был достигнут оптимальный баланс между двумя классическими метриками: скоростью и искажением. Однако в общении между людьми такой баланс требует дополнительной метрики, то есть качества человеческого восприятия посредством оценки с участием человека. В исследовании 6G нам также необходимо изучить новую метрику для межмашинного взаимодействия, то есть аналогичный показатель машинного восприятия, оценка которого может быть реализована с помощью глубокой нейросети;
- **многополюсный JSCC:** классическая теория информации рассматривает кодирование многополюсного источника на основе теоремы Вольфа, которая предполагает, что два входных потока независимы и одинаково важны. Однако во многих практических сценариях 6G данные с датчиков могут не быть независимыми

или одинаково важными. Например, два датчика могут измерять один и тот же показатель физического мира, но делать это из разных мест или с использованием разных технологий. Кроме того, один датчик может собирать больше информации, чем другой, испытывать больше помех канала или зависеть от коэффициента использования канала другого датчика. Набирают популярность исследования многопользовательского JSCC, особенно в приложениях для построения глубокой нейросети, где мультисенсорные большие данные сжимаются для обучения машинной модели.

Глубокая нейросеть поддерживает многополюсные входы. Один такой вход должен быть назначен как основной вход для машинного обучения. Во время процесса обучения другие входные данные будут автономно взвешиваться и объединяться нейронами. В этом случае нейросеть будет отбрасывать неподходящую информацию, поступившую с датчика, присвоив ему низкий вес;

- **JSCC, зависящий от задачи:** традиционное кодирование источника фокусируется на искажении сквозного канала для удовлетворения требований человеческого восприятия. Однако требования машинного восприятия могут значительно различаться в таких аспектах, как обнаружение, классификация и реконструкция объекта. Достаточно очевидно, что обнаружение или классификация цели допускает большее сжатие информации, чем реконструкция цели. Посредством обучения энкодеров и декодеров глубокой нейросети могут быть реализованы различные задачи машинного обучения.

23.3.3. Принципы проектирования JSCC для 6G

По сравнению с SSCC, схема JSCC более эффективна при кодировании блоков конечной длины и ограничениях задержки. Многие существующие системы оптимизируют структуру JSCC на основе схемы сжатия (например, сжатие Хаффмана [49, 53]) или в зависимости от типа источника данных (например, изображение [61, 62] или текст [65, 66]). Однако в 6G есть три возможности нового применения JSCC. Первая – это базовая межмашинная беспроводная связь с короткими блоками и малой задержкой. Вторая – это высокоскоростная видеосвязь для виртуального мира, где гибридные аналоговые и цифровые JSCC могут обеспечить низкую задержку и низкое энергопотребление благодаря простой схеме сжатия, особенно для связи на малых расстояниях. В машинном обучении данные могут быть сжаты с помощью методов автокодирования, основанных на информационно узком месте. Это третья новая область применения – коммуникации между объектами ИИ в 6G.

Чтобы повысить эффективность JSCC и его применений в 6G, необходима обобщенная схема. Кроме того, для извлечения интересующей семантики в коммуникациях межмашинного типа могут использоваться методы машинного обучения. Некоторые соображения по разработке и направлениям дальнейших исследований перечислены ниже:

- желательно разработать обобщенную схему для JSCC. Такая схема должна быть применима к различным сценариям и не сильно зависеть от схемы сжатия источников;
- для соответствия требованиям различных сценариев нужна повышенная масштабируемость. С этой целью можно провести исследования JSCC на битовом уровне и на уровне сообщений, например для защиты от ошибок на битовом уровне и уровне сообщения соответственно;
- для JSCC необходимы новые опорные метрики проектирования, особенно в межмашинной связи. Такие метрики должны учитывать передачу с потерями, своевременность информации, управление вниманием, память источников и каналов и т. д.;
- для сценариев с большим количеством подключений, особенно для межмашинных взаимодействий, информация, совместно используемая разными пользователями, может сильно коррелировать. Эту корреляцию можно использовать для дальнейшего повышения эффективности JSCC;
- кодирование источника и обмен данными с использованием машинного обучения открывают более широкое пространство для оптимизации JSCC по сравнению с традиционными подходами. Разработки в [62, 66] заложили фундамент для новых подходов к разработке, но необходимы дальнейшие исследования. А для защиты конфиденциальности пользователей желательна система, основанная на «черном ящике» контента и «белом ящике» канала.

23.4. Сетевое кодирование на физическом уровне

23.4.1. Основы

Теорема Шеннона рассматривает связь «точка–точка» по сквозному каналу без памяти AWGN, тогда как сетевое кодирование²⁴ имеет дело с кодированием для многоточечных сетей. В этом случае сетевое коди-

²⁴ Сетевое кодирование – раздел теории информации, изучающий вопрос оптимизации передачи данных по сети с использованием техник изменения пакетов данных на промежуточных узлах.

рование на физическом уровне обеспечивает выигрыш в показателях за счет лучшего кодирования и разнесения узлов, особенно в случае достаточно большого перечня применений 6G, имеющих ограничения допустимой задержки. Из-за потенциального увеличения показателей представляет интерес дальнейшее исследование сетевого кодирования, в том числе даже для аналоговых сигналов (т. е. аналогового сетевого кодирования). Сетевое кодирование было впервые представлено Алсведе, Каем, Ли и Йунгом в 2000 году в их основополагающей статье [67]. Узел обработки кодирует входящие информационные потоки для повышения эффективности использования спектра. Чтобы выполнить это сетевое кодирование, узел кодирования принимает и объединяет несколько пакетов в объединенный пакет и передает его. Входные потоки могут поступать от одного или нескольких узлов, называемых внутри-сессионными или межсессионными соответственно. Более длинный кодовый блок означает более высокий выигрыш от кодирования, а разные пути передачи представляют некоторый выигрыш от разнесения. В статьях [68–70] обсуждаются несколько типов сетевого кодирования на физическом уровне, которые кратко перечислены ниже.

- **Линейное сетевое кодирование:** узел кодирования выполняет линейную комбинацию входных данных. Это может быть такая простая операция, как исключающее ИЛИ (XOR). Линейное сетевое кодирование относительно просто и широко используется для вещания и ретрансляции. Случайное линейное сетевое кодирование работает аналогично, но использует случайный коэффициент для преодоления неизвестных топологий или дистрибутивной связи.
- **Аналоговое сетевое кодирование (ANC):** оно работает непосредственно с аналоговыми сигналами, накладывая электромагнитные волны от S_1 и S_2 на релейном узле. ANC обычно используется в двусторонней релейной связи, как показано на рис. 23.4в.

Различные сценарии применения сетевого кодирования на физическом уровне

- **Сетевое кодирование для широковещательной/групповой/одноадресной связи:** в 5G появилась групповая передача с HARQ для связи по прямому каналу для повышения надежности. Однако эффективность и задержка HARQ могут быть дополнительно улучшены, особенно в сценариях широковещательной и групповой передачи, как показано на рис. 23.4а. Поскольку одна и та же часть информации передается множеству пользовательских устройств, вероятность потери информации высока, особенно для устройств в плохих условиях канала. Эта вероятность может

быть уменьшена путем использования сетевого кодирования (например, с XOR для повторной передачи HARQ) для объединения потерянных пакетов от разных приемников или одного и того же приемника [71, 72]. Во время одноадресной передачи информация, отправленная одному пользовательскому устройству, может быть подслушана другим устройством, то есть может быть разрешена повторная передача HARQ на основе сетевого кодирования [73].

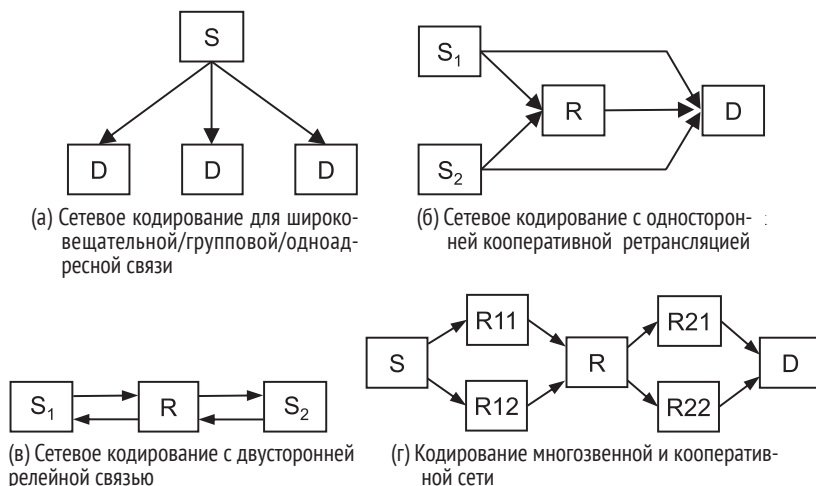


Рис. 23.4. Иллюстрации сценариев применения для сетевого кодирования

- **Сетевое кодирование с односторонней кооперативной ретрансляцией:** односторонняя релейная кооперативная связь повысит надежность. На рис. 23.4б и S_1 , и S_2 отправляют информацию к D в ортогональных ресурсных единицах в первой фазе, а ретранслирующий узел R пересылает информацию, полученную от S_1 и S_2 , в других ортогональных ресурсных единицах во второй фазе. Сетевое кодирование может поддерживать тот же порядок разнесения с аналогичным пространственным трактом, но с меньшим количеством ресурсов по сравнению со схемой с ортогональными единицами ресурсов во второй фазе [74]. На втором этапе узел R может выполнить операцию XOR между битами, принятыми от S_1 и S_2 . В то время как традиционная схема занимает четыре единицы ресурса, представленная схема кодирования сети на физическом уровне требует только трех для достижения того же порядка разнесения.
- **Сетевое кодирование с двусторонней релейной связью:** двусторонняя релейная связь позволит двум источникам, например S_1 и S_2 на рис. 23.4в, связываться друг с другом через релейный узел.

Если на втором этапе используется цифровое сетевое кодирование, требуются три единицы ресурса, тогда как если используется аналоговое сетевое кодирование, требуются только две единицы ресурса. Одним из типов аналогового сетевого кодирования является сетевое кодирование физического уровня, предложенное в работах [68, 69, 75]: на первом этапе два источника одновременно передают сигналы на ретранслятор, предполагая адекватную синхронизацию. После приема ретранслятор сначала выполняет канальное декодирование наложенного сигнала, затем преобразует наложенную информацию в информацию сетевого кодирования и наконец применяет канальное кодирование для широковещательной передачи двум источникам. Поскольку источникам неизвестны помехи, узлы-источники выполняют операции XOR, чтобы получить ожидаемый сигнал из широковещательного сигнала. В отличие от сетевого кодирования на физическом уровне, ретранслятор с аналоговым сетевым кодированием [70] просто усиливает и широковещательно передает полученный наложенный сигнал сетевого кодирования обоим источникам. Затем каждый узел-источник может декодировать свой ожидаемый сигнал на основе известного сигнала.

- **Кодирование многозвенной и кооперативной сети:** кооперативная многозвенная связь – это разновидность передачи в ячеистой сети, как показано на рис. 23.4г. Однако компромиссы между пропускной способностью, задержкой и надежностью для ячеистых сетей до конца не изучены. Одним из примеров схемы является адаптивное случайное сетевое кодирование [76, 77]. Было рассмотрено сетевое кодирование при многозвенной совместной передаче. Исходя из обратной связи и количества новых пакетов, используется как априорное, так и постериорное сетевое кодирование с прямой коррекцией ошибок. Между соседними переходами предлагается согласование скорости для выравнивания скорости тракта и максимизации пропускной способности [77].

23.4.2. Принципы проектирования сетевого кодирования 6G

Сетевое кодирование дает два существенных преимущества: выигрыш от кодирования и выигрыш от разнесения. С точки зрения теории информации, дополнительная информация поможет декодеру нейтрализовать известные радиопомехи. Таким образом, мы можем расширить область применения сетевого кодирования от широковеща-

тельной/групповой передачи до кооперативной и ретрансляционной передачи в 6G.

- **Мгновенно декодируемое сетевое кодирование:** для удовлетворения требований к сверхнизкой задержке и стоимости в 6G были предложены схемы мгновенно декодируемого сетевого кодирования (instantly decodable network coding, IDNC) [78, 79]. Кодеры и декодеры сетевого кодирования используют только простой двоичный XOR для уменьшения задержки и потребления памяти, несмотря на потерю информации. Прогрессивный декодер IDNC выгоден для связи, чувствительной к задержкам, но требует наличия достаточной обратной связи для получения статуса декодирования от приемника.
- **Совместное канальное и сетевое кодирование:** сетевое кодирование, применяемое на физическом уровне, рассматривается как «внешний код». Для повышения производительности канальное и сетевое кодирование можно разрабатывать вместе, разделяя между ними информацию об избыточности. Часть информации об избыточности включается в кодовое слово канала, а остальная часть включается в сетевое кодовое слово и передается по другому пути через узел ретрансляции [80]. Узел декодирования может использовать последовательную отмену для декодирования информации из обоих кодовых слов. Ожидается, что в 6G значительную долю пользователей будут составлять машины, которые могут работать вместе, формируя адаптивное сетевое кодирование [81]. Перевешивают ли выгоды от кодирования и разнесения преимущества избыточности, зависит от структуры дополнительной информации и разделения между совместно работающими пользователями.
- **Аналоговое сетевое кодирование (ANC):** ожидается, что ANC обеспечит меньшую задержку и более высокую эффективность. Схема ANC типа 1 (ANC-1) предложена в 2006 г.; были проведены как теоретические работы, так и проверка прототипа [69, 82, 83]. ANC-1 требует, чтобы две линии связи с релейным узлом были точно синхронизированы с точки зрения амплитуды, фазы, времени и частоты; однако достичь этого сложно [83], и может потребоваться более высокая нагрузка на калибровку и квитирование связи. ANC-2 [70] снижает требования к синхронизации, при этом обеспечивая низкую задержку, но усиливает как сигналы, так и шум. В сценариях с высоким отношением сигнал/шум желательно использовать ANC-2, потому что высокое отношение сигнал/шум может быть дополнительно подкреплено планированием лучей.

23.5. Источники

- [1] C. Shannon and W. Weaver, *The mathematical theory of communication*. University of Illinois Press, 1949.
- [2] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, *Deep learning enabled semantic communication systems*, arXiv preprint arXiv:2006.10685, 2020.
- [3] R. W. Hamming, *Error detecting and error correcting codes*, The Bell System Technical Journal, vol. 29, no. 2, pp. 147–160, 1950.
- [4] M. J. Golay, *Notes on digital coding*, Proc. IEEE, vol. 37, p. 657, 1949.
- [5] D. E. Muller, *Application of boolean algebra to switching circuit design and to error detection*, Transactions of the IRE Professional Group on Electronic Computers, no. 3, pp. 6–12, 1954.
- [6] E. Prange, *Cyclic error-correcting codes in two symbols*. Air Force Cambridge Research Centre, 1957.
- [7] A. Hocquenghem, *Codes correcteurs derreurs*, Chiffres, vol. 2, no. 2, pp. 147–56, 1959.
- [8] R. C. Bose and D. K. Ray-Chaudhuri, *On a class of error correcting binary group codes*, Information and Control, vol. 3, no. 1, pp. 68–79, 1960.
- [9] I. S. Reed and G. Solomon, *Polynomial codes over certain finite fields*, Journal of the Society for Industrial and Applied Mathematics, vol. 8, no. 2, pp. 300–304, 1960.
- [10] R. Silverman and M. Balser, *Coding for constant-data-rate systems*, Transactions of the IRE Professional Group on Information Theory, vol. 4, no. 4, pp. 50–63, 1954.
- [11] M. Sudan, *Decoding of reed solomon codes beyond the error-correction bound*, Journal of Complexity, vol. 13, no. 1, pp. 180–193, 1997.
- [12] M. P. Fossorier and S. Lin, *Computationally efficient soft-decision decoding of linear block codes based on ordered statistics*, IEEE Transactions on Information Theory, vol. 42, no. 3, pp. 738–750, 1996.
- [13] P. Elias, *Coding for noisy channels*, IRE Convention Record, vol. 3, pp. 37–46, 1955.
- [14] C. Berrou, A. Glavieux, and P. Thitimajshima, *Near Shannon limit error-correcting coding and decoding: Turbo-codes*, in Proc. ICC'93 – IEEE International Conference on Communications, vol. 2. IEEE, 1993, pp. 1064–1070.
- [15] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, *Design of capacity-approaching irregular low-density parity-check codes*, IEEE Transactions on Information Theory, vol. 47, no. 2, pp. 619–637, 2001.

- [16] E. Arıkan, *Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels*, IEEE Transactions on Information Theory, vol. 55, no. 7, pp. 3051–3073, 2009.
- [17] Y. Polyanskiy, H. V. Poor, and S. Verdú, *Channel coding rate in the finite blocklength regime*, IEEE Transactions on Information Theory, vol. 56, no. 5, pp. 2307–2359, 2010.
- [18] H. Zhang, R. Li, J. Wang, S. Dai, G. Zhang, Y. Chen, H. Luo, and J. Wang, *Paritycheck polar coding for 5G and beyond*, in Proc. 2018 IEEE International Conference on Communications (ICC). IEEE, 2018, pp. 1–7.
- [19] X. Liu, Q. Zhang, P. Qiu, J. Tong, H. Zhang, C. Zhao, and J. Wang, *A 5.16 Gbps decoder ASIC for polar code in 16nm FinFET*, in Proc. 2018 15th International Symposium on Wireless Communication Systems (ISWCS). IEEE, 2018, pp. 1–5.
- [20] J. Tong, H. Zhang, X. Wang, S. Dai, R. Li, and J. Wang, *A soft cancellation decoder for parity-check polar codes*, arXiv preprint arXiv:2003.08640, 2020.
- [21] H. Zhang, J. Tong, R. Li, P. Qiu, Y. Huangfu, C. Xu, X. Wang, and J. Wang, *A flipsyndrome-list polar decoder architecture for ultra-low-latency communications*, IEEE Access, vol. 7, pp. 1149–1159, 2018.
- [22] B. Li, H. Zhang, and J. Gu, *On pre-transformed polar codes*, arXiv preprint arXiv:1912.06359, 2019.
- [23] T.-Y. Chen, K. Vakilinia, D. Divsalar, and R. D. Wesel, *Protograph-based raptor-like LDPC codes*, IEEE Transactions on Communications, vol. 63, no. 5, pp. 1522–1532, 2015.
- [24] M. C. Coşkun, G. Durisi, T. Jerkovits, G. Liva, W. Ryan, B. Stein, and F. Steiner, *Efficient error-correcting codes in the short blocklength regime*, Physical Communication, vol. 34, pp. 66–79, 2019.
- [25] D. J. Costello, L. Dolecek, T. E. Fuja, J. Kliewer, D. G. Mitchell, and R. Smarandache, *Spatially coupled sparse codes on graphs: Theory and practice*, IEEE Communications Magazine, vol. 52, no. 7, pp. 168–176, 2014.
- [26] D. G. Mitchell, M. Lentmaier, and D. J. Costello, *Spatially coupled LDPC codes constructed from protographs*, IEEE Transactions on Information Theory, vol. 61, no. 9, pp. 4866–4889, 2015.
- [27] Z. Si, M. Andersson, R. Thobaben, and M. Skoglund, *Rate-compatible LDPC convolutional codes for capacity-approaching hybrid ARQ*, in Proc. 2011 IEEE Information Theory Workshop. IEEE, 2011, pp. 513–517.
- [28] Wikipedia, *B5G wireless Tb/s FEC KPI requirement and technology gap analysis*. <https://epic-h2020.eu/downloads/EPIC-D1.2-B5G-Wireless-Tbs-FEC-KPI-Requirement-and-Technology-Gap-Analysis-PU-M07.pdf>.

- [29] C. Kestel, M. Herrmann, and N. Wehn, *When channel coding hits the implementation wall*, in Proc. 2018 IEEE 10th International Symposium on Turbo Codes & Iterative Information Processing (ISTC). IEEE, 2018, pp. 1–6.
- [30] A. Süral, E. G. Sezer, Y. Ertuğrul, O. Arikan, and E. Arikan, *Tera-bits-per-second throughput for polar codes*, in Proc. 2019 IEEE 30th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC Workshops). IEEE, 2019, pp. 1–7.
- [31] X. Wang, H. Zhang, R. Li, J. Tong, Y. Ge, and J. Wang, *On the construction of G Ncoset codes for parallel decoding*, in Proc. 2020 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2020, pp. 1–6.
- [32] L. Huang, H. Zhang, R. Li, Y. Ge, and J. Wang, *AI coding: Learning to construct error correction codes*, IEEE Transactions on Communications, vol. 68, no. 1, pp. 26–39, 2019.
- [33] L. Huang, H. Zhang, R. Li, Y. Ge, and J. Wang, *Reinforcement learning for nested polar code construction*, in Proc. 2019 IEEE Global Communications Conference (GLOBECOM). IEEE, 2019, pp. 1–6.
- [34] S. A. Hashemi, C. Condo, and W. J. Gross, *Fast simplified successive-cancellation list decoding of polar codes*, in Proc. 2017 IEEE Wireless Communications and Networking Conference Workshops (WCNCW). IEEE, 2017, pp. 1–6.
- [35] S. B. Korada, E. Şaşıoğlu, and R. Urbanke, *Polar codes: Characterization of exponent, bounds, and constructions*, IEEE Transactions on Information Theory, vol. 56, no. 12, pp. 6253–6264, 2010.
- [36] A. J. Ferris, C. Hirche, and D. Poulin, *Convolutional polar codes*, arXiv preprint arXiv:1704.00715, 2017.
- [37] H. Saber, Y. Ge, R. Zhang, W. Shi, and W. Tong, *Convolutional polar codes: LLRbased successive cancellation decoder and list decoding performance*, in Proc. 2018 IEEE International Symposium on Information Theory (ISIT). IEEE, 2018, pp. 1480–1484.
- [38] P. Trifonov, *On construction of polar subcodes with large kernels*, in Proc. 2019 IEEE International Symposium on Information Theory (ISIT). IEEE, 2019, pp. 1932–1936.
- [39] N. Presman, O. Shapira, and S. Litsyn, *Polar codes with mixed kernels*, in Proc. 2011 IEEE International Symposium on Information Theory. IEEE, 2011, pp. 6–10.
- [40] R. Mori and T. Tanaka, *Non-binary polar codes using Reed-Solomon codes and algebraic geometry codes*, in Proc. 2010 IEEE Information Theory Workshop. IEEE, 2010, pp. 1–5.

- [41] E. Arıkan, *From sequential decoding to channel polarization and back again*, arXiv preprint arXiv:1908.09594, 2019.
- [42] T. Richardson, *Error-floors of ldpc codes*, in Proc. 41st Annual Conference on Communication, Control and Computing, 2003, pp. 1426–1435.
- [43] Huawei, *Details of the polar code design*, 3rd Generation Partnership Project (3GPP), Technical Report R1-1611254, Nov. 2016, 3GPP TSG RAN WG1 #87 Meeting.
- [44] B. Li, H. Shen, and D. Tse, *Parallel decoders of polar codes*, arXiv preprint arXiv:1309.1026, 2013.
- [45] N. Van Huynh, D. T. Hoang, X. Lu, D. Niyato, P. Wang, and D. I. Kim, *Ambient backscatter communications: A contemporary survey*, IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 2889–2922, 2018.
- [46] M. Fresia, F. Perez-Cruz, H. V. Poor, and S. Verdú, *Joint source and channel coding*, IEEE Signal Processing Magazine, vol. 27, no. 6, pp. 104–113, 2010.
- [47] V. Kostina and S. Verdú, *Fixed-length lossy compression in the finite blocklength regime*, IEEE Transactions on Information Theory, vol. 58, no. 6, pp. 3309–3338, 2012.
- [48] V. Kostina and S. Verdú, *Lossy joint source-channel coding in the finite blocklength regime*, IEEE Transactions on Information Theory, vol. 59, no. 5, pp. 2545–2575, 2013.
- [49] A. Guyader, E. Fabre, C. Guillemot, and M. Robert, *Joint source-channel turbo decoding of entropy-coded sources*, IEEE Journal on Selected Areas in Communications, vol. 19, no. 9, pp. 1680–1696, 2001.
- [50] L. Pu, Z. Wu, A. Bilgin, M. W. Marcellin, and B. Vasic, *LDPC-based iterative joint source-channel decoding for JPEG2000*, IEEE Transactions on Image Processing, vol. 16, no. 2, pp. 577–581, 2007.
- [51] M. Grangetto, P. Cosman, and G. Olmo, *Joint source/channel coding and MAP decoding of arithmetic codes*, IEEE Transactions on Communications, vol. 53, no. 6, pp. 1007–1016, 2005.
- [52] S. Lonardi, W. Szpankowski, and M. D. Ward, *Error resilient LZ’77 data compression: Algorithms, analysis, and experiments*, IEEE Transactions on Information Theory, vol. 53, no. 5, pp. 1799–1813, 2007.
- [53] Y. Wang, M. Qin, K. R. Narayanan, A. Jiang, and Z. Bandic, *Joint source-channel decoding of polar codes for language-based sources*, in Proc. 2016 IEEE Global Communications Conference (GLOBECOM). IEEE, 2016, pp. 1–6.
- [54] J. Garcia-Frias and J. D. Villasenor, *Joint turbo decoding and estimation of hidden Markov sources*, IEEE Journal on Selected Areas in Communications, vol. 19, no. 9, pp. 1671–1679, 2001.

- [55] G.-C. Zhu and F. Alajaji, *Joint source-channel turbo coding for binary Markov sources*, IEEE Transactions on Wireless Communications, vol. 5, no. 5, pp. 1065–1075, 2006.
- [56] J. Garcia-Frias and W. Zhong, *LDPC codes for compression of multi-terminal sources with hidden Markov correlation*, IEEE Communications Letters, vol. 7, no. 3, pp. 115–117, 2003.
- [57] K. Bhattad and K. R. Narayanan, *A decision feedback based scheme for Slepian–Wolf coding of sources with hidden Markov correlation*, IEEE Communications Letters, vol. 10, no. 5, pp. 378–380, 2006.
- [58] S. Jakubczak and D. Katabi, *Softcast: One-size-fits-all wireless video*, in Proc. ACM SIGCOMM 2010 Conference, 2010, pp. 449–450.
- [59] B. Tan, J. Wu, H. Cui, R. Wang, J. Wu, and D. Liu, *A hybrid digital analog scheme for MIMO multimedia broadcasting*, IEEE Wireless Communications Letters, vol. 6, no. 3, pp. 322–325, 2017.
- [60] F. Liang, C. Luo, R. Xiong, W. Zeng, and F. Wu, *Hybrid digital–analog video delivery with Shannon–Kotelnikov mapping*, IEEE Transactions on Multimedia, vol. 20, no. 8, pp. 2138–2152, 2017.
- [61] O. Y. Bursalioglu, G. Caire, and D. Divsalar, *Joint source-channel coding for deep-space image transmission using rateless codes*, IEEE Transactions on Communications, vol. 61, no. 8, pp. 3448–3461, 2013.
- [62] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, *Deep joint source-channel coding for wireless image transmission*, IEEE Transactions on Cognitive Communications and Networking, vol. 5, no. 3, pp. 567–579, 2019.
- [63] C.-H. Lee, J.-W. Lin, P.-H. Chen, and Y.-C. Chang, *Deep learning-constructed joint transmission-recognition for internet of things*, IEEE Access, vol. 7, pp. 76 547–76 561, 2019.
- [64] M. Jankowski, D. Gündüz, and K. Mikolajczyk, *Deep joint source-channel coding for wireless image retrieval*, in Proc. 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020, pp. 5070–5074.
- [65] N. Farsad, M. Rao, and A. Goldsmith, *Deep learning for joint source-channel coding of text*, in Proc. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018, pp. 2326–2330.
- [66] M. Rao, N. Farsad, and A. Goldsmith, *Variable length joint source-channel coding of text using deep neural networks*, in Proc. 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2018, pp. 1–5.
- [67] R. Ahlswede, N. Cai, S.-Y. Li, and R. W. Yeung, *Network information flow*, IEEE Transactions on Information Theory, vol. 46, no. 4, pp. 1204–1216, 2000.

- [68] S. T. Başaran, G. K. Kurt, M. Uysal, and T. Altunbaş, *A tutorial on network coded cooperation*, IEEE Communications Surveys & Tutorials, vol. 18, no. 4, pp. 2970–2990, 2016.
- [69] S. Zhang, S. C. Liew, and P. P. Lam, *Hot topic: Physical-layer network coding*, in Proc. 12th Annual International Conference on Mobile Computing and Networking, 2006, pp. 358–365.
- [70] S. Katti, S. Gollakota, and D. Katabi, *Embracing wireless interference: Analog network coding*, ACM SIGCOMM Computer Communication Review, vol. 37, no. 4, pp. 397–408, 2007.
- [71] D. Nguyen, T. Tran, T. Nguyen, and B. Bose, *Wireless broadcast using network coding*, IEEE Transactions on Vehicular Technology, vol. 58, no. 2, pp. 914–925, 2008.
- [72] Z. Zhang, T. Lv, X. Su, and H. Gao, *Dual XOR in the air: A network coding based retransmission scheme for wireless broadcasting*, in Proc. 2011 IEEE International Conference on Communications (ICC). IEEE, 2011, pp. 1–6.
- [73] H. Zhu, B. Smida, and D. J. Love, *Optimization of two-way network coded HARQ with overhead*, IEEE Transactions on Communications, vol. 68, no. 6, pp. 3602–3613, 2020.
- [74] Y. Chen, S. Kishore, and J. Li, *Wireless diversity through network coding*, in Proc. 2006 IEEE Wireless Communications and Networking Conference, vol. 3. IEEE, 2006, pp. 1681–1686.
- [75] S. Zhang and S.-C. Liew, *Channel coding and decoding in a relay system operated with physical-layer network coding*, IEEE Journal on Selected Areas in Communications, vol. 27, no. 5, pp. 788–796, 2009.
- [76] A. Cohen, G. Thiran, V. B. Bracha, and M. Médard, *Adaptive causal network coding with feedback for multipath multi-hop communications*, in Proc. 2020 IEEE International Conference on Communications (ICC). IEEE, 2020, pp. 1–7.
- [77] A. Cohen, D. Malak, V. B. Brachay, and M. Medard, *Adaptive causal network coding with feedback*, IEEE Transactions on Communications, vol. 68, no. 7, pp. 4325–4341, 2020.
- [78] A. Douik, S. Sorour, T. Y. Al-Naffouri, and M.-S. Alouini, *Instantly decodable network coding: From centralized to device-to-device communications*, IEEE Communications Surveys & Tutorials, vol. 19, no. 2, pp. 1201–1224, 2017.
- [79] M. S. Karim et al., *Instantly decodable network coding: From point to multi-point to device-to-device communications*, Ph.D. dissertation, Australian National University, March 2017.

- [80] C. Hausl and P. Dupraz, *Joint network-channel coding for the multiple-access relay channel*, in Proc. 2006 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks, vol. 3. IEEE, 2006, pp. 817–822.
- [81] X. Bao and J. Li, *A unified channel-network coding treatment for user cooperation in wireless ad-hoc networks*, in Proc. 2006 IEEE International Symposium on Information Theory. IEEE, 2006, pp. 202–206.
- [82] S. C. Liew, S. Zhang, and L. Lu, *Physical-layer network coding: Tutorial, survey, and beyond*, Physical Communication, vol. 6, pp. 4–42, 2013.
- [83] Y. Tan, S. C. Liew, and T. Huang, *Mobile lattice-coded physical-layer network coding with practical channel alignment*, IEEE Transactions on Mobile Computing, vol. 17, no. 8, pp. 1908–1923, 2018.

Глава 24

Новый множественный доступ

24.1. Почему нужен новый множественный доступ

В беспроводной связи данные передаются от одного устройства к другому через радиоресурсы. Основными физическими радиоресурсами являются время и частота. Проблема *множественного доступа* (multiple access, MA) возникает, когда необходимо обслужить несколько пользователей, используя ресурсы с ограниченными степенями свободы. Схема MA не может состоять только из одного модуля. Вместо этого она обычно включает в себя несколько модулей обработки сигналов на физических каналах, таких как кодирование, модуляция, предварительное кодирование, картирование ресурсов, управление мощностью и даже формы сигналов. В коммуникациях типа «точка–точка» пропускная способность одного сквозного канала максимальна. Подход множественного доступа отличается от этого тем, что некоторые модули могут быть спроектированы специально для совместной работы нескольких пользователей или так, чтобы максимизировать пропускную способность системы в целом. Это особенно актуально, когда количество пользователей велико и доступные степени свободы являются общими для всех пользователей.

В 6G, как описано во второй части книги, будут обслуживаться разнообразные клиенты с жесткими и экстремальными требованиями. Это влечет за собой необходимость в масштабируемой структуре MA, при которой необходимо учитывать следующие соображения.

- **Все размеры пакетов:** с учетом вариантов использования разнообразных услуг, описанных в части II этой книги, в 6G механизм MA должен быть адекватно масштабируемым, чтобы иметь дело со всеми типами размеров пакетов. Они могут варьироваться от

огромных голограмм в иммерсивных службах XR до нескольких байтов обновлений статуса промышленных датчиков или датчиков мониторинга состояния здоровья. Разные размеры пакетов соответствуют разной продолжительности передачи и разным уровням диверсификации.

- **Все категории устройств:** как обсуждалось в главе 18, смартфоны 6G больше не будут единственными доминирующими устройствами. К одной и той же сети будут иметь доступ различные категории устройств, от высокоинтеллектуальных автомобилей до роботов и чрезвычайно недорогих датчиков. Огромные различия в возможностях этих устройств по обработке сигналов, временной и частотной синхронизации, параметрам оборудования и средствам питания могут предъявлять очень разные требования к МА. Ожидается, что количество подключений устройств частного, совместного и промышленного применения увеличится в десять раз за 10 лет. Эта перспектива заставляет нас учитывать потребность в сверхмассивных подключениях при проектировании множественного доступа в сетях 6G.
- **Все типы трафика:** разнообразие услуг также приводит к разнообразию типов трафика. Это отражается в различных схемах поступления трафика, а также в разнообразных требованиях к скорости передачи данных, задержке и надежности. Схема МА в сети 6G должна быть достаточно масштабируемой, чтобы эффективно обслуживать пользователей с длинными пакетами, периодическими и спорадическими моделями трафика. Эффективное мультиплексирование пользователей с различными крайними требованиями к производительности – еще одна проблема, которая до сих пор полностью не решена в 5G даже с использованием трафика eMBB, URLLC и mMTC. Попытка возложить всю нагрузку на планировщик MAC не будет достаточно эффективным решением, поэтому следует рассмотреть некоторые межуровневые схемы.
- **Все сценарии развертывания:** как обсуждалось в главах 6 и 20, сеть 6G будет обеспечивать трехмерное покрытие с NAPS, БПЛА и VLEO как часть RAN. Расширение потенциальных сценариев развертывания поставит новые задачи для проектирования 6G МА. Одной из типичных проблем для интегрированного неземного доступа является потенциально большое рассогласование по времени и ограниченный бюджет канала из-за больших расстояний распространения. Дополнительной новой областью для исследования станут механизмы МА, включающие совместную передачу между несколькими уровнями 3D RAN.

24.2. Обзор существующих решений

С точки зрения мультиплексирования ресурсов методы МА делятся на две категории: *ортогональный множественный доступ* (orthogonal multiple access, ОМА) и *неортогональный множественный доступ* (non-orthogonal multiple access, NOMA). Кроме того, существуют две схемы, связанные с процедурами передачи с точки зрения планирования ресурсов: передачи со служебной информацией (grant-based, GB) и передачи без служебной информации или *безгрантовой передачи* (grant-free, GF). Эти методы МА и схемы передачи привлекли внимание исследователей и разработчиков стандартов 5G [1]. Более того, оба метода МА применимы как к передачам типа GB, так и GF. Далее мы описываем достижения в этих областях.

24.2.1. Ортогональный множественный доступ

В предыдущих поколениях ортогональный множественный доступ был доминирующей схемой. «Ортогональный» указывает, что существует независимая степень свободы, принадлежащая каждому пользователю, по крайней мере в одной из областей, включая частоту, время, код и пространство. Это означало, что на передачу пользователя не влияют другие пользователи, обслуживаемые одновременно. Исходя из ресурсов, которые разделяют, чтобы различать множество пользователей, можно выделить следующие схемы ОМА: множественный доступ с частотным разделением (FDMA), множественный доступ с временным разделением (TDMA), множественный доступ с кодовым разделением (CDMA), множественный доступ с пространственным разделением каналов (SDMA) и множественный доступ с ортогональным частотным разделением каналов (OFDMA) [2].

- **FDMA:** для подключения нескольких пользователей к сети система FDMA выделяет неперекрывающиеся полосы частот каждому пользователю или потоку данных в качестве выделенных каналов. Чтобы полностью устранить помехи для соседних пользователей, следует использовать защитные полосы. Это снижает эффективность использования спектра системы в целом. Таким образом, FDMA недостаточно эффективен и не позволяет огромному количеству пользователей получить доступ к сети.
- **TDMA:** система TDMA делит временную область на множество временных интервалов (тайм-слотов). Передатчик использует эти слоты для передачи разных сигналов разным пользователям. Биты полезной нагрузки каждого пользователя разделяются по времени и отправляются пакетами, когда доступны соответствующие временные интервалы. Таким образом, в сети можно размес-

тить только ограниченное количество пользователей или потоков данных.

- **CDMA:** в системе CDMA и временные, и частотные ресурсы могут одновременно использоваться несколькими пользователями или потоками данных. Различные пользователи или потоки данных различаются ортогональными или полуортогональными (также известными как почти ортогональные) кодами расширения. Очень длинные последовательности необходимы для достижения хорошего выигрыша при обработке, а ортогональность нужна для размещения большого количества пользователей. Эти требования делают его недостаточно гибким для работы с большой полосой пропускания и массивным MIMO.
- **OFDMA:** этот стандарт был разработан на основе сигналов OFDM. Он обеспечивает жесткую и ортогональную упаковку поднесущих в частотной области с разнесением поднесущих, обратным длительности символа. В системе OFDMA временная и частотная плоскость делятся на двумерный растр. Минимальная единица раstra называется ресурсным элементом (resource element, RE), состоящим из одной поднесущей (или тона) в частотной области и одного символа во временной области. Каждый RE передает модулированный символ, принадлежащий одному пользователю или потоку данных. Таким образом достигается максимальная гибкость ресурсов, которые можно легко комбинировать с такими функциями, как массивный MIMO. Однако стандарт предъявляет высокие требования к временной и частотной синхронизации, обслуживание которой может быть слишком дорогостоящим для очень дешевых и маломощных устройств.
- **SDMA:** с развитием современных технологий MIMO появился стандарт SDMA – вариант для использования с массивными антенными решетками. В системе SDMA аналоговый или цифровой формирователь луча может создавать несколько сфокусированных пространственных лучей для выделенных запланированных пользователей. В сочетании с идеальной информацией о состоянии канала лучи для запланированных пользователей могут быть почти ортогональными друг другу, как это требуется для минимизации межпользовательских помех. Однако ортогональность SDMA во многом зависит от точности информации о канале. К сожалению, ее становится трудно отслеживать из-за растущего числа пользователей и спорадического трафика.

Хотя, как показано выше, для улучшения пропускной способности системы, удобства пользователей и увеличения количества подключений были разработаны различные схемы на основе ОМА, все еще остаются

ограничения и пробелы, которые необходимо устранить, чтобы удовлетворить разнообразные и очень жесткие требования к услугам в 6G. В частности, можно выделить следующие проблемы:

- **ограничение количества одновременно обслуживаемых пользователей:** количество пользователей или потоков данных строго ограничено количеством ортогональных каналов в системе ОМА;
- **высокие накладные расходы на служебные данные и ресурсы для гарантии ортогональности:** ортогональность в системе ОМА обычно гарантируется предоставлением ресурсов, отправляемых базовыми станциями пользователям до того, как может начаться какая-либо передача в любом направлении, включая восходящие, нисходящие или прямые линии связи. Когда количество пользователей увеличивается, накладные расходы на обработку сигналов увеличиваются пропорционально, уменьшая пропускную способность системы в целом. Более того, для определенных услуг IoT динамическая обработка грантов может либо вводить недопустимую задержку, либо занимать более 50 % полезной нагрузки при каждой передаче. Ни тот, ни другой случаи неприемлемы с точки зрения эффективности использования спектра и энергоэффективности. Ортогональность может быть гарантирована с помощью предварительно сконфигурированных ресурсов, но накладные расходы на ресурсы все равно остаются большими, особенно когда трафик ведет себя непредсказуемо и спорадически;
- **сильная зависимость от информации о канале:** в реальных сценариях развертывания производительность многопользовательской системы MIMO с обратной связью (MU-MIMO) и координированной многоточечной передачи/приема (coordinated multipoint, CoMP) все еще далека от теоретических границ, особенно для пользователей высокоскоростной мобильной связи. Основная причина этого – их сильная зависимость от точного знания информации о канале для предварительного кодирования с обратной связью. Однако если случаются нарушения сети, получение точной информации невозможно, и качество работы систем MU-MIMO или CoMP с обратной связью соответственно ухудшается. Это может произойти в случаях старения канала, задержки обратной связи или выраженных межсотовых помех, которые обычно встречаются на практике.

24.2.2. Неортогональный множественный доступ

Чтобы преодолеть ограничения ОМА, была разработана технология неортогонального множественного доступа (NOMA), которая терпима

к конфликту ресурсов в ортогональных каналах. С точки зрения теории сетевой информации [3, 4] как для восходящего канала MA [5], так и для нисходящего широкополосного канала [6] совокупная пропускная способность для нескольких пользователей NOMA может быть больше по сравнению с параллельной передачей для одного и того же набора пользователей на ортогонально разделенных ресурсах. Доказано, что такие методы, как последовательное подавление помех [5] и «кодирование на грязной бумаге» (предварительное кодирование Коста) [6] на базовых станциях, позволяют достичь оптимальных значений пропускной способности для канала MA и канала BC соответственно. Более того, NOMA может поддерживать дальнейшую суперпозицию сигналов в ортогональном подпространстве при передаче MU-MIMO. Одно недавнее исследование [7] показало, что, интегрируя кодовую часть схемы NOMA в передачу MU-MIMO восходящей линии связи, можно получить дополнительный выигрыш в пропускной способности даже для системы с 64 приемными антеннами.

Во многих сценариях применений важна не пропускная способность для одного пользователя, а количество пользователей, которые могут быть обслужены с гарантированной целевой скоростью. В этом смысле NOMA можно рассматривать как механизм, который обменивает пиковую скорость передачи данных одного соединения на большее количество соединений [8]. Это достигается за счет использования приемопередатчиков NOMA для подавления межпользовательских помех. Это чрезвычайно полезно в IoT, где есть много устройств, но при этом каждое устройство не предъявляет очень высокие требования к скорости передачи данных. В сочетании с MU-MIMO NOMA больше способствует увеличению количества обслуживаемых пользователей, чем простое наращивание количества передающих или приемных антенн.

Таким образом, при хорошей конструкции приемопередатчика NOMA может обеспечить следующие неотъемлемые преимущества перед OMA:

- **увеличение пропускной способности многопользовательской системы:** как доказано теорией сетевой информации, NOMA дополнительно увеличивает пропускную способность системы, не используя дополнительные спектральные ресурсы или антенны;
- **поддержка перегруженной передачи:** NOMA дополнительно увеличивает общее количество подключений за счет введения допустимой коллизии символов в ортогональных каналах. В таких случаях система считается *перегруженной*. Благодаря подходящей схеме многопользовательского детектирования (MUD) и кодовым книгам NOMA, которые могут включать многомерные комбина-

ции [9, 10], сигнатуры расхождения [11] и шаблоны конфликтов ресурсов [12, 13], может быть достигнут коэффициент перегрузки выше 300 %. Здесь коэффициент перегрузки определяется как соотношение между количеством пользователей, к которым одновременно осуществляется доступ, и количеством ортогональных RE;

- **обеспечение надежной безгрантовой передачи:** как будет описано ниже, чтобы избежать больших накладных расходов на сигнальный механизм и уменьшить задержку установления связи, основанной на использовании служебной информации, была предложена безгрантовая передача, особенно подходящая для услуг с небольшими размерами пакетов и спорадическим трафиком. Обладая устойчивостью к конфликтам символов, безгрантовый механизм с поддержкой NOMA может реализовать более агрессивное совместное использование ресурсов между большим количеством пользователей, сохраняя при этом тот же уровень целевой надежности;
- **обеспечение надежного MU-MIMO без обратной связи:** чтобы преодолеть узкое место, связанное с получением информации о канале в реальном времени при массивных передачах в системе MIMO с петлей обратной связи, NOMA предоставляет альтернативу MU-MIMO без обратной связи. Поскольку этот механизм не полагается на точную информацию о канале, он может быть более устойчивым к изменениям сети, таким как старение канала и мобильность пользователей. Этот механизм без обратной связи может быть дополнительно расширен для случаев с несколькими базовыми станциями, в которых вместо сбора точной информации о канале от целевых пользователей и последующей отправки этой информации на каждую партнерскую базовую станцию для совместного предварительного кодирования каждая из партнерских базовых станций может выбирать неортогональный код и совместно передавать потоки данных без обмена информацией о канале с пользователями или другими базовыми станциями. Это особенно полезно, когда точная информация недоступна или ее получение обходится слишком дорого;
- **обеспечение гибкого мультиплексирования услуг:** традиционным подходом ОМА для эффективного обслуживания диверсифицированного трафика является динамическое планирование ресурсов. Этот подход требует дополнительных служебных данных и может оказаться недостаточно быстрым для соответствия требованиям к минимальной задержке в некоторых сценариях. Благодаря суперпозиционной природе NOMA небольшие пакеты

с малой задержкой могут быть наложены поверх больших пакетов для совместной передачи, тем самым улучшая задержку и одновременно сокращая накладные расходы.

NOMA в нисходящем канале

Если говорить о NOMA в нисходящем канале, базовые станции передают сумму сигналов для нескольких пользователей, которые занимают одинаковые частотные и временные ресурсы. Обычно базовые станции назначают в одну группу NOMA нескольких пользователей с существенно разными отношениями сигнал/шум. Например, если взять двух пользователей, которым запланированы одни и те же ресурсы и назначены разные сигнатуры NOMA, то один пользователь находится рядом с базовой станцией, а другой – далеко от нее; передатчик будет выделять высокую мощность передаче дальнему пользователю и низкую мощность – расположенному поблизости. На приемной стороне дальний пользователь может рассматривать сигналы ближнего пользователя как шум, мощность которого намного ниже, чем мощность его собственных сигналов; ближний пользователь может отличать свои собственные сигналы от сигналов удаленного пользователя, детектируя и восстанавливая последние.

В 3GPP исследование нисходящего канала NOMA было начато в R13 в рамках исследовательской темы под названием «многопользовательская суперпозиционная передача» (multiuser superposition transmission, MUST) [14]. MUST делится на три категории:

- **MUST категория 1:** закодированные биты двух или более совместно запланированных пользователей независимо отображаются в символы составных комбинаций, но для составной комбинации не выполняют сопоставление символов по Грей;
- **MUST категория 2:** закодированные биты двух или более совместно запланированных пользователей совместно отображаются в символы составных комбинаций, а затем для составной комбинации выполняют сопоставление символов по Грей;
- **MUST категория 3:** закодированные биты двух или более совместно запланированных пользователей напрямую отображаются в символы составной комбинации.

Для большинства существующих механизмов нисходящего канала NOMA в каждой группе NOMA есть только два пользователя, и общая пропускная способность по-прежнему считается основным критерием для разработки или выбора проекта. Потенциальное добавление в NOMA кодовой области может привести к уменьшению необходимого разрыва уровней сигнал/шум между пользователями, что еще больше

упростит группирование пользователей. Некоторые начальные подходы к разработке можно найти в [15]. Более того, при разработке новых версий NOMA для сетей 6G, вероятно, придется учитывать новые показатели, такие как задержка при планировании нисходящего канала для более разнообразных типов трафика.

NOMA в восходящем канале

В случае восходящего канала связи NOMA несколько пользователей передают свои сигналы на базовую станцию через одни и те же временные и частотные ресурсы [16]. Схема работы восходящего канала NOMA более сложна, поскольку до того, как несколько сигналов данных от разных пользователей мультиплексируются вместе, каждый пользователь получает случайный канал. Эта особенность исключает предварительное проектирование составной комбинации с наложением, как в категориях MUST 2 и 3. Кроме того, при проектировании необходимо учитывать случайный характер каналов конкретных пользователей. Исследование восходящего канала NOMA для 5G NR было начато в релизе 3GPP R14. Для передачи по восходящему каналу было предложено множество различных схем NOMA с поддержкой массового соединения и надежных GF-передач. Затем рассмотрение этого вопроса было продолжено в релизе R15 в качестве отдельного предмета для изучения [17]. На сегодняшний день ключевые аспекты NOMA в восходящем канале следующие:

- **унифицированная структура приемопередатчика:** исследование NOMA охватывало схемы передачи на предмет того, как уменьшить межпользовательские помехи, и схемы приема на предмет того, как справляться с оставшимися межпользовательскими помехами, чтобы поддерживать большее количество одновременных передач. Унифицированная структура приемопередатчика [18] для NOMA восходящего канала на основе сигналов OFDM изображена на рис. 24.1. Выполняемые на стороне передатчика специфические для пользователя операции на уровне битов и/или символов предназначены для облегчения декодирования наложенных многопользовательских данных на стороне приемника с разумной сложностью. На стороне приемника применяются усовершенствованные многопользовательские технологии приема для лучшего подавления межпользовательских помех с учетом затрат на реализацию и задержки;
- **сигнатура NOMA на стороне передатчика:** некоторые типичные примеры сигнатуры NOMA – это множественный доступ с разреженным кодом (SCMA) [9, 10], многопользовательский совместный доступ (MUSA) [11], множественный доступ с

распределением ресурсов (RSMA) [19], множественный доступ с разделением по шаблону (PDMA) [12], множественный доступ с чередованием сетки (IGMA) [13] и множественный доступ с разделением-чередованием (IDMA) [20]. Все эти схемы могут быть охарактеризованы различными операциями на уровне битов и символов, которые в исследовании NOMA определены как *сигнатуры* (signature).

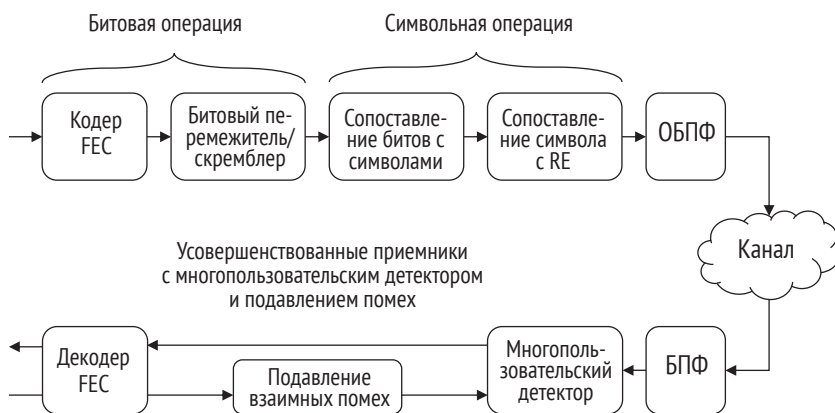


Рис. 24.1. Унифицированная структура приемопередатчика восходящего канала NOMA на основе сигналов OFDM

Операции на уровне битов включают в себя специфичные для пользовательского оборудования кодирование на уровне битов, скремблирование и перемежение. IDMA использует в качестве сигнатуры МА функцию перемежения на уровне битов, чтобы различать других пользователей. Операции на уровне символов включают в себя расширение, модуляцию, скремблирование и перемежение на уровне символов, специфичное для пользовательского оборудования.

Важным типом сигнатуры МА являются расширяющие последовательности на уровне символов со свойствами низкой взаимной корреляции или низкой плотности. Для NOMA предлагаются различные схемы расширяющих последовательностей на уровне символов, такие как последовательности с равенством оценке Уэлча, комплексные последовательности с квантованными элементами и грассмановы последовательности. Кроме того, для повышения эффективности расширение и модуляция на уровне символов могут быть разработаны совместно, например в SCMA;

- **реализация NOMA на стороне приемника:** из-за неортогональных свойств систем NOMA неизбежны межпользовательские

помехи на стороне приемника таких систем. Конструкция приемника NOMA является ключом к подавлению межпользовательских помех. Унифицированная итерационная схема приемника NOMA [21] состоит из двух частей: (1) алгоритм многопользовательского детектирования (MUD) и (2) структура итеративного подавления помех (IC). Алгоритмы MUD с низкой сложностью и малой задержкой, а также структуры подавления помех очень важны для широкого применения NOMA, поскольку обеспечивают баланс между производительностью и затратами на реализацию.

Что касается структур подавления помех, недавние исследования [22] показали, что структура гибридной параллельной интегральной схемы (PIC) обеспечивает наилучший компромисс между производительностью и сложностью. В гибридной структуре PIC существуют итерации внешнего цикла между MUD и канальными декодерами, а программная и жесткая части интегральной схемы работают параллельно. Это означает, что для пользователей, чьи информационные потоки успешно декодированы, их передаваемые сигналы восстанавливаются с использованием декодированных информационных битов и исключаются из всех принятых сигналов. Для пользователей, чьи информационные потоки не полностью декодированы, внешние LLR возвращаются в качестве входных данных в MUD в качестве отправной точки для детектирования следующего цикла.

В академических кругах существует множество алгоритмов MUD, таких как алгоритм передачи сообщений (message passing algorithm, MPA) [23], алгоритм распространения математического ожидания (expectation propagation algorithm, EPA) [24], линейная минимальная среднеквадратичная ошибка (linear minimum mean square error, LMMSE) [25] и элементарная оценка сигнала (elementary signal estimator, ESE) [20]. Эти алгоритмы MUD состоят из следующих элементов:

- **детектор MPA:** условные вероятности обновляются и передаются туда и обратно между ребрами. На факторном графе, соответствующем данной схеме NOMA, они соединяют функциональный узел, представляющий ресурсный элемент, в пару с переменным узлом, представляющим уровень данных или пользователя. После ряда итераций (внутренних циклов) вычисляются LLR для кодированных битов и затем вводятся в канальный декодер. Детектор MPA может достичь точности, близкой к оптимальному детектору максимального правдоподобия, но с гораздо меньшей сложностью, особенно когда факторный граф разреженный. Его встроенная функция «разделяй и властвуй» позволяет реализовать высокий уровень параллелизма;

- **детектор ЕРА:** в нем используется классический метод приближенного байесовского вывода [26]. Он проецирует истинное апостериорное распределение переданных символов в семейство гауссовых распределений путем итеративного сопоставления средних и дисперсий с истинным апостериорным распределением. В каком-то смысле ЕРА представляет собой гауссово приближение к МРА, но также с учетом негауссова распределения вероятностей переданных символов. Прямым преимуществом этого приближения является линейная сложность в отношении размера модуляции и количества пользовательских устройств, при сохранении почти такой же производительности, как у МРА в большинстве представляющих интерес сценариев применения [24];
- **детектор LMMSE:** он аппроксимирует априорное распределение сигнала как гауссово, либо со средним значением и дисперсией, вычисленными из мягких LLR, возвращаемых канальным декодером, либо с нулевым средним значением и дисперсией, масштабируемыми мощностью сигнала, если мягкая обратная связь недоступна. LMMSE можно рассматривать как частный случай ЕРА, когда несколько антенн обрабатываются совместно без внутренних итераций между FN и VN;
- **детектор ESE:** просто аппроксимирует межсимвольные помехи как гауссов шум. Детектор ESE должен полагаться на итерации внешнего цикла (мягкая обратная связь от канального декодера) для достижения приемлемых характеристик детектирования. Если имеется большое количество пользователей NOMA, количество итераций внешнего цикла может быть очень большим, как и задержка декодирования.

Совсем недавно в статье [21] были выявлены связи между различными алгоритмами MUD с точки зрения унифицированного вариационного вывода (variational inference, VI) и предложены унифицированные приемники NOMA на его основе.

Хотя в этой статье авторы не пришли к выводу о том, как определить схему NOMA в NR, различные компании выполнили всестороннее моделирование на уровне каналов и системы, чтобы оправдать преимущество NOMA над OFDMA, базовой линией OMA. Исследователи пришли к соглашению, что для оцененных сценариев наблюдаются значительные преимущества NOMA с точки зрения общей пропускной способности на уровне восходящего канала и способности к перегрузке. Кроме того, можно наблюдать преимущества в увеличении пропускной способности системы с точки зрения поддерживаемой скорости

поступления пакетов (packet arrival rate, PAR) на заданном уровне сбоя системы, например при 1%-ной скорости отбрасывания пакетов (packet drop rate PDR). Затем PAR можно использовать для оценки поддерживаемых соединений с учетом заданных шаблонов трафика для каждого пользователя.

Таким образом, NOMA является многообещающим решением для увеличения пропускной способности системы за счет обслуживания большего количества пользователей с теми же радиоресурсами. Однако у существующих схем NOMA все еще остаются проблемы, которые необходимо решить. Во-первых, многие приемники NOMA используют итерационные операции, что делает их более сложными, чем приемники OMA. Во-вторых, схема NOMA в сочетании с MIMO все еще требует дальнейшего изучения в плане повышения общей производительности. Наконец, соответствующие процедуры передач NOMA, такие как HARQ, выбор сигнатуры и адаптация линии связи, требуют дальнейшего изучения.

24.2.3. Безгрантовый множественный доступ

Базовые станции выполняют передачу служебной информации (grant broadcast, GB) через динамически планируемые ресурсы. В отличие от GB-передачи, GF-передача выполняется через предварительно сконфигурированные или полустатически сконфигурированные ресурсы для пользователей восходящего или нисходящего канала связи. Ресурсы, сконфигурированные таким образом, могут совместно использоваться более чем одним пользователем GF, что может привести к конфликтам передачи между пользователями GF, и поэтому GF-передача также называется *конкурентной передачей*. Благодаря полустатическому механизму конфигурации схема GF-передачи, которую можно охарактеризовать фразой «пришел и ушел», подходит для услуг и применений, требующих малой задержки, таких как трафик URLLC. Более того, схема GF-передачи имеет особое значение для поддержки аperiodического и чувствительного к задержке трафика, время прибытия которого непредсказуемо, но при этом трафик должен быть передан сразу после его прибытия.

Схема GF-передачи также может значительно сэкономить мощность и снизить накладные расходы на сигнализацию для передачи по восходящей линии связи благодаря тому факту, что ей не нужно отправлять запрос планирования на базовую станцию по прибытии трафика и что она может избежать процесса детектирования в получении контрольной информации. Экономия энергии в результате предотвращения квитирования запросов планирования и детектирования управляющих сигналов имеет первостепенное значение для чувствительных к мощ-

ности устройств, таких как интеллектуальные датчики (которым требуется длительный срок службы батареи).

Однако из-за того, что схема основана на конкуренции, когда несколько пользователей могут совместно использовать временные и частотные ресурсы, безгрантовая передача по восходящему каналу может вызывать определенные конфликты. Это, в свою очередь, может привести к неизбежным повторным передачам и опасениям по поводу надежности. К счастью, коллизийный характер GF-передачи можно преодолеть с помощью преимуществ NOMA. С помощью комбинации GF-передачи (восходящий канал связи) и NOMA каждому пользователю GF может быть полустатически назначена тщательно разработанная или случайно выбранная сигнатура NOMA из предварительно сконфигурированного пула сигнатур. Межпользовательские помехи, вызванные коллизиями передачи, затем могут быть устранены усовершенствованным приемником NOMA, который имеет возможность с высокой надежностью разделять перекрывающиеся сигналы. В результате схема GF плюс NOMA является ключевой технологией, обеспечивающей надежную, быструю и эффективную (восходящую) передачу данных в беспроводной сети.

Как обсуждалось выше, схема GF с поддержкой NOMA способна уменьшить задержку доступа, объем служебных данных управления сигнальным механизмом и потребление энергии пользовательскими устройствами, особенно при передаче малых пакетов и чувствительных к задержке передач для экономичных и энергоэффективных устройств. Однако у этой схемы все еще есть некоторые проблемы, которые необходимо решить, такие как поиск оптимальной формы пилот-сигнала и разрешение коллизий в NOMA, оптимальный формат обратной связи GF HARQ, компромиссы между обратной связью GF и повторной передачей, а также компромиссы между достижимыми скоростями, длиной пакета и надежностью (с точки зрения теории информации). В последнее время изучение этих тем привлекает внимание исследователей, и академический обзор по этим темам можно найти в [1]. Обзор [1] также охватывает раннюю работу по решению проблемы доставки коротких пакетов с точки зрения теории информации, а еще недавнюю работу по оценке характеристик схемы GF с NOMA.

Ход стандартизации безгрантовой связи

GF-передача, как многообещающая схема передачи для уменьшения объема служебных сигналов и сокращения задержки передачи, обсуждалась и была представлена в первом выпуске стандарта NR (R15) для передач как по восходящему, так и по нисходящему каналу. При GF-передаче по восходящему каналу ресурсы пользователя полустатически конфигурируются базовой станцией в пределах одного конфигурируе-

мого периода, при этом предполагается, что такие сконфигурированные ресурсы являются периодическими во времени. После настройки такие ресурсы могут быть немедленно использованы пользователем, что гарантирует передачу с малой задержкой. При передаче по нисходящему каналу ресурсы пользователя могут использоваться только после того, как они активированы динамическим сигналом от базовой станции, что помогает избежать ненужного детектирования передачи по нисходящему каналу на стороне пользователя и снизить его потребление энергии.

Кроме того, пользователь может иметь несколько GF-конфигураций одновременно для обслуживания различных типов трафика с разнообразными требованиями к услугам, что может дополнительно улучшить как задержку, так и надежность. Чтобы уменьшить энергопотребление и задержку, стандарты NR R16 и R17 дополнительно охватывают передачу по прямому каналу (от устройства к устройству), особенно в сценариях V2X, и передачу в неактивном состоянии (промежуточное состояние между ожиданием и активным), чтобы уменьшить энергопотребление и время ожидания.

Безгрантовый доступ и сжатая выборка

Из-за того, что трафик в беспроводной сети по своей природе спорадический, обнаружение активного пользователя является проблемой при GF-передаче для множественного доступа. При GF-передаче по восходящему каналу связи базовая станция должна обнаруживать активных пользователей перед декодированием и восстановлением сигналов. Обычно процесс обнаружения основан на пилотных сигналах или сигнатурах передачи данных. Следовательно, форма пилотных сигналов или сигнатур и соответствующие алгоритмы детектирования имеют решающее значение для GF-передачи. Во многих работах [27, 28] было доказано, что задача детектирования сигнала может быть смоделирована как задача сжатой выборки²⁵ (compressed sensing, CS), когда существующие результаты теории сжатой выборки могут быть использованы для облегчения проектирования схем безгрантового доступа.

В 2009 г. для решения задачи CS-детектирования был предложен алгоритм *приближенной передачи сообщений* (approximate message passing, AMP) [29]. С тех пор AMP использовался в качестве стандартного ал-

²⁵ Для этого понятия на английском языке существует несколько терминов – compressed sensing, compressive sensing, compressive sampling, compressed sampling и sparse sampling, – что вызывает изрядную путаницу и множество вопросов даже среди носителей языка, не говоря уже о русском переводе. В русских публикациях этот термин иногда оставляют без перевода. В этой книге мы остановились на переводе «сжатая выборка», чтобы не вступать в противоречие с предложенной авторами аббревиатурой CS. – Прим. перев.

горитма детектирования в задаче сжатой выборки, и для различных сценариев применения было предложено множество алгоритмов AMP. В 2011 году в [30] был предложен алгоритм обобщенной приближенной передачи сообщений (generalized approximate messaging passing, GAMP) для работы с более сложными моделями с нелинейными отношениями между входами и выходами. В 2017 г. в [27] был предложен улучшенный алгоритм AMP на основе шумоподавителя с пороговым значением MMSE для обнаружения активных пользователей в сценариях GF-передачи с массовым подключением. Кроме того, этот улучшенный алгоритм AMP может использоваться для безгрантового доступа в сочетании с конфигурациями массивного MIMO [28]. Опять же, в 2017 году в [31] был предложен алгоритм ортогональной приближенной передачи сообщений (orthogonal approximate message passing, OAMP), который, как утверждалось, обеспечивает достижение оптимальных по Байесу результатов с более высокой скоростью сходимости, чем традиционный алгоритм AMP.

Некогерентный безгрантовый доступ

Хотя безгрантовый доступ в сочетании с NOMA может смягчить проблему коллизии сигналов данных, для большинства схем NOMA по-прежнему требуется хорошая оценка канала. Поэтому для GF на основе NOMA требуется разработка ортогонального или почти ортогонального пилот-сигнала. Это приведет к большим накладным расходам на пилот-сигнал в сценариях массового подключения, особенно при передаче пакетов небольшого размера. В академических и промышленных учреждениях исследователи предложили возможные решения данной проблемы. Например, в 2017 году в [32] для решения вышеуказанной проблемы была предложена структура некогерентной передачи с общей кодовой книгой. Там же было дано некоторое теоретическое обоснование этого подхода. Данный подход получил название *произвольного доступа без источника* (unsourced random access) – решение без какого-либо планирования или пилотных сигналов и, таким образом, позволяющее избежать проблемы коллизии пилотных сигналов, – и нацелен на массовое подключение с небольшой длиной блока. Позднее были предложены некоторые улучшенные схемы в рамках этой структуры, например [33, 34]. Посредством таких схем, как разделение субблоков и межблочное кодирование, или тензорная модуляция [35], концепция произвольного доступа без источника приводит нас к реализуемым решениям безгрантового доступа. Кроме того, произвольный доступ без источника в сочетании с алгоритмами на основе сжатой выборки на стороне приемника может стать одним из кандидатов на решение для крупномасштабного доступа в будущем.

24.3. Ожидания от новых разработок и потенциальные направления исследований

24.3.1. Множественный доступ для услуг URLLC большой емкости

Услуги типа URLLC были введены в 5G в основном для поддержки вертикальных приложений, таких как управление движением в промышленности 4.0 и автономное вождение в V2X. Благодаря стандартам, определенным в самых последних спецификациях NR, одновременно могут быть достигнуты сверхвысокая надежность и низкая задержка. Однако из-за использования ортогональных ресурсов и консервативной адаптации линии связи за каждой обслуживающей базовой станцией может быть одновременно закреплено только очень ограниченное количество пользователей. В 6G такие услуги продолжают развитие в направлении расширения списка применений с еще более высокими требованиями и большим количеством устройств, таких как совместные роботы и дистанционное управление. В этой связи механизм множественного доступа 6G и связанные с ним процедуры будут дополнительно усовершенствованы, при этом ожидается, что ключевую роль сыграет GF- или GB-передача в сочетании с NOMA.

Когда дело доходит до надежности передачи данных на физическом уровне, ее обычно можно повысить за счет использования более высоких степеней разнесения для борьбы с неопределенностями, вызванными явлением замирания в беспроводных каналах. Ее также можно улучшить за счет более продвинутого кодирования для борьбы с шумом в каналах. Что касается этих способов повышения надежности, стоит упомянуть некоторые потенциальные направления для будущего совершенствования NOMA:

- **NOMA с совместной многоресурсной схемой:** в настоящее время сигнатуры NOMA разрабатываются в основном с акцентом на одну или две конкретные области ресурсов, такие как временная/частотная область, кодовая область, пространственная область или область мощности. Чтобы обеспечить лучшее разнесение и допускать большее количество межпользовательских помех, в сигнатурах NOMA следует совместно использовать ресурсы из разных областей, тем самым улучшая характеристики каналов с замираниями и увеличивая общую пропускную способность;
- **NOMA с совместным канальным кодированием:** в итеративной структуре приемника NOMA, описанной в разделе 24.2, уже существует итерация внешнего цикла между MUD и канальны-

ми декодерами, которая использует преимущества кодирования для подавления межпользовательских помех. Ожидается, что надежность многопользовательской системы доступа можно дополнительно повысить, если разрабатывать сигнатуры NOMA совместно с канальным кодированием. Это будет особенно полезно для коротких пакетов MA с требованиями высокой надежности.

24.3.2. Множественный доступ для дешевых и маломощных устройств

Как мы упоминали в разделе 24.1, множественный доступ недорогих устройств будет важным сценарием использования сети 6G. В настоящее время для подключения недорогих и маломощных устройств в 5G используются стандарты NB-IoT и eMTC. В 6G, поскольку целевой срок службы батареи для устройств IoT сенсорного типа, как ожидается, будет удвоен (до примерно 20 лет), а количество устройств IoT возрастет как минимум в десять раз, существует постоянная потребность в дальнейшем снижении энергопотребления и стоимости этих устройств. В настоящее время и в NB-IoT, и в eMTC по-прежнему требуется синхронизация времени/частоты. Фактически она и составляет большую часть энергопотребления этих устройств, в сопровождении нечастой и небольшой полезной нагрузки.

В 6G постоянная синхронизация может быть неприемлемой для некоторых сценариев использования, особенно для чрезвычайно дешевых и маломощных устройств. Однако дешевые компоненты в радиочастотных цепях этих устройств могут также вызвать проблемы с фазовым шумом в системе 6G, влияние которого больше, если устройства работают на средних частотах, а не на низких. При проектировании множественного доступа в сети 6G следует учитывать следующие направления исследований с прицелом на чрезвычайно дешевые и маломощные устройства:

- **множественный доступ с устойчивостью к сдвигу времени/частоты и фазовому шуму:** асинхронная передача нарушит ортогональность между пользователями, даже если применяется OMA. В этом случае будут уместны сигнатуры и приемники NOMA, устойчивые к сдвигу времени/частоты и фазовому шуму. Кроме того, асинхронные сигналы могут быть разработаны совместно с преемопередатчиками NOMA;
- **множественный доступ и сигналы с низким PAPR:** в 5G были изучены схемы доступа с использованием сигналов с низким PAPR; такие схемы включают SCMA на основе DFT и одното-

ный SCMA [36]. С точки зрения чрезвычайно низких затрат, множественный доступ надо изучать вместе с новыми формами сигналов с низким PAPR в сети 6G. Иными словами, схемы OMA и NOMA должны быть оптимизированы для поддержки передачи с низким PAPR.

24.3.3. Множественный доступ и сверхмассивное подключение

Сверхмассивное подключение – это незаменимая способность, необходимая для системы 6G во многих применениях, особенно для передачи коротких пакетов со спорадическими моделями трафика. Чтобы обеспечить сверхмассивное подключение в рамках ограниченных радиоресурсов, необходим безграничный доступ с поддержкой NOMA. Однако, чтобы реализовать сверхмассивное подключение, нам необходимо решить следующие проблемы, присущие текущим применениям GF-передачи и NOMA:

- **коллизия данных:** из-за ограниченных общих радиоресурсов и отсутствия координации между передачами при массовом подключении неизбежна коллизия данных. Чтобы разместить большее количество одновременно активных устройств, необходимо провести исследования по улучшению схем передачи NOMA и алгоритмов MUD;
- **коллизия пилотных сигналов:** хотя проблему коллизии данных можно уменьшить с помощью передовых методов NOMA, коллизия пилотных сигналов остается нерешенной проблемой для безграничного множественного доступа. Поэтому дальнейшего изучения требуют схемы расширения пилот-сигнала или даже некогерентные схемы без пилот-сигнала. В некоторых случаях схемы необходимо рассматривать в рамках NOMA, чтобы вместе решать проблемы коллизии пилот-сигналов и проблемы коллизии данных, которые являются двумя критическими проблемами для сверхмассивного подключения;
- **сверхмассивные антенны:** проблему сверхмассивного подключения в конфигурациях с массивными антеннами на стороне базовой станции нельзя рассматривать как просто увеличенную версию исходной проблемы. Она может быть особенно актуальна, когда количество антенн очень велико, антенны больше не размещены вместе или используются новые компоненты, такие как RIS. Некоторые особые свойства канала в случае применения сверхмассивных антенн могут быть применены для решения проблем коллизии как для данных, так и для пилот-сигнала.

24.3.4. Множественный доступ и надежное формирование луча

В 6G диапазоны миллиметровых волн и (суб)ТГц должны использоваться в сочетании с массивным MIMO, особенно в применениях с большой пропускной способностью. С увеличением количества антенных элементов и переходом в более высокочастотный диапазон сигнальные лучи MIMO станут чрезвычайно узкими. Следовательно, будет сложно генерировать несколько очень узких лучей, точно направленных на нескольких пользователей одновременно, потому что во многих сценариях будет трудно получить точную информацию о состоянии канала. В таких случаях для надежного формирования луча могут быть разработаны улучшенные схемы NOMA совместно с многопользовательским предварительным кодированием. Располагая возможностями NOMA, MU-MIMO может использовать модифицированную схему предварительного кодирования. Вместо формирования очень узких и точных лучей, нацеленных на каждого пользователя отдельно, прекодер MU-MIMO теперь может генерировать более широкие лучи для нацеливания на группу пользователей, которые дополнительно мультиплексируются с помощью NOMA. Увеличение ширины луча делает его более устойчивым к изменениям параметров канала, например вызванным мобильностью пользователя или задержкой измерения и обратной связи. Межпользовательские помехи внутри группы будут дополнительно подавляться приемником NOMA с помощью переданных сигнатур NOMA. Аналогичные идеи могут быть дополнительно применены при совместной передаче с несколькими базовыми станциями.

24.3.5. Множественный доступ с поддержкой ИИ

Как обсуждалось в главе 19, при разработке радиоинтерфейса в 6G будут широко применяться технологии машинного обучения и искусственного интеллекта. Методика искусственного интеллекта может облегчить разработку многих традиционных модулей в каналах связи. Если говорить о множественном доступе в сети 6G, то и передатчик, и приемник могут использовать методы ИИ, как описано ниже.

- **Передатчик с искусственным интеллектом:** перед нами стоит задача разработки схемы передачи множественного доступа с низкой стоимостью, низким PAPR, малой задержкой, высокой надежностью и широкими возможностями подключения. Традиционная методология проектирования на основе классической модели может оказаться недостаточной для соответствия результатам такому количеству ключевых показателей. Возможным ва-

риантом решения проблемы может быть нейронная сеть, обученная на имеющихся данных. Например, нейронную сеть можно использовать для разработки сигнатур NOMA совместно со многими другими модулями, такими как форма сигнала и предварительное кодирование MIMO.

- **Приемник с искусственным интеллектом:** методы искусственного интеллекта / машинного обучения могут сыграть роль в облегчении проектирования MUD для NOMA. Упомянутый выше алгоритм EPA MUD – это один из примеров, полученных из методики вариационного вывода, которая широко использовалась во многих задачах ИИ, связанных с байесовской оценкой. Техника вариационного вывода обеспечивает прирост характеристик за счет изучения априорных распределений в качестве входных данных MUD вместо простого предположения, что они являются гауссовыми. В сети 6G с множественным доступом, как отмечалось ранее, многие традиционные модели могут больше не работать из-за таких факторов, как асинхронная передача или фазовый шум. Методы искусственного интеллекта могут быть альтернативным решением для автоматического сопоставления сценариев. Более того, с помощью нейронных сетей могут быть совместно разработаны различные модули приемника, например для обнаружения активности пользователя, оценки канала и MUD.

24.4. Источники

- [1] M. Vaezi, Z. Ding, and H. V. Poor, *Multiple access techniques for 5G wireless networks and beyond*. Springer, 2019.
- [2] J. G. Proakis and M. Salehi, *Digital communications*. McGraw-Hill, 2007.
- [3] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.
- [4] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.
- [5] D. N. C. Tse, P. Viswanath, and L. Zheng, *Diversity-multiplexing tradeoff in multiple access channels*, IEEE Transactions on Information Theory, vol. 50, no. 9, pp. 1859–1874, 2004.
- [6] S. Vishwanath, N. Jindal, and A. Goldsmith, *Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels*, IEEE Transactions on Information Theory, vol. 49, no. 10, pp. 2658–2668, 2003.
- [7] K. Senel, H. V. Cheng, E. Björnson, and E. G. Larsson, *What role can NOMA play in massive MIMO?* IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 3, pp. 597–611, 2019.

- [8] Y. Liu, Z. Qin, and Z. Ding, *Non-orthogonal multiple access for massive connectivity*. Springer, 2020.
- [9] H. Nikopour and H. Baligh, *Sparse code multiple access*, in Proc. 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). IEEE, 2013, pp. 332–336.
- [10] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, *Scma codebook design*, in Proc. 2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall). IEEE, 2014, pp. 1–5.
- [11] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang, and J. Xu, *Multi-user shared access for internet of things*, in Proc. 2016 IEEE 83rd Vehicular Technology Conference (VTC-Spring). IEEE, 2016, pp. 1–5.
- [12] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, *Pattern division multiple access – a novel nonorthogonal multiple access for fifth-generation radio networks*, IEEE Transactions on Vehicular Technology, vol. 66, no. 4, pp. 3185–3196, 2016.
- [13] Q. Xiong, C. Qian, B. Yu, and C. Sun, *Advanced NoMA scheme for 5G cellular network: Interleave-grid multiple access*, in Proc. 2017 IEEE Globecom Workshops. IEEE, 2017, pp. 1–5.
- [14] J. M. Meredith, *Study on downlink multiuser superposition transmission for LTE*, in Proc. TSG RAN Meeting, vol. 67, 2015.
- [15] H. Nikopour, E. Yi, A. Bayesteh, K. Au, M. Hawryluck, H. Baligh, and J. Ma, *SCMA for downlink multiple access of 5G wireless networks*, in Proc. 2014 IEEE Global Communications Conference. IEEE, 2014, pp. 3940–3945.
- [16] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, *Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends*, IEEE Communications Magazine, vol. 53, no. 9, pp. 74–81, 2015.
- [17] 3GPP, *Study on non-orthogonal multiple access (NOMA)*, 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.812, Dec. 2018, version 16.0.0. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3236>.
- [18] Y. Chen, A. Bayesteh, Y. Wu, B. Ren, S. Kang, S. Sun, Q. Xiong, C. Qian, B. Yu, Z. Ding et al., *Toward the standardization of non-orthogonal multiple access for next generation wireless networks*, IEEE Communications Magazine, vol. 56, no. 3, pp. 19–27, 2018.
- [19] Qualcomm Incorporated, *RSMA*, 3rd Generation Partnership Project (3GPP), RAN1 (R1) 164688, May 2016. <https://www.3gpp.org/DynaReport/TDocExMtg--R1-85--31662.htm>.

- [20] L. Ping, L. Liu, K. Wu, and W. K. Leung, *Interleave division multiple-access*, IEEE Transactions on Wireless Communications, vol. 5, no. 4, pp. 938–947, 2006.
- [21] X. Meng, L. Zhang, C. Wang, L. Wang, Y. Wu, Y. Chen, and W. Wang, *Advanced NOMA receivers from a unified variational inference perspective*, IEEE Journal on Selected Areas in Communications, 2020. (Early Access) <https://ieeexplore.ieee.org/abstract/document/9181630>.
- [22] X. Meng, Y. Wu, C. Wang, and Y. Chen, *Turbo-like iterative multi-user receiver design for 5G non-orthogonal multiple access*, in Proc. 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall). IEEE, 2018, pp. 1–5.
- [23] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, *Factor graphs and the sum-product algorithm*, IEEE Transactions on Information Theory, vol. 47, no. 2, pp. 498–519, 2001.
- [24] X. Meng, Y. Wu, Y. Chen, and M. Cheng, *Low complexity receiver for up-link SCMA system via expectation propagation*, in Proc. 2017 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2017, pp. 1–5.
- [25] X. Wang and H. V. Poor, *Iterative (turbo) soft interference cancellation and decoding for coded CDMA*, IEEE Transactions on Communications, vol. 47, no. 7, pp. 1046–1061, 1999.
- [26] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.
- [27] Z. Chen and W. Yu, *Massive device activity detection by approximate message passing*, in Proc. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2017, pp. 3514–3518.
- [28] Z. Chen, F. Söhrabi, and W. Yu, *Sparse activity detection for massive connectivity*, IEEE Transactions on Signal Processing, vol. 66, no. 7, pp. 1890–1904, 2018.
- [29] D. L. Donoho, A. Maleki, and A. Montanari, *Message-passing algorithms for compressed sensing*, Proceedings of the National Academy of Sciences, vol. 106, no. 45, pp. 18 914–18 919, 2009.
- [30] S. Rangan, *Generalized approximate message passing for estimation with random linear mixing*, in Proc. 2011 IEEE International Symposium on Information Theory Proceedings. IEEE, 2011, pp. 2168–2172.
- [31] J. Ma and L. Ping, *Orthogonal amp*, IEEE Access, vol. 5, pp. 2020–2033, 2017.
- [32] Y. Polyanskiy, *A perspective on massive random-access*, in Proc. 2017 IEEE International Symposium on Information Theory (ISIT). IEEE, 2017, pp. 2523–2527.

- [33] R. Calderbank and A. Thompson, *Chirrup: A practical algorithm for unsourced multiple access*, Information and Inference: A Journal of the IMA, vol. 9, no. 4, pp. 875–897, Dec. 2020.
- [34] A. Fengler, P. Jung, and G. Caire, *Sparcs for unsourced random access*, arXiv preprint arXiv:1901.06234, 2019.
- [35] A. Decurninge, I. Land, and M. Guillaud, *Tensor-based modulation for unsourced massive random access*, IEEE Wireless Communications Letters, 2021.
- [36] 3GPP, *Discussion on the design of NOMA transmitter*, 3rd Generation Partnership Project (3GPP), RAN1 (R1) 1812187, Nov. 2018. <https://www.3gpp.org/DynaReport/TDocExMtg--R1-95--18807.htm>.

Глава 25

Ультрамассивный MIMO

25.1. Почему нужен ультрамассивный MIMO

Беспроводная связь с несколькими передающими антеннами, превосходящая возможности фазированных антенных решеток [1], стала популярной темой исследований в конце XX века. В течение этого периода появилось множество новых схем передачи и приема, таких как пространственно-временные блочные коды (space-time block codes, STBC) [2], код Аламоути [3], многослойное пространство-время Bell Laboratories (BLAST) [4] и сферическое декодирование [5]. Это значительно улучшило пропускную способность канала и надежность нескольких передающих антенн.

Вскоре эти новые технологии были внедрены в 3GPP. Его первая версия (1G LTE) поддерживала до четырех портов передающей антенны на нисходящем канале, которые могли быть идентифицированы пользователями с помощью опорных сигналов соты (cell reference signal, CRS). Чтобы удовлетворить постоянно растущий спрос на плотность пользователей, пропускную способность и надежность соединения, вместе с новыми версиями LTE постепенно улучшалась и поддержка MIMO. Позже, когда на сцену вышел новый радиointерфейс, массивный MIMO изначально поддерживался через архитектуру MIMO на основе лучей, особенно в диапазонах TDD, которые облегчили массовое развертывание MIMO (из-за более простого получения информации о канале).

Для достижения строгих KPI сети 6G должны подвергаться уплотнению, что позволяет размещать более широкую полосу пропускания в более узких диапазонах и использовать MIMO еще более высокого порядка. Это приведет к появлению распределенного радиодоступа в сеть 6G с плотно размещенными удаленными радиоголовками (remote radio head, RRH), ELAA и большим количеством антенных панелей. Вдобавок каждая антенная панель будет иметь больше антенных элементов в миллиметровых диапазонах, конструкция которых впоследствии будет модифицирована для ТГц-диапазона, чтобы использовать гораздо более широкий спектр. Помимо огромного количества частотных ресур-

сов, терагерцовый MIMO также будет способствовать достижению других ключевых показателей эффективности 6G, таких как локализация и сканирование. В этой главе мы изучаем текущее состояние развертывания MIMO в сетях радиодоступа, а также потенциальные улучшения MIMO в контексте 6G.

25.2. Обзор существующих решений

Несмотря на то что в NR был достигнут значительный прогресс в плане повышения надежности и пропускной способности каналов MIMO, дальнейшее усовершенствование для создания еще более быстрых и надежных каналов MIMO является желательным и достижимым.

В настоящее время NR имеет широкую поддержку MIMO как в низко-, так и в среднечастотном диапазоне ниже 7,125 ГГц, известном как частотный диапазон 1 (FR1) [6], а также в диапазонах высоких частот от 24,250 до 52,6 ГГц (известных как FR2) [7]. Однако дальнейшая поддержка MIMO в эпоху 6G сопряжена с рядом проблем. Например, по мере увеличения количества антенных элементов, развернутых в каждом узле, и лучей передачи/приема также увеличиваются издержки измерения радиосигнала, управления лучом и получения информации о состоянии канала.

Мы представляем существующие технологии MIMO, принятые в NR для FR1 и FR2, в разделах 25.2.1 и 25.2.2. В разделе 25.2.3 обсуждаются совместные решения MIMO и рассматривается поддержка NR для сотрудничества между различными узлами.

25.2.1. Технологии MIMO для FR1

Технология MIMO была ключевым техническим компонентом систем беспроводной связи с момента появления 4G, то есть стандартов долгосрочного развития (LTE) и усовершенствованного LTE (LTE-A). На самых ранних этапах стандартизация LTE производилась исходя из предположения, что базовые станции имеют только несколько антенных портов, сопоставленных с несколькими секторными антеннами, где антенные элементы использовались для формирования диаграммы направленности и ее электронного наклона. Передача по нескольким антеннам в LTE/LTE-A использовала разные режимы передачи для получения улучшенного разнесения, мультиплексирования и массивности, причем каждый режим оптимизирован для разного набора конфигураций системы, таких как антенные порты, FDD, TDD, несущая частота, мобильность и т. д. Большое количество режимов передачи усложняло реализацию и ограничивало достижимый выигрыш в более новых режимах, особенно если принять во внимание обязательную обратную

совместимость с более ранними устройствами с ограниченными возможностями. Система 5G NR объединила все режимы передачи в единую схему передачи с пространственным мультиплексированием. По мере развития LTE до R13 и R14 массовая технология MIMO уже была достаточно зрелой для развертывания в коммерческих сетях. Таким образом, начиная с начальных этапов стандартизации 5G NR технология массивного MIMO была основным фактором при проектировании системы NR, которая включает в себя несколько активных антенных решеток, содержащих десятки или сотни антенных элементов на базовых станциях. Эти решетки способны формировать различаемые по пилот-сигналу лучи для сотен антенных портов, а количество антенных элементов намного больше, чем количество используемых радиочастотных цепей.

Получение информации о состоянии канала на стороне передатчика всегда было ключевым компонентом MIMO. Базовая схема измерения канала в системах LTE/LTE-A и 5G NR обычно требует определенных опорных сигналов. Например, это может быть опорный сигнал состояния канала (CSI-RS) в нисходящей линии связи или зондирующий опорный сигнал (SRS) в восходящей линии связи, на основе которого приемник может измерять и оценивать характеристики интересующего канала. Например, станция 5G обычно передает пользователю CSI-RS, который затем выполняет измерение канала, квантование и дает обратную связь. Схема опорного сигнала в NR очень гибкая и может быть подстроена под каждого пользователя, включая периодичность, время, частоту и пространственное разрешение, чтобы оптимизировать поддержку различных сценариев использования, таких как eMBB, URLLC и mMTC. Как обсуждалось в главе 1, сети 6G будут иметь еще больше сценариев использования и будут содержать еще больше антенных элементов, чем 5G. Простое расширение механизма NR для получения информации состояния канала 6G приведет к увеличению служебных данных и более длительной задержке на конфигурацию, передачу опорного сигнала и обратную связь.

Ключевым техническим компонентом для получения информации о состоянии канала в NR была кодовая книга. Стандарт 5G NR поддерживает различные кодовые книги, известные как кодовые книги типа I и типа II [8]. Кодовая книга NR основана на предположении о регулярных и однородных плоских и кросс-поляризованных антенных решетках на gNB (базовые станции 5G NR), где векторы на основе дискретного преобразования Фурье хорошо структурированы, чтобы соответствовать векторам управления таких антенных решеток. В целом кодовые книги типа I относительно просты. Они предназначены для выбора одного луча из большого набора направлений луча на основе дискретного

преобразования Фурье. В отличие от кодовых книг типа I, кодовые книги типа II нацелены на сценарии MU-MIMO и предназначены для определения результирующего луча как содержащего линейную комбинацию нескольких доминирующих лучей. Следовательно, кодовые книги типа II более обширны и могут обеспечивать более высокое разрешение характеристик канала.

Поскольку сотовые сети становятся все более плотными, взаимные сетевые помехи становятся важным фактором, влияющим на характеристики сети. Как следствие измерение взаимных помех стало важным аспектом получения информации о состоянии канала. Как правило, gNB настраивает пользовательское устройство на выполнение измерения взаимных помех на основе определенных предположений. По мере увеличения сложности сети необходимо искать более эффективные схемы ограничения помех, выходящие за рамки существующих решений.

25.2.2. Технологии MIMO для FR2

В NR поддерживается связь через FR2 (от 24,25 до 52,6 ГГц). 3GPP планирует расширить поддержку частотных диапазонов до 71 ГГц [9] и улучшить технологии передачи сигналов для FR2. В отличие от связи через FR1 (0,41–7,125 ГГц), которая может работать с MIMO как используя лучи, так и без них, в диапазоне FR2 связь происходит только с использованием лучей и массивов антенных элементов. Это следствие необходимости борьбы с потерями на трассе и малой апертуры антенны из-за небольшой длины волны. Такой массивный MIMO использует узкие лучи, которые занимают только наиболее значимый путь (или пути) беспроводного канала. Следовательно, функция управления лучом в NR предназначена для формирования и удержания подходящей пары лучей; в частности, это луч передатчика и соответствующий луч приемника, которые при совмещении обеспечивают хорошее качество связи. В целом функция управления лучом включает в себя различные компоненты: начальное формирование луча, развертку/подстройку и отслеживание луча, указание луча и восстановление после потери луча.

В лучевой системе MIMO передача данных, опорные сигналы измерения канала, физические каналы управления и каналы синхронизации – все они формируются за счет настраиваемой диаграммы направленности. Для расширения охвата данных и управления как в восходящем, так и в нисходящем направлениях используются иерархический поиск луча, настройка формы и восстановление после сбоя через все более узкие лучи. Стороны, занятые в формировании канала связи, не сообщают друг другу форму и направление лучей в явном виде. Лучи неявно идентифицируются по соответствующим опорным сигналам. Пользовательское устройство настроено на постоянный мониторинг несколь-

ких лучей. Затем оно может выбрать один или несколько доступных лучей и сразу сообщить об отказе канала, если он должен быть обнаружен через канал с произвольным доступом (random-access channel, RACH).

Если следовать описанной выше процедуре управления лучом, то по мере продвижения 6G к еще более высоким полосам миллиметрового диапазона и увеличения количества антенных элементов и лучей либо на стороне передатчика, либо на стороне приемника накладные расходы на измерения и обратную связь также увеличиваются.

Другой важной функцией управления лучом для FR2 является указание луча: gNB может указывать пользователю лучи Tx/Rx для каналов передачи/приема сигналов. Указание луча в NR реализовано через информацию о квазилокализации (QCL) и пространственных соотношениях. Например, после измерения луча и получения отчета gNB может указать пользователю, что запланированный общий физический канал нисходящей линии связи (PDSCH) будет следовать предположению QCL, с конкретным опорно-измерительным сигналом и определенным положением относительно пространственных параметров Rx. Тем самым он сообщает пользователю, что для передачи PDSCH пространственная фильтрация может быть такой же, как и для соответствующего приема опорно-измерительного сигнала. Механизм указания луча в NR основан на указании конфигурации передачи (transmission configuration indication, TCI); другими словами, в каждом измерении должен применяться действующий опорный сигнал. Данный сигнальный механизм снижает накладные расходы, но плохо применим к более узким лучам, поскольку такое измерение предполагает, что исходная пара лучей для формирования линии связи гарантированно была обнаружена.

25.2.3. Совместный MIMO

Совместная работа сетевых узлов изучается с первых выпусков LTE. Во время стандартизации NR в исследованиях совместной работы стремились раскрыть потенциал для огромных улучшений производительности. Несмотря на потенциальные успехи в разработке теории, на практике совместная работа сопряжена с трудностями и ограничениями. Однако NR включает в себя некоторые способы преодоления этих проблем. Кроме того, совместная работа с конечными пользователями также может значительно улучшить опыт взаимодействия с пользователем. В этом разделе мы изучаем схемы взаимодействия сетей и пользователей, а также их роли в NR.

Многоточечная передача

В традиционных сотовых сетях пользователь одновременно подключается только к одной точке приема-передачи (transmit-receive point,

TRP); каждая точка имеет независимое планирование и передачу. В совместном MIMO несколько точек координируют свои передачи таким образом, что пользователь может получать от них сигналы одновременно. Ниже мы рассмотрим несколько схем совместного MIMO.

Первая схема – это динамический выбор точки (dynamic point selection, DPS). В DPS пользователь обслуживается одной точкой, которая динамически выбирается в кластере точек в интервале времени передачи (transmission time interval, TTI). Сеть может динамически выбирать точку, которая либо имеет лучшее качество канала, либо выгружает большую часть трафика от пользователя.

Вторая схема – это когерентная совместная передача (coherent joint transmission, CJT), когда несколько точек передают пользователю одни и те же потоки данных с когерентным предварительным кодированием. Однако для реализации CJT необходима поддержка транзитных сетей с высокой пропускной способностью, таких как Cloud-RAN.

Третья схема – это некогерентная совместная передача (non-coherent joint transmission, NCJT), когда несколько TRP передают пользователю разные потоки данных с некогерентным предварительным кодированием. NCJT не полагается на жесткую синхронизацию или высокоточные измерения канала и может применяться как для идеальной, так и для неидеальной координации транзитных соединений с несколькими точками. NCJT также может применяться для повышения надежности URLLC. Текущий стандарт NR поддерживает только идеальную транспортную сеть с несколькими TRP URLLC на основе централизованного планирования. Для NR будет полезно поддерживать передачу URLLC в неидеальных транзитных сетях.

Стоит отметить, что несколько TRP также могут взаимодействовать с целью приема и декодирования данных от обслуживаемых пользователей. Это позволит значительно увеличить покрытие сети и увеличить пропускную способность в восходящем направлении. Однако совместный прием на стороне сети может зависеть от реализации. Поскольку большинство методов не полагаются на конкретную процедуру или сигнальный механизм пользователя, они могут быть реализованы посредством организации сети для многоточечного приема.

Ориентированная на пользователя сеть

Соты NR могут содержать различные TRP и RRH, каждая из которых, возможно, обслуживает несколько лучей передачи и приема. Радиointерфейсы NR следуют так называемому принципу проектирования UCNC, чтобы гарантировать, что пользователи могут свободно перемещаться между различными лучами и сотами без прерывания связи. Эти лучи могут принадлежать одной или разным TRP в одной и той же соте

NR. Следовательно, для обеспечения такого уровня гибкости требуется учитывать множество различных аспектов радиointерфейсов NR. Например, ожидается, что пользовательские устройства будут измерять несколько блоков синхросигналов (synchronization signal blocks, SSB), выполнять развертку луча для RACH и поддерживать различные состояния и типы QCL. Более того, все специфические для пользователя физические каналы управления и данных после завершения начального доступа могут быть сконфигурированы для использования специфических для пользователя последовательностей и скремблеров, которые не связаны с конкретной TRP. Такая гибкость позволяет сети со множеством виртуальных сот NR совместно использовать TRP и RRH. Ожидается, что в сети 6G эта концепция радиointерфейса получит дальнейшее развитие и сможет гарантировать, что экстремальное уплотнение сети и связь на основе лучей предоставят пользователю действительно прозрачный хэндовер (передачу обслуживания), по крайней мере на физическом уровне.

В [10] авторы обсуждают бессотовую концепцию массивного MIMO с синхронизированными по фазе TRP. Авторы показывают, что по мере роста числа участвующих TRP затухание, шум и помехи исчезают, и остается только проблема загрязнения пилот-сигнала. 6G может еще больше снизить зависимость от соты, тем самым улучшив взаимодействие с пользователем и общие характеристики сети.

Концепция бессотовой системы MIMO в [10] основана на некоторых строгих предположениях. Для решения практических и теоретических вопросов и успешного развертывания сетей бессотовой связи требуются дальнейшие исследования и эксперименты. Авторы [11] исследовали некогерентную бессотовую систему MIMO, цель которой – уменьшить зависимость от жесткой синхронизации между TRP. Влияние сбоя оборудования изучается в [12], а в [13] авторы утверждают, что эффект упрочнения канала бессотовой системы MIMO сильно зависит от количества антенн на TRP и статистических свойств среды распространения.

Совместная пользовательская передача

Хотя массивный MIMO разрабатывается для новых сетей, которые используют как FR1, так и FR2, некоторые конечные пользователи могут столкнуться с ограничениями при доступе к FR2 из-за своих физических ограничений. Однако конечные пользователи, имеющие прямые линии связи, могут сотрудничать, объединяя свои антенны и ресурсы для формирования распределенной системы MIMO. Сотрудничество с пользователями дает несколько преимуществ, включая выигрыш от разнесения, выигрыш от мультиплексирования, выигрыш от агрегирования мощности, расширение покрытия и экономию энергии. В ти-

пичном решении для взаимодействия с пользователем предполагается, что несколько пользовательских устройств могут слышать сообщения от или для определенного пользователя. Посредством обмена информацией пользовательские устройства образуют виртуальную многоантенную решетку, способствуя исходной передаче аналогично MIMO.

Пользовательское взаимодействие может использовать мультиплексирование, при котором исходные данные разделяются на несколько частей по множеству устройств, и каждый пользователь передает только часть данных. В этом случае ожидается гораздо более высокая пропускная способность, особенно для услуг eMBB. Другие схемы взаимодействия направлены на повышение надежности и доступности за счет введения совместного разнообразия, как описано в [14–16]. Один из простых методов совместного разнесения – выбрать лучшего пользователя для помощи в передаче. Этот метод можно рассматривать как распределенную версию выбора антенны MIMO. Существует также значительное количество исследований, предлагающих использовать распределенное пространственно-временное кодирование для совместного разнесения.

Для большинства методов совместной работы одна из ключевых проблем – получение информации о состоянии канала до начала совместной передачи. Например, при совместном мультиплексировании для оптимального разделения данных может потребоваться глобальная информация. Что касается совместного разнесения, перед началом совместной передачи требуется либо глобальная информация о состоянии канала, либо информация точки передачи. Другой ключевой вопрос, требующий дальнейшего исследования, – это синхронизация между несколькими взаимодействующими узлами, которая является обязательной для некоторых схем.

По этим причинам существующий механизм MIMO в NR имеет унифицированную инфраструктуру получения информации о канале и управления лучом. При разработке многопользовательского MIMO принимали во внимание схему кодовой книги. Совместная передача учитывает ограничение транспортной сети. Однако по мере того, как 6G движется к еще более высокочастотным диапазонам миллиметровых волн, ожидается улучшение, связанное с измерением состояния канала и управлением лучом.

25.3. Новые технологии MIMO

MIMO – одна из ключевых технологий, которая позволит 6G достичь целевых показателей эффективности с точки зрения пользовательской и сетевой пропускной способности, надежности, маневренности и энергосбережения. Успех MIMO в 6G зависит не только от усовершенствования

ний существующих решений 5G, но и от новых решений. В частности, ожидается, что будущие системы MIMO будут включать передовые достижения в области радиочастотных технологий, а также новые материалы, архитектуры антенн и методы обработки сигналов. В этом разделе мы кратко рассмотрим новые технологии MIMO в 6G и исследуем как возможности, так и проблемы, которые они несут.

Терагерцовый MIMO

Постоянно растущий спрос на беспроводные данные, особенно передаваемые со скоростями порядка Тбит/с на короткие расстояния, может быть удовлетворен только за счет использования обширного (но широко не используемого) субТГц-спектра. Использование ТГц-спектра влияет на множество различных аспектов устройства сети. Изучению этих воздействий посвящено много исследований [17, 18]. Полоса частот ТГц-диапазона занимает пространство между миллиметровыми волнами и инфракрасным светом, которые связаны с радиочастотными и оптическими устройствами соответственно. Как обсуждалось в главе 16, последние достижения в области радиочастотных компонентов, обработки сигналов и антенной связи повлияют на различные аспекты систем MIMO, включая модуляторы, формы сигналов и конструкции приемопередатчиков. В новых трансиверах могут использоваться электронные, фотонные или комбинированные технологии [17, 19, 22]. Новая антенная технология может использовать решетки активных и пассивных элементов для улучшения формирования диаграммы направленности и покрытия.

Между миллиметровыми волнами и ТГц-диапазоном есть много общего, что делает развертывание MIMO в этих частях спектра похожим. Поэтому было бы естественно распространить решения и архитектуры миллиметрового диапазона на терагерцовый. Однако существуют важные характеристики и практические соображения относительно различий этих двух диапазонов. К ним относятся: характеристики канала, конструкция устройства и генерация сигналов, а также антенная технология [17, 22].

Терагерцовые волны страдают от высоких потерь на трассе. Показано, что потери на трассе увеличиваются квадратично с увеличением несущей частоты в свободном пространстве (см. формулу передачи Фрииса) [23]. Помимо этих огромных потерь в тракте передачи, с ТГц-волнами связаны и другие проблемы, такие как низкая эффективность усилителей мощности и проблемы, вызванные погодными условиями. Природные помехи, такие как молекулярное поглощение или сцинтилляция из-за тепла и влажности, ограничивают использование ТГц-диапазона локальными сценариями, такими как внутренние сети, умные

офисы и умные фабрики [17, 21]. Более того, поскольку беспроводные каналы ТГц-диапазона практически не испытывают рассеяния, их охват ограничен физическими препятствиями, такими как стены и потолки. Кроме того, из-за незначительной дифракции волн беспроводные каналы состоят в основном из путей прямой видимости и, возможно, из нескольких путей отражения, создаваемых стенами или мебелью.

Очень широкая полоса пропускания, доступная на ТГц-частотах, сопряжена со специфическими проблемами. Во-первых, мощность теплового шума линейно пропорциональна используемой полосе пропускания, что приводит к очень низкому соотношению сигнал/шум даже на средних расстояниях Tx-Rx [17, 21]. Во-вторых, широкополосные радиочастотные цепи более дороги и потенциально рассеивают больше энергии, что приводит к еще более низкой энергоэффективности терагерцовых трансиверов [17, 21]. В-третьих, структура решетки из антенных подрешеток (array of sub-arrays, AoSA) ТГц-трансиверов наряду с очень широкой полосой пропускания приводит к явлению отклонения или расщепления луча [24]. Это явление возникает из-за необходимости поддерживать приемлемую апертуру антенны. Размеры антенной панели не могут линейно уменьшаться с изменением полосы частот, в то время как длительность импульсов сокращается по мере увеличения доступной полосы пропускания. В результате разные антенные элементы (принадлежащие либо к одной, либо к разным антенным подрешеткам) испытывают разницу во времени прохождения дистанции, сравнимую с длительностью импульса или превышающую ее. Эта разница во времени приведет к тому, что лучи в частях используемой полосы пропускания будут указывать в немного разных направлениях, как показано (с увеличением) на рис. 25.1. Это явление отклонения приводит к тому, что чрезвычайно узкие передающие и приемные лучи указывают в разных (и нежелательных) направлениях. Возможным решением этого явления является использование предварительного кодирования с задержкой [24].

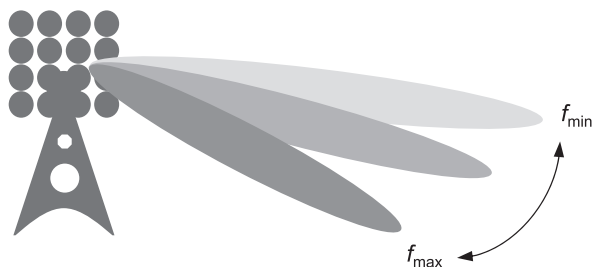


Рис. 25.1. Отклонение луча в ТГц-диапазоне

Другой проблемой для терагерцовых линий является непроизводительная нагрузка пилот-сигнала и обратной связи, необходимая для

формирования и поддержания пар лучей между передатчиком и приемником.

Поиск луча требует обширного сканирования на обоих концах беспроводной связи. Незначительное движение или поворот с любой стороны может привести к потере луча [20–22, 24]. Следовательно, гибкое управление лучом с низкими накладными расходами и высокой надежностью является ключом к успеху развертывания сети в ТГц-диапазоне. Хотя последовательное формирование диаграммы направленности остается возможным решением для низкозатратного поиска узких лучей, оно страдает двумя недостатками: очень низким бюджетом канала на этапе поиска широкого луча и большой задержкой из-за увеличенного количества измерений на каждом шаге. Такая задержка и накладные расходы, если они не будут устранены, ограничат использование и преимущества связи в ТГц-диапазоне. Учитывая проблемы с управлением лучом и другие характеристики ТГц-канала, для протоколов более высокого уровня, таких как протоколы MAC, требуются более эффективные решения [18, 21].

25.3.2. Реконфигурируемые интеллектуальные поверхности

Реконфигурируемые интеллектуальные поверхности (RIS) относительно недавно стали многообещающей парадигмой проектирования беспроводных сетей и режимов беспроводной передачи. Они также могут создавать интеллектуальные радиосреды (или интеллектуальные радиоканалы), т. е. распространением радиоволн в окружающей среде можно управлять для создания персонализированного канала связи [25]. В обобщенной модели, изображенной на рис. 25.2, сеть RIS формируется между несколькими TRP для создания крупномасштабных интеллектуальных радиоканалов, обслуживающих нескольких пользователей. В отсутствие управляемой среды архитектура беспроводной системы и режим передачи могут быть оптимизированы только в соответствии со статистическими свойствами физических каналов и/или информацией, возвращаемой от приемника к передатчику. В управляемой среде RIS сначала воспринимают данные среды и возвращают их в систему. Исходя из этих данных, система оптимизирует режим передачи и параметры RIS по интеллектуальным радиоканалам на стороне передатчика, канала и приемника.

Благодаря поддержке формирования луча, связанной с RIS, использование интеллектуальных радиоканалов может значительно улучшить качество связи, производительность системы, покрытие соты и качество связи на границе соты в беспроводных сетях; это было доказано многочисленными результатами моделирования в различных сценариях

[26, 27]. Стоит отметить, что не все панели RIS имеют одинаковую структуру. Это связано с тем, что они обладают различными возможностями регулировки фазы, которые варьируются от плавного управления фазой до дискретного управления только с несколькими уровнями. Исследования показали, что при большом количестве блоков RIS ограниченных регулируемыми фазами на блок достаточно для значительного улучшения общих характеристик системы [26, 28].

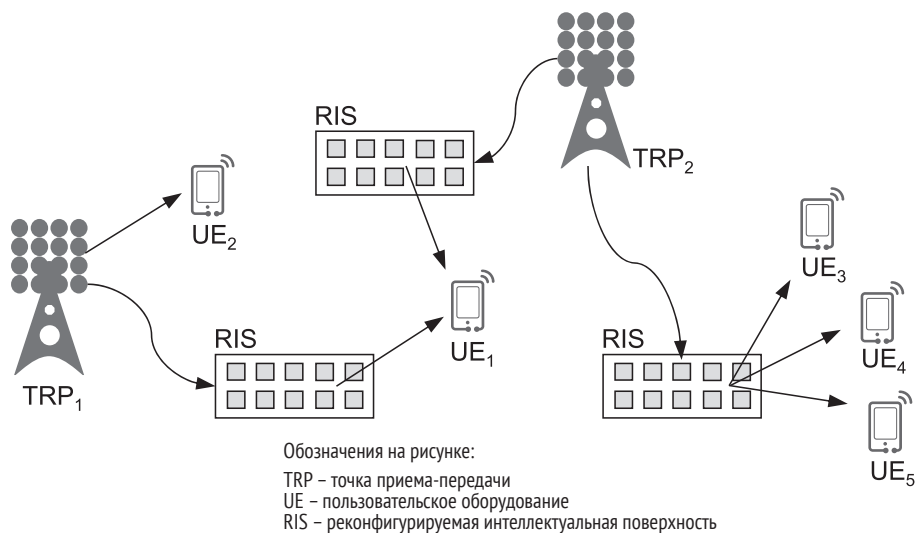


Рис. 25.2. Сценарий беспроводной передачи с интегрированными RIS

Другое применение RIS – это передатчики, которые напрямую модулируют свойства падающих радиоволн, такие как фаза, амплитуда, поляризация и частота, без необходимости в активных компонентах (как того требуют радиочастотные цепи в традиционных передатчиках MIMO) [29, 30]. Например, недавно была опробована модуляция 256QAM на основе RIS [31]. Передатчики на основе RIS обладают многими достоинствами, такими как простая аппаратная архитектура, низкая аппаратная сложность, низкое энергопотребление и высокая эффективность использования спектра [32, 33]. Таким образом, RIS представляют собой новое направление в разработке передатчиков в радиосистемах.

MIMO с поддержкой RIS имеет другие потенциальные применения, чем те, что мы уже обсуждали раньше. Например, его можно использовать для помощи в быстром формировании луча с использованием точного позиционирования или для преодоления эффектов блокировки за счет сбора данных о каналах в системе миллиметровых волн [34]. Его также можно использовать в NOMA, чтобы повысить надежность при очень низком отношении сигнал/шум, разместить больше пользователей и задействовать схемы с модуляцией более высокого порядка [35].

RIS также потенциально применимы для организации передачи с нативной физической безопасностью, беспроводной передачи энергии (или одновременной передачи данных и беспроводной передачи энергии) и гибких голографических экранов [36–38].

25.3.3. Антенные решетки со сверхбольшой апертурой

За последние 10 лет академическими исследователями и промышленными инженерами были разработаны и протестированы различные устройства, алгоритмы и системные компоненты, расширяющие границы применения MIMO. Что касается массовых исследований MIMO 6G, использование решеток со сверхбольшой апертурой (extremely large aperture array, ELAA) является многообещающим направлением для сверхмассивных MIMO.

В [39] ELAA определяют как решетку, состоящую из сотен антенных элементов, которые совместно обслуживают распределенных пользователей. Конечная цель ELAA состоит в том, чтобы все конечные пользователи использовали взаимно ортогональные каналы с пропускной способностью для каждого пользователя, аналогичной каналу с аддитивным белым гауссовым шумом [40, 41]. В дополнение к расширению услуг мобильной широкополосной связи, которые составляют большую часть беспроводного трафика [42], большое пространственное разрешение ELAA также можно использовать для пространственного мультиплексирования большого количества устройств связи.

Большое количество исследований было сосредоточено на поиске новых алгоритмов, удовлетворяющих особым требованиям ELAA. В [39, 43, 44] авторы обсуждают, как методы обработки MMSE могут использоваться для подавления межсотовых помех, следовательно, ELAA может эффективно работать при развертывании в сотах.

Исследование ELAA обнаружило у них уникальные свойства, которые значительно отличаются от наблюдаемых в традиционных массивных MIMO:

- **переход от дальнего поля к ближнему:** с увеличением размера апертуры решетки в ELAA волновой фронт становится ощутимо сферическим и, следовательно, не может быть аппроксимирован как плоская волна. Следовательно, при распространении электромагнитного излучения начинает преобладать так называемое *излучательное ближнее поле*. Используя сферическую форму волнового фронта, антенная решетка может определять не только пространственные углы волны, то есть угол прихода (AoA) и угол выхода (AoD), но и пространственную глубину, которую прошла волна. Эта новая функция дополнительно позволяет применять новые схемы мультиплексирования, используя информацию о

пространственной глубине для разделения потоков или пользовательского оборудования посредством предварительного кодирования;

- **переход от стационарного канала к нестационарному:** когда апертура решетки охватывает сотни длин волн, беспроводной канал может стать нестационарным. Это означает, что рассеиватели видны только частично или их вклад в мощность может значительно варьироваться по всей решетке. Нестационарность привносит в систему характерную разреженность каналов, когда сигнал от любого пользователя появляется только в небольшой части решетки. Подобная разреженность каналов способствует упрощению схем обработки сигналов, таких как подавление помех от подрешетки, и методов передачи сообщений.

Чтобы улучшить показатели системы, в будущем при разработке ELAA необходимо учитывать указанные выше уникальные особенности.

25.3.4. MIMO с искусственным интеллектом

Как обсуждалось в главе 19, благодаря своей исключительной способности извлекать признаки из данных машинное обучение в последнее время стало многообещающим решением многих прикладных задач, включая проблемы передачи на физическом уровне [45, 46]. В целом машинное обучение можно использовать для повышения производительности существующих алгоритмов, чтобы приблизиться к оптимальной производительности алгоритма с меньшей сложностью. Его даже можно использовать для поиска решений путем моделирования задач, которые не могут быть решены другими способами.

Машинное обучение использовалось в широком спектре приложений MIMO. Некоторые типичные примеры включают управление мощностью MIMO [47, 48], детектирование информации о состоянии канала [49, 50], оценку канала [51, 52], предварительное кодирование MIMO [53, 54] и детектирование [55, 56].

В машинном обучении искусственный интеллект изучает основные характеристики целевого признака на основе данных, которые ему предоставили, и затем может применить эти знания к другой задаче. Это можно рассматривать как обновление традиционных подходов к решению проблем, когда некоторые параметры оптимизируются на основе упрощенной математической модели целевого признака. Подобные традиционные подходы ограничены точностью и сложностью математических моделей, которые часто не могут точно учесть основные характеристики рассматриваемого признака. Как показывают текущие исследования, схемы MIMO с искусственным интеллектом всег-

да превосходят обычные схемы. Ниже приведены некоторые типичные примеры MIMO с поддержкой ИИ.

- **Получение информации о состоянии канала с помощью ИИ:** в [49] авторы использовали пары сетей кодер/декодер для получения более эффективной обратной связи о состоянии канала. Вместо того чтобы использовать случайную проекцию, кодер изучает преобразование (от исходных данных каналов к сжатым представлениям) с помощью обучающих данных. Между тем декодер таким же образом изучает обратное преобразование (из сжатых представлений в исходные данные каналов). Согласно результатам тестирования, этот основанный на данных метод может получить более точную оценку канала, чем обычные методы. Дальнейшего повышения качества обучения можно достичь, если MIMO с искусственным интеллектом дополнительно будет использовать моделирование алгоритмов в процессе обучения. Интегрируя стратегии, основанные на моделях, в обычное обучение на основе данных, можно добиться более быстрой сходимости и более точного вывода.
- **Схема предварительного кодирования с помощью ИИ:** в статье [53] предложена основанная на глубоком обучении структура MIMO миллиметрового диапазона для эффективного гибридного предварительного кодирования. Этот подход направлен на то, чтобы выйти за фундаментальные ограничения существующих гибридных схем предварительного кодирования. Эти ограничения включают высокую сложность, а также плохое использование пространственной информации. В данном исследовании каждый из прекодеров, необходимых для получения оптимизированного декодера, рассматривается как отображение в глубокую нейросеть. Результаты показывают, что этот подход способен минимизировать BER и повысить спектральную эффективность массивного MIMO миллиметрового диапазона, при этом существенно уменьшая вычислительную сложность.
- **Детектирование MIMO с помощью машинного обучения:** лучшие результаты также наблюдаются при детектировании MIMO с помощью машинного обучения. В [55] авторы предложили сеть глубокого обучения для высококачественного детектирования MIMO. Обычный детектор максимального правдоподобия является оптимальным с точки зрения минимизации вероятности ошибки при одновременном детектировании символов, но при этом имеет очень высокую сложность. Предлагаемая глубокая нейросеть получена с использованием метода прогнозируемого градиентного спуска. Моделирование показывает, что эта

сеть обеспечивает почти оптимальные показатели детектирования – точность, аналогичную детектору максимального правдоподобия, но при этом более эффективна и надежна; она работает как минимум в 30 раз быстрее.

- **Оценка канала MIMO с помощью ИИ:** быстрая и качественная оценка канала становится все более сложной задачей для сотовых сетей с массивным MIMO по мере увеличения числа антенн. В статье [52] предложен метод оценки канала на основе глубокого обучения для многосотовых систем MIMO с ограниченными помехами. Оценщик использует специально разработанную глубокую нейросеть. Предлагаемый метод включает всего два шага: во-первых, шумоподавление принятого сигнала, а во-вторых, оценку канала с помощью обычного алгоритма наименьших квадратов. Моделирование доказывает, что этот модуль оценки может приблизиться к минимальной среднеквадратичной ошибке (MMSE) для сигналов большой размерности, избегая при этом сложных обращений каналов и не требуя знания ковариационной матрицы канала. Эта оценка также устойчива к загрязнению пилотного сигнала и при определенных условиях может полностью устранить его.

Исходя из значительных улучшений показателей системы, зафиксированных в ходе текущих исследований, мы ожидаем, что ИИ станет хорошим кандидатом для обеспечения более надежного и эффективного MIMO в 6G. Однако прежде чем ИИ можно будет широко применять в системах MIMO, необходимо решить различные проблемы. Во-первых, существующие проекты MIMO с искусственным интеллектом могут учиться на входных данных, но плохо обобщаются на модифицированное распределение или нестационарность. Недостаточное использование усвоенных знаний – еще один ключевой недостаток текущих моделей MIMO с использованием искусственного интеллекта. Кроме того, неадекватное извлечение данных может ограничить эффективность и точность алгоритмов MIMO с использованием искусственного интеллекта. Все вышеперечисленные факторы затрудняют достижение баланса между затратностью обучения и точностью вывода в MIMO широкого применения с использованием ИИ.

25.3.5. Другие потенциальные технологии MIMO

Орбитальный угловой момент (ОАМ)

ОАМ – это угловой момент электромагнитного луча, который зависит от пространственного распределения поля, но не от поляризации. В 1992 году Аллен и др. были первыми, кто объединил концепцию ОАМ

с идеей оптического вихря [57]. В 2007 году Бо Тиде и др. творчески расширили концепцию ОАМ на радиочастоты, что привело к появлению радиочастотного ОАМ (далее именуемого просто ОАМ) [58].

Способы генерации мод ОАМ ранее обсуждались в главе 15. Формы электромагнитных волн, несущие моды ОАМ, называемые *лучами ОАМ*, можно разделить на несколько типов, например гауссов луч, луч Лагерра–Гаусса, луч Бесселя и луч Бесселя–Гаусса. Что касается беспроводной связи в области радиочастот, лучи ОАМ увеличивают нестационарность беспроводных каналов и вносят спиральную фазовую структуру в радиальное направление распространения. Предыдущие исследования подтвердили осуществимость ОАМ как с теоретической [59], так и с практической [60] точки зрения.

Классические методы приема на основе ОАМ-мультиплексирования перечислены в табл. 25.1. Были представлены несколько примеров связи на миллиметровых ОАМ-волнах, например [61] в Ка-диапазоне, [62] в Е-диапазоне и [63] в D-диапазоне. Для адаптации ОАМ к большим расстояниям и медленной мобильности были предложены методы на основе двумерных круговых решеток с управляемым лучом [64], метод псевдодоплеровской интерполяции [65], метод отпечатков электромагнитного поля [66] и метод модульного устойчивого фазового градиента [67].

Все текущие исследования указывают на ОАМ как на эффективный метод использования пространственной области, особенно когда передатчик и приемник находятся в ближнем поле, то есть ближе, чем *расстояние Рэлея* (определяемое в данном случае как осевое расстояние с разностью фаз $\pi/8$ между осевым и краевыми лучами) [68]. Для получения выигрыша от пространственного разнесения в условиях динамического канала с использованием унифицированного оборудования предложена интегрированная структура ОАМ плюс MIMO, называемая также многорежимно-многопространственной (multi-mode–multi-spatial, MOMS) [69].

Таблица 25.1. Классические методы приема на основе ОАМ-мультиплексирования

Метод приема	Апертура приемной антенны	Источник
Скалярно-фазовый градиент	Полная	[70]
Спектральный анализ	Полная	[71]
Пространственное сопоставление на основе нейросети	Полная	[72]
Скалярно-фазовый градиент	Частичная	[73]
Векторно-фазовый градиент	Частичная	[66]
Псевдодоплеровский	Частичная	[74]

Однако другие исследования [69, 75–79] выявили недостатки использования лучей OAM. Они не могут принести новую независимую степень свободы в системы беспроводной связи и не помогают выйти за рамки традиционной пропускной способности MIMO с той же канальной средой. Следовательно, OAM не может повысить эффективность использования спектра систем MIMO в классических сценариях беспроводной связи, таких как сети сотового радиодоступа. Тем не менее в некоторых благоприятных сценариях (например, в микросотах и малых сотах внутри помещений) OAM может обеспечить более высокую эффективность использования спектра, чем системы MIMO с простыми и неоптимальными декодерами. Это связано с тем, что множественные моды OAM улучшают позиционное число (condition number) канала.

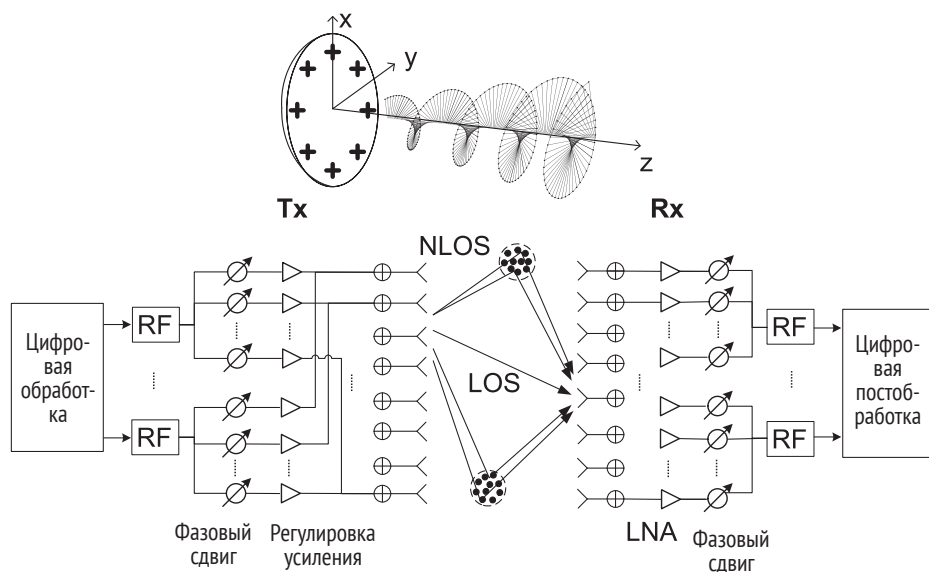


Рис. 25.3. Система беспроводной связи на основе OAM

Несмотря на упомянутые недостатки, существует двойная конечная выгода от внедрения передачи OAM. Во-первых, когда канал удовлетворяет параксиальному ограничению и условию LOS, информация, переносимая ортогональными лучами OAM, может быть эффективно демультиплексирована без сложной постобработки сигнала, как показано на рис. 25.3. Эта низкая сложность обработки имеет большое значение для сверхвысокой скорости беспроводной связи (например, пропускная способность Тбит/с в миллиметровых и ТГц-диапазонах). Во-вторых, как упоминалось ранее, лучи OAM можно рассматривать как специальный метод пространственной фильтрации, который можно комбинировать с обработкой сигналов MIMO для получения гибкого

компромисса между мультиплексированием и выигрышем от разнесения в более общих условиях канала [69, 80]. Эта комбинированная схема особенно подходит для антенных решеток миллиметрового диапазона на малых расстояниях.

На практике ОАМ можно использовать для высокоскоростной передачи путем введения мультиплексирования доступа с разделением мод. Например, его можно использовать при транзитной передаче между базовыми станциями, передаче от устройства к устройству (D2D) и связи со спутника на спутник, где относительно легко получить требуемое выравнивание оси.

Голографический MIMO

Теоретическая емкость систем массивных MIMO растет с увеличением количества антенн. Это означает, что теоретически всегда предпочтительнее иметь большее количество антенн, а оптимальное количество – бесконечно много [39]. Однако на практике возникают проблемы с постоянным увеличением количества антенн. Например, одна проблема заключается в том, как реализовать очень большое количество антенн в ограниченном пространстве в форме пространственно непрерывной апертуры передачи/приема. Другой проблемой является ограничение пропускной способности MIMO при рассмотрении электромагнитного эффекта между плотно расположенными антеннами. Исследования, направленные на решение таких проблем, относятся к категории исследований *голографического MIMO* (Holo-MIMO). Однако подобное исследование получило несколько других названий, включая исследования *голографического радио*, *голографических радиочастотных систем* [81–84] или даже *голографического массивного MIMO* и *голографического формирования луча* [85–88].

Holo-MIMO определяется как единственная форма пространственно-ограниченной системы со множеством антенн, в которой количество антенн может неограниченно расти при сохранении ограниченного размера апертуры (то есть приблизительно непрерывной апертуры). Эта конструкция основана на принципах, противоположных конструкции ELAA, где высокая глубина резкости достигается за счет увеличенной апертуры антенной решетки. Одним из потенциальных преимуществ системы Holo-MIMO является то, что непрерывная апертура может обеспечить более высокое пространственное разрешение в беспроводной связи, чем очень большое количество обычных дискретных антенн. Еще одно преимущество состоит в том, что непрерывная апертура позволяет создавать и принимать электромагнитные волны с произвольными пространственно-частотными компонентами без нежелательных боковых лепестков.

Теоретические исследования степеней свободы сплошных и ограниченных апертур восходят к 2005 году [89]. В этом исследовании было показано, что глубина резкости линейно зависит от произведения размера апертуры и углового разброса согласно случайному распространению вне прямой видимости (NLOS). Более того, результаты [87, 88] показывают, что объемные апертуры могут асимптотически приводить к двукратному увеличению доступной глубины резкости по сравнению с плоскими апертурами, независимо от того, насколько тонкими являются объемные апертуры. В [86] было показано, что для достижения выраженного эффекта сверхразнесения может быть использовано взаимное связывание близко расположенных антенн с помощью предварительного кодирования.

Назначением голографического радио является создание активной непрерывной электромагнитной апертуры, которая обеспечит пространственное мультиплексирование на уровне голографических изображений сверхвысокой плотности и пикселизации со сверхвысоким разрешением. Пространственный синтез и модуляция поля радиочастотной волны позволяют получить трехмерное электромагнитное поле, структурированное на уровне пикселей (пространство мультиплексирования с высокой плотностью в голографическом радио), которое отличается от разреженного пространства луча, связанного с традиционным массивным MIMO. Более того, формирование изображения с голографической интерференцией может быть использовано для получения спектральной голограммы радиопередающих источников (пользовательского оборудования). В таком случае есть возможность избежать необходимости в обычных процедурах передачи пилот-сигнала и оценки состояния канала.

Одной из проблем голографических радиосистем является стоимость реализации активных непрерывных апертур. Возможное решение – объединить большое количество антенных элементов в компактное пространство в виде метаповерхности. Однако этот метод ограничен только пассивным отражением из-за свойства метаповерхности. Поэтому он не подходит для активных решеток. Альтернативным (и более многообещающим) решением является использование жестко связанной решетки (tightly coupled array, TCA) широкополосных антенн. Эта технология основана на токовой пластине фотодетекторов с односторонним движением носителей (UTC-PD), где уплотненная высокочастотная питающая сеть заменена оптоволоконными линиями, преимуществами которых являются низкая стоимость и малое энергопотребление [90].

Другой проблемой голографического радио является сложность обработки сигналов, возникающая вследствие огромного количества

данных, генерируемых почти бесконечным числом антенных элементов внутри непрерывной апертуры. Одно из решений – преобразовать радиосигналы в оптические и обрабатывать их непосредственно в оптической области, где возможны более высокая скорость обработки и меньшее энергопотребление. Преобразование в оптические сигналы может быть достигнуто с помощью оптического быстрого преобразования Фурье [90].

Точные преимущества упаковки большего количества антенн в апертуры ограниченного размера остаются неясными. Текущие исследования показывают, что глубина резкости все еще ограничена размером апертуры. В [82, 84] было указано, что из-за отсутствия моделей голографическое радио нуждается в особой теории и технике моделирования для сближения теорий связи и электромагнетизма. Более того, оценка показателей связи Holo-MIMO требует специальных вычислений в области электромагнетизма, таких как алгоритмы и инструменты моделирования, связанные с вычислительным электромагнетизмом и компьютерной голографией.

25.4. Ожидания от новых разработок и потенциальные направления исследований

Предполагается, что в 6G произойдет существенное расширение диапазона спектра, в котором может работать технология радиодоступа. Спектральный диапазон будет охватывать волны от менее 1 ГГц до миллиметровых и терагерцовую часть спектра. Также в 6G появятся различные новые типы наземных и неназемных сетевых узлов, включая спутники, NAPS, БПЛА и даже RIS. С учетом этих изменений MIMO в 6G будет использовать архитектуру наземного/воздушного вещания, охватывающего все части отведенного спектра (как показано на рис. 25.4) по сравнению с традиционным наземным MIMO, установленным на стационарных опорах.

В будущем TRP будут включать разные типы антенн, которые могут быть активными или пассивными, фиксированными или подвижными. Все антенны образуют виртуальную большую антенную решетку, которая сможет интеллектуально и гибко обслуживать мобильных пользователей, удовлетворяя требованиям KPI 6G. Такая гибкая и универсальная сеть расширяется не только за счет достижений в существующих технологиях, но и за счет многочисленных новых технологий, таких как ELAA, RIS, искусственный интеллект, сканирование, новые материалы, а также новые конструкции и структуры антенн.

Эти новые технологии предоставят универсальный набор инструментов для увеличения пропускной способности сети. Чтобы воспользо-

зоваться этими технологиями с максимальной отдачей, мы должны принять новые принципы проектирования и критерии оптимизации структуры MIMO в 6G, которые будут рассмотрены ниже.

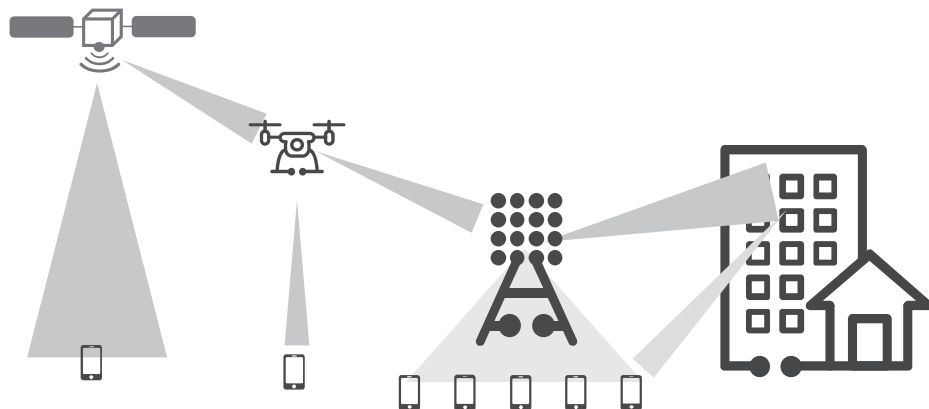


Рис. 25.4. Структура полноспектрального 3D M-MIMO

25.4.1. MIMO со сканированием

Ожидается, что 6G объединит возможности сканирования и связи. Опираясь на возможности искусственного интеллекта, сетевые узлы 6G и пользовательские устройства будут сотрудничать, чтобы воспользоваться мощными возможностями сканирования и сделать сеть 6G осведомленной о своем окружении и ситуации.

Осведомленность о ситуации (situation awareness, SA) – это развивающаяся парадигма связи, в которой сетевое оборудование принимает решения на основе знаний о среде распространения, моделях пользовательского трафика, поведении мобильного пользователя и погодных условиях. Если сетевое оборудование знает местоположение, ориентацию, размер и структуру основного кластера, взаимодействующего с электромагнитной волной в окружающей среде, оно может получить более точную картину состояния канала, такую как направление луча, затухание и потери при распространении, уровень помех и их источник для повышения пропускной способности и надежности сети. Например, радиокарту можно использовать для управления лучом и детектирования CSI со значительно меньшими затратами ресурсов, чем бесцельное сплошное сканирование лучом. В следующих параграфах мы исследуем, как сканирование может помочь в получении CSI и управлении лучом.

- **Получение CSI в реальном времени:** серьезной проблемой для инфраструктуры MIMO в 6G является необходимость быстрого и точного получения CSI. Традиционные методы получения CSI,

используемые в 4G и 5G, увеличивают накладные расходы в области временных и частотных ресурсов. Накладные расходы еще больше возрастают по мере увеличения количества антенн. При использовании традиционных методов увеличение количества антенн также увеличивает задержку измерения и старение CSI. Это серьезная проблема, поскольку она может сделать полученную CSI бесполезной из-за чрезмерного старения, особенно при наличии узконаправленной связи, которая более чувствительна к ошибке CSI. Без интеллектуальной схемы получения CSI в реальном времени рано или поздно все сведется к тому, что измерение CSI и обратная связь будут потреблять все временные/частотные ресурсы. Одно из решений состоит в том, чтобы использовать методы обнаружения и позиционирования, способствующие определению подпространства канала и идентификации лучей-кандидатов. Это решение может уменьшить пространство поиска луча при одновременном снижении энергопотребления как для пользователя, так и для сети. Вдобавок сканирование позволяет отслеживать и прогнозировать беспроводные каналы в реальном времени, что приводит к снижению затрат на поиск луча и получение CSI. Более того, обратная связь CSI 6G должна быть обобщена, чтобы не зависеть от структуры антенны.

Чтобы сделать детектирование CSI более эффективным и менее дорогостоящим, в 6G должны использоваться характеристики каналов терагерцовых линий, а также доступные данные сканирования. Терагерцовый канал еще более разрежен в угловой и временной областях, чем канал миллиметрового диапазона, в то время как доступная полоса пропускания и антенная решетка в этом диапазоне еще больше улучшают временное и угловое разрешение. В результате приемопередатчики ТГц-диапазона могут различать пути распространения с меньшим количеством измерений, чем приемопередатчики миллиметрового диапазона, в зависимости от количества антенных элементов. Кроме того, данные сканирования можно использовать для компенсации воздействия движения и вращения или для прогнозирования возможных направлений приходящих волн. Этот прогноз возможен благодаря знанию местоположения и ориентации точек доступа и конечных пользователей, а также расположения возможных отражателей, таких как стены, потолки и мебель.

- **Упреждающее ориентированное на пользователя управление лучом:** ожидается, что MIMO 6G будет использовать увеличенное количество антенных элементов для передачи и приема, что делает радиointерфейс 6G преимущественно лучевым. Что-

бы упростить технологии MIMO для достижения целей сетей 6G, необходима надежная, гибкая, проактивная система управления лучом с малыми накладными расходами. Для достижения этой цели система управления лучом должна соответствовать определенным принципам проектирования.

Система упреждающего управления лучом прогнозирует и обнаруживает отказ луча, а затем устраняет его. Система должна обеспечивать быстрое восстановление луча, а также автономное отслеживание, подстройку и регулировку лучей. Для реализации упреждающей способности интеллектуальный выбор луча на основе данных должен сопровождаться данными сканирования окружающей среды и данными о местоположении, собранными через радиоинтерфейсы. В качестве альтернативы 6G может использовать другие источники данных, чтобы обеспечить мобильность без передачи обслуживания путем использования управляемых лучей, зависящих от пользователя.

25.4.2. Управляемый радиоканал и топология сети

Возможность управлять средой и топологией сети посредством стратегического развертывания RIS, БПЛА и других неназемных и управляемых узлов представляет собой важный сдвиг парадигмы в MIMO эпохи 6G. Такая управляемость контрастирует с традиционной парадигмой связи, где передатчики и приемники адаптируют свои методы связи для достижения максимальной пропускной способности, предсказываемой теорией информации для данного беспроводного канала. Вместо этого, управляя средой и топологией сети, MIMO 6G стремится изменять беспроводной канал и адаптироваться к условиям сети, чтобы увеличить пропускную способность *сети*.

Одним из способов управления средой является адаптация к топологии сети по мере того, как распределение пользователей и структура трафика меняются с течением времени. Этот подход предполагает использование при необходимости NAPS и БПЛА. Эти неназемные точки доступа тоже имеют свои достоинства и недостатки, как описано в главе 20.

MIMO с поддержкой RIS использует их для повышения характеристик MIMO путем создания интеллектуальных радиоканалов. Чтобы полностью раскрыть потенциал MIMO с помощью RIS, необходимо тщательно изучить новые системные архитектуры и более эффективные схемы или алгоритмы. По сравнению с традиционным формированием диаграммы направленности как на стороне передатчика, так и на стороне приемника, MIMO с поддержкой RIS имеет большую гибкость при реализации улучшенного формирования диаграммы направленности. Эта техноло-

гия также помогает избежать затухания из-за перегораживания трассы между передатчиком и приемником. Поскольку связь между TRP и RIS является общей для всех обслуживаемых пользователей, ее состояние существенно влияет на общие показатели MIMO с поддержкой RIS. Следовательно, необходимо правильно оптимизировать стратегию развертывания индивидуальных и групповых RIS. Обычно накладные расходы на измерения увеличиваются с увеличением количества блоков RIS. Поскольку расстояние между двумя соседними блоками RIS невелико (от одной восьмой до половины длины волны), в высокочастотных диапазонах в любой заданной области решетки существует множество блоков RIS. Использование традиционного сбора данных CSI для оптимизации параметров RIS приведет к очень высоким накладным расходам на измерения для однопользовательской MIMO с поддержкой RIS, не говоря уже о многопользовательской. Чтобы решить эти проблемы, мы должны исследовать гибридные схемы сбора данных CSI, поддерживающие частично активные RIS.

25.4.3. MIMO на FR2 и терагерцовые частоты

В эпоху 6G-технологии миллиметровых волн станут еще более зрелыми и позволят использовать еще более высокочастотный ТГц-диапазон; следовательно, это пойдет на пользу MIMO за счет смягчения его ограничений. На этих более высоких частотах антенные панели становятся плотнее, а лучи имеют тенденцию становиться более узкими. Следовательно, они более чувствительны к движению, старению и ошибкам CSI, ориентации пользователя и сетевого узла и несовпадению калибровки. Более того, радиочастотные устройства и антенны сталкиваются с беспрецедентными проблемами, такими как нелинейность и низкая эффективность. В этом разделе мы изучаем проблемы, связанные с MIMO миллиметрового и ТГц-диапазона. Мы также рассматриваем возможные конструкции MIMO в 6G.

- **Минимальные накладные расходы на управление лучом:** одним из ожидаемых преимуществ 6G является возможность использования большого количества антенн и лучей с ограниченными накладными расходами на измерения и обратную связь. Решение основывается в основном на изначально разреженных каналах во временной и угловой областях, а также на доступе к данным множества датчиков и другой вспомогательной информации. Еще одно ожидаемое преимущество 6G – это возможность измерения луча и наличие обратной связи, которая более эффективно использует радиоресурсы с точки зрения сети, при этом гарантируя, что сеть может обслуживать различные пользовательские устройства с минимальной адаптацией.

- **Естественная поддержка многопанельной связи:** наличие нескольких антенных панелей на стороне пользователя позволяет пользователю поддерживать надежную связь с сетью, а связь является устойчивой к вращению пользователя и блокировке трассы луча. В 5G эта функция реализована как надстройка над базовым вариантом системы, но в 6G ожидается встроенная поддержка многопанельной связи. Сеть и пользователь могут лучше координировать доступные пользовательские панели и лучи. Это станет возможным не только с помощью измерений, но и с помощью данных сканирования, информации гироскопа и, возможно, алгоритмов с поддержкой ИИ. Исходя из опыта 5G, в проекте 6G будет учтено повышенное энергопотребление, необходимое для обслуживания нескольких активных панелей на стороне пользователя. Таким образом, преимущества многопанельной связи будут доступны большему количеству оконечного оборудования.
- **Полное использование доступных данных для управления/указания луча:** для FR1 диаграмма направленности и выбор луча в значительной степени зависят от реализации. Следовательно, поддержка LTE и NR для FR1 не зависит от выбора аналоговых лучей на стороне пользователя. То же самое касается поддержки NR для FR2. Стратегия формирования луча скрыта на обоих концах и неявно представлена посредством идентификации пилот-сигнала без указания точной диаграммы направленности или обновления стратегии. Это приводит к значительным усилиям, направленным на обеспечение взаимодействия. Возможности 6G в области миллиметрового MIMO помогают нам лучше настраивать диаграммы направленности, антенные панели и стратегии формирования диаграммы направленности во времени и пространстве, чтобы оптимизировать лучи с обеих сторон с меньшими накладными расходами и задержкой при более эффективном использовании доступных данных.
- **Надежная компенсация аппаратных ограничений в ТГц-диапазоне:** ожидается, что 6G расширит текущие полосы частот до терагерцового уровня. Такое расширение должно быть тщательно продумано с учетом практических и физических возможностей и проблем, связанных с ТГц-диапазоном, радиочастотными устройствами и обработкой сигналов.

Как и в случае с миллиметровыми волнами, связь в ТГц-диапазоне в значительной степени зависит от механизма формирования луча, в котором используется очень большая антенная решетка из подрешеток. Данная решетка содержит еще большее количество антенных элементов на сторонах передатчика и приемника, чтобы компенсировать бо-

лее серьезные потери на трассе, чем для FR2. Следовательно, антенные элементы становятся очень маленькими и располагаются в непосредственной близости друг от друга. Получившаяся антенная панель должна вмещать не только антенны, но и огромное количество маломощных радиочастотных малошумящих усилителей, фазовращателей, ЦАП и АЦП, как описано в главе 11. Эти компоненты должны работать в очень широком диапазоне частот беспроводной связи. Более того, технологии и конструктивные решения ММО должны учитывать аппаратные ограничения, такие как нелинейность, проблемы калибровки и фазовый шум, которые зачастую сильнее проявляют себя в более широкополосных устройствах. Они также должны учитывать нестабильность луча, вызванную колебаниями физических и электрических свойств компонентов и окружающей среды (например, неблагоприятное воздействие тепла и влажности).

Конструкция антенной панели ТГц-диапазона может столкнуться с дилеммой разрешения АЦП. Цифровое формирование луча высокого порядка может быть реализовано за счет использования нескольких высокоскоростных АЦП, подключенных к различным подрешеткам и радиочастотным цепям. Высокоскоростные АЦП с высоким разрешением дороги и рассеивают много энергии. Следовательно, для многих недорогих и маломощных устройств альтернативой может быть использование нескольких АЦП с низким разрешением (всего 1 бит) или нескольких АЦП с высоким разрешением; возможно, даже их комбинация. Оценка канала и демодуляция данных для АЦП с низким разрешением требуют тщательного проектирования пилотных сигналов, модуляции и формы сигнала для борьбы с высоким шумом квантования.

25.4.4. Антенные решетки со сверхбольшой апертурой

Ожидается, что ELAA значительно улучшит эффективность использования спектра беспроводных радиоинтерфейсов за счет улучшенного пространственного повторного использования, что является преимуществом увеличения количества антенн. Что касается функциональности ММО для 6G, мы ожидаем от разработки ELAA следующего:

- **обобщенное квантование канала и обратная связь:** в 6G могут быть развернуты новые антенные решетки. Из-за явления нестационарности (см. раздел 25.3.3) схема кодовой книги, которая предполагает, что каждый тракт канала виден через все антенные элементы антенной решетки, не подходит для нестационарных каналов. Кроме того, элементы антенны могут иметь нерегулярную и неоднородную компоновку. Следовательно, должен быть разработан обобщенный механизм квантования канала и обрат-

ной связи, чтобы приспособиться к развертыванию различных антенных структур в будущем;

- **MU-MIMO и SU-MIMO более высокого порядка:** в системах 4G и 5G передача MIMO основана на когерентном детектировании. Выделенные опорные сигналы демодуляции предназначены для оценки эффективных каналов и внутри- и межпользовательских помех при выполнении MIMO-коррекции на стороне приемника. Поскольку объем служебных данных опорного сигнала возрастает пропорционально порядку MU-MIMO и однопользовательского MIMO (SU-MIMO), необходимо будет определить компромисс между производительностью и служебными данными в 6G;
- **оценка канала с низкой сложностью:** с увеличением числа антенн, наряду с вышеупомянутыми условиями нестационарного канала в ближней зоне, традиционные методы оценки канала, такие как LMMSE, будут невозможны, если не уменьшить объем служебных данных пилот-сигнала. Это связано не только с возросшим масштабом системы, но и с корректировкой модели физического канала. Методы на основе ИИ могут использовать частичное расположение ненулевых элементов (partial sparsity pattern) для оценки видимых областей рассеивателей. Более того, с помощью управления сферической решеткой мы можем локализовать рассеиватели и идентифицировать видимые области для восстановления нестационарных каналов ближнего поля.

25.4.5. MIMO с поддержкой ИИ

Благодаря своей превосходной способности к обобщениям, точному извлечению признаков и высокому параллелизму ИИ может улучшить показатели систем MIMO.

Но прежде всего ИИ должен способствовать внедрению MIMO в 6G; он поможет справиться с большими размерами решетки и все более разнообразными сетевыми сценариями. Основными проблемами при увеличении размеров решетки в 6G являются высокопроизводительный логический вывод каналов, предварительное кодирование / формирование диаграммы направленности и схемы детектирования. Например, независимо от использования SU-MIMO, MU-MIMO или отслеживания луча работа MIMO зависит от высокоточного детектирования CSI и последующего предварительного кодирования. Однако достижение высокоточного предварительного кодирования становится все более сложной задачей. Отчасти это связано с резким увеличением объема требуемых вычислений. Аналогичные проблемы возникают при детектировании MIMO. Чтобы решить эти проблемы, при проектировании

ММО с поддержкой ИИ следует уделять внимание сбору и использованию данных, развитию теории обучения и новаторским схемам для различных сценариев ММО. Далее мы рассмотрим эти решения более подробно.

Во-первых, развитые и проработанные схемы сбора, проверки и использования данных имеют решающее значение для обеспечения точного усвоения (потенциально) огромных объемов данных, которые предоставляются приемными устройствами и датчиками в 6G. Эти данные позволяют ИИ делать прямые или косвенные выводы и прогнозировать соответствующую информацию для получения CSI, управления лучом, мобильности, передачи обслуживания, кеширования (например, дальность действия устройства, скорость и ориентация), доплеровский сдвиг, AoA и AoD. Если взять в качестве примера получение CSI, благодаря большому количеству данных, предоставляемых устройствами и датчиками, ИИ может извлекать больше полезных признаков из информации об угле и расстоянии, чтобы восстановить или спрогнозировать CSI.

Во-вторых, дальнейшие исследования в области теории обучения и вывода позволят реализовать более точные и эффективные проекты ММО с поддержкой ИИ. Например, изучение интерпретируемости глубоких нейросетей, вероятно, поможет нам разработать методологию выбора правильной модели ИИ и ее последующего эффективного обучения, что послужит фундаментом для создания нативного ИИ. Улучшая способы повторного использования и разбиения на модули предшествующей информации, предоставляемой сканированием, мы можем улучшить обучение и логический вывод ИИ ММО на основе моделей и данных, а в свою очередь обучение на основе моделей и данных улучшит показатели системы ММО с точки зрения точности прогнозирования и накладных расходов на обучение (например, объем обучающих данных и время).

В-третьих, основываясь на внутренних различиях между традиционными решениями ММО на основе моделей и схемами, зависящими от данных, ММО с поддержкой ИИ внесет фундаментальные изменения в традиционную структуру ММО. В качестве наглядного примера: системы на основе автоэнкодеров могут эффективно заменить традиционные сложные каналы. Схема модуля физического уровня для сквозного канала E2E должна быть рассчитана на совместную оптимизацию приемопередатчиков. Учитывая потенциально огромные объемы данных и ожидаемую эволюцию теории машинного обучения, можно предположить, что ММО с поддержкой ИИ станет неотъемлемой чертой 6G. Ниже приведены некоторые принципы проектирования и возможные направления исследований для ММО с поддержкой искусственного интеллекта.

- **Чрезвычайно высокая эффективность:** благодаря использованию искусственного интеллекта MIMO в 6G должен иметь чрезвычайно низкие накладные расходы на измерение канала, обратную связь, управление лучом и демодуляцию данных. Благодаря своим впечатляющим аналитическим способностям ИИ может полностью исследовать внутреннюю структуру многомерных каналов и проводить чрезвычайно эффективные измерения. Более того, используя свою способность к выводам с использованием временной корреляции, ИИ может дополнительно улучшить текущие (или даже будущие) измерения на основе исторических данных измерений. Например, независимо от того, какую форму имеет антенна, вывод канала будет реализован в системах MIMO с чрезвычайно низкими издержками пилот-сигнала. Это возможно благодаря способности ИИ понимать внутреннюю структуру каналов. ИИ также может участвовать в интеллектуальном формировании луча. На основе искусственного интеллекта можно построить более эффективную, инновационную и менее сложную систему управления лучом.
- **Чрезвычайно низкая сложность:** используя сложные логические и вычислительные возможности ИИ, можно эффективно решить несколько традиционных проблем с узкими местами MIMO (например, детектирование MIMO, управление помехами и распределение ресурсов), чтобы приблизиться к теоретическим пределам производительности. Например, детектирование по принципу максимального правдоподобия может обеспечить оптимальную работу и с существующими детекторами; тем не менее по мере увеличения числа переменных, определяющих решение, его сложность возрастает экспоненциально. Несмотря на большой интерес к реализации неоптимальных алгоритмов детектирования, современные алгоритмы страдают от проблем с надежностью и задержкой. Изучая правила детектирования на основе данных, ИИ может предоставить новый подход для достижения почти оптимального детектирования с меньшей сложностью.
- **Чрезвычайно высокая точность и универсальность:** искусственный интеллект хорошо известен своим точным извлечением признаков из различных типов данных. При наличии достаточного количества обучающих данных ИИ может получить более эффективные представления CSI. В качестве примера области, которая может быть улучшена, можно назвать детектирование CSI в NR, которое всегда использует простую модель отображения и квантования, что приводит к относительно плохому использова-

нию разреженных признаков. С другой стороны, ИИ обеспечивает более точное извлечение признаков CSI из данных за счет использования методов, зависящих от данных, для создания более эффективных представлений CSI. Кроме того, благодаря разнообразию данных с внешних устройств также может быть гарантирована универсальность.

Чтобы получить подлинно интеллектуальный массивный MIMO в 6G, мы должны разработать более надежный, более эффективный и простой в реализации MIMO с поддержкой ИИ. Для достижения этой цели следует использовать объяснимый ИИ, чтобы облегчить обучение и повысить его эффективность. Для обеспечения улучшенного и дифференцированного взаимодействия с пользователем потребуется структура, нацеленная на приоритет пользователя. Наконец, для надежного обучения ИИ в сквозном канале E2E все структуры и решения должны обладать двунаправленностью.

25.5. Источники

- [1] J. Spradley, *A volumetric electrically scanned two-dimensional microwave antenna array*, in Proc. 1958 IRE International Convention Record, vol. 6. IEEE, 1966, pp. 204–212.
- [2] V. Tarokh, N. Seshadri, and A. R. Calderbank, *Space-time codes for high data rate wireless communication: Performance criterion and code construction*, IEEE Transactions on Information Theory, vol. 44, no. 2, pp. 744–765, 1998.
- [3] S. M. Alamouti, *A simple transmit diversity technique for wireless communications*, IEEE Journal on Selected Areas in Communications, vol. 16, no. 8, pp. 1451–1458, 1998.
- [4] G. J. Foschini, *Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas*, Bell Labs Technical Journal, vol. 1, no. 2, pp. 41–59, 1996.
- [5] M. O. Damen, H. El Gamal, and G. Caire, *On maximum-likelihood detection and the search for the closest lattice point*, IEEE Transactions on Information Theory, vol. 49, no. 10, pp. 2389–2402, 2003.
- [6] 3GPP, *User equipment (ue) radio transmission and reception; part 1: Range 1 standalone*, 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.101, May 2020, version 16.4.0. http://www.3gpp.org/ftp/Specs/archive/38_series/38.101-3/38101-3-g40.zip.
- [7] 3GPP, *User equipment (ue) radio transmission and reception; part 2: Range 2 standalone*, 3rd Generation Partnership Project (3GPP),

- Technical Specification (TS) 38.101, May 2020, version 16.4.0. http://www.3gpp.org/ftp/Specs/archive/38_series/38.101-3/38101-3-g40.zip.
- [8] 3GPP, *Physical layer procedures for data*, 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.214, July 2020, version 16.2.0. http://www.3gpp.org/ftp/Specs/archive/38_series/38.214/38214-g20.zip.
 - [9] 3GPP, *New SID: Study on supporting NR from 52.6 GHz to 71 GHz*, 3rd Generation Partnership Project (3GPP), Technical Report RP-193259, Dec. 2019. https://www.3gpp.org/ftp/tsg_ran/TSG_RAN/TSGR_86/Docs/RP-193259.zip.
 - [10] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, *Cell-free massive MIMO versus small cells*, IEEE Transactions on Wireless Communications, vol. 16, no. 3, pp. 1834–1850, 2017.
 - [11] Ö. Özdogan, E. Björnson, and J. Zhang, *Downlink performance of cell-free massive MIMO with Rician fading and phase shifts*, in Proc. 2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2019, pp. 1–5.
 - [12] J. Zhang, Y. Wei, E. Björnson, Y. Han, and X. Li, *Spectral and energy efficiency of cellfree massive MIMO systems with hardware impairments*, in Proc. 2017 9th International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2017, pp. 1–6.
 - [13] Z. Chen and E. Björnson, *Can we rely on channel hardening in cell-free massive MIMO?* in Proc. 2017 IEEE Globecom Workshops. IEEE, 2017, pp. 1–6.
 - [14] J. N. Laneman and G. W. Wornell, *Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks*, in Proc. 2002 Global Telecommunications Conference, vol. 1. IEEE, 2002, pp. 77–81.
 - [15] Y. Jing and B. Hassibi, *Distributed space-time coding in wireless relay networks*, IEEE Transactions on Wireless Communications, vol. 5, no. 12, pp. 3524–3536, 2006.
 - [16] A. Bletsas, H. Shin, and M. Z. Win, *Cooperative communications with outage-optimal opportunistic relaying*, IEEE Transactions on Wireless Communications, vol. 6, no. 9, pp. 3450–3460, 2007.
 - [17] L. Bariah, L. Mohjazi, S. Muhaidat, P. C. Sofotasios, G. K. Kurt, H. Yanikomeroglu, and O. A. Dobre, *A prospective look: Key enabling technologies, applications and open research topics in 6G networks*, arXiv preprint arXiv:2004.06049, 2020.
 - [18] S. Ghafoor, N. Boujnah, M. H. Rehmani, and A. Davy, *Mac protocols for terahertz communication: A comprehensive survey*, IEEE Communications Surveys & Tutorials, 2020.
 - [19] Y. Zhao, *A survey of 6G wireless communications: Emerging technologies*, arXiv preprint arXiv:2004.08549, 2020.

- [20] Y. Yifei, Z. Yajun, Z. Baiqing, and P. Sergio, *Potential key technologies for 6G mobile communications*, in SCIENCE CHINA Information Sciences, Springer, vol. 63, pp. 1–19, 2020.
- [21] C. Han, Y. Wu, Z. Chen, and X. Wang, *Terahertz communications (TeraCom): Challenges and impact on 6G wireless systems*, arXiv preprint arXiv:1912.06040, 2019.
- [22] M. H. Alsharif, A. H. Kelechi, M. A. Albream, S. A. Chaudhry, M. S. Zia, and S. Kim, *Sixth generation (6G) wireless networks: Vision, research activities, challenges and potential solutions*, Symmetry, vol. 12, no. 4, p. 676, 2020.
- [23] H. T. Friis, *A note on a simple transmission formula*, Proceedings of the IRE, vol. 34, no. 5, pp. 254–256, 1946.
- [24] J. Tan and L. Dai, *THz precoding for 6G: Applications, challenges, solutions, and opportunities*, arXiv preprint arXiv:2005.10752, 2020.
- [25] M. Di Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin et al., *Smart radio environments empowered by reconfigurable AI metasurfaces: An idea whose time has come*, EURASIP Journal on Wireless Communications and Networking, vol. 2019, no. 1, pp. 1–20, 2019.
- [26] Y. Han, W. Tang, S. Jin, C.-K. Wen, and X. Ma, *Large intelligent surface-assisted wireless communication exploiting statistical CSI*, IEEE Transactions on Vehicular Technology, vol. 68, no. 8, pp. 8238–8242, 2019.
- [27] C. You, B. Zheng, and R. Zhang, *Intelligent reflecting surface with discrete phase shifts: Channel estimation and passive beamforming*, in Proc. 2020 IEEE International Conference on Communications (ICC). IEEE, 2020, pp. 1–6.
- [28] Q. Wu and R. Zhang, *Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts*, IEEE Transactions on Communications, vol. 68, no. 3, pp. 1838–1851, 2019.
- [29] B. Xiong, L. Deng, R. Peng, and Y. Liu, *Controlling the degrees of freedom in metasurface designs for multi-functional optical devices*, Nanoscale Advances, vol. 1, no. 10, pp. 3786–3806, 2019.
- [30] J. Yuan, E. De Carvalho, and P. Popovski, *Wireless communication through frequency modulating reflective intelligent surfaces*, arXiv preprint arXiv:2007.01085, 2020.
- [31] T. J. Cui, *Electromagnetic information computing and intelligent control*, in Proc. Huawei Internal Workshop. Huawei, 2020.
- [32] W. Tang, X. Li, J. Y. Dai, S. Jin, Y. Zeng, Q. Cheng, and T. J. Cui, *Wireless communications with programmable metasurface: Transceiver design and experimental results*, China Communications, vol. 16, no. 5, pp. 46–61, 2019.

- [33] W. Tang, J. Y. Dai, M. Z. Chen, K.-K. Wong, X. Li, X. Zhao, S. Jin, Q. Cheng, and T. J. Cui, *Mimo transmission through reconfigurable intelligent surface: System design, analysis, and implementation*, IEEE Journal on Selected Areas in Communications, 2020.
- [34] Y. Cui and H. Yin, *An efficient CSI acquisition method for intelligent reflecting surface assisted mmwave networks*, arXiv preprint arXiv:1912.12076, 2019.
- [35] V. C. Thirumavalavan and T. S. Jayaraman, *BER analysis of reconfigurable intelligent surface assisted downlink power domain NOMA system*, in Proc. 2020 International Conference on COMMunication Systems & NETworkS (COMSNETS). IEEE, 2020, pp. 519–522.
- [36] Q. Wu and R. Zhang, *Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints*, arXiv preprint arXiv:1910.06220, 2019.
- [37] N. M. Tran, M. M. Amri, J. H. Park, S. I. Hwang, D. I. Kim, and K. W. Choi, *A novel coding metasurface for wireless power transfer applications*, Energies, vol. 12, no. 23, p. 4488, 2019.
- [38] J. W. Wu, Z. X. Wang, L. Zhang, Q. Cheng, S. Liu, S. Zhang, J. M. Song, and T. J. Cui, *Anisotropic metasurface holography in 3D space with high resolution and efficiency*, IEEE Transactions on Antennas and Propagation, 2020.
- [39] E. Björnson, J. Hoydis, and L. Sanguinetti, *Massive MIMO has unlimited capacity*, IEEE Transactions on Wireless Communications, vol. 17, no. 1, pp. 574–590, 2017.
- [40] S. Hu, F. Rusek, and O. Edfors, *Beyond massive MIMO: The potential of data transmission with large intelligent surfaces*, IEEE Transactions on Signal Processing, vol. 66, no. 10, pp. 2746–2758, 2018.
- [41] E. Björnson and E. G. Larsson, *How energy-efficient can a wireless communication system become?* in Proc. 2018 52nd Asilomar Conference on Signals, Systems, and Computers. IEEE, 2018, pp. 1252–1256.
- [42] N. Heuvel et al., *Ericsson mobility report*, Ericsson AB, Technol. Emerg. Business, Stockholm, Sweden, Technical Report EAB-17, vol. 5964, 2017.
- [43] A. Amiri, M. Angelichinoski, E. De Carvalho, and R. W. Heath, *Extremely large aperture massive MIMO: Low complexity receiver architectures*, in Proc. 2018 IEEE Globecom Workshops. IEEE, 2018, pp. 1–6.
- [44] E. Björnson and L. Sanguinetti, *Making cell-free massive MIMO competitive with MMSE processing and centralized implementation*, IEEE Transactions on Wireless Communications, vol. 19, no. 1, pp. 77–90, 2019.

- [45] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, *Machine learning paradigms for next-generation wireless networks*, IEEE Wireless Communications, vol. 24, no. 2, pp. 98–105, 2016.
- [46] Z. Qin, H. Ye, G. Y. Li, and B.-H. F. Juang, *Deep learning in physical layer communications*, IEEE Wireless Communications, vol. 26, no. 2, pp. 93–99, 2019.
- [47] L. Sanguinetti, A. Zappone, and M. Debbah, *Deep learning power allocation in massive MIMO*, in Proc. 2018 52nd Asilomar Conference on Signals, Systems, and Computers. IEEE, 2018, pp. 1257–1261.
- [48] C. D’Andrea, A. Zappone, S. Buzzi, and M. Debbah, *Uplink power control in cell-free massive MIMO via deep learning*, in Proc. 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). IEEE, 2019, pp. 554–558.
- [49] C.-K. Wen, W.-T. Shih, and S. Jin, *Deep learning for massive MIMO CSI feedback*, IEEE Wireless Communications Letters, vol. 7, no. 5, pp. 748–751, 2018.
- [50] J. Wang, Y. Ding, S. Bian, Y. Peng, M. Liu, and G. Gui, *UL-CSI data driven deep learning for predicting DL-CSI in cellular FDD systems*, IEEE Access, vol. 7, pp. 96 105–96 112, 2019.
- [51] P. Dong, H. Zhang, G. Y. Li, I. S. Gaspar, and N. Naderi Alizadeh, *Deep CNN-based channel estimation for mmWave massive MIMO systems*, IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 5, pp. 989–1000, 2019.
- [52] E. Balevi, A. Doshi, and J. G. Andrews, *Massive MIMO channel estimation with an untrained deep neural network*, IEEE Transactions on Wireless Communications, vol. 19, no. 3, pp. 2079–2090, 2020.
- [53] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, *Deep-learning-based millimeterwave massive MIMO for hybrid precoding*, IEEE Transactions on Vehicular Technology, vol. 68, no. 3, pp. 3027–3032, 2019.
- [54] T. Mir, M. Z. Siddiqi, U. Mir, R. Mackenzie, and M. Hao, *Machine learning inspired hybrid precoding for wideband millimeter-wave massive MIMO systems*, IEEE Access, vol. 7, pp. 62 852–62 864, 2019.
- [55] T. D. N. Samuel and A. Wiesel, *Deep MIMO detection*, in IEEE 18th International Workshop on Signal Processing and Advanced Wireless Communications (SPAWC). IEEE, 2017, pp. 1–5.
- [56] H. He, C.-K. Wen, S. Jin, and G. Y. Li, *Model-driven deep learning for MIMO detection*, IEEE Transactions on Signal Processing, vol. 68, pp. 1702–1715, 2020.
- [57] L. Allen, M. W. Beijersbergen, R. Spreeuw, and J. Woerdman, *Orbital angular momentum of light and the transformation of Laguerre–Gaussian laser modes*, Physical Review A, vol. 45, no. 11, p. 8185, 1992.

- [58] B. Thidé, H. Then, J. Sjöholm, K. Palmer, J. Bergman, T. Carozzi, Y. N. Istomin, N. Ibragimov, and R. Khamitova, *Utilization of photon orbital angular momentum in the low-frequency radio domain*, Physical Review Letters, vol. 99, no. 8, p. 087701, 2007.
- [59] S. M. Mohammadi, L. K. Daldorff, J. E. Bergman, R. L. Karlsson, B. Thidé, K. Forozesh, T. D. Carozzi, and B. Isham, *Orbital angular momentum in radio system study*, IEEE Transactions on Antennas and Propagation, vol. 58, no. 2, pp. 565–572, 2009.
- [60] F. Tamburini, E. Mari, A. Sponselli, B. Thidé, A. Bianchini, and F. Romanato, *Encoding many channels on the same frequency through radio vorticity: First experimental test*, New Journal of Physics, vol. 14, no. 3, p. 033001, 2012.
- [61] D. Lee, H. Sasaki, H. Fukumoto, Y. Yagi, T. Kaho, H. Shiba, and T. Shimizu, *An experimental demonstration of 28 GHz band wireless OAM-MIMO (orbital angular momentum multi-input and multi-output) multiplexing*, in Proc. 2018 IEEE 87th Vehicular Technology Conference (VTC-Spring). IEEE, 2018, pp. 1–5.
- [62] NEC Corporation, *NEC successfully demonstrates real-time digital OAM mode multiplexing transmission in the 80 GHz-band for the first time*, Dec. 2018. https://www.nec.com/en/press/201812/global_20181219_02.html.
- [63] NEC Corporation, *NEC successfully demonstrates real-time digital OAM mode multiplexing transmission over 100 m in the 150 GHz-band for the first time*, March 2020. https://www.nec.com/en/press/202003/global_20200310_01.html.
- [64] M. Klemes, H. Boutayeb, and F. Hyjazie, *Orbital angular momentum (OAM) modes for 2-D beam-steering of circular arrays*, in Proc. 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). IEEE, 2016, pp. 1–5.
- [65] M. Klemes, *Reception of OAM radio waves using pseudo-doppler interpolation techniques: A frequency-domain approach*, Applied Sciences, vol. 9, no. 6, p. 1082, 2019.
- [66] R. Ni, Y. Lv, Q. Zhu, and M. Debbah, *Electromagnetic field fingerprint method for circularly polarized OAM*, in Proc. 2020 IEEE International Conference on Communications Workshops. IEEE, 2020, pp. 1–6.
- [67] Y. Lv, Q. Zhu, and R. Ni, *Modulo-based phase gradient method for OAM mode detection*, in Proc. 2020 9th IEEE/CIC International Conference on Communications (ICCC). IEEE, 2020, pp. 1–6.
- [68] J. D. Kraus and R. J. Marhefka, *Antennas for all applications (Third Edition)*. McGraw- Hill, 2001.
- [69] R. Ni, Y. Lv, Q. Zhu, G. Wang, and G. He, *Degrees of freedom of multi-mode-multispatial in-line-of-sight channels*, in Proc. 2020 IEEE GLOBECOM Conference. IEEE, 2020, pp. 1–6.

- [70] S. M. Mohammadi, L. K. Daldorff, K. Forozesh, B. Thidé, J. E. Bergman, B. Isham, R. Karlsson, and T. Carozzi, *Orbital angular momentum in radio: Measurement methods*, Radio Science, vol. 45, no. 4, pp. 1–14, 2010.
- [71] H. Wu, Y. Yuan, Z. Zhang, and J. Cang, *UCA-based orbital angular momentum radio beam generation and reception under different array configurations*, in Proc. 2014 Sixth International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2014, pp. 1–6.
- [72] S. Rostami, W. Saad, and C. S. Hong, *Deep learning with persistent homology for orbital angular momentum (oam) decoding*, IEEE Communications Letters, vol. 24, no. 1, pp. 117–121, 2019.
- [73] S. Zheng, X. Hui, J. Zhu, H. Chi, X. Jin, S. Yu, and X. Zhang, *Orbital angular momentum mode-demultiplexing scheme with partial angular receiving aperture*, Optics Express, vol. 23, no. 9, pp. 12 251–12 257, 2015.
- [74] C. Zhang and M. Lu, *Detecting the orbital angular momentum of electromagnetic waves using virtual rotational antenna*, Scientific Reports, vol. 7, no. 1, pp. 1–8, 2017.
- [75] O. Edfors and A. J. Johansson, *Is orbital angular momentum (oam) based radio communication an unexploited area?* IEEE Transactions on Antennas and Propagation, vol. 60, no. 2, pp. 1126–1131, 2011.
- [76] M. Tamagnone, C. Craeye, and J. Perruisseau-Carrier, *Comment on encoding many channels on the same frequency through radio vorticity: First experimental test*, New Journal of Physics, vol. 14, no. 11, p. 118001, 2012.
- [77] J. Xu, *Degrees of freedom of oam-based line-of-sight radio systems*, IEEE Transactions on Antennas and Propagation, vol. 65, no. 4, pp. 1996–2008, 2017.
- [78] R. Gaffoglio, A. Cagliero, G. Vecchi, and F. P. Andriulli, *Vortex waves and channel capacity: Hopes and reality*, IEEE Access, vol. 6, pp. 19814–19822, 2017.
- [79] A. Omar, *Dependence of beamforming on the excitation of orbital angular momentum (OAM) modes*, IEEE Transactions on Antennas and Propagation, 2020.
- [80] S. Saito, H. Suganuma, K. Ogawa, and F. Maehara, *Performance analysis of OAMMIMO using SIC in the presence of misalignment of beam axis*, in Proc. 2019 IEEE International Conference on Communications Workshops. IEEE, 2019, pp. 1–6.
- [81] D. Prather, *Toward holographic rf systems for wireless communications and networks*, IEEE ComSoc Technology News, 2016.
- [82] N. Rajatheva, I. Atzeni, E. Bjornson, A. Bourdoux, S. Buzzi, J.-B. Dore, S. Erkucuk, M. Fuentes, K. Guan, Y. Hu et al., *White paper on broadband connectivity in 6g*, arXiv preprint arXiv:2004.14247, 2020.

- [83] M. Latva-aho, K. Leppänen, F. Clazzer, and A. Munari, *Key drivers and research challenges for 6g ubiquitous wireless intelligence*, 6G Flagship, University of Oulu, Sept. 2019.
- [84] Y. Yuan, Y. Zhao, B. Zong, and S. Parolari, *Potential key technologies for 6g mobile communications*, Science China Information Sciences, vol. 63, pp. 1–19, 2020.
- [85] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, *Massive mimo is a reality – what is next?: Five promising research directions for antenna arrays*, Digital Signal Processing, vol. 94, pp. 3–20, 2019.
- [86] E. J. Black, *Holographic beam forming and mimo*, Pivotal Commware, 2017. <https://pivotalcommware.com/wp-content/uploads/2017/12/Holographic-Beamforming-WP-v.6C-FINAL.pdf>.
- [87] A. Pizzo, T. L. Marzetta, and L. Sanguinetti, *Degrees of freedom of holographic MIMO channels*, in Proc. 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2020, pp. 1–5.
- [88] A. Pizzo, T. L. Marzetta, and L. Sanguinetti, *Spatial characterization of holographic mimo channels*, arXiv preprint arXiv:1911.04853, 2019.
- [89] A. S. Poon, R. W. Brodersen, and D. N. Tse, *Degrees of freedom in multiple-antenna channels: A signal space approach*, IEEE Transactions on Information Theory, vol. 51, no. 2, pp. 523–536, 2005.
- [90] D. W. Prather, S. Shi, G. J. Schneider, P. Yao, C. Schuetz, J. Murakowski, J. C. Deroba, F. Wang, M. R. Konkol, and D. D. Ross, *Optically upconverted, spatially coherent phased-array-antenna feed networks for beam-space mimo in 5g cellular communications*, IEEE Transactions on Antennas and Propagation, vol. 65, no. 12, pp. 6432–6443, 2017.

Глава 26

Интеграция каналов доступа и прямых каналов связи

26.1. Почему нужна интеграция каналов

Радиодоступ в беспроводных системах в основном построен на архитектуре P2MP, в которой мобильные устройства подключаются к одной или нескольким базовым станциям. С момента внедрения 4G беспроводная связь поддерживает использование *одноранговых прямых каналов* – иначе называемых *боковыми каналами* (sidelink) – для беспроводной связи типа «устройство–устройство» на ближних расстояниях. Связь между транспортными средствами (vehicle-to-vehicle, V2V) как практический пример связи D2D была определена в системах 4G LTE, а улучшенный протокол V2V представлен в системах 5G NR.

В структуре радиоинтерфейса прямые каналы NR заимствуют ключевые функции каналов доступа NR (связь между пользовательскими устройствами и базовыми станциями). Таким образом, прямые каналы стали неотъемлемой частью стандарта NR. В частности, прямые каналы NR наследуют ключевые концепции и базовую структуру от каналов доступа NR, хотя и с некоторыми упрощениями, такими как полоса пропускания, гибкая структура кадра, кодирование, модуляция и формы сигналов. Пользовательские устройства могут быстро устанавливать прямые каналы друг с другом, независимо от того, подключены они к одной и той же соте NR или нет.

Как было сказано в части II, многие новые применения 6G требуют высокой пропускной способности и низкой задержки. Такие применения, включая Ultimate XR, голографический дисплей, тактильную передачу, высокодинамичное управление движением, позиционирование и визуализацию, обычно подразумевают связь на коротком расстоянии между двумя устройствами. Следовательно, чтобы соответствовать требованиям к пропускной способности на уровне Тбит/с и задержке меньше нескольких миллисекунд, в эпоху 6G потребуются новые тех-

нологии ближнего радиуса действия, такие как *прямой суперканал* (super-sidelink, далее просто *суперканал*). Идеальными кандидатами на такую технологию являются терагерцовая связь и оптическая беспроводная связь (optical wireless communication, OWC), которые предлагают чрезвычайно широкую полосу пропускания. Таким образом, ожидается, что прямые суперканалы 6G будут иметь чрезвычайно высокую пропускную способность и чрезвычайно низкую задержку благодаря использованию нового ТГц-диапазона или OWC и смогут поддерживать локальные меш-сети.

Суперканалы 6G и связанные с ними ячеистые сети должны быть неотъемлемой частью общей мобильной системы. Это необходимо для удовлетворения как технических, так и прикладных требований. С точки зрения технических требований, высокая распространенность мобильных устройств делает связь между мобильными устройствами на малых расстояниях более возможной. Кроме того, подход D2D или V2V, представленный в LTE или NR в качестве связи по прямому каналу, разгружает трафик базовых станций и облегчает связь между устройствами на близких расстояниях. В 6G для каналов доступа будут широко использоваться миллиметровые волны, а для связи на короткие расстояния на более высоких частотах будут применяться ТГц-технологии и OWC. С точки зрения требований к приложениям, текущий спектр ИМТ становится слишком переполненным. Как упоминалось в предыдущих главах, 6G свяжет между собой огромное количество интеллектуальных устройств. Кроме того, различным типам устройств потребуются гибкие соединения для обмена контентом, ориентированным на человека, а также для других типов межмашинных коммуникаций, таких как роевой интеллект и управление движением на промышленном уровне. Все это требует огромной пропускной способности. Следовательно, чтобы справиться со стремительно растущим спросом на беспроводной трафик, срочно требуется *интегрированный суперканал*, использующий новые диапазоны беспроводного спектра.

Интегрированные суперканалы имеют несколько преимуществ перед выделенными каналами доступа и прямыми каналами. Например, динамическая локально организованная суперканальная сеть, такая как многозвенная меш-сеть, может ощутимо разгрузить трафик канала доступа, увеличить общую емкость соты, повысить скорость передачи данных пользователя и, возможно, даже уменьшить задержку. Эта интегральная архитектура также важна в сценариях применения V2V. В частности, меш-сеть может агрегировать всю информацию датчиков со всех соседних беспилотных транспортных средств. В этой меш-сети устройства могут обнаруживать и подключаться друг к другу через суперканалы, даже если они не находятся в зоне покрытия соты. Интег-

ральная архитектура позволит меш-сети с суперканалами найти баланс между гибкостью, пропускной способностью и мобильностью, тем самым достигнув превосходного сочетания показателей.

На рис. 26.1 показан пример внутренней ячеистой сети с суперканалами 6G. Сеть включает в себя несколько каналов доступа и суперканалы для покрытия в разных комнатах. Будучи тесно интегрированными с каналами доступа 6G и используя методы ретрансляции, суперканалы могут достигать чрезвычайно высокой пропускной способности и низкой задержки.

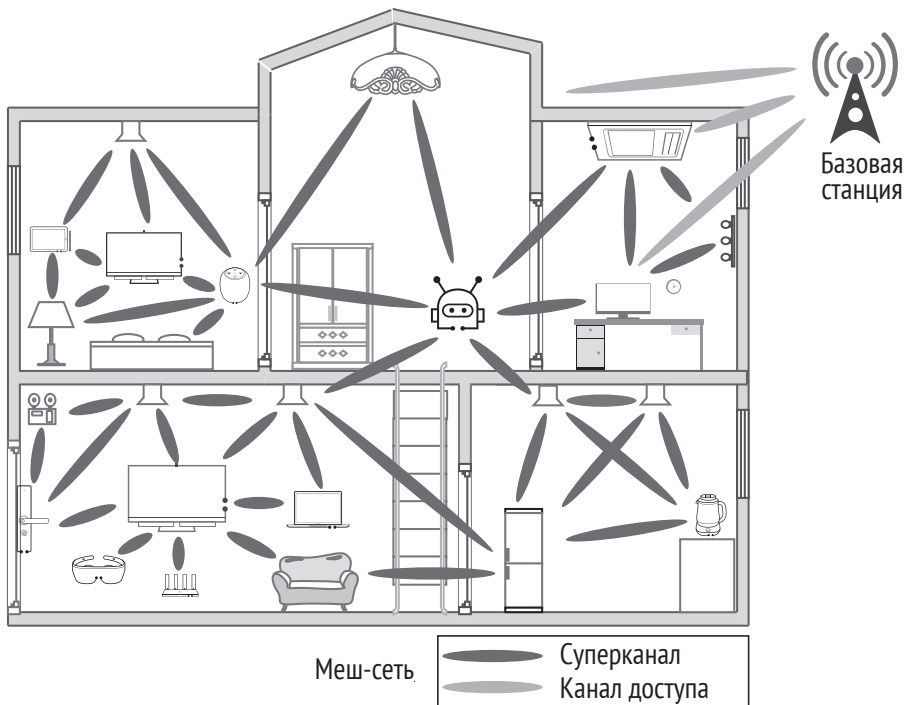


Рис. 26.1. Пример внутридомовой ячеистой сети со связью через суперканалы 6G

На рис. 26.2 представлен еще один пример, демонстрирующий меш-сеть с суперканалами 6G V2V. Эта сеть включает в себя несколько каналов доступа и суперканалы для обмена информацией с датчиков и изображениями с высоким разрешением в реальном времени для автономного вождения. В сценарии V2V критически важно использовать меш-сеть с чрезвычайно высокой пропускной способностью и малой задержкой. Благодаря интеграции каналов доступа и суперканалов можно быстро создавать и настраивать меш-сети.

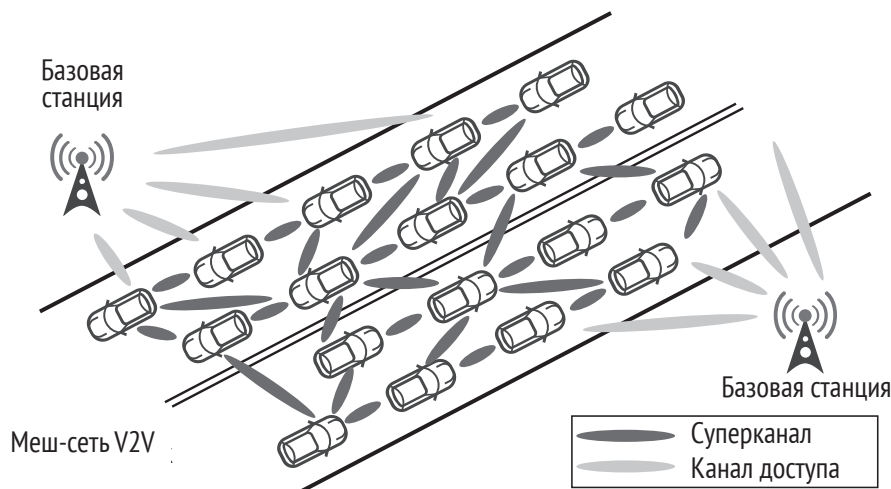


Рис. 26.2. Пример ячеистой сети с суперканалами 6G V2V

Суперканалы, работающие в диапазонах ТГц и оптических частот, имеют несколько преимуществ по сравнению с традиционными технологиями радиочастотной связи. Например, они используют обширный спектр и обладают устойчивостью к помехам. Если вышеупомянутые полосы частот задействованы для организации суперканалов, желательно использовать унифицированную конструкцию физического уровня для ТГц-, оптической и радиочастотной связи.

26.2. Обзор существующих решений

В 5G NR версии R16 прямые каналы наследуют структуру каналов доступа (также известных как каналы Uu в 3GPP). Это означает аналогичное использование той же формы сигнала, кодирования, модуляции, структуры кадра и MIMO. Следовательно, прямые каналы NR и каналы доступа Uu могут быть интегрированы для упрощения реализации пользовательского оборудования и даже повторного использования наборов микросхем, что позволяет значительно сократить расходы для поставщиков оборудования.

Кроме того, прямые каналы связи 5G также используют некоторые новые функции, включая синхронизацию, выделение ресурсов, HARQ и подавление взаимных помех. Эти функции могут жестко или гибко контролироваться мобильной сетью. Ниже мы рассмотрим их подробнее.

- Для синхронизации по прямому каналу связи разные типы источников синхронизации, каждый из которых связан с приоритетом синхронизации, указываются для разных сценариев – в зоне по-

крытия, вне зоны покрытия и частичной зоне покрытия, – чтобы получить одно и то же опорное время. Например, в сценарии «в зоне покрытия» в качестве источника синхронизации может выступать сеть. Или, если одно пользовательское устройство находится в зоне доступа, а другое – нет, то устройство, находящееся в зоне доступа, может действовать как источник синхронизации.

- Для выделения ресурсов прямого канала были определены два режима. Режим 1 применяется к сценариям в зоне покрытия, когда ресурсы прямого канала выделяет сеть. В таких сценариях канал Uu управляет прямым каналом. Режим 2 применяется как к сценариям в зоне покрытия, так и вне зоны покрытия, когда пользовательские устройства автономно выбирают ресурсы для связи по прямому каналу. Эти ресурсы могут быть оперативно сконфигурированы сетью, когда передающее пользовательское устройство находится в зоне обслуживания, или предварительно сконфигурированы, когда оно находится вне зоны.
- Повторная передача (HARQ) по прямому каналу повышает надежность групповой и одноадресной передачи. Показательным примером является расширение услуги NR V2X на все автомобили в беспилотной автоколонне. В данном случае все машины могут получать информацию от головной машины для управления автоколонной. Эта информация позволяет транспортным средствам скоординированно двигаться с меньшим промежутком, чем обычно, при этом все транспортные средства движутся вместе и в одном направлении. В этом случае ресурсы обратной связи и повторной передачи HARQ могут быть оперативно сконфигурированы сетью, когда автомобили находятся в зоне обслуживания, или предварительно сконфигурированы, когда они находятся вне зоны.
- При борьбе со взаимными помехами между внутрисполосными прямыми каналами используется управление мощностью без обратной связи на основе потерь в тракте каналов Uu и прямых каналов. Например, минимальные потери в тракте канала Uu и прямого канала используются для определения мощности передатчика. В режиме 2 выделения ресурсов для прямого канала передатчик пользовательского устройства может полагаться на декодирование SCI и измерение мощности принимаемого опорного сигнала (RSRP) для определения ресурсов-кандидатов, потенциально доступных для использования при автономном выборе, избегая при этом конфликтов с ресурсами, занятыми другими оконечными устройствами. Это помогает устройству бороться с помехами, которые оно может как порождать, так и принимать.

Что касается использования спектра, управления лучом и ретрансляции в случае прямого канала, ниже описаны некоторые сопутствующие ограничения.

- Если говорить о спектре прямых каналов, то до 5G R16 поддерживаются лицензированные диапазоны до 7 ГГц и нелицензированные диапазоны интеллектуальной транспортной системы (intelligent transportation system, ITS). Прямые каналы NR поддерживают как внутрисполосные общие несущие с каналами Uu, так и внеполосные выделенные несущие прямых каналов. Первые требуют более жесткого управления сетью и используют мультиплексирование с временным разделением (TDM) между каналами доступа Uu и прямыми каналами за счет увеличения задержки, тогда как последние предлагают меньшую задержку, но потребляют больше ресурсов спектра.
- Поскольку прямой канал в версии R16 в основном ориентирован на частоты ниже 7 ГГц, передача по прямому каналу с помощью луча еще не полностью поддерживается. Тем не менее в связи с началом интенсивных исследований диапазона миллиметровых волн ожидается, что в будущих версиях будет продолжена оптимизация передачи по прямым каналам на основе луча как в зоне покрытия, так и вне зоны покрытия.
- Относительно ретрансляции прямых каналов в стандарте NR еще не достигнут консенсус по ретрансляции каналов уровня 2 и уровня 3. В настоящее время прямые каналы NR не поддерживают многозвенную ретрансляцию, которая особенно важна, когда плотность пользовательских устройств высока, а расстояние между ними по прямому каналу связи очень мало.

26.3. Ожидания от новых разработок и потенциальные направления исследований

При разработке сети 6G необходимо предусмотреть вспомогательные технологии для интеграции суперканалов и каналов доступа. Ожидания от разработок и потенциальные направления исследований обсуждаются в следующих разделах.

26.3.1. Ключевые технологии суперканалов

Суперканалы сети 6G должны использовать более высокочастотный спектр, особенно для связи на короткие расстояния. Например, уже предлагалось использование сверхвысоких частот, таких как ТГц-диапазон и оптическая беспроводная связь (OWC) [1, 2].

За последние несколько десятилетий терагерцовая связь вызвала большой исследовательский интерес, особенно с точки зрения ТГц-модуляторов, антенн и моделирования и оценки ТГц-каналов [3]. ТГц-диапазон, который находится в области спектра от 100 ГГц до 10 ТГц и находится между миллиметровым и оптическим диапазонами, имеет уникальные характеристики. Например, он предлагает в 100 раз большую полосу пропускания, чем обычный радиочастотный спектр, что означает, что он может обеспечить сверхвысокую пропускную способность. Однако компоненты связи ТГц-диапазона (например, усилители мощности, смесители и антенны) не лишены проблем с точки зрения энергоэффективности и конструктивных ограничений. Следовательно, необходимы дальнейшие исследования для повышения характеристик устройств. Также важно вести разработку на уровне системы (например, совместная разработка физического уровня и формы сигналов для компенсации ограничений устройства).

За последние несколько лет технологии OWC достигли значительного прогресса. Например, связь в видимом свете (visible light communication, VLC) и точное воспроизведение света (light fidelity, Li-Fi) уже либо стандартизированы, либо продемонстрированы [4, 5], а оптика в свободном пространстве (free space optical, FSO) использовалась для дальней связи D2D [6], где источники света напрямую излучают в беспроводные каналы. Технология VLC использует светодиод и фотодиод в качестве источника и приемника соответственно и модулирует свет, изменяя интенсивность, чтобы отображать нули и единицы непосредственно в оптическую волну; эта модуляция должна выполняться достаточно быстро, чтобы избежать эффекта мерцания. Использование технологии VLC в 5G предварительно протестировано в работе [7]. Однако на технологию OWC существенно влияют погодные условия и атмосферная турбулентность. Например, туман ослабляет оптические сигналы, что приводит к недоступности канала OWC. На смягчение этой проблемы направлена гибридная технология RF/OWC, которая сочетает в себе преимущества радиочастотного доступа (надежность) и OWC (емкость) и является приемлемым дополнительным вариантом [8].

26.3.2. Интеграция суперканалов с каналами доступа как единая конструкция

Интеграция суперканалов и каналов доступа подразумевает использование экстремально широкой полосы спектра, что позволяет достичь сверхвысокой пропускной способности и скорости передачи данных. В этом разделе основное внимание уделяется тому, как достичь эффективной интеграции (например, унифицированный радиоинтерфейс) и реализовать сеть с суперканалами (например, меш-сеть).

Суперканалы 6G нуждаются в исследовании унифицированных процедур физического уровня, таких как широкополосная передача, механизмы произвольного доступа, а также измерение и выбор ресурсов. В начале интеграции, возможно, придется заново рассмотреть составные части радиоинтерфейса, включая формы сигналов, модуляцию, мультиплексирование, кодирование и схемы MIMO. За последние несколько лет в этом направлении был достигнут обнадеживающий прогресс. Например, авторы [7] рассматривали интеграцию 5G и VLC, используя VLC для связи между светофорами и транспортными средствами. А в [9, 10] проанализированы показатели систем VLC на основе схемы 5G OFDMA или NOMA-OFDMA.

Как упоминалось ранее, к сетям 6G будет подключено огромное количество устройств. Однако из-за ограниченной зоны покрытия суперканалов связи и препятствий для сигналов (например, стены, мебель и даже люди могут препятствовать прохождению сигнала) важным фактором при разработке суперканалов связи является многозвенная меш-сеть D2D. Выяснение того, как объединить новую конструкцию радиоинтерфейса и сетевые технологии (такие как каналы доступа и меш-сети прямых каналов), также является важной областью исследований.

Также следует учитывать распределение и совместное использование спектра различными суперканалами и каналами доступа. Из-за узкого луча и короткого расстояния ТГЦ или OWC передачи относительно легко разделить один и тот же спектр или канал между разными суперканалами и каналами доступа. Тем не менее это повышает требования к подавлению взаимных помех между суперканалами с использованием технологии направленного луча. Интегрированная конструкция должна максимизировать пропускную способность сети за счет использования персонализированной конструкции радиоинтерфейса канала доступа, включая различные возможности дуплексирования и инвариантное направление канала. При проектировании суперканалов 6G также может потребоваться использование нелицензируемого спектра или измерения с помощью лицензированной части спектра.

Ожидается, что технология суперканалов 6G будет поддерживать формирование лучей как для однозвенной, так и для многозвенной передачи. Поскольку на сверхвысокой частоте суперканалов лучи могут быть очень узкими, необходимы исследования в области управления лучом с низкими издержками и минимальным влиянием на мобильность. Исследование также должно учитывать тот факт, что передача луча между передатчиком и приемником может быть асимметричной, особенно в OWC, из-за ограничений мощности, размера и сложности устройств [2]. Также необходимы исследования интеграции суперканала и канала доступа с точки зрения управления радиочастотным, тера-

герцовым и оптическим лучом, чтобы решить проблемы с асимметричной передачей луча и блокировкой трассы для определенных каналов. При этом также следует учитывать интеграцию нескольких суперканалов и каналов доступа. В главе 25 более подробно рассмотрены вопросы, связанные с лучом.

В отчете [11] обсуждается ретрансляция уровней 2 и 3, но ни один из подходов не применим для передачи трафика или многозвенной передачи в сценариях, чувствительных к задержке. Соответственно, проводятся исследования по использованию гибких технологий пересылки, таких как декодирование и пересылка (DF), сжатие и пересылка (CF), а также ретрансляция с усилением и пересылкой (AF), для раздельного или комбинированного использования в меш-сети. Кроме того, кодирование сети РНУ может использоваться в ячеистой сети для повышения эффективности и надежности, а также снижения задержки. Этот вопрос был подробно рассмотрен в главе 23.

Если удастся эффективно реализовать радиоинтерфейс и сетевые технологии, мы считаем, что интеграция суперканала и канала доступа 6G приведет к смене парадигмы. С учетом гибкой топологии сети и возможности связи на скорости порядка Тбит/с, которые также обеспечивают меш-сети с суперканалами 6G, перед нами откроется потенциально новое пространство применений разнообразных устройств.

26.4. Источники

- [1] Y. Corre, G. Gougeon, J.-B. Doré, S. Bicaïs, B. Miscopein, E. Faussurier, M. Saad, J. Palicot, and F. Bader, *Sub-THz spectrum as enabler for 6G wireless communications up to 1 Tbit/s*, in Proc. 6G Wireless Summit, Mar. 2019, Levi Lapland, Finland.
- [2] M. Z. Chowdhury, M. Shahjalal, M. Hasan, Y. M. Jang et al., *The role of optical wireless communication technologies in 5G/6G and IoT solutions: Prospects, directions, and challenges*, Applied Sciences, vol. 9, no. 20, p. 4367, 2019.
- [3] Z. Chen, X. Ma, B. Zhang, Y. Zhang, Z. Niu, N. Kuang, W. Chen, L. Li, and S. Li, *A survey on terahertz communications*, China Communications, vol. 16, no. 2, pp. 1–35, 2019.
- [4] A. Bensky, *Short-range wireless communication*. Newnes, 2019.
- [5] D. Tsonev, S. Videv, and H. Haas, *Light fidelity (Li-Fi): Towards all-optical networking*, in Proc. Conference on Broadband Access Communication Technologies VIII, vol. 9007. International Society for Optics and Photonics, 2014, p. 900702.
- [6] A. K. Majumdar, *Advanced free space optics (FSO): A systems approach*. Springer, 2014.

- [7] F. Nizzi, T. Pecorella, S. Caputo, L. Mucchi, R. Fantacci, M. Bastianini, C. Cerboni, A. Buzzigoli, A. Fratini, T. Nawaz et al., *Data dissemination to vehicles using 5G and VLC for smart cities*, in Proc. 2019 AEIT International Annual Conference (AEIT). IEEE, 2019, pp. 1–5.
- [8] M. Z. Chowdhury, M. T. Hossan, M. K. Hasan, and Y. M. Jang, *Integrated RF/optical wireless networks for improving QoS in indoor and transportation applications*, Wireless Personal Communications, vol. 107, no. 3, pp. 1401–1430, 2019.
- [9] H. Marshoud, P. C. Sofotasios, S. Muhaidat, G. K. Karagiannidis, and B. S. Sharif, *On the performance of visible light communication systems with non-orthogonal multiple access*, IEEE Transactions on Wireless Communications, vol. 16, no. 10, pp. 6350–6364, 2017.
- [10] B. Lin, W. Ye, X. Tang, and Z. Ghassemlooy, *Experimental demonstration of bidirectional NOMA-OFDMA visible light communications*, Optics Express, vol. 25, no. 4, pp. 4348–4355, 2017.
- [11] 3GPP, *Study on NR sidelink relay*, 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.836, Sept. 2020, version 17.0.0. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specification-id=3725>.

Резюме части V

В этой части мы обсудили потенциальные функции радиоинтерфейса 6G и потенциальные вспомогательные технологии, на которые будут опираться универсальные сценарии использования, более строгие KPI и более широкий диапазон спектра в 6G. Технологии искусственного интеллекта и машинного обучения сделают радиоинтерфейс 6G настраиваемым и интеллектуальным. Это, в свою очередь, позволит радиоинтерфейсу поддерживать настраиваемую оптимизацию параметров передачи, чтобы обеспечить максимальное удобство работы пользователя без ущерба для пропускной способности сети.

Что касается классических модулей физического уровня систем беспроводной связи, они нуждаются в фундаментальных улучшениях, направленных на максимизацию показателей каналов 6G и удовлетворение новых экстремальных требований к услугам. Следуя этому направлению, мы обсудили новые ожидания и потенциальные направления исследований для кодирования, формирования сигналов, модуляции, множественного доступа и сверхмассивного MIMO. 6G объединяет несколько элементов: высокоинтегрированное сканирование и связь, интегрированные наземные и неназемные сети, интегрированные пассивные и активные узлы и интегрированные суперканалы. Таким образом, требуется всестороннее изучение и тщательное проектирование связи, использующей сканирование, и сканирования с помощью средств связи, совместных действий с несколькими соединениями, быстрой коммутации кросс-соединений и интеллектуальной управляемой радиосреды.

Нативное энергосбережение и сверхгибкое использование спектра также являются жизненно важными аспектами проекта 6G, который, как ожидается, обеспечит режим энергосбережения по умолчанию и унифицированный механизм для различных дуплексных режимов и поддержит сосуществование пользовательских устройств с различными дуплексными возможностями.

Как и в случае с предыдущими поколениями, технический прогресс, необходимый для следующего поколения, должен преодолеть множество проблем. Мы считаем, что интеллектуальный характер радиоинтерфейса 6G открывает широкий спектр возможностей для решения этих проблем.

Часть VI

Новые подходы к проектированию сетевой архитектуры 6G

Введение в часть VI

Несмотря на необычайный успех мобильных сетей, многие основные принципы проектирования и исходные предпосылки, которые хорошо служили нам в последние десятилетия, не ушли в прошлое, а лишь выиграли от быстрой адаптации к новым технологиям и услугам. С точки зрения эволюции, мобильные сети эволюционировали в сторону большей пропускной способности и предоставления разнообразных услуг для удовлетворения постоянно растущих потребностей. Значительный прогресс наблюдается в области беспроводной связи, поддерживающей сверхвысокую скорость передачи данных, малую задержку и массовые подключения. Однако архитектура мобильной сети продолжает развиваться не столь быстрыми темпами. В частности, постепенные изменения сопровождаются проблемами обратной совместимости, требованиями к взаимодействию сервисов, разукрупнением сетевых функций и введением дополнительных функций, поддерживающих новые сценарии применения. Сложность изменения сетевой архитектуры вытекает из проблем, связанных с затратами и стабильностью системы, а вовсе не из-за традиционного мышления или догм телекоммуникационной отрасли. Возьмем, к примеру, IP-сеть. Переход с IPv4 на IPv6 – нелегкая задача. Что касается системы мобильной связи, которая используется для обслуживания абонентов и услуг по всей стране, незначительные модификации могут оказать большое влияние на всю систему.

Мобильная сеть изначально была разработана для предоставления голосовых услуг. С тех пор как архитектура, так и прикладная реализация мобильных сетей следовали централизованной и иерархической парадигме, которая отражает природу голосового трафика и интернет-приложений. С каждым новым поколением мобильных сетей появлялись новые возможности и услуги. Именно в этом контексте считается, что распространение интеллекта, совместная связь и сканирование станут новыми возможностями, которые лягут в основу будущих сценариев применения, когда на сцене появится 6G. Также стоит отметить, что появляются новые требования к защите конфиденциальности, глобальному покрытию и управлению данными и параллельно проводятся обширные исследования новых применений и новых устройств. Новые технологии, такие как ИИ / машинное обучение и блокчейн, также будут включены в новую архитектуру в качестве инструментов реализации.

Помимо предоставления обычных услуг подключения, системы 6G также могут служить распределенными платформами для выполнения пользовательских вычислений во всех отраслевых сценариях. Это возможно благодаря тому, что сеть 6G будет основана на децентрализованной и ориентированной на пользователя архитектуре с возможностями нативного ИИ.

6G сместит традиционные парадигмы в сторону революционной архитектуры, которая отвечает новым требованиям и обладает новыми возможностями, основанными на передовых технологиях.

Право собственности на данные: от контроля третьих лиц к самосуверенитету

Данные, возможно, являются самым ценным активом в цифровом обществе, и ключ к решению наших проблем конфиденциальности лежит в том, как мы контролируем данные. В современных системах мобильной связи пользовательские данные распределены по многочисленным сетевым объектам, но централизованно принадлежат и контролируются одной третьей стороной. По этой причине конечные пользователи не контролируют свои собственные данные. Несмотря на то что системы мобильной связи оснащены механизмами защиты конфиденциальности пользователей в соответствии с законами и постановлениями, личная информация может быть легко передана сторонним поставщикам услуг без ведома пользователя. Кроме того, операторам связи также сложно монетизировать свои многочисленные источники данных.

Ключевые требования к конфиденциальности и защите данных отражены в Общем регламенте по защите данных (GDPR) в Европе, который, возможно, является первым в мире законодательным актом, призванным защитить потребителей от кражи и неправомерного использования данных. Он выступает за *самостоятельное управление личными данными*, что означает, что право собственности на данные должно быть возвращено конечным пользователям без участия каких-либо посредников.

Развертывание: от наземного 2D к глобальному 3D покрытия

Одна из наших целей при развертывании мобильных сетей – постоянно расширять широту и глубину покрытия. Двухмерное покрытие – это модель развертывания мобильных сетей, которая фокусируется на предоставлении услуг наземным абонентам мобильной связи. Ожидается, что 6G перейдет к трехмерной модели глобального покрытия всей земной поверхности, которая помогает смягчить «цифровой разрыв» и эффективно поддерживает новые сценарии приложений, такие как автомобильные сети, а также сети датчиков, работающие через спутники, NAPS или дроны, как упоминалось в главе 6.

Сетевая услуга: смещение парадигмы от оператора к пользователю

Подход, при котором во главу угла ставят оператора, приводит к проектированию сетей, которые должны поддерживать большое количество служб (например, на региональном или национальном уровне). Это привело к появлению монолитных сетевых объектов, таких как объекты управления мобильностью (mobility management entity, MME), обслуживающий шлюз (serving gateway, SGW) и сетевой шлюз пакетной передачи данных (packet data network gateway, PGW) в 4G, и сетевых функций, таких как функции управления доступом (access management function, AMF) и функции уровня пользователя (user plane functions, UPF) в 5G, которые, как ожидается, будут одновременно обслуживать большое количество конечных пользователей.

С функциональной точки зрения сеть управляет состоянием каждого пользовательского устройства или конечного пользователя. В этом смысле сеть, по сути, представляет собой большой распределенный конечный автомат, т. е. она поддерживает согласованные состояния для различных сетевых функций. За этим стоит сложный обмен служебными сообщениями, который ограничивает степень, в которой могут быть улучшены характеристики сети (например, задержка), а также может привести к увеличению количества точек атаки. По мере увеличения числа подключенных устройств/пользователей монолитные сетевые функции (как физические, так и виртуальные) также становятся потенциальными источниками серьезных узких мест. Учитывая тот факт, что сеть по своей сути управляет состоянием каждого устройства или пользователя, мы можем понять, почему необходима концепция сети, где во главу угла поставлены потребности каждого устройства и каждого пользователя. Если говорить более конкретно, ориентированная на пользователя система способна предоставлять виртуальную частную сеть (virtual private network, VPN) для каждого пользователя, а сетевая услуга, предоставляемая через подобную VPN, на уровне пользователя содержит такие функции, как управление мобильностью, управление политиками, управление сеансом и управление личными данными. Кроме того, служебные данные и соответствующие параметры сети могут быть оптимизированы на уровне пользователя.

Сеть: от «общей» сети к «своей» сети

Парадигма, ориентированная на пользователя, меняет структуру предоставления сетевых услуг – вместо централизованных услуг, доступных всем пользователям, будут услуги, доставляемые каждому пользователю, т. е. сетевые услуги будут переориентированы на пользователей. Чтобы предоставлять персонализированные услуги, архитек-

тура сети должна избегать универсального подхода, а также позволять пользователям иметь и самостоятельно управлять своими собственными мобильными сетями, как в настоящее время они поступают со своими домашними сетями и Wi-Fi. Таким образом, конечные пользователи смогут выбирать сетевые ресурсы для создания своих собственных VPN. Подход, ориентированный на пользователя, должен выходить за рамки сетевых функций и стать предпосылкой для претворения в жизнь концепции «своей сети».

Подключенность: от систем, ориентированных на подключение, к системам, ориентированным на задачи

Обычные системы связи ориентированы на установление соединения и изначально предназначались для передачи голоса, а затем данных. Начальная и конечная точки связи четко определены конечными пользователями и службами, которые они намереваются использовать, или другими пользователями, с которыми они планируют общаться. Таким образом, весь механизм связи, например управление сеансом и управление мобильностью, должен быть спроектирован так, чтобы обеспечивать достаточную поддержку этой модели связности.

Ожидается, что система мобильной связи 6G будет состоять из множества распределенных узлов (например, оконечных устройств, узлов радиодоступа и сетевого оборудования) с интеллектуальными функциями (для обеспечения собственной поддержки интеллектуальных услуг или использования интеллекта для самосовершенствования). Две ключевые уникальные услуги, которые будет предоставлять 6G, – это искусственный интеллект и сканирование. Для предоставления этих услуг одна и та же задача может выполняться на множестве распределенных узлов при надлежащей их координации; это называется *связью, ориентированной на задачу*. Технологии беспроводной связи в будущем должны поддерживать различные типы устройств и изменяющиеся во времени топологии для достижения оптимальных параметров связи, ориентированной на выполнение конкретных задач.

Оконечное устройство: от раздельной архитектуры TPC к интегрированной

Пользователь для доступа к службам или приложениям использует три ключевых компонента – это оконечное устройство (terminal), сетевой канал (pipe) и приложения в облаке (cloud), которые в совокупности обозначаются аббревиатурой TPC. В предыдущих поколениях беспроводных сетей каждый из этих компонентов представлял собой отдельную технологию со своей собственной экосистемой. По мере того как сети и приложения становятся все более виртуализированными, и то, и

другое можно реализовать в виде виртуальных сущностей, работающих в облаке. С появлением требовательных к данным и ресурсоемких приложений, таких как AR/VR и голографическая связь, ожидается, что скорость передачи данных достигнет уровня Тбит/с (как описано в главе 1). Однако это приведет к тому, что узким местом конечных устройств станет быстродействие, поскольку их вычислительные возможности ниже, чем у граничных облаков²⁶ (edge cloud). Следовательно, необходимо продумать распределение нагрузки между конечными устройствами и сетями (особенно граничными облаками). Конечные устройства можно гибко разрабатывать в соответствии с различными сценариями использования и возможностями. Например, большее количество пользовательских задач может быть перенесено с устройств в облако или, наоборот, системные и сетевые задачи могут быть перенесены на конечные устройства – для этого нет концептуальных ограничений. Еще одно преимущество состоит в том, что можно будет отделить идентификаторы пользователей от конечных устройств благодаря устранению физических границ между ТРС.

Искусственный интеллект: от облачного ИИ к сетевому

В современных сетях службы ИИ остаются на уровне приложений и располагаются в центральном облаке. Базовая сеть 5G имеет ограниченную область использования, даже несмотря на то, что она предназначена для предоставления интеллектуальной поддержки с помощью нового типа сетевой функции (т. е. функции анализа сетевых данных), и в ее архитектуре искусственный интеллект не является нативным. В эпоху 6G сетевая архитектура и ИИ будут идти рука об руку. Иными словами, поддержка искусственного интеллекта будет одним из фундаментальных факторов, способствующих инновациям в сетевой архитектуре. Таким образом, глубокая конвергенция коммуникационных и вычислительных ресурсов вкупе с полностью распределенной архитектурой приведет к переходу от облачного ИИ к сетевому. Преимущества этого перехода выходят за рамки достижения превосходных характеристик связи (например, сверхнизкой задержки), достигаемых за счет приближения услуг ИИ к конечным пользователям, поскольку проблемы конфиденциальности также могут быть решены локально. Это один из основных мотивов развития сетевой архитектуры 6G, на который существенно повлияют требования к конфиденциальности и управлению данными, такие как GDPR. Это особенно верно с точки зрения реализации функций ИИ в реальном времени, использования больших данных для машинного обучения и выполнения логических выводов ИИ, ко-

²⁶ Часть архитектуры современной системы беспроводного доступа, где выполняются граничные вычисления (edge computing). – *Прим. перев.*

торые неэффективны в централизованной облачной архитектуре ИИ и будут включены в сетевую архитектуру.

Безопасность: от простой безопасности к надежности и доверию

Безопасность 5G рассматривается совершенно отдельно от других сетевых сервисов и реализуется через автономную структуру. В 6G один из основных сдвигов парадигмы – от классического понимания безопасности к более широкому понятию нативной надежности. Этот сдвиг включает в себя обеспечение безопасности на конструктивном уровне и широкий круг тем для исследований, таких как модель доверия и защитные признаки (security thread) на основе многообещающих квантовых вычислений и применение новых технологий (например, ИИ и машинного обучения) при разработке решений по обеспечению безопасности. Чтобы разработать надежную архитектуру 6G и определить границы ее возможностей, следует опираться на новые сценарии использования, которые приводят к новым требованиям, а также на новые вспомогательные технологии.

Глава 27

Технологии для сетевой архитектуры искусственного интеллекта

27.1. История вопроса

Мобильная связь вступила в новую захватывающую эпоху, когда персональные устройства, такие как смартфоны и планшеты, становятся основной вычислительной платформой для многих приложений. Поскольку эти устройства активно используются в нашей повседневной жизни (от профессиональной деятельности до досуга и от обучения до развлечений), они имеют доступ к беспрецедентному объему персональных и прочих данных. Поэтому компании, специализирующиеся на искусственном интеллекте, ведут активные исследования и разработки в этой области, чтобы предоставлять конечным пользователям более удобные и полезные услуги.

Ожидается, что применение ИИ принесет пользу не только бизнесу в потребительском секторе, но также коснется всех секторов экономики и затронет все аспекты жизни общества. Вне зависимости от типа приложения или службы искусственный интеллект должен собирать и анализировать различные типы данных, а затем применять соответствующие выводы для выполнения определенного действия или набора действий. Большинство текущих реализаций ИИ основаны на централизованном обучении, при котором данные собираются со всей системы, но обучение выполняется в одном месте, обычно в энергоемком центре обработки данных, предназначенном для этого типа вычислений (например, со специальным оборудованием). В настоящее время ИИ не включен в архитектуру сети и в основном является облачным ИИ, т. е. базовая сеть просто используется для передачи данных в облако, где размещается основной интеллект, выполняющий обра-

ботку и вывод данных. Нам еще предстоит стать свидетелями слияния сети и искусственного интеллекта. Например, когда 6G предоставляет широкому кругу пользователей искусственный интеллект и сканирование в качестве новых услуг, возникает проблема с развертыванием безопасной и надежной сетевой инфраструктуры для передачи данных из источников информации с целью дальнейшего применения метода обучения, размещенного в облаке, или других типов доступных ресурсов. Поэтому крайне важно прийти к единому пониманию сетевой архитектуры, способной обеспечить поддержку нативного ИИ. Для этого нам также необходимо выделить три столпа, на которых держится вся структура ИИ, и сформулировать обусловленные ими потенциальные требования к системе мобильной связи.

- **Данные** являются ключевым активом индустрии искусственного интеллекта, и их называют «новой нефтью». Поскольку первая волна услуг с ИИ больше ориентирована на применение в категории «бизнес для потребителя» (business-to-consumer, B2C), прямыми источниками данных служат конечные пользователи. Данные дополнительно монетизируются с помощью применений в категории «бизнес для бизнеса» (business-to-business, B2B) в промышленных отраслях, которые влекут за собой различные варианты использования, бизнес-модели и технические требования. Исходя из ограничений пропускной способности, а также соображений безопасности и конфиденциальности, мы ожидаем, что отраслевые данные будут обрабатываться на границе системы, которая обычно находится в помещениях предприятия. Таким образом, услуги ИИ, ориентированные на промышленность, будут предоставляться более распределенным образом, с учетом фактора локальности. Эта тенденция приведет к дискуссиям относительно требований к управлению данными, их обработке, принадлежности права собственности на данные и выводы и т. д. Система мобильной связи, которая может удовлетворить эти требования, а также полностью соответствует политикам управления данными, наверняка получит высокую оценку потребителей.
- **Вычислительные ресурсы** – основа индустрии искусственного интеллекта. Чем мощнее приложение ИИ, тем больше ему требуется вычислительных ресурсов. Учитывая это, наверняка рано или поздно окажется, что даже мощному вычислительному пулу, расположенному в облаке, не хватает масштабируемости для удовлетворения запросов пользователей. Если мы заикливаемся только на традиционной вертикальной модели обработки данных, то запуск приложений ИИ лишь в центральном облаке в определенных сценариях может не работать. Поэтому новая

синергия перемещает ИИ из облака вглубь системы мобильной связи, представляющей собой сверхвысокопроизводительную инфраструктуру, которая может эффективно управлять разнородными ресурсами с масштабируемостью и эластичностью. Эта область исследований одновременно увлекательна и сложна, поскольку может привести к полной реорганизации традиционных системных архитектур и философии проектирования.

- **Алгоритмы** определяют тип интеллекта, предоставляемого приложениями ИИ, а также тип данных и объем вычислительной мощности, которые приложения ИИ используют и потребляют соответственно. От инфраструктуры не зависит, как определяются алгоритмы ИИ, но она может обеспечить качественную поддержку для запуска этих алгоритмов. Например, реализация глубокого обучения (такого как федеративное обучение) зависит от связи, т.е. ее пропускной способности и задержки. Поэтому сетевая системная архитектура может влиять на обучение ИИ и его логический вывод.

27.2. Соображения и принципы проектирования

Новая системная архитектура, нацеленная на распространение интеллекта путем нативной поддержки, должна учитывать все три перечисленных выше конструктивных требования. В результате облачный ИИ будет преобразован в сетевой.

27.2.1. Основные требования

Управление данными

Система 6G в силу самой своей природы будет генерировать огромное количество данных, связанных с работой сети и управлением сетью, с деятельностью пользователей, сканированием окружающей среды и работой оконечных устройств. Поскольку эти данные будут поступать из совершенно разных областей, актуальные задачи проектирования новой системы 6G включают в себя эффективную организацию данных и управление ими с учетом защиты конфиденциальности.

Передовая инфраструктура

В эпоху 6G технологии данных, информации и связи будут полностью переплетены, образуя передовую инфраструктуру. Чтобы распространение ИИ стало реальностью, необходимо обеспечить обучение в реальном времени и последующий вывод на основе практически все-

общей подключенности и распределенных вычислительных ресурсов. Следовательно, система мобильной связи должна обладать чрезвычайно высокими характеристиками (например, сверхнизкой задержкой и сверхвысокой скоростью передачи данных), особенно при сборе и обработке данных. Вдобавок ко всему для достижения оптимальных характеристик при условии соблюдения ограничений конфиденциальности и безопасности требуется глубокая конвергенция всех типов вычислительных ресурсов от всех типов заинтересованных сторон. Сервисы ИИ могут использовать эту передовую инфраструктуру, чтобы полностью раскрыть потенциал трех перечисленных выше основ (т. е. данных, вычислительных ресурсов и алгоритмов).

Гибкое развертывание ИИ

Простая процедура развертывания важна для привлечения услуг от внешних партнеров, особенно когда это требует определенного уровня сетевых навыков и знаний в области ИТ. Это требование в основном относится к разработке на уровне платформы. Иными словами, собственные и партнерские службы ИИ должны легко развертываться в 6G (на периферии или централизованно), и платформа не должна ограничивать миграцию служб.

Открытость экосистемы

Предоставляя встроенную поддержку искусственного интеллекта, 6G охватит гораздо более широкую экосистему, чем системы связи предыдущего поколения. Помимо традиционных игроков в сфере электросвязи (например, поставщиков, операторов, абонентов и т. д.), корпоративных клиентов будут привлекать новые типы услуг ИИ, предоставляемые системой мобильной связи. Предполагается, что 6G будет построена на многопользовательской экосистеме, которая является основой основ новой мобильной сети, поэтому открытость системы (с точки зрения сотрудничества как на уровне бизнеса, так и на техническом уровне) должна быть простой, гибкой и заслуживающей доверия. Несмотря на то что сеть 5G уже стоит в начале пути к широкой открытости экосистемы (например, за счет раскрытия сетевых возможностей [1] и интерфейсов прикладных программ (API) [2]), она по-прежнему полагается на традиционные подходы к телекоммуникациям, основанные на стандартизованных сетевых функциях, процедурах и интерфейсах. В 6G эти соображения следует рассматривать как внутреннюю причину архитектурной революции.

27.2.2. Ключевое различие между 5G и 6G

Разработчики сети 5G предусмотрели определенную поддержку ИИ, особенно на уровне ядра сети. Для этого предназначена функция ана-

лиза сетевых данных (network data analytics function, NWDAF) [1], основная цель которой – облегчить сбор данных и аналитику в 5G. Иными словами, это сетевая функция 5G, специально разработанная для сбора данных и предоставления аналитической информации. Например, она предоставляет аналитическую информацию для других сетевых функций и тем самым способствует предоставлению сетевых услуг. NWDAF поддерживает сбор данных от сетевых функций 5G и от OA&M [3]. Для этого он также предоставляет услугу регистрации и предоставления метаданных соответствующим сетевым функциям.

Однако наличие NWDAF не означает нативную поддержку ИИ в 5G по следующим причинам:

- **ограниченный источник данных:** NWDAF в основном поддерживает сбор данных и аналитику на основе данных, полученных от функций сети 5G. Однако она не учитывает данные инфраструктуры, окружающей среды, оконечных устройств, датчиков и т. д.;
- **отсутствие защиты конфиденциальности данных:** данные, задействованные в работе сети 5G, в основном относятся к самой сети; поэтому защита конфиденциальности данных не рассматривается на фундаментальном уровне;
- **отсутствие поддержки внешних служб ИИ:** NWDAF является функцией ядра сети 5G, а внешние службы ИИ не могут подключаться непосредственно к ядру сети 5G или RAN;
- **отсутствие решений на уровне инфраструктуры:** ключевые особенности архитектуры 5G (например, разделение сети, URLLC и mMTC) предназначены для удовлетворения вертикальных требований с точки зрения производительности, функциональности и эксплуатации. Обеспечению поддержки ИИ для 5G (например, с точки зрения управления данными и распределенной архитектуры) не уделялось особого внимания;
- **отсутствие управления данными:** ИИ представляет из себя нечто большее, чем просто сбор и анализ данных. Чтобы обеспечить нативную поддержку ИИ, следует предусмотреть отдельную функцию управления данными, и это выходит за рамки 5G.

Следовательно, нативная поддержка ИИ – одно из фундаментальных архитектурных различий между 5G и 6G.

27.3. Особенности архитектуры

27.3.1. Общий обзор

Важнейшая роль 6G заключается в создании инфраструктуры со сверхвысокой производительностью, в которой вычислительные ре-

судсы очень близки к конечным пользователям. Эта инфраструктура будет соответствовать высоким отраслевым стандартам устойчивости. Система 6G будет свободно управлять всеми типами ресурсов, предоставляя подключения, достаточные для полной конвергенции вычислений и связи; обобщенная схема архитектуры показана на рис. 27.1.

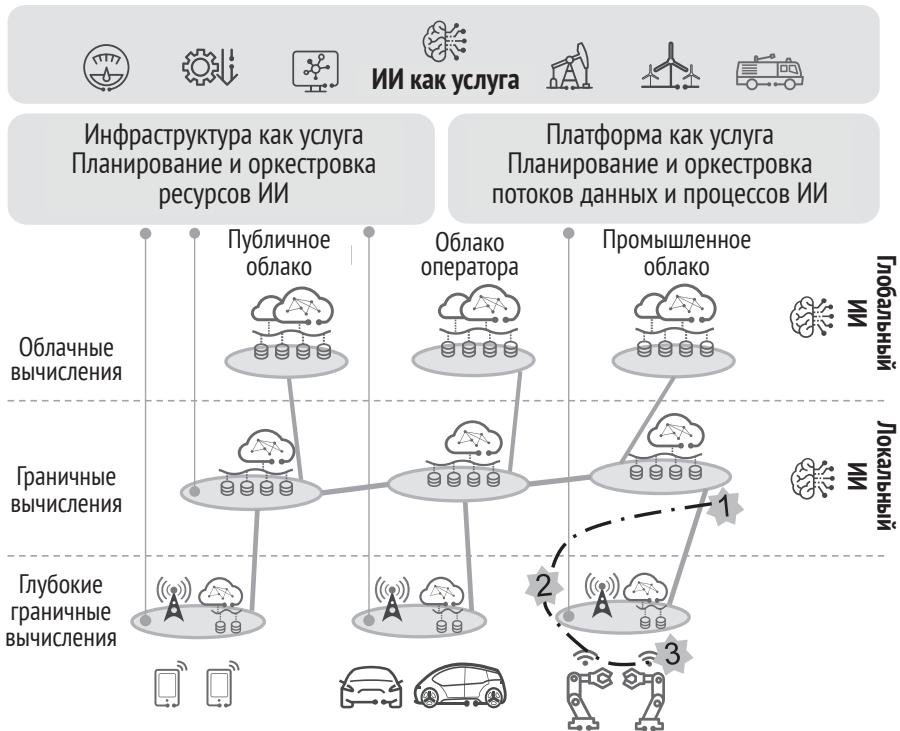


Рис. 27.1. Объем проектирования системной архитектуры для обеспечения поддержки ИИ

6G обеспечит поддержку функций нативного ИИ, которые могут найти воплощение в трех бизнес-моделях, как показано на рис. 27.1: инфраструктура как услуга, платформа как услуга и ИИ как услуга, которые фактически заимствованы у облачных сервисов.

- **Инфраструктура как услуга (IaaS):** в этой модели система мобильной связи управляет вычислительными возможностями и ресурсами связи (например, ресурсами облачных и граничных вычислений, а также вычислительными ресурсами базовых станций и устройств), необходимых для работы служб ИИ. Основная цель – сделать доступными возможности высокопроизводительной инфраструктуры, отвечающие требованиям служб искусственного интеллекта. При этом, помимо скоростной передачи большого количества данных, наличие такой высокопроизводи-

тельной инфраструктуры подразумевает глубокую интеграцию служб ИИ в систему и обеспечение их бесперебойной работы. Пока до конца не ясно, будет ли это обновленной концепцией периферийных вычислений или станет неким фундаментальным прорывом. На этот вопрос еще предстоит найти ответ.

- **Платформа как услуга (PaaS):** в этой модели инфраструктура 6G сама по себе служит платформой искусственного интеллекта и может управлять рабочими процессами искусственного интеллекта, управлять данными и выполнять другие задачи. Например, для реализации услуги ИИ могут потребоваться данные из различных источников, и агрегирование данных из разных технических областей может оказаться основной трудностью для поставщиков услуг ИИ. В этом случае модель PaaS может использовать опыт игроков информационно-коммуникационной отрасли в области эксплуатации и управления инфраструктурой (в частности, опыта, полученного в традиционных сценариях с участием нескольких поставщиков и операторов).
- **Искусственный интеллект как услуга (AIaaS):** в этой модели инфраструктура 6G предоставляет услуги искусственного интеллекта для внешних клиентов. Примеры таких услуг могут включать в себя высокоточную локализацию с использованием ИИ, предсказание поведения пользователей и т. д.

Эти три модели вполне могут сосуществовать, поскольку они охватывают разнообразные требования к услугам ИИ от самых разных клиентов. Сервисы искусственного интеллекта, работающие в этой продвинутой инфраструктуре, принесут множество преимуществ, а именно:

- **переход от глобального ИИ к локальному:** с точки зрения общенациональной сети централизованное обучение обходится слишком дорого, поскольку оно включает сбор и отправку данных по всей сети в центральный объект. Тем не менее локальный ИИ или комбинация локального и глобального ИИ способны снизить энергопотребление, тем самым снизив затраты. Как бы то ни было, традиционный метод сбора данных и централизованное обучение быстро приведут к возникновению узких мест и единых точек отказа в центральном местоположении или вокруг него (например, доступность центра обработки данных, используемого для обучения, применение критических путей обмена данными с дата-центром и доступность узлов на этих критических путях). Более того, использование личных данных с помощью централизованного метода обучения может оказаться невозможным по причине отсутствия разрешения владельца данных;

- **переход от автономного ИИ к работающему в реальном времени:** кроме всего прочего, традиционные решения машинного обучения в значительной степени полагаются на автономное обучение, которое включает доступ к обучающим данным или моделирование и симуляцию абстрактной версии среды. Однако недостатком этого процесса является то, что он упрощает многие детали, а также игнорирует множество важных показателей, что, в свою очередь, снижает качество модели. К счастью, скоростная распределенная инфраструктура открывает новые возможности для создания *ИИ в реальном времени* (включая обучение и вывод), особенно для сценариев использования со строгими требованиями к задержке (например, управление с обратной связью в промышленности).

27.3.2. Связь, ориентированная на задачи

Один из ключевых сдвигов парадигмы 6G связан с ориентированной на задачи моделью связи, которая предназначена для поддержки новых услуг, таких как искусственный интеллект и сканирование. Все традиционные услуги мобильной связи (например, голос и данные) ориентированы на предоставление соединения, то есть ресурсы связи распределяются на основе предварительно поданных запросов от конечных пользователей и ожидаемых услуг, независимо от того, намерены ли пользователи разговаривать с другими пользователями или они хотят подключиться к серверу в облаке. В результате соединения устанавливаются на основе намерений пользователей.

Для сравнения: связь, ориентированная на задачи, основана на совершенно иной философии, когда установленное соединение зависит от того, какую задачу в этот момент решает пользователь. Например, задача ИИ будет заключаться в сборе и обработке данных по определенной географической области в соответствии с мобильностью пользователей, распределением населения в реальном времени или интенсивностью использования оконечных устройств. Компании общественного транспорта могли бы использовать эту информацию для отслеживания мобильности жителей, особенно в часы пик. Поэтому компании будут использовать службу ИИ в 6G, не вникая в то, как данные собираются от конкретных пользователей для получения определенной информации.

В другом примере задача сканирования может заключаться в получении трехмерной картины окружающей среды на разных частотах. Благодаря наличию такой картины становятся возможными новые варианты использования, например сканирование в реальном времени с высокоточной локализацией и отслеживанием перемещений, а также распознавание жестов и действий. Такие задачи выполняются путем

координации нескольких базовых станций, подключенных клиентских устройств и вычислительных ресурсов, как показано на рис. 27.2.

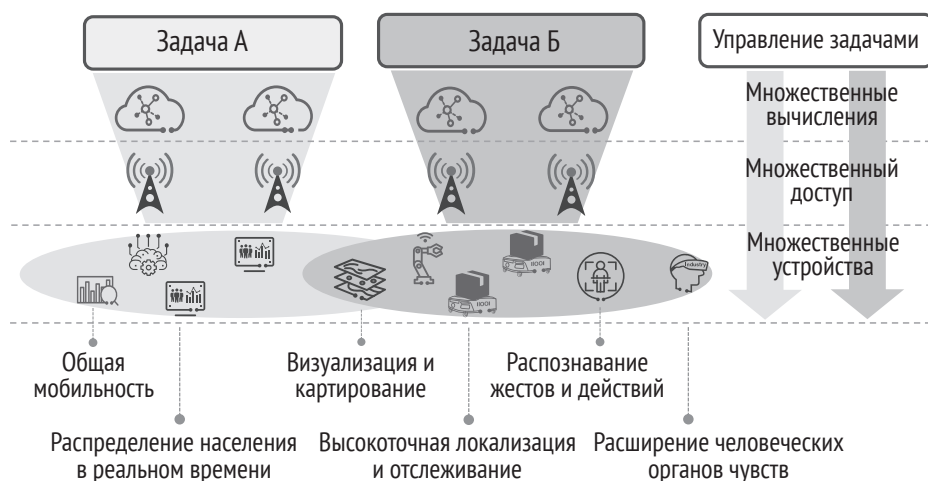


Рис. 27.2. Концепция подключения, ориентированного на задачи

Разработка ориентированного на задачи коммуникационного решения охватывает четыре основных аспекта: управление задачами, управление ресурсами / планирование во время работы, управление данными и управление соединениями.

Управление задачами

С точки зрения архитектуры, для управления задачами могут быть введены новые сетевые службы и API, реализующие определение, выполнение и управление задачами на протяжении всего их жизненного цикла. Функциональный модуль управления задачами может разбить задачу на функциональные подзадачи, которые необходимо решить различными методами из-за разнородного типа доступа и форматов оконечных устройств. Благодаря наличию глубоко конвергентной вычислительной и коммуникационной платформы некоторые подзадачи могут выполняться локально.

Управление ресурсами / планирование во время работы

В традиционной системе мобильной связи управление ресурсами (особенно на стороне радиодоступа) происходит на основе упреждающих запросов на установление соединения и ресурсы планируются на основе информации о подписке устройства и типах доступных ему услуг.

Для выполнения задачи этой системе необходимо иметь возможность доступа и координации вычислительных ресурсов различного оборудо-

вания связи в различных географических точках. С точки зрения ресурсов, определенная задача может быть разбита на вычислительную задачу и коммуникационную задачу. Задачи первого типа могут быть распределены в разных облаках, граничных облаках, базовых станциях и устройствах, тогда как задачи второго типа могут быть распределены среди различных устройств в нескольких сотах. Таким образом, концепция новой системы 6G значительно отличается от концепции традиционной системы связи, поскольку управление ресурсами или планирование будет основываться на поставленных перед системой задачах, а не на запросах конечных пользователей. Вместо резервирования радиоресурсов для установления сеанса связи двух сторон друг с другом вычислительные и коммуникационные ресурсы планируются с прицелом на выполнение задачи в регионе X.

Ориентированное на задачу управление ресурсами или планирование должны обеспечивать целостное представление о ресурсах в зависимости от географического положения и рабочего объема задачи. Если задача выполняется в рамках одной технической или бизнес-области, можно использовать весь существующий опыт управления сетью. Что касается многопользовательских сценариев или систем, которые распространяются в большом масштабе, управление ресурсами E2E и планирование во время работы станут слишком сложными (например, с точки зрения установления доверия, механизмов начисления платы и аудита), и, вероятно, появятся новые методы. В главе 31 более подробно рассматриваются эти потенциально возможные новые методы.

Управление данными

В контексте выполнения задач необходимо изучить некоторые новые области, такие как сбор данных, защита конфиденциальности, хранение информации, возможности сети и раскрытие знаний. Управление данными – это специальная технология в составе архитектуры для управления данными в более широком контексте, представленная в главе 30, и ее можно использовать для поддержки нативного ИИ.

Управление соединениями

Для выполнения задачи могут быть установлены многочисленные соединения, например между устройствами и датчиками окружающей среды, а также коммуникационным и вычислительным оборудованием. Что касается критериев эффективности выполнения задач, они включают гибкое и эффективное взаимодействие между пользователями и оконечными устройствами, множественный доступ и множественные вычисления. По существу, традиционные механизмы (такие как управление сеансом и переключениями) могут нуждаться в фундаментальных изменениях.

Архитектура сети радиодоступа

Для более эффективного выполнения задач, возможно, будет иметь смысл разделение функций управления задачами и функций работы с пользователями на уровне сетевых служб, чтобы управление задачами могло выполняться на большей территории. Таким образом, базовая станция 6G может быть логически разделена на две части: узел управления (обозначенный как cNB) и несколько узлов обслуживания (обозначенные как sNB). cNB обеспечивает функцию управления задачами через несколько sNB, расположенных на большой территории. Такая структура может снизить накладные расходы на общую управляющую информацию. sNB обеспечивает связь между различными оконечными устройствами и сетевым оборудованием. cNB и sNB – это логические объекты, выходящие за рамки формата физического оборудования. Используя этот подход, ресурсы RAN можно легко сгруппировать в виртуальные кластеры, каждым из которых можно управлять, например при выполнении определенных задач. Преимущество такого кластера состоит в том, что он контролирует все соответствующие ресурсы на управляемом наборе узлов для лучшего управления ресурсами. Размер области кластера должен быть достаточным для выполнения текущих задач.

27.3.3. Глубококонвергентные вычисления и граничные коммуникации

Ожидается, что система 6G будет обслуживать отраслевые сценарии, требующие оптимальной производительности и обработки данных на локальном уровне; следовательно, *граничные узлы* (edge nodes) станут инновационной платформой для будущих сетей мобильной связи. Мобильные граничные вычисления (edge computing), или так называемые граничные вычисления с множественным доступом (multi-access edge computing, MEC), – это давно известное понятие, которое приобретает все большее значение в области мобильной связи. Оно смещает вычислительные ресурсы от централизованного расположения в сторону границ системы мобильной связи, тем самым обеспечивая гибкость с точки зрения развертывания и эксплуатации сетевых услуг, а также улучшенные характеристики сети (например, задержку). Однако многочисленные практические затруднения ограничивают развертывание граничных узлов. Например, когда установлен шлюз безопасности, развертывание граничного узла между RAN и шлюзом сопряжено с большими трудностями. Это связано с тем, что сеансы IPsec не могут быть прерваны, как показано на рис. 27.3. Вдобавок к этому мы не можем сдвинуть границу вычислений глубоко в RAN из-за ограничений в архитектуре радиодоступа и соответствующей логике связи.

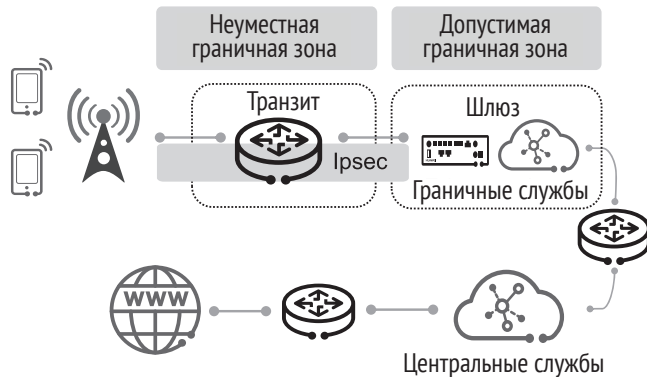


Рис. 27.3. Определение граничной области при смещении в сторону RAN

В 6G глубокая конвергенция между вычислениями и связью на уровне RAN является фундаментальным фактором, который обязательно следует учитывать, и мы называем его «глубокими граничными вычислениями» в следующем смысле: глубокие граничные вычисления расположены близко к источникам данных, т. е. в глубине сети, благодаря чему службы ИИ могут быть изначально интегрированы в архитектуру RAN.

Глубокие граничные вычисления включают в себя нечто большее, чем просто запуск облачного сервера на границе мобильной сети. Нам также необходимо подумать о реконструкции архитектуры RAN, чтобы сместить граничную область еще глубже, минимизировать влияние на стеки протоколов и улучшить общую масштабируемость системы. Исследование глубоких граничных вычислений могло бы включать перечисленные ниже направления.

Новая вычислительная схема

В настоящее время уже можно использовать возможности искусственного интеллекта на уровне RAN. Например, мы можем оптимизировать планирование ресурсов и уменьшить помехи. Тем не менее это основная задача, для которой сегодняшние мобильные сети используют ИИ как инструмент, вместо того чтобы нативно поддерживать его, а это две разные концепции (то есть ИИ для сетей и сети для ИИ). Давайте поговорим об этом подробнее.

Для глубокой интеграции граничных вычислений в RAN может потребоваться новый тип радиооборудования, которое мы будем называть *радиовычислительным узлом* (radio computing node, RCN), чтобы отличать его от узла радиодоступа, как показано на рис. 27.4. Внутри RCN может быть введен независимый вычислительный уровень для различных задач хоста (например, ИИ и сканирование) или для уровня управления (control plane, CP), пользовательских уровней (user plane,

UP) и т. д. Этот тип вычислительных ресурсов не только смещает граничную область в узел радиодоступа, но также требует пересмотра общей архитектуры RAN, что, в свою очередь, позволит беспрепятственно объединить вычислительные ресурсы с возможностями связи в RAN. Вероятно, на этом уровне придется внедрить локальный контроллер для управления ресурсами беспроводной связи, функциями и конвейерами ИИ.

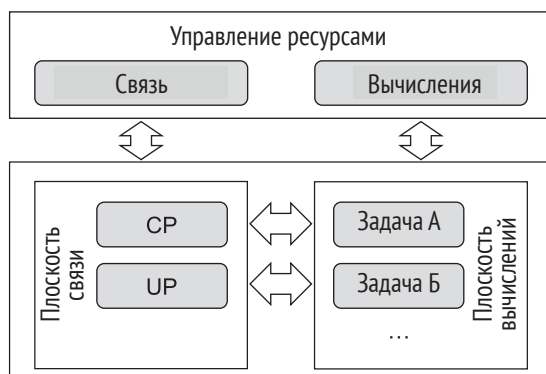


Рис. 27.4. Глубокая конвергенция коммуникационных и вычислительных возможностей в RCN

Более того, для поддержки динамического развертывания и эффективной связи между службами, развернутыми на глубоких границах, в стандарте должны быть указаны соответствующие интерфейсы, применяемые процедуры и протоколы. Такие службы могут воспользоваться преимуществами уменьшения задержки, транспортных расходов, потенциальных рисков безопасности и проблем конфиденциальности.

Новый стек протоколов

Сдвиг границ вычислений также может привести к пересмотру стека протоколов, который, например, можно упростить за счет выработки высокопроизводительных решений, поддерживающих ориентированные на задачи сервисы (например, рабочие процессы ИИ). В системе 5G соединения для передачи данных устанавливаются между пользовательским оборудованием и серверами в облаке через базовые станции и UPF в базовой сети. В случае глубокой конвергенции вычислительных и коммуникационных ресурсов на уровне RAN соединения устанавливаются напрямую между пользовательским оборудованием и Pod-модулями или контейнерами, размещенными на вычислительном уровне глубокой границы, что позволяет значительно упростить механизм связи. Таким образом, можно обеспечить высокоэффективную связь, которая поддерживает критически важные службы в реальном времени.

Повышенная масштабируемость системы

Системы мобильной связи используются как для общенационального покрытия (например, услуги типа B2C), так и для покрытия местного масштаба (например, услуги типа B2B). Следовательно, количество узлов с поддержкой глубоких граничных вычислений может исчисляться миллионами, и все они не должны быть изолированными ресурсами. Вероятно, появится возможность создать крупномасштабную сеть, основанную на огромном количестве независимых, но управляемых узлов с глубокими граничными вычислениями, которые потенциально могут служить точками соприкосновения различных технических и бизнес-решений для повсеместного распространения интеллекта. С одной стороны, это будет высокопроизводительная сеть, с другой – масштабируемая, открытая и прозрачная. Эффективная синхронизация данных и услуг, планирование времени выполнения, выбор подходящих ресурсов и определение локальных (например, промышленных) приложений являются важными аспектами проектирования архитектуры системы.

Формируя крупномасштабную сеть посредством узлов с глубокими границами, можно будет использовать комбинированный локально-глобальный ИИ, в отличие от чисто глобального ИИ. Как прямое следствие цифровая трансформация отрасли может проходить с намного большим размахом и охватом.

27.3.4. Эксплуатация и управление службами искусственного интеллекта

Эксплуатация и управление службами ИИ – еще один важный аспект, который необходимо учитывать при разработке архитектуры с поддержкой нативного ИИ. Эта структура используется для упрощения интеграции и развертывания служб ИИ (особенно от внешних поставщиков).

- **Эксплуатация служб ИИ:** сеть 6G будет построена на многопользовательской экосистеме, которая по своей сути опирается на несколько технических и экономических аспектов. Что касается эксплуатации службы ИИ, эта экосистема должна работать бесперебойно и эффективно, особенно в сценариях с несколькими операторами и поставщиками. Чтобы установить рабочие отношения даже в ситуациях нулевого доверия, для эксплуатации ИИ могут потребоваться такие технологии, как блокчейн.
- **Управление службами ИИ:** включает управление рабочими процессами ИИ и данными, а также управление вычислительными и коммуникационными ресурсами. Чтобы управлять службами ИИ в различных областях, как показано на рис. 27.5, важно различать

ключевые интерфейсы, используемые в одной области управления и в разных областях. В главе 31 подробно рассказывается о создании платформы, поддерживающей многопользовательскую коллаборацию.

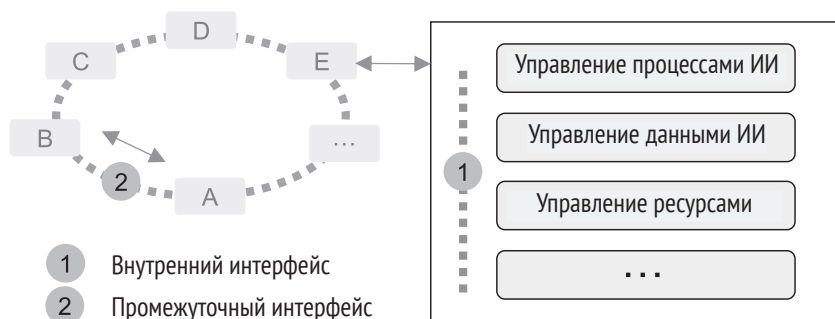


Рис. 27.5. Структура управления ИИ в различных областях

Традиционный метод построения интерфейсов в системе мобильной связи, вероятно, будет стандартизирован. Тем не менее создание новой структуры, полностью основанной на этой стандартизации, не рекомендуется. Вместо этого можно было бы использовать передовые подходы в данной области, например сам ИИ на современных компонентах с открытым исходным кодом [4, 5].

Модуль эксплуатации и управления услугами ИИ – это новый компонент поверх традиционной архитектуры системы мобильной связи. По существу, он является характерным примером нового типа сетевых функций и принципов разработки, например следующих:

- **взаимодействие ИИ с другими службами:** в целях защиты конфиденциальности данных модули безопасности в составе архитектуры могут препятствовать доступу к информации со стороны ИИ, что неизбежно скажется на его работе. Тем не менее в архитектуре будет предусмотрена поддержка работы службы ИИ для смягчения этих ограничений;
- **управление ресурсами:** услуги ИИ нуждаются в значительных вычислительных и коммуникационных ресурсах. Следовательно, управление ресурсами включает в себя правильную координацию как вычислительных, так и сетевых ресурсов для устройств, серверов (или точек обслуживания, например центров обработки данных) и сети. Ошибочное управление может привести к высокому энергопотреблению, а также к снижению характеристик сети и качества услуг (например, увеличению задержки);
- **развертывание служб ИИ:** новая архитектура облегчит развертывание служб ИИ, включая определение мест развертывания и

соответствующих промежуточных соединений в соответствии с описанием уровня обслуживания.

27.4. Источники

- [1] 3GPP, *System architecture for the 5G system*, 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 23.501, Aug. 2020, version 16.5.1. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144>.
- [2] 3GPP, «*Common API Framework for 3GPP northbound APIs*», 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 23.222, July 2020, version 17.1.0. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3337>.
- [3] 3GPP, *Architecture enhancements for 5G System (5GS) to support network data analytics services*, 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 23.288, July 2020, version 16.4.0. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3579>.
- [4] *Kubeflow*. <https://www.kubeflow.org/>.
- [5] *Kubernetes*. <https://kubernetes.io/>.

Глава 28

Архитектура, ориентированная на пользователя

28.1. Обзор технологии UCN

В предыдущих системах мобильной связи пользовательские устройства, которые обращаются к оператору сети с запросом на обслуживание, не имеют никакого контроля над самой сетью. *Ориентированные на пользователя сети* (user-centric networks, UCN), как ключевая особенность архитектуры 6G, относятся к системам с особой структурой, которая позволяет им подстраиваться под пользователя и быть управляемой пользователем. Как следует из названия, «ориентированная на пользователя» означает, что пользователи смогут определять (с помощью ИИ) перечень услуг сети, которые они хотели бы получать, а также способы их использования и управления. Для работающих служб пользователи смогут настраивать политики использования ресурсов, а также оконечные устройства, принадлежащие одному и тому же пользовательскому домену²⁷. Пользователи контролируют все данные, которые ими создаются или принадлежат им, а также соответствующие права процесса (например, идентификация, авторизация доступа и информация о статусе пользователя). Пользователи могут быть физическими лицами-потребителями, предприятиями или отраслевыми пользователями, а услуги будут состояться интеллектуальным и персонализированным образом в пользовательском домене на основе профиля, поведения и предпочтений пользователей.

Идеология UCN изменит способ взаимодействия пользователей, сетевых сервисов и приложений, принятый в настоящее время, что повлияет на соответствующий доступ к сети, управление мобильностью, безопасность и владение личными цифровыми активами. В соответствии с текущей тенденцией к модульности и облачности сетевых функ-

²⁷ Под пользовательским доменом (user domain) мы здесь подразумеваем совокупность пользовательских задач, оборудования, вычислительных ресурсов и предоставляемых ему услуг сети. – *Прим. перев.*

ций основанная на программном обеспечении телекоммуникационная услуга позволяет развертывать основные сетевые функции (такие как пересылка, управление сеансами и управление политиками) без ограничений местоположения. Конструкция сети для каждого пользователя будет модульной с общим контекстом, что устранил обмен сообщениями между традиционными сетевыми функциями.

Одна из основных тенденций, как мы ранее определили, – это растущий интерес к локальной/частной связи в ограниченной области. В свете этого предпосылкой для построения UCN является полностью распределенная и взаимосвязанная граничная платформа. У нее есть свои важные преимущества. Например, она по своей сути будет обладать устойчивостью к атакам благодаря децентрализованности, которая ограничивает масштаб ущерба. Более того, UCN поддерживает модель «системы систем», в которой различные системы и пользователи с соответствующими пулами ресурсов взаимодействуют для получения окончательных требуемых результатов, в то время как каждый пользователь организует свои собственные ресурсы посредством различных взаимодействий. На основе этого принципа можно легко улучшить как отказоустойчивость системы, так и эффективность обработки трафика.

В UCN пользовательский домен не будет ограничен физическими пользовательскими устройствами, поскольку сетевой домен будет служить его естественным расширением, в котором область развертывания служб и приложений определяется на основе различных критериев, например услуг, требующих большей вычислительной мощности или более критичной пропускной способности сети.

И последнее, но не менее важное: UCN будет оптимальным методом предоставления действительно персонализированных услуг для каждого пользователя, что, в свою очередь, будет способствовать смене парадигмы с «сети для всех» на «собственную сеть». Иными словами, UCN превратит 6G из сети простого доступа в среду предоставления услуг.

28.2. Соображения и принципы проектирования UCN

28.2.1. Уроки, извлеченные из существующих сетей

Система мобильной связи развивалась на протяжении нескольких поколений. Отличительные особенности и поддерживаемые услуги каждого поколения определяют фундаментальную философию проектирования системной архитектуры. Перед переходом с 5G на 6G важно осознать все ограничения текущей системы с учетом потенциальных вариантов использования и нашего видения сети 6G.

Философия системы, ориентированной на сетевые функции (NF)

С функциональной точки зрения мобильная сеть состоит из *уровня управления* и *уровня пользователя*. Уровень управления в основном «руководит» уровнем пользователя, обеспечивая выделение сетевых ресурсов для предоставления услуг в соответствии с запросами пользователя, тогда как уровень пользователя в основном следует указаниям и реализует их. В действующей сегодня мобильной связи применялся принцип проектирования, основанный на сетевых функциях (network function, NF), что привело к созданию монолитных сетевых объектов. Уровень управления состоит из набора NF, ответственных за аутентификацию и авторизацию доступа, мобильность, управление сеансом, управление данными (например, репозиторий NF, контекст пользовательского оборудования и политику) и управление операциями (например, мобильность, сеанс и QoS). В свою очередь, уровень пользователя состоит из набора пересылочных элементов, которые транспортируют пакеты данных пользователя от источника к месту назначения и сообщают метрики потока обратно на уровень управления.

Когда количество абонентов исчисляется миллионами, а предоставляемые услуги сосредоточены на передаче голоса и данных, такая конструкция имеет очевидные преимущества с точки зрения развертывания и затрат на эксплуатацию и управление. Тем не менее постоянно возрастающая сложность этой конструкции интуитивно ощущается в быстро растущем числе абонентов, количестве сетевого оборудования, протоколов, интерфейсов и взаимосвязей/взаимодействий между сетевым оборудованием. Текущая архитектура сотовой сети использует множество протоколов, которые задействуют несколько сетевых функций и многочисленные обмены сообщениями. Безупречность любой реализации в лучшем случае сомнительна, потому что сообщения, передаваемые в этих процедурах связи, сложны. Поскольку протоколы предназначены для поддержания согласованности состояний пользовательского оборудования в сетевых функциях, большое количество состояний, которые необходимо поддерживать в различных сетевых функциях, становится препятствием для масштабирования.

Проблемы с развертыванием в централизованной архитектуре

Архитектура сотовой сети была в первую очередь разработана для сеансов с длительным сроком жизни (например, голосовых) с энерго- и ресурсоемким пользовательским оборудованием. Все предыдущие поколения архитектуры сотовой сети в значительной степени опирались на концепцию архитектуры голосовой сети, т. е. на высокоцентрализо-

ванный дизайн и иерархически развернутую сеть. Одним из принципов проектирования централизованной архитектуры является NF-ориентированность, в результате чего монолитные сетевые функциональные объекты реализуются либо в физическом оборудовании, либо в виртуальных объектах, обслуживающих большое количество конечных пользователей, таких как управление мобильностью, данные устройства/абонента, управление сеансом, аутентификация, пересылка пакетов данных и применение политик. С точки зрения функциональности сотовая сеть управляет состоянием пользовательского оборудования или конечного пользователя; по сути, это большой распределенный конечный автомат, то есть сеть в целом должна поддерживать согласованные состояния для различных сетевых функций путем обмена сообщениями.

Несмотря на продолжающиеся дискуссии о преимуществах и недостатках распределенных сетей по сравнению с централизованными, очевидно, что когда количество подключенных пользователей/устройств многократно увеличивается и устройства становятся как разнородными, так и более интеллектуальными, возникают проблемы безопасности и масштабируемости:

- **единая точка отказа и уязвимость для атак типа «отказ в обслуживании» (DoS):** из-за открытой архитектуры на основе IP сотовые беспроводные системы уязвимы для обычных атак через интернет. В качестве примера можно привести базовую мобильную сеть, которая уязвима для DoS-атак из-за ее централизованного характера. Как обычные, так и распределенные DoS-атаки на централизованные сетевые функциональные модули влекут за собой серьезные последствия, поскольку мобильные сети являются критически важной инфраструктурой, обеспечивающей стабильную жизнедеятельность всего общества;
- **проблемы масштабируемости:** начиная с 4G/LTE иерархическая сеть на основе шлюзов стала пользовательским уровнем базовых сетей 4G и 5G. Она предназначена для управления сеансами каждого пользовательского устройства, т. е. всякий раз, когда устройство требует доступа к данным, должна произойти последовательность обмена сообщениями между различными компонентами ядра и RAN, прежде чем может быть передан любой пакет данных уровня приложения. Поскольку большинство подключенных устройств – это смартфоны, которые поддерживают длительные сеансы подключения без жестких требований к задержке и ограничений обработки пакетов управления, текущая архитектура работает, если разумное количество этих устройств подключено через базовую сеть. Однако как только количество подключенных устройств резко возрастет и устройства станут

более разнообразными (например, датчики в интеллектуальном транспортном средстве, промышленные роботы, а также вездесущие носимые и имплантируемые устройства), в текущей сетевой архитектуре могут возникнуть критические узкие места.

Трудности с монетизацией данных

Система мобильной связи сталкивается с дилеммой в отношении данных. С одной стороны, она содержит множество источников данных, которые операторы связи пытаются монетизировать, особенно данные, прямо или косвенно связанные с конечными пользователями из-за правил конфиденциальности. С другой стороны, конечные пользователи не могут контролировать свои собственные данные, а это означает, что они должны быть уверены, что поставщики услуг не будут ими злоупотреблять. Кроме того, пользователям нелегко выявить и отследить нарушения частной жизни из-за недостаточной осведомленности и знаний.

Проблемы с преобладающей моделью, ориентированной на оператора

Хотя 5G лучше реализует вертикальную структуру клиента за счет поддержки нелицензированного доступа и кампусных сетей, конструкция и архитектура базовой сети почти исключительно созданы для операторов. Это означает, что не уделяется особого внимания функциям, необходимым для различных потребителей, обладающих собственной вертикальной иерархией (например, промышленные предприятия). Скорее, в логике 5G, ориентированной на операторов, подобные потребители должны спрашивать операторов связи, кто сможет их лучше обслуживать. Это проблематично для всех игроков, включая самих операторов связи: поддержка такого огромного количества неочевидных функций и дополнительных требований к услугам – большая проблема. Недавно для лучшего удовлетворения этих требований были созданы новые промышленные форумы, например 5GAA и 5G-ACIA для автомобильной и обрабатывающей промышленности соответственно. Однако даже при хорошем понимании этих требований операторам связи придется принимать непростые решения, особенно потому, что в условиях классического рыночного давления не все, что технически выполнимо, является экономически целесообразным.

28.2.2. Основные требования

Требования к архитектуре, ориентированной на пользователя

Традиционные сети были спроектированы по принципу «каждая функция на всю сеть», как показано на рис. 28.1. Это привело к появ-

лению нескольких монолитных сетевых функций, обслуживающих большое количество пользователей/устройств, при этом каждая сетевая функция, такая как AMF и функция управления сеансом (session management function, SMF) в 5G [1], теперь имеет очень специфические задачи. Таким образом, можно с уверенностью сказать, что принцип NF-ориентированности естественным образом приводит к централизованной архитектуре.

Архитектура, ориентированная на пользователя, означает, что каждый пользователь будет иметь свою собственную выделенную сеть, объединяющую все нужные ему функции для предоставления услуг, как показано на рис. 28.1. Согласно этому принципу проектирования, UCN отвечает за управление мобильностью, политиками, сеансами и персональными данными (с помощью децентрализованных технологий). Одним из преимуществ UCN является резкое сокращение обмена сигналами и следующее из этого уменьшение задержки. Кроме того, для связи требуется меньшее количество протоколов, что, в свою очередь, снижает сложность и, следовательно, приводит к менее дорогостоящим и более простым в управлении реализациям. Ориентированная на пользователя архитектура естественно и идеально сочетается с децентрализованной технологией, которая неуязвима для DDoS-атак и единых точек отказа (single point of failure, SPOF).

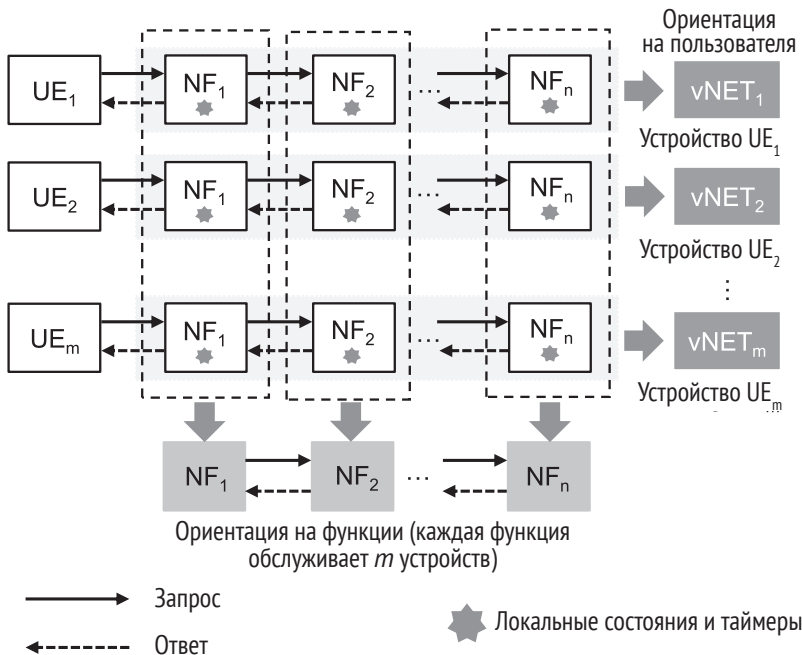


Рис. 28.1. Системы, ориентированные на пользователя, и системы, ориентированные на сетевые функции

Требования к децентрализации

- **Доступность:** доступность и стабильность сервиса должны быть обеспечены за счет использования как децентрализованных, так и распределенных облачных технологий, которые помогают равномерно распределять нагрузку на систему между сетевыми узлами.
- **Однородность:** количество типов сетевых узлов должно быть сведено к минимуму, чтобы улучшить масштабируемость, сложность и надежность сети, а также в конечном итоге снизить стоимость реализации сети.
- **Неуязвимость:** децентрализованная сетевая архитектура изначально неуязвима для атак. Поскольку реализация всех услуг мобильной сети является распределенной, когда некоторые узлы или сеть повреждены или атакованы, остальная часть сети не страдает.
- **Самоорганизация:** в децентрализованной сети сетевые узлы могут взаимодействовать друг с другом с помощью определенных алгоритмов, например распределенных хеш-таблиц [2] или блокчейна.

Требования к ресурсоэффективности

В каждой архитектуре важно эффективное использование доступного пула ресурсов системы связи для уменьшения затрат и эксплуатационных расходов при предоставлении услуг потребителям. Если рассматривать низкую общую стоимость владения (total cost of ownership, TCO) в качестве одной из основных целей проектирования и учитывать принятое ныне активное стремление к экологически устойчивым ИТ, то можно сделать вывод, что роль эффективного общесистемного управления ресурсами в будущем значительно возрастет. Это приведет к появлению нескольких новых проблемных областей, которые в настоящее время не изучаются, недооцениваются или полностью игнорируются как промышленностью, так и академическими кругами. Например, распространенные сегодня смешанные среды вычислений/хранения/сети требуют подходящих решений для согласования ресурсов: неоднородность ресурсов затрудняет использование единых механизмов, поскольку в разных областях применяются разные подходы, и зачастую они замыкаются внутри своей области, не проецируясь наружу. Кроме того, уникальные подходы, скорее всего, не будут соответствовать требованиям разнородных ресурсов.

В этом контексте планирование обслуживания во время работы имеет первостепенное значение, поскольку оно позволяет обеспечить пре-

восходные функциональные свойства поддерживаемых распределений (так называемых «срезов») и снизить совокупную стоимость владения. Динамическое назначение ресурсов, как стремление к более эффективному совместному использованию инфраструктуры (включая вычислительные, сетевые и энергетические ресурсы), труднодостижимо из-за разнородной, частичной или устаревшей информации, ее привязки к работе в реальном времени и отсутствия какого-либо центрального руководящего органа либо механизма.

Решение проблемы планирования работы в крупных сетевых системах требует обширных фундаментальных исследований. Например, нам нужно узнать, как использовать существующие решения, полученные в результате исследования работы центров обработки данных, и применять их в масштабе сети с поддержкой нескольких арендаторов и параллелизма. Для этого потребуются подходящие механизмы разрешения конфликтов, особенно если мы стремимся к гарантированному предоставлению услуги. Как показывают выводы, полученные в результате исследования распределенных систем, главной целью должно быть повышение эффективности, а не оптимальности.

Требования к новым границам между разными сферами

Во всех предыдущих архитектурах систем мобильной связи границы всегда были очерчены относительно физических объектов (между пользовательским оборудованием и сетью, между радиодоступом и ядром и т. д.), что приводило к отчасти неожиданным ограничениям предоставления услуг. В будущей системе границы должны пролегать в пространстве услуг: должны быть разделены области выполняемых услуг. Для этого услуги обычно должны быть отделены от инфраструктуры, и эта тенденция уже нашла подтверждение в сфере ИТ-услуг, где знание инфраструктуры уже не столь необходимо с точки зрения пользователя. Однако в этом сценарии за необходимую устойчивость (т. е. безопасность и надежность) отвечает среда выполнения службы. Мы утверждаем, что это единственный разумный путь развития для многоцелевых инфраструктур ICT, поскольку он ведет к внутренне адаптируемым услугам с минимальными требуемыми уровнями доверия. Это означает, что можно запускать одну и ту же службу в разных условиях с практически одинаковыми свойствами без необходимости проецировать требования к надежности службы (которые не обсуждаются) непосредственно в надежность инфраструктуры (что может повлечь непомерные затраты). Переопределение физических границ также будет способствовать смене парадигмы от раздельной архитектуры TPC к интегрированной.

Очевидным следствием из предыдущих рассуждений является необходимость устранить противоречие между оптимальностью и разме-

ром результирующего сетевого взаимодействия многопользовательских ресурсов. Фактически несколько заинтересованных сторон будут задействованы в качестве просьюмеров, которые независимо поставляют ресурсы и потребляют услуги. Это несовместимо с глобальным подходом к полномочиям, и поэтому система должна быть сильно децентрализована. Стоит отметить, что, учитывая потенциально крупный масштаб новой системы, она может выглядеть неоптимальной по сравнению с существующей централизованно управляемой инфраструктурой операторов. Однако централизованные архитектуры сами по себе масштабируются не лучше: учитывая их «удаленность» от уровня ресурсов, они, возможно, даже меньше подходят для быстрой адаптации во время работы. Сетевые операторы обычно используют децентрализованное управление, или полагаются на возросшую интеллектуальность ресурсов, или располагаются ближе к ним (например, иерархическое управление на основе политик). Сегодня они в основном используются для оптимальной предварительной инициализации; тем не менее, когда речь идет о смешанных услугах, предварительная инициализация сталкивается с ограничениями. Кроме того, нерешенной проблемой остается многопользовательский характер системы: не ясно, как объединить различные политики, которые будут применяться в одном и том же пуле ресурсов. По этим причинам при объединении автономности и искусственного интеллекта с концепциями распределенных систем может потребоваться сочетание согласованности (необходимой для оптимальности) с сегментацией (необходимой для масштабирования). Например, это решение можно применить в сценариях, где для оптимальности необходимо использовать адаптацию в ходе выполнения, при этом сохраняя доступность за счет возможного избыточного выделения ресурсов.

Требования к владению цифровыми активами

Мы уделяем все больше внимания вопросам конфиденциальности и владения данными, которые считаются «новой нефтью» цифровой эпохи. Другими словами, пользователи должны иметь возможность полностью контролировать свою личную цифровую идентичность, а также монетизировать свои собственные данные. Таким образом, они должны иметь возможность выбирать, какими данными делиться с другими сторонами, и иметь уверенность, что эти данные не используются без согласия. Более того, пользователи должны иметь возможность изолировать себя для предотвращения утечки данных. Что не менее важно, они должны иметь возможность отозвать доступ третьих лиц к своим данным. Это способы, которыми архитектура системы 6G обеспечивает полный контроль пользователей над своими данными.

28.3. Особенности архитектуры

28.3.1. Децентрализованная архитектура, ориентированная на пользователя

Общий обзор

Благодаря гибкости программно-конфигурируемых сетей (software-defined network, SDN) и виртуализации сетевых функций (network functions virtualization, NFV) ориентированная на пользователя базовая сеть позволяет каждому клиентскому устройству получить свой собственный «частный» виртуальный экземпляр базовой сети через децентрализованную структуру виртуальных экземпляров, реализованных в виде виртуальных машин или контейнеров, как показано на рис. 28.2. Эти виртуальные экземпляры функций можно разделить на две категории: узлы обслуживания сетевого уровня (network-level service node, NSN) и узлы обслуживания уровня пользователя (user-level service node, USN). Они могут реализовать типичные функции базовой сети, такие как управление мобильностью, управление политиками и управление абонентами/устройствами. NSN, который может быть распределенной или централизованной упрощенной версией функций базовой сети, служит первой точкой подключения пользовательского устройства к сети и отвечает за его аутентификацию во время процедур регистрации доступа (то есть присоединения и установления направленного канала по умолчанию). USN, которые могут быть полностью распределены и самоорганизованы в соответствии с виртуальной сетью, предназначены для одного UE и обрабатывают все функции базовой сети, включая функции уровня пользователя и уровня управления.

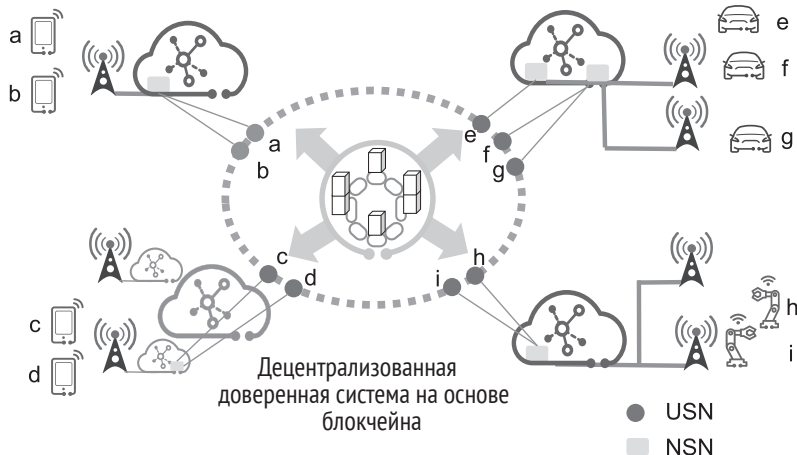


Рис. 28.2. Децентрализованная и ориентированная на пользователя архитектура

UCN превосходят современную архитектуру во многих отношениях. В частности, благодаря своему распределенному и персонализированному характеру они не позволяют ботнетам выполнять DDoS-атаки и одновременно дают возможность применять настраиваемые политики для каждого пользовательского устройства (например, безопасность и QoS). Децентрализованное управление пользователями и данными позволяет конечным пользователям полностью контролировать свои цифровые активы.

Еще один важный момент, который стоит принять во внимание, – это тот факт, что UCN предназначены для выхода за физические границы обычного TPC, где пространство пользователя ограничено физическими оконечными устройствами. Это означает, что в соответствии с концепцией UCN пользовательская область может быть расширена до любого объема ресурсов узлов обслуживания пользовательского уровня (USN) в системе мобильной связи.

Проще говоря, идеология архитектуры UCN заключается в том, чтобы быть фундаментальной сетевой архитектурой, которая является гибкой, безопасной и надежной, но при этом имеет простую структуру с полным набором функций. Эти достоинства делают сеть 6G заслуживающей доверия.

Эволюция или революция

Одной из основных технологических особенностей системы 6G, более всего с точки зрения проектирования базовой сети, является смещение акцента с сетевых функций на пользователя. Исходя из положения дел в сегодняшней базовой сети 5G, для реализации этого видения можно использовать механизмы срезов сети (slice), называемые однопользовательскими срезами. Однако это усовершенствование по-прежнему основано на фундаментальной архитектуре, ориентированной на функции. В соответствии с методом срезов каждый пользователь может воспринимать сеть как собственный набор (срез) сетевых служб, состоящих из сетевых функций, который, возможно, и не нуждается в радикальном изменении, но при этом оставляет нерешенные вопросы, такие как владение цифровыми активами и т. д.

Фундаментальная архитектура может выиграть от совершенно нового подхода и ориентации на пользователя. Сравнение концепций архитектуры системы 5G и UCN приведено на рис. 28.3. В архитектуре 5G сетевые функции полностью определяются на сетевом уровне, который в сети 6G превращается в UCN за счет разработки сетевых функций, реализуемых на уровне пользователя. Взаимодействие между NSN и USN или внутри домена NSN по-прежнему нуждается в стандартизации, но службы будут располагаться исключительно внутри домена USN.

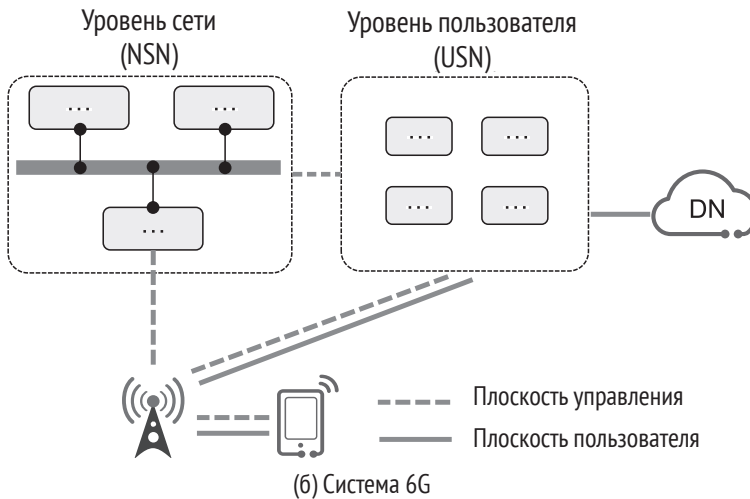
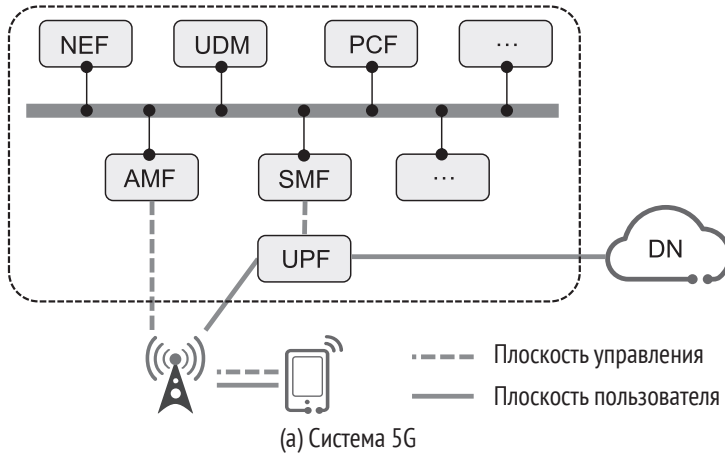


Рис. 28.3. Возможная эволюция архитектуры

По сути, архитектура UCN не диктует, как должны разрабатываться сетевые службы (например, управление мобильностью и сеансами). Это связано с тем, что в будущих системах связи эти службы будут, по сути, обычными компьютерными приложениями. Если создана надежная инфраструктура с эффективной архитектурой, службы верхнего уровня, естественно, будут использовать традиционные преимущества удобного развертывания, эксплуатации и масштабируемости. Таким образом, UCN не будет радикально отличаться от традиционных сетевых сервисов, а вместо этого представит их в децентрализованной форме, которая является более надежной, безопасной и отказоустойчивой. В то же время новая архитектура обеспечит соответствие управления конфиденциальностью данных подписчиков нормативным требованиям.

28.3.2. Слияние физического и кибернетического миров

Отражение физических сущностей в цифровом мире

Цифровой двойник стал популярной концепцией, применяемой во многих отраслях промышленности. Цифровой двойник представляет собой цифровую или виртуальную копию физического объекта, которым могут быть активы, процессы, системы и т. д. Впервые этот подход был использован в моделировании производственных операций [3], когда для снижения общих затрат, повышения производительности и обеспечения качества продукции при проектировании систем, моделировании, а также управлении рисками решили использовать цифровые аналоги физических объектов. Концепция цифрового двойника эволюционировала с течением времени, и в настоящее время ее использование в значительной степени ориентировано на IoT. Она специально разработана с учетом соединения вещей через беспроводные технологии и сети, а также для использования облачных вычислений. В эпоху 6G слияние физического и кибернетического миров станет основной тенденцией, которая приведет к переходу цифрового двойника на новый уровень в качестве вспомогательной технологии, потенциально способствующей разработке архитектуры, ориентированной на пользователя.

Термин «отражение», представленный в этой главе, включает основу концепции цифрового двойника, но выходит за рамки чисто цифрового двойника. Являясь виртуальным представителем любого физического объекта в системе 6G, отражение действует как «умный цифровой агент», то есть не только копирует поведение физических объектов, но и проактивно инициирует действия в зависимости от различных обстоятельств. Например, отражение может следить за перемещениями пользователей, приближаться к пользователям, переходя из центрального облака в пограничное облако, или выделять соответствующие ресурсы для пользователей. Оно имеет такие функции, как управление идентичностью, аутентификация, авторизация доступа, управление цифровыми активами пользователей, мобильностью, сеансами и взаимодействием с приложениями.

Отражение может представлять сущности очень разного масштаба – от физических объектов до целой организации. Оно не только включает в себя цифровую модель или описание физических объектов, но также выступает как автономный агент, живущий в сети. Кроме того, оно управляет соответствующими элементами ресурсов и взаимодействует с другими отражениями. Следовательно, мы можем рассматривать отражение как интеллектуальную сущность, которая управляет всеми пользовательскими ресурсами, чтобы формировать, поддерживать и контролировать составленный из них пул; и оно также komponует

пользовательские сервисы, которые интеллектуально планирует и развертывает в пуле ресурсов. Рисунок 28.4 иллюстрирует эту концепцию и представляет функции, которые могут существовать в домене отражения пользователя Б. Этот домен развертывается в системе мобильной связи в соответствии с профилем пользователя и его поведением, чтобы предоставить пользователю Б полностью настраиваемый и ориентированный на его задачи оптимизированный сервис. Сюда входят сетевые функции, которые могут быть подобраны для конкретного пользователя (например, USN) и созданы, а также обновлены отражением пользователя Б. Более того, возможности ИИ платформы 6G могут быть использованы для реализации проактивного предоставления сетевых услуг.



Рис. 28.4. Концепция пользовательского отражения, работающего в сети

Заметим, что развитое ядро пакетной коммутации (evolved packet core, EPC) сети 4G или базовую сеть 5G можно рассматривать как некий определенный тип отражения, ориентированный на потребности традиционных операторов мобильной сети. По мере развития сети 5G мы наблюдаем диверсификацию типов пользователей. Многие из этих пользователей представляют собой вертикальные отрасли, у которых есть как ресурсы, так и навыки, но при этом они имеют совершенно другие потребности по сравнению с операторами мобильной связи. Следовательно, концепция отражения стремится обобщить понятие

функциональности базовой сети эволюционным путем, поддерживая унаследованные от прежних поколений потребности, а также обеспечивая стандартизированную поддержку новых категорий.

В рамках концепции отражения службы управляются непосредственно отражениями и развертываются на соответствующих элементах ресурсов полностью децентрализованным образом. Для управления данными цифровых активов любого формата, которые хранятся в распределенной базе данных, может использоваться система доступа к распределенным данным. Что касается распределенной системы построения отображений, ее можно использовать для обеспечения обновления почти в реальном времени и выполнения запросов для сопоставления между идентификаторами, именами и адресами. Кроме того, для эксплуатации отражений и управления ими можно использовать систему управления жизненным циклом.

Соображения относительно будущей архитектуры

С точки зрения архитектуры, концепция отражений может быть поддержана углубленным изучением следующих аспектов:

- **базовые службы отражений:** в основном они включают управление местными ресурсами и построение отражений. Между ресурсом и его отражением устанавливается соединение для связи. Это соединение работает через интерфейс отражение–устройство (reflection-to-device, R2D), как показано на рис. 28.4. Профили и текущие состояния ресурса отправляются через соединение в отражение и используются для создания, обновления и расширения самого отражения. В обратном направлении отражение отправляет сигналы для управления ресурсом, который будет выполнять определенные действия для реализации и запуска служб;
- **расширенные службы отражения:** в основном это понятие включает составление служб и развертывание с помощью политик. Отражение должно быть спроектировано так, чтобы отвечать на запросы вызова службы из внутреннего домена и области взаимодействия между отражениями. Отражение проверяет доступность ресурсов и выполнимость запрошенных услуг. В результате службы развертываются на ресурсах (устройствах) через интерфейс между физическим объектом и его отражением. Всякий раз, когда требуется ресурс взаимодействия отражений, задействованные отражения будут общаться, например через интерфейс отражение–отражение (reflection-to-reflection, R2R), как показано на рис. 28.4. Следовательно, интерфейс R2R реализует децентрализованный уровень управления, который поддерживают различные домены отражений. В этом случае службы раз-

вертываются совместно через отдельные элементы управления соответствующих доменов отражений, но, возможно, в пересекающихся путях ресурсов. Службы обнаружения и развертывания ресурсов также поддерживаются как часть служб отражений;

- **реализация отражения:** домен отражения будет создан совместно с уровнем данных (об этом пойдет речь в главе 30), опираясь на функции управления данными, предоставляемые системой 6G;
- **детализированное управление автономными ресурсами:** учитывая, что граница проводится в пространстве служб, а не между физическими ресурсами, несколько независимо работающих отражений обычно будут предоставлять услуги из общего пула ресурсов, созданного всеми устройствами на уровне ресурсов. Используя программируемость устройств и инфраструктуры, отражения могут запускать сложные службы на уровне инфраструктуры или даже в большем масштабе. Без дополнительных системных механизмов и координации на уровне ресурсов это привело бы к классической проблеме «разделения мозга»²⁸, низкой надежности обслуживания или огромному избыточному выделению ресурсов, поскольку локальный анализ не позволяет оптимизировать использование ресурсов на глобальном уровне. Причина этого – неопределенность в отношении запросов и операций других отражений.

28.3.3. Управление цифровыми активами

Опираясь на подход UCN, мы можем сформировать цифровой домен пользователя с четко выраженной границей, содержащий ресурсы данных, такие как профиль пользователя, профиль службы и профиль ресурсов (включая как сетевые ресурсы, так и ресурсы окончательного оборудования). Это ценные цифровые активы, принадлежащие пользователю, и ими следует надлежащим образом управлять в полном соответствии с нормативными требованиями к безопасности и конфиденциальности. Подход UCN может гарантировать, что право собственности на данные в сети 6G будет принадлежать пользователю, что, в свою очередь, означает, что именно пользователи будут контролировать использование данных и распространение информации во всем цифровом мире. Таким образом, сама конструкция UCN обеспечивает нативную поддержку управления цифровыми активами, снижая риск злоупотребления данными и нарушения конфиденциальности. Более того, она позволяет пользователям монетизировать свои собственные данные. Этот механизм может быть основан на DLT и других технологиях, которые мы подробно рассмотрим в главе 31.

²⁸ Возникновение двух и более ведущих в схеме с ведущим и ведомыми. – Прим. перев.

Пользователи в цифровом обществе нуждаются в узнаваемом цифровом идентификаторе в качестве адреса для обмена информацией. При наличии цифрового идентификатора становится известной связанная с ним информация, такая как статус, поведение или транзакции. Перенос персонализированной информации на цифрового двойника дает множество преимуществ. Например, это обеспечивает удобные и безопасные транзакции, снижает затраты и повышает эффективность работы. Отсутствие в настоящее время надежной схемы управления цифровыми удостоверениями привело ко многим проблемам, таким как нарушение конфиденциальности и злоупотребление данными. Следовательно, нам необходимо создать фонд развития UCN и вместе разработать схему цифровой идентификации, которая отвечает требованиям подтверждения личности, авторизации, управления цифровыми активами и т. д.

28.4. Источники

- [1] 3GPP, *System architecture for the 5G system*, 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 23.501, Aug. 2020, version 16.5.1. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144>.
- [2] I. Stoica, R. Morris, D. Liben-Nowell, D. R. Karger, M. F. Kaashoek, F. Dabek, and H. Balakrishnan, *Chord: A scalable peer-to-peer lookup protocol for internet applications*, IEEE/ACM Transactions on Networking, vol. 11, no. 1, pp. 17–32, 2003.
- [3] M. Grieves, *Digital twin: Manufacturing excellence through virtual factory replication*, White paper, vol. 1, pp. 1–7, 2014.

Глава 29

Механизмы обеспечения благонадежности

29.1. Понятие благонадежности системы

29.1.1. От философии к обществу

Слово *благонадежность* (trustworthiness) сразу же вызывает у нас философские мысли. Оно служит фундаментом *доверия* (trust), на котором строятся эффективные сообщества как у людей, так и у животных. В наши дни создание развитых обществ, к которым мы стремимся, было бы невозможно без доверия.

На протяжении многих лет люди определяли доверие по-разному. Например, доверие рассматривается как ситуация, когда «одна сторона (доверитель) готова полагаться на действия другой стороны (доверенного лица) [1]; ситуация устремлена в будущее». «Благонадежность – это характеристики или поведение одного человека, которые вызывают у другого положительные ожидания» [2]. В словаре Мерриам–Вебстера благонадежность определяется как качество человека, позволяющее на него положиться, и перечислено множество синонимов [3].

Характеризуя уровень доверия к чему-либо, люди чаще всего пытаются описать его с помощью инженерного языка. В электронной коммерции описаны четыре общих индекса доверия: сертификат конфиденциальности от третьей стороны, гарантии конфиденциальности, сертификат безопасности от третьей стороны и функции безопасности [4]. Среди них функции безопасности играют самую большую роль в обеспечении доверия потребителей к электронной коммерции.

Изучение рисков началось после того, как благонадежность начали рассматривать как критерий безопасности. При анализе благонадежности какого-либо объекта учитывают следующие три фактора: источник информации, саму информацию и получателя. Эти три фактора также включают субфакторы, например точность, достоверность и т. д. [5].

Как показывают результаты вышеупомянутых исследований, благонадежность может быть определена и измерена только при наличии определенного сценария, на который можно опираться. Даже при наличии такого сценария без существования четкого профиля надежности отношение к соответствующим факторам зависит от отдельных лиц. Более того, «важность этих факторов может оцениваться по-разному в зависимости от человека, принимающего решение о доверии, а также от ситуации; такова собственно субъективная природа доверия» [6].

29.1.2. От общества к промышленности

Поскольку технологическое развитие вступает в фазу сосуществования физического и цифрового миров, благонадежность инфраструктуры информационных и коммуникационных технологий (ICT) становится еще более важной. Она в значительной мере влияет на желание пользователей подписывать контракты на услуги, предоставляемые инфраструктурой. Таким образом, все заинтересованные стороны в сфере ICT нуждаются в благонадежности.

Рекомендация ITU-T X.509 дала определение доверия в области ICT. «Как правило, можно сказать, что объект “доверяет” второму объекту, когда первый объект делает предположение, что второй объект будет вести себя точно так, как ожидает первый объект» [7]. ITU проводит стандартизацию доверия в телекоммуникационной отрасли с 2015 года, особенно в области интернета вещей, последовательно публикуя технические отчеты, документы и стандарты рекомендаций [8–12]. В документах предлагается концепция доверия и его классификация как с архитектурной, так и с технической точек зрения. Следом за предложениями по стратегии обеспечения доверия в системе ICT были представлены рекомендации по анализу доверия в сети, кибербезопасности и IoT.

Исследование благонадежности началось с интернета вещей и теперь охватывает всю экосистему связи. В 2018 году новый рабочий объект – структура безопасности, основанная на доверительных отношениях в экосистеме 5G, – получил предварительное одобрение Сектора стандартизации электросвязи (ITU-T). По мнению заинтересованных сторон, для них настало время определить роли, отношения и обязанности, связанные с безопасностью, и достичь консенсуса по экосистеме 5G [13].

В 2017 году Национальный институт стандартов и технологий (NIST) опубликовал серию отчетов под заголовком «Структура киберфизических систем» (cyber-physical systems, CPS). Общественная рабочая группа по киберфизическим системам определяет благонадежность киберфизической системы как «доказуемую вероятность того, что система будет работать в соответствии с заданным поведением при любом на-

боре условий, что подтверждается ее характеристиками, включая, но не ограничиваясь, безопасностью, защищенностью, конфиденциальностью, надежностью и устойчивостью». Это был первый случай, когда организация по стандартизации определила надежность CPS [14].

Все отраслевые исследования, определения и справочные понятия в конечном счете направлены на обеспечение пользователей надежной коммуникационной средой, и поэтому объективные и субъективные факторы неоднократно становились предметом споров. Продолжаются исследования, направленные на достижение благонадежности в области ICT.

Было разработано много моделей, включая основанную на онтологии методику, предложенную для анализа взаимозависимостей и конфликтующих требований на протяжении всего жизненного цикла CPS [15]. Кроме того, были проведены исследования по практическому применению модели измерения доверия на основе теории игр, описывающей надежность коммуникации радарной диаграммы, подхода к оценке рисков на основе облачной модели и различных других решений.

Поставщики и интеграторы решений приложили немало усилий для удовлетворения и даже превышения потребностей пользователей и постоянно корректируют требования к надежности своих продуктов, чтобы не отставать от запросов пользователей. Например, требования потребителей значительно расширились, и теперь недостаточно уделять внимание доверию только на этапе разработки; потребитель ожидает непрерывного внедрения лучших отраслевых практик и стандартов вплоть до активного участия разработчика в создании здоровой экологической среды в отрасли.

Как мы можем определить благонадежность с точки зрения коммуникации? Какие научные доказательства мы должны представить, чтобы помочь людям, принимающим решения, сделать правильный вывод о том, заслуживает ли доверия 6G? Другими словами, каких качеств, свидетельствующих о благонадежности, мы ожидаем от 6G?

Мы можем составить представление о благонадежности сети 6G, рассматривая, как она наследует превосходные возможности и передовой опыт сетей предыдущего поколения, а также принимая во внимание различные дисциплины и социальные проблемы.

29.2. Благонадежность сложных систем связи

Без сомнения, мобильная связь как важнейшая инфраструктура представляет собой мост, соединяющий физический и цифровой миры. Она продолжает передавать информацию между людьми, между вещами, а также между людьми и вещами. С быстрым развитием коммуникационных технологий возможности передачи становятся все более про-

двинутыми и разнообразными. Люди привыкли к информационному комфорту, который обеспечивают коммуникационные сети: семейное общение, телеконференции, ведение бизнеса и другие дела, тесно связанные с людьми, а также исследование неизведанного во Вселенной.

Людей все больше заботит благонадежность сетей связи, и поэтому мы считаем, что мнение пользователей является единственным адекватным критерием. Частные пользователи ожидают, что благонадежные сети будут предоставлять стабильные услуги, защищать личную информацию и не причинять физических травм. В свою очередь, корпоративные клиенты оценивают благонадежность своих сетей с точки зрения стабильности работы, высокого качества связи, отсутствия утечки конфиденциальной информации, отсутствия злонамеренных вторжений и т. д.

Рассуждая о благонадежности сети 6G, мы должны учитывать все эти факторы. Иными словами, с чего начать определение благонадежности в соответствии с ожиданиями пользователей и как этого добиться? Мы считаем, что благонадежность системы связи – многогранное понятие.

Благонадежность подразумевает наличие множества аспектов

К аспектам благонадежности можно отнести использование технологий, соблюдение законов и постановлений, консенсус и совместные действия, а также позитивные убеждения пользователей. Каждый из этих аспектов имеет свою характеристику и особенности.

- **Технология:** служит связующим звеном для других показателей. Следует использовать все технологии, способствующие созданию надежной сети, например криптографию, анализ данных, машинное обучение и оценку безопасности.
- **Законы и постановления:** поскольку «правила создаются и соблюдаются посредством социальных или правительственных институтов» [16], законы и постановления постоянно влияют на стабильность рынка. С того момента, как 5G начала набирать обороты, разные страны начали активно развивать и дорабатывать свои законы и постановления в области кибербезопасности, безопасности связи и защиты конфиденциальности. В индустрию связи были внедрены соответствующие стратегии и меры по тестированию, оценке и проверке оборудования, устройств и целых сетей. Постоянная разработка этих законов определяет обязанности заинтересованных сторон, а также постепенно готовит всю технологическую цепочку к созданию более надежной сети [17–19].

- **Сотрудничество:** организации, занимающиеся разработкой глобальных стандартов, являются важной движущей силой сотрудничества. За исторический период между 2G и 5G сеть связи перешла от простых телефонных звонков и служб коротких сообщений к предоставлению различных отраслевых услуг. Этого было бы невозможно достичь без сотрудничества всех заинтересованных сторон экосистемы. ITU, 3GPP, Ассоциация GSM (GSMA) и Альянс мобильных сетей следующего поколения (NGM-N) – широко известные сообщества, которые предоставляют заинтересованным сторонам возможности свободно обсуждать технологии и бизнес. Опираясь на знания и вдохновение, полученные от этих платформ, коммуникационная сеть постепенно переходит от закрытой структуры к открытой. Благодаря открытости и доступности сети начинает развиваться межотраслевое сотрудничество, и различные отрасли промышленности сотрудничают между собой, чтобы предоставлять пользователям индивидуализированные и высококачественные услуги.

В качестве примера можно привести спецификации 3GPP R16, которые поддерживают два важных промышленных сектора (автомобилестроение и промышленную автоматизацию). Версия R16 также помогает «другим отраслям, таким как транспорт (например, будущая система мобильной связи для железных дорог) и СМИ (например, широкополосное вещание в сети 5G)» [20].

Сообщество безопасности тоже стремится развивать сотрудничество. Это связано с тем, что сектор безопасности может более адекватно и эффективно противостоять атакам благодаря анализу угроз и обмену технологиями, а также совместной работе членов сообщества. В связи с этим девиз всемирно известного мероприятия по безопасности RSA Conference 2014 звучал так: «делиться, учиться и защищать». Такое сотрудничество даст людям возможность «извлечь выгоду из идей и взаимоотношений, которые могут сформировать будущее информационной безопасности» [21].

- **Доверие:** это наиболее эфемерный фактор. Когда пользователи судят о том, заслуживает ли коммуникационная сеть доверия, их соображения преобладают над всеми остальными показателями. Мы нуждаемся в обширных научных и технических знаниях, которые помогут нам достичь консенсуса в отношении понимания технологии и соответствующих методик. Другими словами, если мы прислушиваемся друг к другу и принимаем мнения друг друга, мы можем достичь согласия и наладить сотрудничество. Конвергенция большего числа умов и рук будет в большей степени

способствовать развитию усовершенствованных коммуникационных инноваций в будущем.

Благонадежность – понятие относительное

Оценивающие, оцениваемые, метод оценки и неопределенные или неожиданные события – все это факторы, влияющие на достоверность оценки благонадежности.

Благонадежность жизненно важна

Поскольку человеческое общество непрерывно меняется, коннотации благонадежности и методы, принятые для ее достижения, остаются устойчивыми, если они непрерывно оптимизируются, развиваются и обновляются.

29.3. Правила достижения благонадежности

29.3.1. Основные принципы

В основе благонадежности 6G лежат два принципа.

Принцип 1: изначальная благонадежность

- **Возможности обеспечения благонадежности должны соответствовать разнообразию услуг 6G:** люди ожидают получить от 6G самые разнообразные услуги, начиная от сканирующей сети и заканчивая дистанционной системой здравоохранения с тактильной обратной связью и низкоорбитальными спутниками. Разнообразная структура сети 6G, ее услуги и требования пользователей сделают эту сеть исключительной во всех смыслах этого слова. Столь же разнообразными должны быть и воплощения благонадежности. Сеть 6G, охватывающая различные технические и деловые аспекты, может потребовать обширного набора возможностей обеспечения благонадежности, которые можно использовать как для строгого и централизованного управления доступом в централизованной части сети, так и для распознавания пользователей, а также их авторизации в части автономной периферии.
- **Благонадежность должна развиваться на протяжении всего жизненного цикла 6G, включая ее проектирование, разработку и эксплуатацию:** в данный момент мы можем определить требования благонадежности, лишь приблизительно оценив их характеристики на этапе проектирования 6G. Следовательно, при разработке продуктов 6G написание программного кода и производство оборудования будут соответствовать проектным требованиям предыдущего этапа. Поэтому механизм доверия в

среде 6G следует эксплуатировать и развертывать с регулярными и непрерывными улучшениями на каждом этапе.

Принцип 2: баланс

Чтобы добиться благонадежности 6G, необходимо учитывать три фактора: первоначальное доверие к клиентам, стоимость атак и скорость восстановления деятельности.

Если разработчики 6G изначально мало доверяют клиентам сети, то, даже если они надеются, что стоимость атак для злоумышленника будет высокой/или восстановление будет быстрым, им придется предусмотреть больше контрмер и строгих ограничений.

- **Клиенты 6G:** это могут быть приложения для операций с ресурсами в сети 6G, включая доступ к сети для пользовательских устройств, доступ к базе данных приложениями, связь между функциональными модулями, доступ к журналу и т. д.
- **Стоимость атак:** злоумышленники атакуют только тогда, когда их выгода превышает стоимость атаки.
- **Восстановление:** механизмы, используемые для возобновления нормальной работы или обслуживания, должны быть способны быстро нейтрализовать атаки, как динамически, так и непрерывно.

На практике надежные архитектуры обычно разрабатываются вместе с определенным первоначальным доверием к клиентам, исходя из чего предлагаются решения по управлению доступом на самых ранних этапах процесса проектирования. Напротив, очень сложно предсказать стоимость атак и возможности восстановления.

Например, в некоторых сценариях предъявляются критические требования (например, высокий уровень безопасности при обмене данными с малой задержкой), а это означает, что злоумышленники должны приложить больше усилий, чтобы добиться успеха, тогда как в других сценариях это просто обычные телефонные звонки. Очевидно, что эти сценарии по-разному привлекают злоумышленников, следовательно, соотношение стоимости атаки и выгоды также отличается. Кроме того, скорость, с которой может быть восстановлено нормальное обслуживание при атаке, в этих двух сценариях тоже различается.

По этим причинам сбалансированное доверие может показаться простым, но на самом деле соответствовать всем требованиям сложно.

29.3.2. Цели

С технологической точки зрения тремя основными характеристиками благонадежности являются безопасность, приватность и отказо-

устойчивость (которые реализуются с помощью технологий криптографии и защиты). Для наглядности мы называем их *тремя столпами благонадежности*, которые опираются на *десять блоков* (три аспекта безопасности, два аспекта конфиденциальности и пять аспектов устойчивости), как показано на рис. 29.1. Три столпа и десять блоков служат фундаментом, на котором будет построена благонадежность 6G.

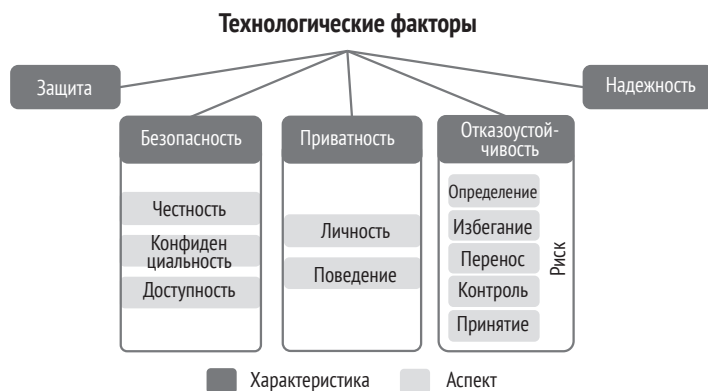


Рис. 29.1. Технологические факторы обеспечения благонадежности

В соответствии с принципами достижения благонадежности 6G нам нужна *нативная архитектура благонадежности*, соответствующая характеристикам безопасности, приватности и отказоустойчивости и основанная на инклюзивной модели доверия.

Эта архитектура должна охватывать весь жизненный цикл 6G, не оставляя пробелов. Мы определили три цели в отношении трех вышеупомянутых столпов и десяти блоков.

Цель 1: сбалансированная безопасность

Так называемая триада AIC – доступность, честность и конфиденциальность (availability, integrity, confidentiality) – представляет собой фундаментальные компоненты безопасности. Баланс – это один из принципов нативной благонадежности, т. е. для разных защищаемых активов/собственности может потребоваться разный уровень защиты или разный вес каждого аспекта в соответствии с различными сценариями.

Цель 2: постоянная защита конфиденциальности

«Конфиденциальность обычно понимается как право людей контролировать или влиять на то, какая информация, связанная с ними, может собираться и храниться, а также кем и кому эта информация может быть раскрыта» [22]. Чтобы защитить информацию о личности и поведении пользователей, в их распоряжение предоставляются такие технологии,

как криптография, дабы гарантировать, что только стороны, уполномоченные пользователями, могут интерпретировать содержание информации, передаваемой между ними.

Цель 3: умная устойчивость

Согласно широко принятому определению, предоставленному Википедией [23], «устойчивость – это способность сети обеспечивать и поддерживать приемлемый уровень обслуживания перед лицом различных сбоев и проблем, связанных с нормальной работой». В данном случае аспекты устойчивости – это набор способностей и инструментов, позволяющих определять и измерять риски. Для обнаружения рисков могут использоваться ситуационная осведомленность и аналитика больших данных, а затем следуют действия, позволяющие избежать рисков. Мы можем передать все или некоторые риски другим сторонам, тем самым облегчая восстановление сети. В противном случае мы можем постараться минимизировать последствия. Наконец, мы принимаем остаточные риски, которые не наносят вреда сетям [16, 24, 25].

Стоит подчеркнуть, что безопасность необходимо постоянно измерять, чтобы сохранить за собой способность быстро реагировать на угрозы и риски. Как указано в отчете ITU-T [25], уровень доверия можно измерить с помощью количественного метода, аналогичного тому, который используется для оценки качества обслуживания (quality of service, QoS) или качества опыта (quality of experience, QoE). Независимо от используемого подхода, определение конкретного уровня доверия зависит от связанных служб и приложений. Уровень доверия – это однозначная, поддающаяся количественной оценке мера доверия к персоне, способностям, свойствам или истинности в отношении кого-то или чего-то [26].

Мы можем подойти к анализу рисков количественно или качественно. Первый подход применяется для присвоения денежного или просто числового значения всем элементам процесса анализа рисков. Второй использует рейтинговую систему для просмотра различных сценариев с разными возможными рисками и ранжирования серьезности угроз и обоснованности различных возможных контрмер на основе субъективных взглядов людей [27].

По результатам анализа рисков можно оценить эффективность существующих механизмов безопасности и необходимость принятия контрмер для снижения общего риска до приемлемого уровня.

29.4. Технологии безопасности

Как было сказано в предыдущих разделах, три столпа и десять блоков являются основой архитектуры безопасности 6G, и на основе инклю-

живной модели доверия мы должны разработать нативную архитектуру благонадежности. Начав с создания модели доверия и затем изучив соответствующие технологии, можно построить надежную сеть 6G.

29.4.1. Модель многостороннего доверия

Доверие пользователей к существующей сети формируется операторами мобильных сетей (mobile network operators, MNO), которые покупают и развертывают сетевое оборудование, уже прошедшее всевозможные тесты и проверки. Для управления пользователями операторы мобильных сетей используют централизованные средства аутентификации и авторизации.

Эта модель доверия может столкнуться со множеством проблем на пути к достижению трех целей благонадежности 6G, которые обсуждались в конце предыдущего раздела. Например, нелегко обеспечить детальный централизованный контроль доступа, который соответствовал бы как высокому уровню безопасности, необходимому для централизованных финансовых сценариев, так и облегченному механизму безопасности для сбора локализованных данных сканирования. Кроме того, трудно установить мгновенное доверие между сотрудничающими партнерами, и нам не хватает эффективного решения для управления цифровыми идентификационными данными.

Поскольку нативность имеет отношение к различным сервисам 6G, это первый принцип, связанный с формированием благонадежности. В данном случае используется *модель многостороннего доверия*, отражающая различные варианты доверия, включая три режима: мост, консенсус и одобрение, как показано на рис. 29.2, и ее основная особенность – децентрализованный многосторонний консенсус с одновременным централизованным доверием и сторонним одобрением. Эти три режима следующие:

- **мост:** это режим установки доверенного моста между объектами через структуру авторизации с центральной точкой, такой как центр политики безопасности или управление безопасностью профилей пользователей. Доверие будет получено из текущей модели доверия;
- **консенсус:** процесс создания более устойчивых и интеллектуальных систем связи в конечном счете сводится к доверию между сторонами. Эти стороны могут быть компонентами сети, разными сторонами цепочки поставок или ролями промышленной экосистемы. В этом режиме транзакции подтверждаются и ответственность распределяется между несколькими сторонами. Основными характеристиками данного режима являются высокая

эффективность и масштабируемость, соответствующие гибким и адаптированным требованиям доступа 6G;

- **подтверждение:** это режим, в котором авторитетная третья сторона измеряет и оценивает надежность сети. Например, как показано на рис. 29.2, сторона В может пригласить третью сторону, чтобы та сказала, заслуживает ли доверия сторона А, и третья сторона может подтвердить надежность стороны А.

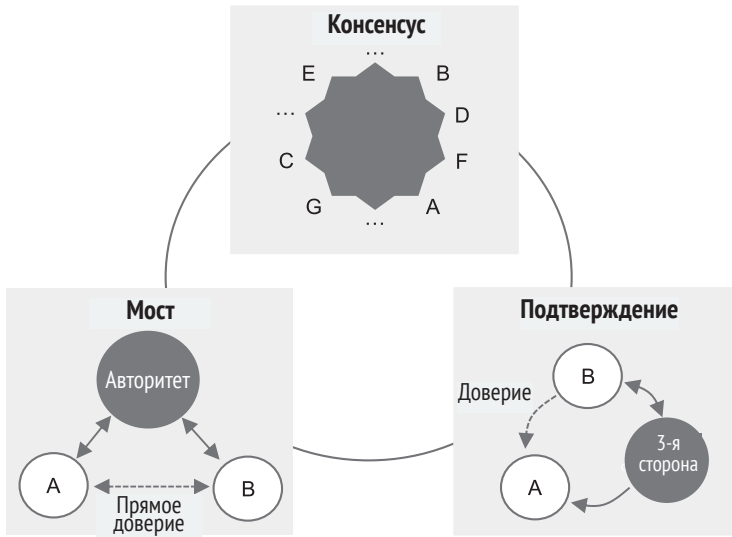


Рис. 29.2. Модель многостороннего доверия

Данные или информация о транзакции, которые нуждаются в одобрении третьей стороны, могут быть быстро и справедливо записаны и подтверждены в режиме консенсуса. Кроме того, консенсус служит ключевым дополнением к взаимному доверию в сценариях граничных автономных услуг.

Три вышеупомянутых режима взаимосвязаны и могут совместно использовать ключевые технологии, такие как криптографические алгоритмы и протоколы, которые должны быть адаптированы к новым сценариям и архитектуре 6G. Более того, распределение и полная интеграция сетевых и вычислительных ресурсов также требуют гибкой и легкой архитектуры безопасности с нулевым доверием, которая применима к децентрализованным сетям.

Для каждой стороны этой тройственной модели (т. е. моста, консенсуса и одобрения) должны быть надлежащим образом достигнуты все цели (сбалансированная безопасность, постоянная защита конфиденциальности и интеллектуальная устойчивость). Однако новые технологии, такие как искусственный интеллект и постквантовая крипто-

графия, меняют глубинные основы благонадежности, построенной на традиционных механизмах. Чтобы обеспечить благонадежность 6G, необходимо знать новые методы атаки и защиты.

29.4.2. Технология распределенного реестра

В основанном на криптографии блокчейне каждый блок цепочки содержит криптографический хеш предыдущего блока, метку времени и данные транзакции. В отличие от него, *распределенный реестр* (distributed ledger) не нуждается в такой цепочке или доказательстве работы. Технология распределенного реестра (distributed ledger technology, DLT) подходит для многих сценариев, сетевых архитектур и приложений (более подробная информация представлена в главе 31). Эта технология, по сути, представляет собой базу данных активов, которая может использоваться несколькими сайтами, учреждениями или географическими регионами [28]. *Блокчейн* – это одна из разновидностей распределенного реестра; также существуют таблицы распределенного реестра, не связанные с блокчейном, распределенная криптовалюта и особая архитектура базы данных [29].

Несмотря на определенные преимущества, сегодняшний блокчейн все еще имеет ограничения:

- **недостаточная гибкость:** блокчейн изначально спроектирован так, чтобы быть неизменяемым, благодаря чему может быть достигнута как высокая достоверность доказательств несанкционированного доступа, так и высокая устойчивость. Во многих сценариях эта неизменяемость обеспечивает надежность оказания финансовых услуг; однако в многочисленных сценариях использования связи в реальном времени некоторые данные должны быть изменяемыми по разным причинам, например изменяемый или программируемый параметр в протоколе связи или виртуальной вычислительной схеме. Вместо того чтобы считаться недействительными в реестре блокчейна, при определенных обстоятельствах данные будут считаться действительными и заноситься в распределенный реестр;
- **низкое быстродействие:** в отличие от некоторых устаревших систем обработки транзакций, которые могут обрабатывать десятки тысяч транзакций в секунду, цепочка блоков биткойнов может обрабатывать только от трех до семи транзакций в секунду. Соответствующий показатель для блокчейна Ethereum [30] составляет всего пятнадцать транзакций в секунду, и поэтому многие наблюдатели не считают технологию блокчейна подходящей для крупномасштабных приложений из-за ее относительно низкого быстродействия;

- **неполные модели безопасности:** многие современные блокчейн-технологии подвержены различным типам атак. Некоторые проекты полагаются на определенные консенсусные алгоритмы (рекурсивные вызовы), которые, как было доказано, содержат уязвимости, такие как атаки децентрализованной автономной организации (decentralized autonomous organization, DAO) в смарт-контракте Ethereum [30].

Другая проблема модели безопасности блокчейна – это появление квантовых вычислений [31], которые потенциально могут сделать уязвимыми лежащие в основе блокчейна криптографические механизмы и протоколы.

Блокчейн-подобная технология для 6G

Консенсус может оказаться наиболее важным режимом в модели многостороннего доверия, поскольку сетевая архитектура 6G будет иметь тенденцию к распределенному характеру. В настоящее время блокчейн-подобная технология, вероятно, достигла зрелости, хотя различные новые применения блокчейна все еще продолжают появляться. Основная проблема заключается в том, как обеспечить соответствие DLT следующим требованиям 6G:

- **сверхвысокая пропускная способность и сверхнизкая задержка:** в отличие от традиционного блокчейна, в котором консенсус достигается медленно, новые алгоритмы и архитектура блокчейна смогут удовлетворить эти требования путем введения новых консенсусных и криптографических алгоритмов;
- **высокая доступность и надежность:** для достижения высокой операционной эффективности сети 6G нуждаются в простых механизмах управления и функционирования. При этом показатели доступности и надежности не только не должны пострадать, но и должны быть улучшены. В связи с этим мы можем повысить доступность, внедрив блокчейн-технологии. Другими словами, операции сети 6G должны быть в состоянии выдержать определенный уровень сбоев или злонамеренных атак, не требуя значительного вмешательства человека;
- **надежная защита конфиденциальности и цифровой суверенитет:** когда дело доходит до защиты конфиденциальности, мы должны соблюдать законы и правила защиты персональных данных (например, GDPR), которые могут различаться в разных странах или регионах. Для этого мы можем использовать технологию распределенного реестра. Тем не менее большинство законов и правил предусматривают право личности удалять данные о себе из системы («право на цифровое забвение»), что противоречит

обязательной неизменности блокчейна [17]. Сегодняшние структуры блокчейнов основаны на цепочках хеширования, и значение хеш-функции чрезвычайно сложно изменить после записи транзакции в цепочку блоков. Если изменения критически необходимы, сегодняшнее решение проблемы представляет собой запуск хардфорка²⁹ [32], который создает уязвимости в системе безопасности. Хардфорк может повлиять на многие существующие записанные блоки, и на повторную проверку транзакции уходит много времени. Поэтому необходимо исследовать возможность изменения или редактирования транзакций, не влияя на другие блоки.

29.4.3. Постквантовая криптография

Квантовые компьютеры могут эффективно решать сложные математические задачи (например, NP-трудные), то есть целочисленную факторизацию или дискретное логарифмирование, «тем самым делая все криптосистемы с открытым ключом, основанные на предположении о сложности вычислений, бесполезными» [33].

Но еще до появления реальных квантовых компьютеров необходимо рассмотреть многие уязвимости традиционных схем. Например, злоумышленник может сохранить сегодняшние сообщения с обменом ключами и взломать их в 2035 году. Получается, что обмен ключами с использованием метода Диффи–Хеллмана (DH) уже уязвим. Когда будут построены крупномасштабные квантовые компьютеры? Как мы можем противостоять как квантовым, так и классическим компьютерным противникам [33]? NIST начал свою деятельность по постквантовой стандартизации в декабре 2016 года. Его усилия направлены на разработку квантовобезопасных стандартов криптографии с открытым ключом, включая схемы для шифрования/создания ключей и цифровые подписи [34].

Промышленная деятельность в области постквантовой криптографии (PQC)

Если говорить о промышленности, то здесь основная деятельность в области постквантовой криптографии осуществляется технологическими гигантами. Например, Google экспериментировал с реализацией «новой надежды» [35] в версии Canary браузера Chrome в течение нескольких месяцев, чтобы оценить влияние гибридных схем, а Thales реализовал ведущие алгоритмы-кандидаты и добавил их в популярные криптографические приложения с открытым исходным кодом [36].

²⁹ Ответвление (форк) от блокчейна, когда продолжать записи в текущую ветку по какой-либо причине невозможно, сопровождаемое изменениями алгоритма и кода.

Кроме того, Microsoft реализует PQC в рамках открытого VPN для проекта с открытым исходным кодом [37], который включает три различных протокола PQC: Frodo-KEM, SIKE и Picnic.

PQC для сети 6G

Согласно аналитическому отчету RAND Corporation, «квантовые компьютеры, способные выполнять криптографические приложения, ожидаются в среднем примерно через 15 лет – около 2033 г. Однако, по оценкам экспертов, это может случиться как раньше, так и позже» [38]. В эпоху 6G, когда станут доступны крупномасштабные квантовые вычисления, придется использовать и соответствующую архитектуру безопасности. Следовательно, уже сегодня нужно проанализировать характеристики алгоритмов PQC и оценить возможности их адаптации к протоколам 6G. Один из возможных способов решения этой проблемы заключается в разработке специализированных алгоритмов PQC, чтобы гарантировать, что они обеспечивают адекватную гибкость для применения в 6G и в то же время применимы к существующим платформам. Более того, алгоритмы и протоколы PQC должны поддерживать гибкую структуру 6G, такую как гибкие уровни безопасности, размеры кода и подписи, которые можно адаптировать к установленным протоколам. Эти алгоритмы также должны обеспечивать более гибкое управление ключами, чтобы соответствовать разным размерам ключей для разных протоколов.

29.4.4. Автономная безопасность

Автономная безопасность будет одной из ключевых функций при создании надежной архитектуры 6G. Вместо того чтобы защищать системы и конечных пользователей от атак злоумышленников по мере их возникновения, эта архитектура будет использовать более проактивный подход. В сочетании с архитектурой системы автономные технологии обеспечения безопасности могут образовать динамическую интеллектуальную защиту, и такие технологии включают машинное обучение, искусственную иммунную систему [39] и другие решения. Можно предвидеть широкое распространение технологий на основе ИИ и машинного обучения для решения проблем безопасности.

ИИ способен не только улучшить обнаружение проблем безопасности, но и предлагать рекомендации аналитикам, тем самым сокращая время отклика на инцидент с сотен часов до секунд и повышая продуктивность работы аналитиков с одного или двух инцидентов до тысяч в день. После развертывания в смоделированной сети 6G модель сможет непрерывно обучаться с использованием актуальных данных и улучшать свои показатели. Кроме того, выполняя имитацию атак,

автономные системы безопасности могут помочь операторам своевременно создавать и настраивать эффективные политики безопасности.

Однако мы не должны забывать о проблемах, связанных с новыми технологиями. Первая проблема – техническая надежность. Надежность глубокой нейронной сети (DNN) напрямую влияет на оценку систем искусственного интеллекта. Более того, поскольку DNN недостаточно прозрачна, ее работа иногда приводит к несправедливым и неточным или не поддающимся толкованию результатам и даже может нарушать соответствующие законы или постановления (например, GDPR). Для обучения ИИ требуются массивные наборы реальных данных, что может привести к утечке, подделке, краже и неправомерному использованию данных из-за отсутствия надлежащей защиты. Вторая проблема – это применение ИИ в обществе. Инциденты в области безопасности и конфиденциальности могут происходить из-за неправильного управления и контроля целей ИИ, проблем с качеством данных и недостаточных знаний разработчиков. Третья проблема – это юридические требования и обязанности: ни в одном законе или нормативном акте четко не определены права и обязанности заинтересованных сторон [40].

Проще говоря, стандартного определения KPI благонадежности не существует. Но если мы перестанем рассуждать о субъективных факторах и сосредоточимся на технических показателях, которые можно использовать для научных исследований, KPI благонадежности будет намного проще определить. Когда-нибудь в ближайшем будущем мы с этим разберемся, измерим и наконец достигнем цели.

29.5. Источники

- [1] Wikipedia, *Trust (social science)*, 2020. [https://en.wikipedia.org/wiki/Trust_\(social_science\)](https://en.wikipedia.org/wiki/Trust_(social_science)).
- [2] Wikipedia, *Trustworthiness*, 2020. <https://simple.wikipedia.org/wiki/Trustworthiness>.
- [3] Merriam–Webster dictionary, *trustworthy*, 2020. <https://www.merriam-webster.com/dictionary/trustworthy>.
- [4] F. Belanger, J. S. Hiller, and W. J. Smith, *Trustworthiness in electronic commerce: The role of privacy, security, and site attributes*, Journal of Strategic Information Systems, vol. 11, no. 3–4, pp. 245–270, 2002.
- [5] J. R. Nurse, S. Creese, M. Goldsmith, and K. Lamberts, *Trustworthy and effective communication of cybersecurity risks: A review*, in Proc. 2011 1st Workshop on Socio-Technical Aspects in Security and Trust (STAST). IEEE, 2011, pp. 60–68.
- [6] J. R. Nurse, I. Agraftotis, S. Creese, M. Goldsmith, and K. Lamberts, *Communicating trustworthiness using radar graphs: A detailed look*, in

- Proc. 2013 11th Annual Conference on Privacy, Security and Trust. IEEE, 2013, pp. 333–339.
- [7] ITU-T Recommendation, *Information technology – open systems interconnection – the directory: Public-key and attribute certificate frameworks*, 2000.
 - [8] G. Lee and H. Lee, *Standardization of trust provisioning study*, Technical Report, International Telecommunication Union, 2005.
 - [9] ITU-T, *Future social media and knowledge society*, 2015.
 - [10] ITU-T, *Trust provisioning for future ICT infrastructure and services*, 2016.
 - [11] ITU-T Recommendation Y.3501, *The basic principles of trusted environment in ICT infrastructure*, 2017.
 - [12] ITU-T Recommendation Y.3502, *Overview of trust provisioning in ICT infrastructures and services*, 2017.
 - [13] ITU-T, Draft Recommendation X.5Gsec-t, *Security framework based on trust relationship for 5G ecosystem*, 2019.
 - [14] E. R. Griffor, C. Greer, D. A. Wollman, and M. J. Burns, *Framework for cyber-physical systems: Vol. 2, working group reports*, NIST Special Publication, 1500-202, 2017.
 - [15] M. Balduccini, E. Griffor, M. Huth, C. Vishik, M. Burns, and D. Wollman, *Ontology based reasoning about the trustworthiness of cyber-physical systems*, IET Journal of IoT, June 2018.
 - [16] Wikipedia, *Law*, 2020. <https://en.wikipedia.org/wiki/Law>.
 - [17] E. D. P. Board, *General data protection regulation*, 2020. <https://gdpr-info.eu/>.
 - [18] The European Network and Information Security Agency, *EU cyber security act, Europe*, 2020. <https://ec.europa.eu/digital-single-market/en/eucybersecurity-act>.
 - [19] The European Network and Information Security Agency, *A trust and cyber secure Europe*, 2020. <https://www.enisa.europa.eu/publications/corporate-documents/a-trusted-and-cyber-secure-europe-enisa-strategy>.
 - [20] Huawei, *Position paper – 5G applications*, 2020. https://www-file.huawei.com/-/media/corporate/pdf/public-policy/position_paper_5g_applications.pdf.
 - [21] Youtube, *RSA Conference. 2014: Share. Learn. Secure*, 2014. <https://www.youtube.com/watch?v=ti07Ufl1gsk>.
 - [22] ITU-T, *Security in telecommunications and information technology*, 2003.
 - [23] Wikipedia, *Resilience (network)*, 2020. [https://en.wikipedia.org/wiki/Resilience_\(network\)](https://en.wikipedia.org/wiki/Resilience_(network)).
 - [24] P. Trimintzios, *Measurement frameworks and metrics for resilient networks and services*, Technical Report, European Network Information Security Agency, p. 109, 2011.

- [25] ITU-T, *Trust in ICT*, 2017. <https://www.itu.int/pub/T-TUTTRUST-2017>.
- [26] ITU-T Recommendation, *Baseline identity management terms and definitions*, 2010.
- [27] S. Harris and F. Maymi, *Cissp all-in-one exam guide*. McGraw-Hill, 2016.
- [28] UK Government Chief Scientific Adviser, *Distributed ledger technology: Beyond block chain*, 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/492972/gs-16-1-distributedledger-technology.pdf.
- [29] Wikipedia, *Distributed ledger*, 2020. https://en.wikipedia.org/wiki/Distributed_ledger#cite_note-UKscienceOffice201601-1.
- [30] V. Buterin et al., *A next-generation smart contract and decentralized application platform*, White Paper, vol. 3, no. 37, 2014.
- [31] E. Rieffel and W. Polak, *An introduction to quantum computing for non-physicists*, ACM Computing Surveys (CSUR), vol. 32, no. 3, pp. 300–335, 2000.
- [32] Ethereum, *The whole forking history*, 2020. <https://blockchain.news/news/ethereum-the-whole-forking-history>.
- [33] L. Chen, S. Jordan, Y.-K. Liu, D. Moody, R. Peralta, R. Perlner, and D. Smith-Tone, *NISTIR 8105 report on post-quantum cryptography*, National Institute of Standards and Technology, p. 10, 2016.
- [34] NIST, *The whole forking history*, 2020. <https://csrc.nist.gov/Projects/post-quantum-cryptography/post-quantum-cryptography-standardization>.
- [35] E. Alkim, L. Ducas, T. Pöppelmann, and P. Schwabe, *Post-quantum key exchange a new hope*, in Proc. 25th USENIX Security Symposium, 2016, pp. 327–343.
- [36] Thales, *Thales esecurity*, 2020. <https://github.com/thales-esecurity?language=c>.
- [37] Microsoft, *PQCrypto-VPN*, 2020. <https://github.com/Microsoft/PQCrypto-VPN>.
- [38] M. J. Vermeer and E. D. Peet, *Securing communications in the quantum computing age: Managing the risks to encryption*, RAND Corporation, 2020. https://www.rand.org/pubs/research_reports/RR3102.html.
- [39] E. Guillen and R. Paez, *Artificial immune systems – AIS as security network solution*, in Proc. International Conference on Bio-Inspired Models of Network, Information, and Computing Systems. Springer, 2010, pp. 680–681.
- [40] Huawei, *Thinking ahead about AI security and privacy protection*, 2019. http://www-file.huawei.com/-/media/CORPORATE/PDF/trust-center/Huawei_AI_Security_and_Privacy_Protection_White_Paper_en.pdf.

Глава 30

Архитектура управления данными в сети 6G

30.1. Почему необходимо управление данными

Данные имеют ключевое значение в современном цифровом обществе, и ожидается, что в будущем в системе 6G будут создаваться, собираться и перемещаться огромные объемы данных. Эти данные будут применяться для различных целей эксплуатации и управления, включая мониторинг показателей, настройку и противодействие сбоям. Они также будут использоваться для обмена знаниями с другими системами и бизнес-секторами, способствуя созданию более широкого диапазона потребительской ценности. Только при этих условиях системы мобильной связи станут ключевым фактором дальнейшего развития других отраслей, например промышленности.

В понятие управления данными может быть заложен разный смысл в зависимости от области его использования, которая простирается от экономики до технологии. Концепция управления данными включает сбор данных, их обработку и хранение, а также развитие соответствующей инфраструктуры для получения и использования высококачественных данных в качестве ключевых активов организации с помощью соответствующих процессов и технологий [1]. В настоящее время данные, генерируемые в системе мобильной связи каждого оператора мобильной сети, изолированы и хранятся отдельно в соответствии с техническими областями, например RAN, базовая сеть (core network, CN), транспортная сеть (transport network, TN), OA&M и оконечные устройства. Данным, принадлежащим различным сетевым сегментам и игрокам, не хватает открытости и прозрачности, что приводит к появлению разрозненных хранилищ, которые являются серьезным препятствием для сбора и обмена данными. С другой стороны, крупные компании, оказы-

вающие услуги поверх сетей мобильной связи, накопили большой опыт в области управления данными и стратегии монетизации (например, в области хранения данных, аналитических услуг, богатых интерфейсов прикладного программирования), которые намного более развиты, чем сектор телекоммуникаций.

Схема управления данными в системе 6G будет иметь важное значение для обеспечения надежной поддержки ИИ и сервисов сканирования; следовательно, ожидается появление новых подходов и системных функций.

30.2. Идеи и принципы проектирования системы управления данными

Управление данными выходит далеко за рамки обычного сбора и хранения данных. В целом при проектировании системы управления данными необходимо учитывать четыре аспекта, представленных на рис. 30.1.

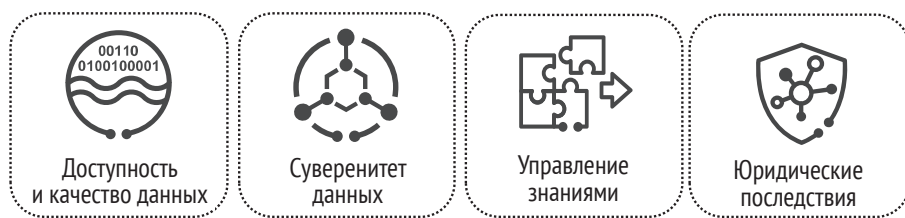


Рис. 30.1. Ключевые аспекты системы управления данными

Доступность и качество данных

Это одна из самых серьезных проблем при применении сервисов искусственного интеллекта в различных отраслях. Повышение доступности данных предполагает сбор данных не только из одного сегмента одной системы, но и из нескольких сегментов разных систем. Следовательно, возникает фундаментальный вопрос: как мы можем преодолеть физические границы (например, между разными типами оборудования, разными операторами и разнородными отраслями), чтобы образовался единый *океан данных*?

После того как разрозненные и изолированные данные собраны и доступны для использования, встает вопрос о том, как контролировать и улучшать качество данных. Доступность огромных объемов данных не означает, что данные являются качественными или пригодными для использования. Следовательно, необходимо реализовать эффективную обработку данных при одновременном снижении сложности вычислений и энергопотребления.

Суверенитет данных

По мере цифровизации общества важность суверенитета, безопасности и конфиденциальности данных становится беспрецедентной. Что касается управления данными, многие страны приняли законы или постановления о защите конфиденциальности. Поставщики услуг также обновляют свои схемы защиты конфиденциальности. Кроме того, правительства развитых стран по всему миру разрабатывают или публикуют правила управления данными. Например, Европейский союз в 2018 году официально выпустил GDPR, регулирующий использование данных на уровне ЕС [2]. Следом в 2019 году Китай выпустил Правила обеспечения безопасности данных, которые вместе с Законом о кибербезопасности, принятым в 2016 году, представляют собой китайский эквивалент GDPR [3]. Соединенные Штаты также вводят в действие законы о конфиденциальности, такие как Закон Калифорнии о конфиденциальности потребителей, который вступил в силу в январе 2020 года [4]. (В России в 2006 г. принят закон 152-ФЗ «О персональных данных», устанавливающий в том числе правила сбора, обработки и хранения персональных данных граждан. Последние поправки в этот закон внесены в 2011 г. – *Прим. перев.*) По каким критериям мы можем оценить внутреннюю значимость данных, чтобы обеспечить предоставление услуг надлежащего качества, а также защитить конфиденциальность с целью соблюдения суверенитета данных? Сегодня это горячая тема для дискуссий. При проектировании системы 6G следует учитывать существующую нормативную неопределенность, особенно в различных географических регионах.

Управление знаниями

Знания можно рассматривать как обработанные данные, имеющие определенное назначение и ценность; их могут напрямую использовать физические и виртуальные объекты, действующие в различных технических и деловых областях. Управление знаниями охватывает создание, обновление и раскрытие знаний. При получении и обновлении знаний следует тщательно анализировать источник и качество данных, а также принимать меры предосторожности против ненадежных или даже злонамеренных источников некачественных или вредоносных данных. Раскрытие знаний зависит от соответствующей платформы и устройства интерфейса.

Правовые последствия

Собираемые и используемые данные становятся все более сложными и конфиденциальными и часто включают прямую трансляцию информации с датчиков и других источников. Нарастающее количество разнообразных источников привело к появлению новых потоков данных

и новых типов контента, вызывающих политическую и юридическую озабоченность по поводу возможных злоупотреблений: организации или правительства могут использовать эти возможности для контроля над обществом. Точно так же новые технологические возможности мешают обычным людям различать настоящий и мошеннический технологический контент, такой как подлинное видео и дипфейк (deep fake; изображение или звук, изготовленные при помощи глубоких нейросетей). Поэтому перед разработчиками уже сегодня стоит задача поддерживать хрупкий баланс между сохранением социальных выгод от передовых технологий и риском нежелательного превращения этих новых технологических возможностей в инструменты для контроля над обществом и ограничения свободы. Чтобы выявлять изощренное мошенничество и предотвращать злоупотребление передовыми технологическими возможностями, необходимо разработать более радикальные правовые и политические инструменты.

30.3. Особенности архитектуры управления данными

Ключевой особенностью систематической структуры управления данными является *независимый уровень данных* (рис. 30.2), который реализует общие подходы к данным, обеспечивая прозрачную, эффективную, внутреннюю безопасность и защиту конфиденциальности для внутренних и внешних операций сети 6G. В следующих разделах представлены основные концепции и соответствующие сетевые функции и услуги.

30.3.1. Независимый уровень данных

Независимый уровень данных предназначен для реализации схемы управления данными в системе 6G. На уровень данных не просто стекаются данные от всех мыслимых источников: здесь протекает полный цикл управления данными, включая создание/сбор, хранение, обработку и аналитику данных, а также предоставление готовых к использованию данных. Следовательно, независимый уровень данных может предоставлять услуги передачи данных для внешних бизнес-структур, таких как вертикальные отрасли (например, автомобилестроение, производство потребительских товаров и электронное здравоохранение), а также для самой системы 6G (например, уровень управления, уровень пользователя и уровень администрирования), для автоматизации и оптимизации сети. Данные могут включать в себя конфигурации, состояния и журналы, связанные с сетевыми операциями, личными данными пользователей, данными датчиков и служебными данными, предоставляемыми другими участниками.

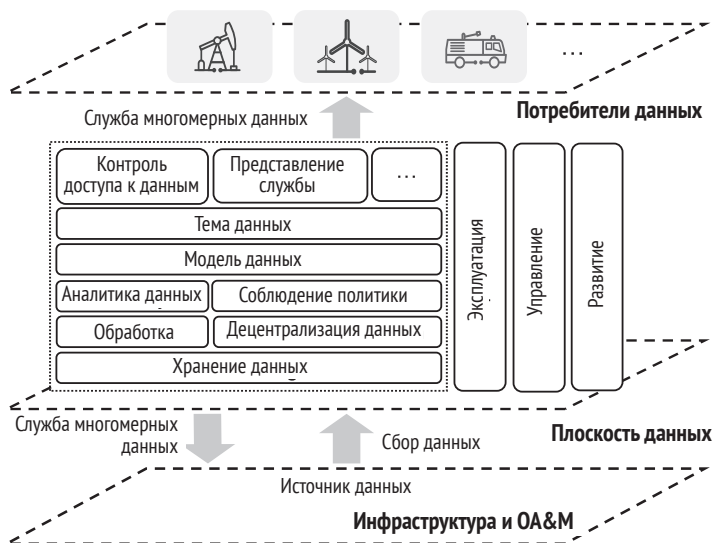


Рис. 30.2. Независимый уровень данных, необходимый для полной управляемости

Собранные данные образуют обширный ресурс, который может быть организован по распределенному принципу. Данные обычно необходимо предварительно обработать (например, выполнить анонимизацию, приведение к заданному формату, шумоподавление, преобразование и извлечение признаков), прежде чем их можно будет использовать, поскольку прямое использование необработанных данных для таких приложений, как ИИ и сканирование, проблематично.

Для обеспечения целостности и законности использования данных во время их обработки необходимо по умолчанию применять установленные законами и сопутствующими нормами политики в отношении данных, охватывающие географические ограничения, национальные или региональные правила конфиденциальности и другие правила. Все операции на уровне данных должны соответствовать правам и обязанностям пользователя данных, указанным в контракте на данные. Кроме того, уровень данных должен обеспечивать десенсибилизацию данных, которая является ключевым условием обеспечения защиты конфиденциальности данных.

Все услуги, предоставляемые упомянутым выше уровнем данных, обслуживаются и управляются автономной системой ОА&М.

Другой важный аспект уровня данных – получение знаний на основе сбора, обработки и согласования данных. Этот процесс должен выполняться в соответствии с требованиями контракта, чтобы организовать юридически безупречную обработку и передачу соответствующих данных из нескольких источников.

Сама структура управления данными может постоянно развиваться и обогащаться за счет новых источников, моделей и тем данных, которые поддерживаются и используются клиентами. Следовательно, она допускает разработку инструментария в реальном времени параллельно с эксплуатацией имеющихся инструментов.

Поскольку уровень данных является логической концепцией, он может быть реализован с централизованной иерархической архитектурой. В качестве альтернативы он может быть реализован как логическая функция и распределен по граничным узлам. В следующих разделах обсуждаются некоторые ключевые компоненты уровня данных.

30.3.2. Многопользовательские роли в управлении данными

Экосистема управления данными включает в себя двойственные роли: заказчик данных / поставщик данных и владелец данных / распорядитель данных. Эти роли могут принадлежать одним и тем же или разным хозяйствующим субъектам. Следовательно, в типичном сценарии управления данными в 6G участвуют несколько игроков, которые могут быть потенциальными потребителями данных, т. е. потреблять данные или знания, предоставленные системой 6G, или могут быть поставщиками данных для системы 6G. Сеть 6G может иметь свою собственную структуру управления данными, но, с другой стороны, 6G также может реализовать структуру управления данными вместе с другими участниками отрасли на основе их собственных знаний в предметной области. Возможны разные пути эволюции (или революции). Поэтому важно еще на этапе разработки установить, как права на данные будут распределяться между различными бизнес-объектами на этапе эксплуатации. Этого можно достичь с помощью децентрализованных технологий, таких как блокчейн. Данный вопрос более подробно обсуждается в главе 31. Тем не менее, являясь новым компонентом систем мобильной связи, независимые уровни данных могут потребовать выработки стандартизированных требований к функциям и интерфейсам.

30.3.3. Ресурс данных

Ресурс данных представляет собой все разнообразие данных, включая структурированные или неструктурированные данные, а также преобразованные, постобработанные или необработанные данные.

Эффективный сбор данных (например, состояния сети и поведения пользователей, таких как модели их мобильности) из беспроводной среды является первоосновой управления данными. Затем можно использовать ИИ для анализа данных и предоставления полученных зна-

ний внутренним или внешним клиентам. Поэтому важно понимать характер и свойства источника данных.

На рис. 30.3 показаны некоторые из основных категорий источников данных в системе 6G.

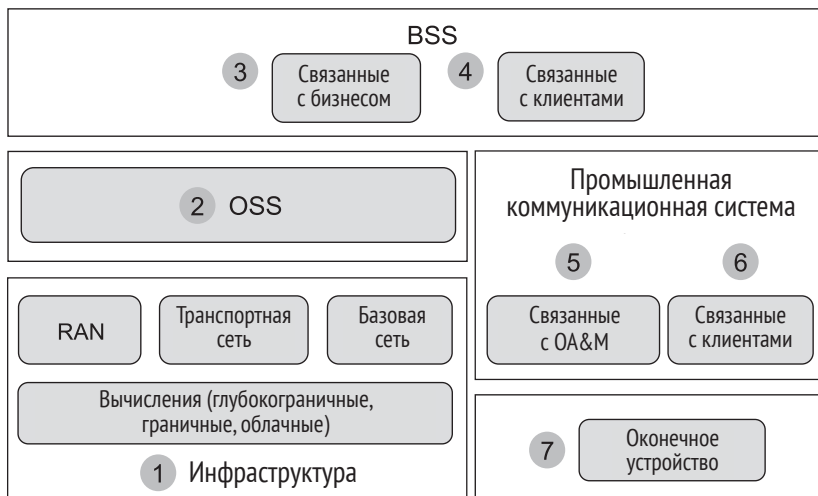


Рис. 30.3. Основные категории источников данных

- **Инфраструктура:** это понятие охватывает систему связи и включает все типы физических и виртуальных ресурсов, такие как RAN, TN и CN. Сюда также входят вычислительные ресурсы, включая облачные, граничные и глубокие граничные вычисления. Данные, относящиеся к инфраструктуре, в основном генерируются внутри инфраструктуры, включая вычислительные ресурсы, ресурсы связи (например, состояние сетевой услуги), информацию сканирования (например, из RAN) и определенные профили пользователей (например, информацию о мобильности, местоположении и связанный контекст).
- **Система поддержки эксплуатации (operation supporting system, OSS):** этот уровень содержит все данные, связанные с эксплуатацией и обслуживанием (OA&M), такие как состояние физического оборудования, информацию о работе системы и информацию о предоставлении услуг.
- **Система поддержки бизнеса (business supporting system, BSS):** этот уровень содержит все данные, связанные с бизнес-логикой, например информацию об управлении отношениями с клиентами и партнерами. Что еще более важно, он также содержит данные о подписках как для потребителей, так и для корпоративных клиентов. Клиенты должны полностью владеть такими данными и, следовательно, иметь полный контроль над ними.

- **Отраслевая система связи:** когда промышленный сектор внедряет 6G, собранные данные могут включать информацию, относящуюся к промышленному варианту использования OA&M. Они также могут содержать такие данные, как профили промышленных пользователей (например, параметры трафика и мобильности) и бизнес-данные / данные об услугах, хранящиеся в облаке. Промышленные заказчики должны полностью владеть данными этого типа.
- **Оконечное оборудование:** данные со стороны конечного оборудования включают вычислительные и коммуникационные ресурсы устройства, профили использования услуг и информацию о сканировании. Конечные пользователи должны полностью владеть данными этого типа.

30.3.4. Сбор данных

В 6G одна из основных ролей управления данными заключается в предоставлении подходящих методов для создания ресурсов данных, что требует наличия соответствующей архитектуры, а также поддержки сетевых функций. Первый шаг к достижению этого – сбор данных. Ключевые шаги сбора данных заключаются в следующем:

- установление соглашения (например, авторизация на право сбора данных) и установление безопасного соединения с источником данных;
- получение требований к сбору данных; принятие решения, какие данные собирать, а также где, когда и как собирать данные в соответствии с требованиями;
- уведомление источника данных об атрибутах данных;
- сбор данных из источника и сохранение их в базе данных;
- управление и поддержание работоспособности базы данных.

30.3.5. Аналитика данных

Управление ресурсом данных открывает возможность предоставления аналитики данных в качестве услуги для различных типов клиентов. Могут применяться четыре типа аналитики:

- **описательная аналитика** собирает статистическую информацию об исторических данных, чтобы обеспечить понимание сети, например быстрое действие сети, профиль трафика, состояние канала и количество пользователей;
- **диагностическая аналитика** позволяет автономно обнаруживать сетевые сбои и нарушения обслуживания, выявляет основ-

ные причины сетевых аномалий и в конечном итоге повышает надежность и безопасность сети;

- **прогностическая аналитика** использует данные для прогнозирования будущих событий, таких как модели трафика, расположение пользователей, поведение и предпочтения пользователей, доступность ресурсов и даже сбои;
- **предварительная аналитика** использует прогностическую аналитику для предоставления предложений по распределению ресурсов, размещению контента и т. д.

Услуги по анализу данных будут генерировать знания, полученные из уровня данных. Результат может включать проактивные знания (например, рекомендации по действиям) и пассивные знания (например, обмен информацией и действия, предпринятые клиентами).

Такой анализ данных может быть запрошен клиентами и адаптирован к их требованиям. Уровень данных должен обеспечивать многомерное предоставление услуг и данных по запросу. В табл. 30.1 перечислены некоторые примеры услуг, которые могут быть предоставлены клиентам. Можно ожидать, что типы клиентов будут гораздо более разнообразными, чем примеры, перечисленные в таблице, и что их требования и варианты использования результатов анализа данных будут отличаться.

Таблица 30.1. Примеры услуг, предоставляемых уровнем данных

Пользователь	Пример услуги
Контроллер сетевой инфраструктуры	Наблюдение за показателями сетевой инфраструктуры
	Наблюдение за потреблением ресурсов сетевой инфраструктуры
	Наблюдение, конфигурирование и оптимизация сетевой инфраструктуры
Контроллер коммуникационной службы	Наблюдение за показателями (например, QoS), анализ, прогнозирование и поддержка
	Оптимизация нагрузки на службу
	Наблюдение за безопасностью службы, анализ, прогнозирование и поддержка
Контроллер подключения устройств	Локализация и различение устройств
	Настройка статуса активности устройства и его отслеживание
	Наблюдение за состоянием мобильности устройства, анализ и прогнозирование

Пользователь	Пример услуги
Контроллер передачи контента	Буферизация данных и оптимизация схемы доставки
Сторонний пользователь	Анализ поведения пользователя (например, степень заинтересованности определенной услугой), локализация и распределение пользователей и т. д.
	Анализ социального контекста (например, социальное окружение пользователя, его настроение и мобильность окружения)

30.3.6. Десенсбилизация данных

Сбор и хранение конфиденциальных данных сопряжены с риском нарушения конфиденциальности и обязанностью защищать конфиденциальность. Десенсбилизация данных необходима для решения проблем конфиденциальности, а также для соблюдения правовых норм, особенно для поддержки ИИ и задач распознавания в рамках 6G.

В частности, для задач ИИ можно рассмотреть структуру с многопрофильным доступом к данным (дифференцированный доступ). В последнее время было проведено значительное количество исследований дифференцированной конфиденциальности в области ИИ [5, 6], сосредоточенных на том, как анонимизировать обучающие данные отдельных устройств.

Десенсбилизация данных во время обучения модели и вывода ИИ является важным критерием при проектировании 6G. Подходы, которые могут быть использованы для обеспечения дифференцированной конфиденциальности, включают введение шума в обучающие данные без ущерба для их статистических свойств, чтобы обученная модель по-прежнему фиксировала признаки в исходном наборе данных [7], и применение криптографических методов, чтобы обучение основывалось на зашифрованных (а не расшифрованных) данных [8]. Альтернативный подход заключается в том, что устройства отправляют в сеть параметры модели, а не обучающие данные. Два примера этого подхода – федеративное обучение [9] и раздельное обучение [10].

Одна из ключевых проблем состоит в том, что злонамеренный инсайдер с полным знанием режима обучения может создать информацию, аналогичную обучающим данным, используя постепенную конвергенцию модели [11]. Например, при федеративном обучении это может привести к утечке информации на злонамеренные устройства. Следовательно, жизненно важно знать, как следует обращаться с различными типами методов обучения, а также учитывать их ограничения, не снижая при этом универсальность десенсбилизации данных.

30.4. Источники

- [1] *Data governance*, wikipedia. https://en.wikipedia.org/wiki/Data_governance#Data_governance_organizations.
- [2] European Parliament, *Regulation (eu) 2016/679 of the European Parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation)*, Official Journal of the European Union L, vol. 119, pp. 1–88, 2016.
- [3] *Measures for data security management (draft for comments)*, May 2019. http://www.cac.gov.cn/2019-05/28/c_1124546022.htm.
- [4] *AB-375 Privacy: Personal information: businesses*. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375.
- [5] P. Vepakomma, T. Swedish, R. Raskar, O. Gupta, and A. Dubey, *No peek: A survey of private distributed deep learning*, arXiv preprint arXiv:1812.03288, 2018.
- [6] C. Dwork, *Differential privacy: A survey of results*, Proc. International Conference on Theory and Applications of Models of Computation. Springer, 2008, pp. 1–19.
- [7] C. Dwork, F. McSherry, K. Nissim, and A. Smith, *Calibrating noise to sensitivity in private data analysis*, in Proc. Theory of Cryptography Conference. Springer, 2006, pp. 265–284.
- [8] M. Minelli, *Fully homomorphic encryption for machine learning*, Ph. D. dissertation, 2018.
- [9] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, *Communication efficient learning of deep networks from decentralized data*, in Proc. Conference on Artificial Intelligence and Statistics. PMLR, 2017, pp. 1273–1282.
- [10] O. Gupta and R. Raskar, *Distributed learning of deep neural network over multiple agents*, Journal of Network and Computer Applications, vol. 116, pp. 1–8, 2018.
- [11] B. Hitaj, G. Ateniese, and F. Perez-Cruz, *Deep models under the gan: Information leakage from collaborative deep learning*, in Proc. 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 603–618.

Глава 31

Архитектура многосторонней экосистемы

31.1. Почему нужна новая архитектура экосистемы

До появления 5G у предыдущих поколений мобильной связи была относительно закрытая экосистема, в которой различные субъекты бизнеса взаимодействовали друг с другом только в случае необходимости. Например, роуминг требует сотрудничества нескольких операторов мобильной связи для устранения технических ограничений. Такое сотрудничество обычно предполагало кропотливые закрытые переговоры по контрактам, которые охватывали широкий спектр технических, деловых и юридических аспектов. Целью 5G было расширение экосистемы в сторону вертикальных отраслей. Поэтому было решено построить систему, привлекающую различные типы клиентов путем раскрытия возможностей сети и соответствующих интерфейсов, например сосредоточив внимание исключительно на сетевых функциях.

Ожидается, что 6G станет цифровой платформой следующего поколения, которая будет предлагать цифровые услуги с полным охватом. Чтобы унифицировать широкий спектр услуг, необходимо будет привлечь игроков из разных областей, включая отрасль ИСТ, а также все другие вертикальные отрасли. Этот процесс объединения уже идет, хотя и находится в зачаточном состоянии, и будет развиваться в будущем.

Новые игроки, наряду с обычными операторами мобильной связи, сформируют разнообразную экосистему 6G. Эта экосистема должна быть открытой: фундаментальное требование к архитектуре – для обеспечения технического и делового сотрудничества в многопользовательской среде. Кроме того, такое сотрудничество должно быть эффективным, прозрачным и заслуживающим доверия. Открытая экоси-

стема также может стимулировать появление новых бизнес-моделей многопользовательского сотрудничества.

Система 6G должна быть открытой для широкого круга игроков, чтобы они в равной степени участвовали в развитии, управлении и эксплуатации сети и могли получать доход от своих бизнесов. Это означает, что существующая базовая парадигма *многопользовательского взаимодействия*, основанная на контрактах или взаимодействии сетевых функций, не подойдет для 6G. Вместо этого 6G разовьется до более глубокой парадигмы *многостороннего участия* и будет представлена как цифровая платформа следующего поколения для всего общества. Но для этого необходимо решить несколько ключевых проблем, которые мешают всей отрасли ICT:

- **врожденная граница между системами:** сетевая система по своей сути в настоящее время является закрытой системой с частной базой данных, политикой управления и логикой работы служб. Если взять в качестве примера управление идентификацией пользователей, то учетные данные пользователей – это частные данные, которые поставщики услуг размещают в хранилищах данных. Это препятствует возможности обмена информацией;
- **дорогостоящее сотрудничество и взаимодействие:** сотрудничество возможно только после того, как обе стороны обсудят взаимоприемлемые условия и подпишут контракт, что является трудоемким и длительным процессом. Однако даже с подписанными контрактами по-прежнему возникает множество практических вопросов из-за нечетких критериев и недопонимания;
- **отсутствие доверия между участвующими игроками:** хотя прямое взаимодействие возможно, остаются потенциальные риски, а это означает, что участники соглашения не хотят или не могут полностью доверять друг другу. В таких ситуациях обычно привлекают посредника – доверенную третью сторону. Зависимость от третьей стороны влечет за собой операционные расходы, снижает эффективность, а также предоставляет третьей стороне чрезмерные полномочия.

6G по своей сути будет многопользовательской сетевой средой и бизнес-экосистемой. Необходимо будет обеспечить надежность и безопасность взаимодействия между разными участниками, а также гибкость создания и прекращения безопасных союзов между ними. Следовательно, преодоление вышеуказанных проблем при проектировании архитектуры 6G является фундаментальным требованием для перехода к парадигме многостороннего участия и достижения действительно открытой экосистемы.

31.2. Идеи и принципы проектирования

Для решения новых возникающих проблем потребуются новые принципы проектирования. Поэтому при проектировании системы 6G необходимо учитывать три следующих принципа.

Принцип 1: открытость

Сеть 6G должна быть более открытой с точки зрения обмена сетевой информацией и знаниями, сетевых операций и раскрытия доступа к возможностям многопользовательской совместной работы.

Как сказано в главе 30, изоляция данных представляет собой разбиение данных текущей системы на изолированные домены, где данные хранятся в частном порядке и не используются совместно. И наоборот, открытость принципиально меняет способ обработки данных (как частных, так и общедоступных) участниками мобильной сети. Несколько участников из разных технических и бизнес-областей должны иметь доступ к хранению и обработке данных, что облегчает поток информации через разные границы системы.

Повышение открытости сети гарантирует, что перемещение и обработка информации в одном или нескольких доменах будут более прозрачными. Сегодняшняя сетевая система воспринимается как черный ящик, где внешние игроки имеют доступ только к выводимым результатам. Такая закрытая система вызывает закономерные опасения по поводу потенциальных проблем с конфиденциальностью и безопасностью. Это не только подрывает доверие клиентов к сетевой системе, но и ставит под угрозу готовность различных игроков рынка сетевых услуг участвовать в ней из-за отсутствия информации, необходимой для взаимодействия на равных.

Открытость сохраняет границу между сферами деятельности участников, а не устраняет ее вместе с соответствующими политиками контроля, но способ передачи информации будет радикально изменен. Кроме того, открытость приносит взаимную выгоду за счет общедоступной информации, не делая систему полностью прозрачной.

Принцип 2: высокая функциональная совместимость

В 6G требуется высокая степень взаимодействия между несколькими участниками. Хотя одно мобильное приложение в настоящее время может использовать услуги, предоставляемые несколькими поставщиками, взаимодействие между ними либо затруднительно, либо невозможно из-за недостаточной открытости.

В 6G требуется более высокая степень взаимодействия, чтобы обеспечить ожидаемые показатели большинства сценариев использования в будущем, поскольку многие из них, вероятно, будут предъявлять

строгие требования к качеству услуги, такие как сверхнизкая задержка. Такая совместимость потребует согласованной работы разных игроков для динамической оптимизации нескольких доменов, а это означает, что междоменные операции станут нормой.

Сеть 6G пересмотрит концепцию взаимодействия. Функциональная совместимость в 6G обеспечит бесшовное междоменное использование функций под прямым контролем разных участников, в отличие от сегодняшней реализации, которая предоставляет лишь интерфейсы для обмена параметрами.

Принцип 3: благонадежность

Сеть 6G будет децентрализованной экосистемой, открытой для свободного входа и выхода различных участников. Благонадежность системы 6G подробно описана в главе 29. С точки зрения многостороннего участия благонадежная система 6G должна:

- обеспечить многопользовательскую структуру междоменного управления данными, в которой может быть реализован детальный контроль доступа к данным;
- быть «точкой доверия» при предоставлении информации нескольким игрокам;
- обеспечить возможность аудита предоставляемых услуг, операций и управления, например через системные журналы, которые должны быть аутентичными и неопровержимыми.

Эти три принципа проектирования взаимосвязаны друг с другом.

31.3. Детали многосторонней архитектуры

В этом разделе рассматривается одна из ключевых технологий, задействованных в создании совместной многосторонней экосистемы, – *технология распределенного реестра (DLT)*. Ожидается, что DLT станет основой предложенных принципов проектирования. Поскольку совместная экосистема будет расширяться и охватывать концепции, связанные как с бизнесом, так и с технологиями, она включает в себя все соответствующие сетевые функции, представленные в этой книге, а именно сетевой ИИ, ориентированную на пользователя сеть, нативную благонадежность и интегрированные неназемные сети.

DLT предоставляет ключевые функции, необходимые для построения новой системы мобильной связи на основе открытой, функционально совместимой и надежной платформы. Такая платформа привлечет участников из различных секторов, от правительства и образования до общественного здравоохранения и поставщиков услуг, таких как

финансы и транспорт. В этом разделе мы обсудим ключевые сетевые функции и объекты, на которые влияет DLT.

31.3.1. Технология распределенного реестра

За последнее десятилетие технология блокчейна стала невероятно популярной благодаря криптовалюте и глубокому влиянию, которое она оказала на современное цифровое общество. Основные концепции блокчейн-технологии стали известны после публикации протокола Рахос [1], заложившего основу для создания биткоина. Биткоин, выпущенный разработчиком (или группой разработчиков) под псевдонимом Сатоши Накамото в 2008 году, представляет собой одноранговую систему электронных денег, которая призвана играть роль публичного реестра для транзакций с криптовалютой. В следующем году был создан блокчейн криптовалюты биткоин. Благодаря использованию блокчейна биткоин полностью децентрализован, так что ни один пользователь не может контролировать электронные деньги, и не существует единой точки отказа – это способствует широкому распространению биткоина. Его основное преимущество – возможность прямых транзакций между пользователями без использования доверенной третьей стороны.

Помимо использования в цифровых валютах, таких как биткоин, с 2014 года технология блокчейн применяется в других приложениях, работающих с использованием смарт-контрактов. Блокчейн 1.0, раньше применяемый только для приложений криптовалюты, со временем превратился в блокчейн 2.0 и 3.0. Развитие блокчейн-технологии открыло новые рынки и возможности: например, децентрализованные приложения (decentralized applications, Dapps) были невозможны с блокчейном версии 1.0. Блокчейн в сочетании со смарт-контрактами позволяет встраивать бизнес-логику и механизмы бизнес-процессов в цепочку и помимо биткоинов обмениваться и другими цифровыми активами. Основные преимущества технологии блокчейн впервые были реализованы в индустрии цифровой экономики. Но по мере развития технологии у блокчейна появился ряд преимуществ с точки зрения прозрачной обработки транзакций между различными бизнес-объектами. Эти преимущества могут быть использованы при проектировании архитектуры 6G, например для создания открытой и совместной экосистемы, а также в целях безопасности.

Блокчейн можно использовать как цифровую систему для записи транзакций с активами в нескольких местах одновременно [2, 3]. В отличие от традиционных баз данных, распределенные реестры не имеют централизованного хранилища данных или функций администрирования. Хотя термины блокчейн и DLT часто используются как взаимозаменяемые, они существенно различаются в том смысле, что

блокчейн является лишь подмножеством технологии распределенного реестра.

Ключевые свойства DLT заключаются в следующем.

Распределенный консенсус

DLT фактически использует децентрализованную архитектуру, в которой несколько участников вместе образуют распределенную систему. *Распределенный консенсус* – это ядро системы на основе DLT. В процедуре распределенного консенсуса записи распределенного реестра могут переходить из одного состояния в другое без централизованного контроля. Каждое изменение состояния подлежит демократической процедуре голосования: после того как большинство узлов достигают соглашения, новое состояние принимается и глобально синхронизируется на каждом узле. Это гарантирует, что все узлы в системе на основе DLT всегда имеют глобальное представление о системе.

Неизменяемость и проверяемость

Процедура распределенного консенсуса обеспечивает неизменяемость записей в системе на основе DLT [4]. Протокол распределенного консенсуса (например, Proof-of-Work в биткойне) устроен таким образом, что невозможно с вычислительной точки зрения подделать записи реестра, если только не скомпрометировано большинство узлов. В зависимости от того, какой распределенный протокол используется, доля скомпрометированных узлов, необходимых для обеспечения вычислительной возможности подделки, колеблется от 30 % до 51 % [5, 6].

Процедура распределенного консенсуса также обеспечивает возможность аудита. Реестр содержит все записи исторического состояния, потому что добавление новой записи к уже имеющимся – это единственный способ обновить реестр. Благодаря свойству неизменяемости каждый узел имеет полную копию записей, что позволяет проводить локальный аудит.

Смарт-контракт

Смарт-контракт – это исполняемый двоичный код, который применяет заданную в коде логику обработки к записям реестра. Одно из основных различий между смарт-контрактом и обычным приложением заключается в том, что выполнение смарт-контракта надежно гарантировано и полностью автоматизировано. Другими словами, при соблюдении заранее определенных условий выполнение смарт-контракта не может быть прервано.

Помимо определения логики обработки записей, хранящихся в реестре, смарт-контракт также предоставляет набор API. Пользователи могут вызывать эти API, чтобы запустить смарт-контракт, отправив транзак-

цию в смарт-контракт. После публикации смарт-контракта в системе на основе DLT он фиксируется в распределенном реестре каждого узла системы, после чего становится общедоступным и неизменяемым.

Ключевые свойства DLT могут быть использованы для достижения трех основных принципов проектирования открытой многосторонней экосистемы, а именно открытости, функциональной совместимости и благонадежности.

- **Открытость** обеспечивает простой способ обмена данными и информацией между несколькими доменами в 6G. Использование распределенного консенсуса позволяет обмениваться данными и информацией по всему миру без централизованной третьей стороны. В то же время свойство неизменяемости помогает обеспечить целостность общих данных и информации. Поскольку любая попытка доступа к данным и информации через DLT будет регистрироваться в реестре, свойство проверяемости имеет ключевое значение для реализации принципа открытости.
- **Функциональная совместимость** гарантирует преобразование 6G в многостороннюю экосистему, обеспечивающую совместное предоставление услуг. Использование смарт-контракта позволяет внешним участникам получать доступ через API к совместно используемым данным из другого домена в соответствии с условиями, определенными владельцем данных. Кроме того, запущенный смарт-контракт (API) будет выполняться автоматически, что способствует функциональной совместимости.
- **Благонадежность** направлена на то, чтобы сделать 6G общим гарантом доверия для хранения данных, обмена информацией и ответственности за качество услуг. Свойство неизменяемости помогает повысить безопасность и надежность данных и информации, совместно используемых в нескольких доменах, в то время как распределенный консенсус повышает надежность, поскольку разные игроки должны достичь согласия по вопросам изменения данных и распространения информации. Другими словами, смена состояния системы определяется посредством процедуры демократического голосования, которая надежно соблюдается добросовестными участниками.

Ожидается, что в экосистеме 6G мобильная сеть будет действовать как цифровая платформа, где несколько пользователей из разных доменов взаимодействуют друг с другом для предоставления услуг той или другой стороне. Это требует реализации новых принципов проектирования, описанных ранее. DLT имеет ряд многообещающих свойств, которые наверняка окажутся полезными при реализации новых прин-

ципов проектирования. Интеграция DLT в сеть 6G будет способствовать предоставлению прозрачных и децентрализованных сетевых функций в соответствии с новыми принципами проектирования. Однако, несмотря на свои многочисленные преимущества, DLT также имеет ограничения. Например, в случаях, когда необходимо совместно использовать большие объемы данных, DLT может спровоцировать нежелательные явления, такие как снижение общей пропускной способности системы, высокая задержка и высокое энергопотребление. Поэтому исследователи должны продолжить свою работу по совершенствованию технологии DLT.

В следующих разделах мы покажем, каким образом DLT может преобразовать систему мобильной связи следующего поколения.

31.3.2. Платформа многостороннего пользования

DLT можно использовать для создания новой схемы управления идентификацией и парадигмы управления цифровыми активами, в которой пользователи управляют своими собственными цифровыми активами, а круг пользователей варьируется от частных потребителей до предприятий. Это поможет упростить управление сетью и OA&M. Кроме того, системы обеспечения эксплуатации (operation supporting system, OSS) и системы обеспечения бизнеса (business supporting system, BSS) на основе DLT станут проще, поскольку отпадет необходимость в стороннем координаторе. Для платформы многостороннего пользования все вышеперечисленное – насущная необходимость.

На рис. 31.1 представлена базовая схема DLT-интегрированной архитектуры для 6G. DLT располагается в ядре системы мобильной связи следующего поколения с поддержкой открытой экосистемы. Фактически DLT – это основа, на которой будет построен новый уровень управления, обеспечивающий самодостаточную схему идентификации, так что аутентификация может выполняться беспрепятственно в разных доменах. Уровень управления данными и доступа к данным (data management / data access, DM/DA) создан для управления сетевыми и пользовательскими данными и поддерживает обмен сегментированными данными при сохранении нативной конфиденциальности в нескольких доменах. На основе этих двух уровней могут быть построены следующие уровни управления сетью и OSS/BSS.

Другие участники из разных доменов также могут присоединиться к коалиции мобильных сетей, создав общие уровни данных на основе DLT. Каждый участник определяет свои собственные политики совместного использования данных и предоставляет смарт-контракт для реализации логики автономной обработки общих данных в распределенном реестре или в разных реестрах.

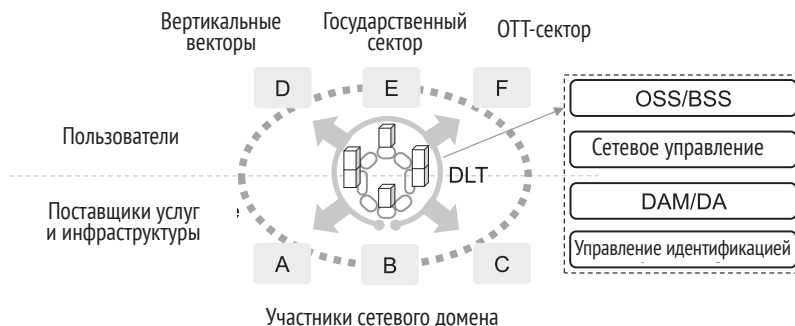


Рис. 31.1. Обобщенная схема DLT-интегрированной архитектуры для 6G

31.3.3. Управление идентификацией

Пользователи хотят контролировать свои данные и общаться анонимно или псевдонимно, чтобы защитить свою личность и конфиденциальность. Для этого в дополнение к криптографическим механизмам нужны передовые технологии, такие как биометрия и DLT.

В обычной системе мобильной связи пользователю требуется SIM-карта для использования сети оператора. При покупке модуля идентификации абонента (т. е. SIM-карты) пользователь предоставляет личные данные, такие как дата рождения, пол, адрес и способ выставления счетов. Это эквивалентно созданию учетной записи (т. е. сохранению учетных данных пользователя) в базе данных оператора. Всякий раз, когда пользователь обращается к сети, информация пользователя сверяется с информацией, хранящейся в базе данных. Пользователю предоставляется доступ к сети только в случае успешной проверки. Каждый оператор хранит такие учетные данные в центральной репозитории пользовательских данных (user data repository, UDR), который изолирован и принадлежит только одному домену; однако ожидается, что в эпоху 6G этот механизм будет работать иначе.

Как показано на рис. 31.2, целью использования DLT является реализация схемы самодостаточной идентификации [7]. В частности, заверяющий орган (например, оператор) после утверждения профиля пользователя выдает ему сертификат. Сертификат пользовательского оборудования содержит открытый ключ пользователя и подпись органа. Орган (например, из сетевого домена А) публикует свой собственный сертификат домена, который содержит открытый ключ органа к реестру. Этот сертификат общедоступен для всех. Чтобы другой орган (например, из сетевого домена Б) мог выполнить аутентификацию, ему нужно только получить сертификат домена, выданный сетевым доменом А. Если подпись сертификата пользователя может быть проверена с помощью полученного открытого ключа, это означает, что сертифи-

кат действительно был выдан заявленным органом-отправителем, и поэтому пользователь прошел аутентификацию.

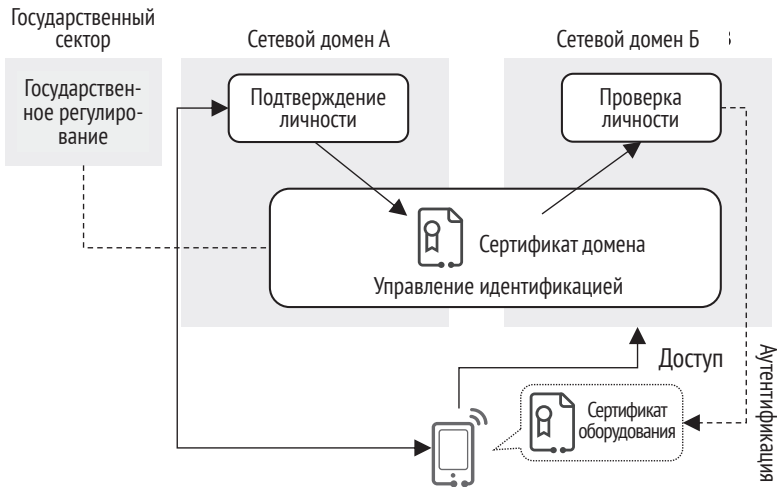


Рис. 31.2. Управление идентификацией на основе DLT

Механизм самостоятельной идентификации открывает новые возможности для управления идентификаторами в мобильной сети. Во-первых, закрытый ключ может быть сгенерирован пользователем, а не создан в базе данных оператора. Поскольку для дешифрования может использоваться только закрытый ключ, суверенитет личности принадлежит владельцу идентификатора (то есть пользователю). Во-вторых, оператор больше не функционирует как создатель идентификатора, а скорее как агент, подтверждающий личность. Это освобождает оператора от необходимости хранить большой объем конфиденциальной информации о пользователях, такой как имена пользователей и пароли. В-третьих, любой оператор может проверить идентификатор в сертификате, получив соответствующий открытый ключ, а это означает, что аутентификация больше не ограничивается отдельными операторами.

Используя схему управления идентификацией на основе DLT, мы можем создать открытую платформу для подтверждения личности. В некоторых случаях может быть удобнее или выгоднее подтверждать личность на уровне государственного или городского органа власти, чем обращаться к операторам, выполняющим эту роль. Чем надежнее индоссант (подтверждающая сторона), тем авторитетнее будет система идентификации.

31.3.4. Управление данными

DLT играет ключевую роль при создании единого надежного уровня данных для мобильной сети 6G. С помощью DLT пользователи могут

получить больший контроль над своими данными, вместо того чтобы позволять другим игрокам (например, поставщикам услуг ОТТ) иметь исключительную власть над ними. Иными словами, все пользователи будут иметь ключи для доступа к своим данным.

Управление данными на основе DLT и доступ к данным цифровых активов пользователей позволяют им управлять своими собственными активами, а также открывают возможности для строительства новой бизнес-модели. Как показано на рис. 31.3, данные могут быть легко и безопасно разделены между разными участниками [8]. В частности, DLT может помочь двумя способами: (1) установить распределенный консенсус по опубликованным данным и (2) помочь напрямую обмениваться данными в автономном режиме, а также гарантировать целостность исходных данных.

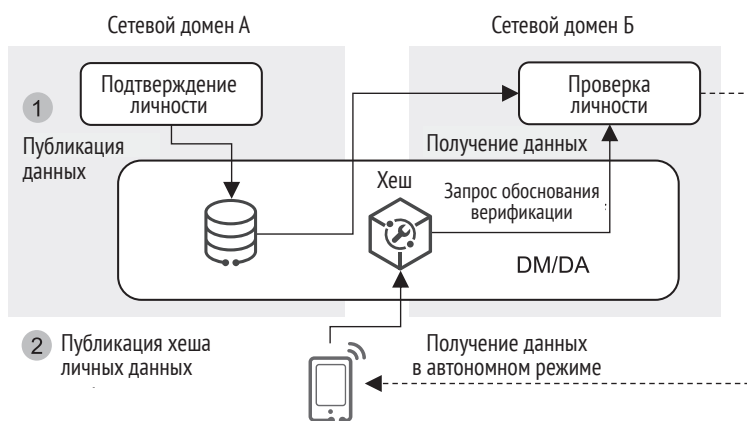


Рис. 31.3. Управление данными на основе DLT

Чтобы установить распределенный консенсус по опубликованным данным, данные напрямую публикуются и проверяются всей сетью с помощью протокола распределенного консенсуса. Каждая часть сети использует одно и то же представление данных реестра, что означает, что данные используются совместно в процессе распределенного консенсуса. Однако, поскольку данные в реестр можно только добавлять, прямая публикация данных в реестре вызывает проблему устойчивости уровня сетевых данных. Другая проблема заключается в том, что, поскольку опубликованные данные реплицируются на каждом узле и общедоступны, конфиденциальность данных не может быть гарантирована. В результате этот подход подходит только для небольшого количества данных, которые являются относительно статичными и нечувствительными к разглашению.

Чтобы пользователи могли напрямую обмениваться данными в автономном режиме, а также гарантировать достоверность исходных

данных, в реестре публикуют проверочную информацию (обычно хеш-значение) исходных данных, а не сами данные. Чтобы проверить достоверность данных, получатель может обратиться к опубликованной проверочной информации. Это связано с тем, что функция неизменяемости DLT гарантирует происхождение доказательства от издателя данных. Кроме того, данные перемещаются в автономном режиме между отправителем и получателем (например, сетевой домен Б принимает данные непосредственно от пользовательского оборудования).

Управление данными на основе DLT формирует новую схему управления и обмена данными в мобильной сети 6G. Во-первых, данные не обязательно должны храниться в частном порядке в одном домене. Вместо этого данные могут быть зашифрованы и распространены по всей сети, поскольку DLT гарантирует их достоверность. Во-вторых, конкретный механизм управления доступом может быть реализован с помощью смарт-контрактов. Любой запрос на доступ к опубликованным данным путем запуска смарт-контракта будет записан, что облегчит будущий аудит. Это особенно важно в отношении данных, которые используются для совместного управления, таких как состояние сети, контекст сеанса и требования QoS.

Управление данными на основе DLT может преобразовать уровень данных в децентрализованный рынок данных, где происходит прямой обмен цифровыми активами между первоначальными владельцами активов и покупателями. Используя DLT, можно организовать полный децентрализованный мониторинг доступа к данным, их передачи и аудита.

31.3.5. Сетевое управление

DLT может найти применение в двух ключевых аспектах управления сетью, а именно в управлении сеансом и управлении доступом и мобильностью.

Управление сеансом

При управлении сеансом маршрут переадресации для пользовательского оборудования на уровне пользователя устанавливается в соответствии с запросом сеанса оборудования. В настоящее время управление сеансом вычисляет маршрут переадресации на основе глобальной информации о сетевом уровне или предопределенных простых правил маршрутизации. Однако в системе 6G управление сеансом могут осуществлять различные распределенные объекты управления в разных доменах. Это затрудняет получение глобальной информации, и может потребоваться, чтобы несколько объектов управления сеансом обменивались сетевой информацией друг с другом, независимо от того, находятся ли они в одном домене.

DLT способствует решению этой проблемы путем ведения общего распределенного реестра, в котором различные управляющие объекты могут публиковать сетевую информацию из своего собственного домена. Чтобы гарантировать, что конфиденциальная информация не будет раскрыта, можно предусмотреть политики публикации. Сеансы могут быть созданы в нескольких доменах с помощью смарт-контрактов, которые публикуются различными объектами управления сеансом. Смарт-контракты определяют условия и входные данные, необходимые для создания маршрута переадресации в определенном домене. После вызова смарт-контракта DLT гарантирует его выполнение, и соответственно развертывается сегмент маршрута. Каждый раз, когда вызывается смарт-контракт, все операции записываются в распределенный реестр для проверки. Эти преимущества делают DLT идеальным выбором для обеспечения совместного использования данных в нескольких доменах.

Управление доступом и мобильностью

Управление доступом и мобильностью (непрерывность сеанса и услуг) сталкивается с проблемой, аналогичной той, с которой сталкивается управление сеансом, когда количество объектов управления увеличивается и объекты распределяются в мобильной сети 6G.

- **Управление доступом:** в текущих реализациях управления доступом пользовательское оборудование аутентифицируется путем запроса службы аутентификации, которая взаимодействует с UDR. В UDR запрос общего ключа генерируется службой аутентификации пользовательского оборудования. Если оборудование передает запрос, служба управления доступом предоставляет разрешение доступа. Механизм запроса общего ключа использует предоставленный ранее предварительный идентификатор оборудования, который дешифруется службой аутентификации, чтобы получить постоянный идентификатор UE.

DLT можно использовать, даже если реализована традиционная схема аутентификации с общим ключом. Запрос общего ключа может быть подготовлен заранее в соответствии с постоянным идентификатором оборудования, хранящимся в UDR. Служба аутентификации подготавливает проверяющий смарт-контракт и публикует его в распределенном реестре, доступ к которому осуществляется набором соответствующих объектов управления доступом. Любой объект управления доступом может извлечь подготовленный запрос и отправить его в пользовательское оборудование. Затем оборудование, запросившее доступ, решает задачу и возвращает свое решение объекту управления досту-

пом, который отправляет решение в смарт-контракт. Если смарт-контракт подтверждает правильность решения, аутентификация оборудования прошла успешно. Это устраняет необходимость в последовательной обработке запроса множеством управляющих объектов, особенно когда сетевая функция должна быть выбрана из множества NF-кандидатов.

Данный подход также работает в междоменном сценарии, когда службы аутентификации в разных доменах подготавливают задачи и соответствующие смарт-контракты проверки. Процесс аутентификации протекает так же, как и в сценарии с одним доменом, без какого-либо междоменного обмена.

Альтернативный и лучший подход состоит в том, чтобы принять парадигму верификации, основанную на самостоятельном управлении идентификатором. Предполагая, что оборудование имеет авторизованный оператором сертификат, оператор публикует его сертификат в распределенном реестре, который распространяется по разным доменам. В этом случае оборудование подключается к объекту управления доступом и запрашивает доступ к сети, одновременно предоставляя свой сертификат. Объект управления доступом анализирует идентификатор домена (то есть смотрит, кто изначально выпустил сертификат оборудования) и извлекает сертификат оператора из распределенного реестра. Затем объект управления доступом проверяет сертификат оборудования. Аутентификация состоялась, если проверка прошла успешно.

Верификация путем предъявления сертификата значительно упрощает процесс аутентификации, хотя необходимо тщательно продумать, как часто оборудование должно подавать заявку на новый сертификат. Особо следует отметить тот факт, что функция аутентификации/авторизации и оператор, выдавший сертификат, не обязательно должны находиться в одном домене.

- **Управление мобильностью:** во время передачи обслуживания (хэндовера) пользовательского оборудования объект управления мобильностью должен использовать контекстную информацию оборудования и, возможно, состояния RAN совместно с объектом управлением сеансом. Это событие запускает последовательное взаимодействие и передачу сигналов между несколькими объектами управления мобильностью и сеансом. В сильно распределенной среде этот подход неэффективен, особенно при частых событиях передачи обслуживания (из-за плотного развертывания ячеек в сети 6G).

DLT может использоваться для реализации стратегии упреждающего хэндовера, согласно которой создается локальный кластер сетевых функций для переключения, так что контекст оборудования в этом кластере может быть синхронизирован заранее. Это сократит время передачи обслуживания, необходимое для отправки и создания нового контекста оборудования. Эффективная синхронизация в данном случае вполне реализуема, поскольку такие кластеры содержат только ограниченное количество локальных пользовательских устройств.

Многие мобильные сетевые функции имеют аналогичные проблемы с администрированием сеанса и управлением доступом и мобильностью. Такие функции тоже можно преобразовать для работы на основе DLT.

31.3.6. Поддержка операционной деятельности и бизнеса

Чтобы стимулировать предстоящую цифровую трансформацию общества и промышленности, необходимо сделать бизнес-среду и экосистему более открытыми с точки зрения разработки, развертывания, эксплуатации, контроля и управления беспроводными сетями. В процессе будут задействованы самые разные участники, играющие разные роли и работающие вместе над созданием совместной экосистемы и разработкой новых бизнес-моделей. В первую очередь это повлияет на структуру обычных OSS и BSS в системе мобильной связи 6G.

Многопользовательская OSS

В обычной OSS информация сначала стекается в одну точку, и лишь затем принимаются решения и вносятся корректировки для последующей реализации в сети. На управляющее решение и его выполнение могут повлиять задержки, сбои в сети, асинхронные состояния сети и другие причины. Хуже того, разные объекты управления могут реализовывать противоречивые решения, что еще больше снижает эффективность управления сетью.

Противоречие между распределенной природой сети и глобально-оптимальным операционным решением является ключевой проблемой, которую нам предстоит решить. Чтобы принимать более обоснованные решения, OSS необходимо иметь более подробную информацию о своей системе. Однако формирование глобального видения всей системы является сложной задачей (например, информация должна быть актуальной, а принятие решений – сложным). С другой стороны, если мы принимаем распределенное решение, информации, доступной на

местном уровне, может быть недостаточно (например, локальные решения изолированных управляющих процессов могут конфликтовать между собой). Кроме того, синхронизация сигнальных механизмов всегда является обременительной и дорогостоящей.

Благодаря механизму распределенного консенсуса для преодоления вышеуказанного противоречия может использоваться DLT. Уровень сетевых данных с поддержкой DLT позволит совместно использовать собранные состояния сети (или проверочную информацию, если требуется конфиденциальность) и синхронизировать их непосредственно на уровне сетевых функций. Что еще более важно, этот уровень данных будет проверять опубликованные данные, достигая распределенного консенсуса для каждого блока информации. Таким образом, уровень сетевых данных с поддержкой DLT обеспечивает OSS надежным источником данных, который предлагает глобальное представление сети для многих рабочих целей, таких как обнаружение сбоев, анализ данных (машинное обучение) и мониторинг QoS.

Что еще более важно, такая OSS на основе DLT устраняет границы между различными участниками, когда можно беспрепятственно совместно использовать информацию/состояния других доменов. Следовательно, каждый домен будет иметь обширную, надежную и синхронизированную информацию для своей деятельности.

Многопользовательская BSS

Польза DLT особенно очевидна в случае обычной BSS, потому что распределенный реестр позволяет использовать междоменную схему ценообразования и упрощает междоменные расчеты. И цены, и условия расчетов могут быть представлены в виде смарт-контрактов, которые полностью заменят бумажные контракты. Условия контракта могут быть согласованы между двумя участниками онлайн или офлайн, и после публикации смарт-контрактов в распределенном реестре полные контракты становятся неизменными.

В схеме назначения стоимости смарт-контракт выполняется автоматически и вступает в силу сразу после его активации покупателем, что приводит к заранее определенному результату. В схеме расчетов соответствующие операции (например, выставление счетов и транзакции) будут выполняться автоматически при выполнении условий без какого-либо вмешательства человека в соответствии с условиями контракта. На протяжении всего процесса не потребуются сторонний координатор.

31.4. Источники

- [1] L. Lamport, *The part-time parliament*, ACM Transactions on Computer Systems, vol. 16, no. 2, pp. 133–169, 1998.

- [2] M. Pilkington, *Blockchain technology: Principles and applications*, in Research handbook on digital transformations. Edward Elgar Publishing, 2016.
- [3] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, *An overview of blockchain technology: Architecture, consensus, and future trends*, in Proc. 2017 IEEE International Congress on Big Data. IEEE, 2017, pp. 557–564.
- [4] F. Hofmann, S. Wurster, E. Ron, and M. Böhmecke-Schwafert, *The immutability concept of blockchains and benefits of early standardization*, in Proc. 2017 ITU Kaleidoscope: Challenges for a Data-Driven Society (ITU K). IEEE, 2017, pp. 1–8.
- [5] M. Castro, B. Liskov et al., *Practical byzantine fault tolerance*, in Proc. Third Symposium on Operating Systems Design and Implementation (OSDI), vol. 99, pp. 173–186, 1999.
- [6] A. Gervais, G. O. Karame, K. Wüst, V. Glykantzis, H. Ritzdorf, and S. Capkun, *On the security and performance of proof of work blockchains*, in Proc. 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016, pp. 3–16.
- [7] O. Jacobovitz, *Blockchain for identity management*, Technical Report, The Lynne and William Frankel Center for Computer Science Department of Computer Science. Ben-Gurion University, Beer Sheva, 2016.
- [8] Y. Zhu, Y. Qin, Z. Zhou, X. Song, G. Liu, and W. C.-C. Chu, *Digital asset management with distributed permission over blockchain and attribute-based access control*, in Proc. 2018 IEEE International Conference on Services Computing (SCC). IEEE, 2018, pp. 193–200.

Глава 32

Интегрированная архитектура неназемных сетей

32.1. Предпосылки к интеграции сетей

Обзор

В отрасли в целом и в частности в отношении сетей, выходящих за рамки 5G, вернулся интерес к интеграции неназемных сетей, таких как спутники, HAPS и дроны, с наземными сетями. Такая интеграция направлена на создание беспроводной сети, охватывающей пока еще не обслуживаемые территории и обеспечивающей трехмерное покрытие, как описано в главе 6. Хотя покрытие беспроводной связи продолжает расти с каждым последующим поколением беспроводной технологии, в некоторых географических регионах все еще могут существовать пробелы в обслуживании, особенно там, где развертывание базовой станции с проводным подключением либо неэкономично, либо невозможно.

Несмотря на то что неназемные сетевые технологии часто считают конкурентами наземных аналогов, в некоторых сетевых структурах они могут восприниматься как функциональное дополнение, и их роль следует рассматривать в зависимости от того, являются они орбитальными или нет. Например, системы HAPS, беспилотные летательные аппараты и воздушные шары не являются орбитальными платформами и с точки зрения архитектуры, как правило, совместимы с существующими сетевыми архитектурами. Эти платформы специально развернуты для покрытия определенного географического региона. С другой стороны, орбитальные платформы, как правило, развертываются в виде глобальной группировки; следовательно, они могут обеспечить повсеместное покрытие. В этой главе мы обсудим, как обеспечить универсальное и повсеместное покрытие в глобальном масштабе, уделяя особое внимание спутниковым технологиям.

В настоящее время ведутся исследования в ряде областей, связанных со спутниковой сетью. Мы надеемся, что эти исследования сделают общедоступной особую функциональность, присущую только спутниковым сетям, что позволит операторам мобильных сетей предлагать новые услуги и создавать новые источники доходов. Опираясь на преимущества спутниковых сетей, интегрированная сеть 6G превзойдет многие возможности других беспроводных и проводных технологий.

Спутниковые сети

Спутниковая технология используется для приложений связи и вещания с середины 1960-х годов. В этом разделе мы уделяем основное внимание спутниковым сетям, используемым для интерактивных услуг, таких как передача голоса и данных. Мы не рассматриваем широко-вещательные приложения (например, радио и телевидение), поскольку они преимущественно предлагаются через спутники GEO; аналогично мы не рассматриваем специальные сети ретрансляции данных для военных целей и космических исследований.

В системах запуска многоразового использования и производстве спутников в настоящее время наблюдается заметный технический прогресс, снижающий стоимость развертывания спутников. Ожидается, что технические инновации и сокращение затрат продолжатся и что недорогие спутниковые технологии и связанные с ними системы запуска будут полностью готовы для развертывания 6G.

Спутниковая технология вызывает все больший интерес как средство построения высокопроизводительных сетей. Однако для минимизации задержек при спутниковой связи требуются спутники LEO или VLEO. Но поскольку эти спутники вращаются на относительно небольшой высоте, каждый из них может обеспечить лишь ограниченное покрытие. Следовательно, требуются более крупные сети или мегасозвездия. Такие компании, как SpaceX и Amazon, и это лишь малая их часть, находятся на этапе использования или приобретения лицензий на развертывание группировок, состоящих из нескольких тысяч спутников, для своих сетей Starlink [1, 2] или Kuiper [3].

В табл. 32.1 приведены типичные примеры спутниковых сетей, которые планируются, предоставляют услуги или развертываются в настоящее время. Эти сети ориентированы на предоставление широкополосных услуг передачи данных с повышенной пропускной способностью по сравнению с услугами, предоставляемыми через геостационарные спутники.

Помимо предоставления выборочных сведений о каждой спутниковой сети, в табл. 32.1 приведена соответствующая конфигурация. Обратите внимание, что Starlink планирует использовать около 40 000 спутников [1, 2] как в полярной, так и в наклонной орбитальной конфигура-

ции, чтобы сформировать сложную группировку. В настоящее время выполняется первая фаза развертывания, включающая только спутники на наклонной орбите (табл. 32.2). Сеть второго поколения (табл. 32.3) планирует задействовать дополнительно 30 000 спутников, включая конфигурации с сильно наклоненной (например, околополярной) орбитой. Напротив, Amazon планирует использовать только конфигурации с наклонной орбитой (см. табл. 32.4), но будет работать в трех орбитальных плоскостях.

Таблица 32.1. Примеры спутниковых сетей (сведения, представленные в этой таблице, обновлены по состоянию на август 2021 г.)

Спутниковая сеть	Состояние	Количество спутников	Высота, км	Конфигурация
Iridium	Работает (голос и данные с низкой скоростью)	66 (шесть орбит)	780	Полярная (86°)
Teledesic	Банкрот	840 (начально)	700 (начально)	Полярная
SpaceX-Starlink	Развертывание (тестирование)	4408 (поколение 1, фаза 1, развернуты)	540–570 328–640	Полярная и наклонная
OneWeb	Развертывание (привлечены новые инвесторы)	648 планируется, к августу 2021 развернуто 254	1200	Полярная
Telesat LEO	Развертывание (тестирование)	298 (117 развернуто)	1000 (полярная) 1200 (наклонная)	Полярная и наклонная
Amazon Kuiper	Планирование	3326	590, 610, 630	Наклонная
O3b/SES	Планирование	20	8062 (средневысотная орбита)	Экваториальная

Таблица 32.2. Подробная информация о группировке SpaceX Starlink (поколение 1, фаза 1) [1]

Группа	Высота, км	Наклонение, град.	Число плоскостей орбит	Спутников на плоскость	Всего спутников
1	550	53	72	22	1584
2	540	53.2	72	22	1584
3	570	70	36	20	720
4	560	97.6	6	58	348
5	560	97.6	4	43	172
Общая численность группировки					4408

Таблица 32.3. Подробная информация о группировке SpaceX Gen2 Starlink (поколение 2) [2]

Группа	Высота, км	Наклонение, град.	Число плоскостей орбит	Спутников на плоскость	Всего спутников
1	328	30	1	7178	7178
2	334	40	1	7178	7178
3	345	53	1	7178	7178
4	360	96.9	40	50	2000
5	373	75	1	1998	1998
6	499	53	1	4000	4000
7	604	148	12	12	144
8	614	115.7	18	18	324
Общая численность группировки					30 000

Таблица 32.4. Подробная информация о группировке проекта Kuiper (Amazon) [3]

Группа	Высота, км	Наклонение, град.	Число плоскостей орбит	Спутников на плоскость	Всего спутников
1	630	51.9	34	34	1156
2	610	42	36	36	1296
3	590	33	28	28	784
Общая численность группировки					3236

32.2. Идеи и принципы проектирования

Консорциум 3GPP проявил значительный интерес к использованию спутников в системах мобильной связи. Например, общая архитектура и связанные проблемы описаны в 3GPP TR38.811 [4, 5], а некоторые потенциальные решения задокументированы в 3GPP TR38.821 [5]. В настоящее время существующие решения предназначены для прямого соединения между пользовательским оборудованием и спутником, а также для подключения к удаленному радиоинтерфейсу и базовому узлу с различными уровнями интеграции протокола радиосвязи непосредственно внутри спутников. Этот единственный на сегодняшний день подход применим, когда спутники участвуют в передаче беспроводного трафика только в базовую сеть RAN. Полная ценность спутниковых сетей раскрывается только тогда, когда спутники поддерживают подключение к RAN, базовой сети и даже к интернету в целом (в данном случае подключение к интернету часто называют «широкополосным доступом»). Такое разнообразное подключение необходимо для того,

чтобы спутниковые сети стали неотъемлемой частью системы 6G, а не оставались в стороне как независимая система, требующая межсетевого взаимодействия.

32.2.1. Спутниковая группировка

Спутниковая сеть находится в непрерывном движении, необходимость которого продиктована фундаментальными законами физики. Одиночный спутник вращается вокруг Земли по эллиптической траектории с центром Земли в одном фокусе эллипса. Большинство спутников, используемых для связи, развернуты на почти круговой орбите. То есть межфокусное расстояние близко к нулю. Положение спутника обычно определяется как прямоугольник, окружающий номинальную точку на круговой орбите. В рамке указаны пределы эксцентриситета орбиты. Орбита обычно характеризуется номинальной высотой спутника и наклоном орбиты. Орбита, пересекающая непосредственно полюса, называется полярной орбитой и имеет наклон 90° ; в противном случае она называется наклонной орбитой.

Наклон орбиты определяет северную и южную крайние точки, покрываемые спутником. Наклон и свойства обращенной к земле антенны определяют максимальные и минимальные широты, которые могут быть покрыты определенной орбитой. Совокупность спутников называется группировкой. Когда несколько спутников образуют сеть, они могут перемещаться по одной или разным орбитам. Геометрия группировки влияет на ее способность обеспечивать глобальное покрытие. Для круговых и почти круговых орбит наиболее важными вариантами группировки являются *полярная* и *дельта-паттерн Уокера* (структура Балларда), как показано на рис. 32.1.

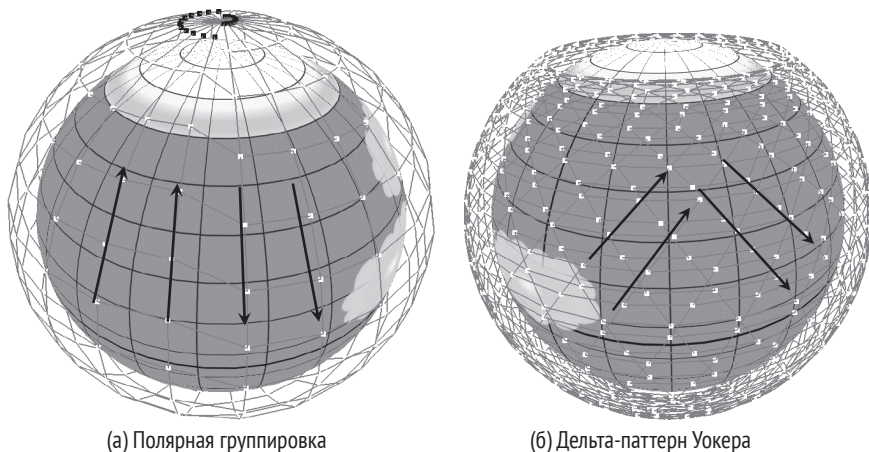


Рис. 32.1. Два наиболее важных варианта группировки с круговой или почти круговой орбитой

Созвездия, состоящие из полярных орбит, обеспечат полный глобальный охват. Однако, хотя большинство спутников будут двигаться в одном направлении, в двух случаях спутники на одной орбите будут двигаться в направлении, противоположном спутникам на соседней орбите. В таких случаях каналы межспутниковой связи вынуждены постоянно реконфигурироваться. Образуется область, называемая *орбитальным швом* (orbital seam), и, поскольку трафик, проходящий через шов, подвергается неоптимальной маршрутизации, эта область оказывает влияние на пропускную способность. Кроме того, сопровождение интерфейса затруднено, т. к. спутники пересекают полюса, а это означает, что межспутниковые линии между орбитами обычно отключены на полюсах.

Группировки Уокера образуются при использовании нескольких наклонных орбит. В таких группировках не возникает швов, потому что спутники остаются почти неподвижными относительно спутников на соседних орбитах, пока не достигнут вершины своих орбит (например, когда спутник, движущийся на север, начинает двигаться на юг). По мере того как спутники движутся на юг, они пересекают другие орбиты, приобретая возможность формировать дополнительную *смежность* (adjacency). Чтобы использовать эту смежность для маршрутизации данных, требуется дополнительный интерфейс (называемый «пятым каналом») с возможностями быстрого отслеживания. Однако использование пятого канала может оказаться невозможным, поскольку это может привести к образованию швов. (Пятым каналом называют канал, который соединяет спутники, движущиеся в противоположных направлениях относительно друг друга в группировке Уокера.)

Поскольку покрытие группировки Уокера ограничено наклоном, оно может быть непригодным, когда для обеспечения покрытия высоких широт требуются спутники на низкой околоземной орбите. Чтобы преодолеть это ограничение, такие компании, как SpaceX (Starlink), развертывают сети, сочетающие в себе оба типа орбитальных группировок, покрывая различные высоты, чтобы обеспечить глобальное покрытие. Например, пользователи в полярных регионах будут получать доступ к спутникам на сильно наклоненных орбитах (например, на околополярных орбитах) вместе с оконечными устройствами, способными работать при малых углах места. Пользователи в более высоких широтах также будут использовать аналогичные устройства.

На рис. 32.2 показан пример группировки, плотность которой аналогична планируемой SpaceX. Хотя в этом примере показана только одна орбитальная оболочка и опущены дополнительные полярные орбиты, обеспечивающие покрытие полярных регионов, он облегчает понимание масштабов этих новых сетей.

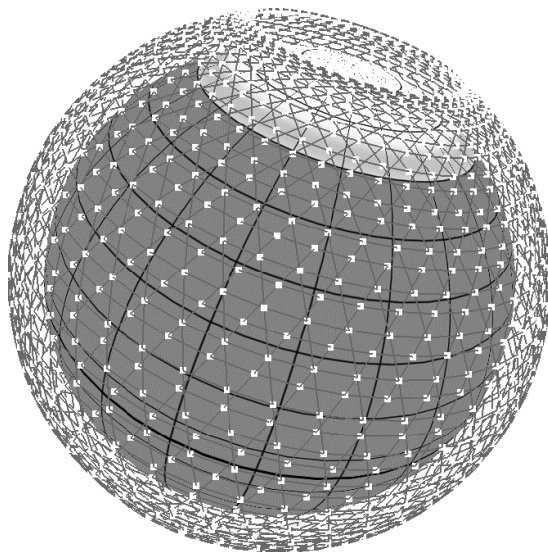


Рис. 32.2. Созвездие высокой плотности

32.2.2. Низкая задержка в глобальном масштабе

В определенных сценариях, особенно на больших расстояниях, группировка спутников на низкой околоземной орбите может обеспечивать сравнимую или даже меньшую задержку, чем наземная оптоволоконная сеть. Это связано с тем, что радиоволны и свет распространяются в свободном пространстве быстрее, чем свет в оптоволокне.

32.2.3. Обеспечение подключения

Для интеграции с системой 6G можно использовать как традиционное соединение с каналом прямой ретрансляции (где спутник действует как простой ретрансляционный узел), так и более продвинутое сетевое соединение (когда трафик ретранслируется между источником и пунктом назначения на больших расстояниях с участием нескольких спутников). Это может быть реализовано с помощью ISL или путем организации спутников для ретрансляции трафика между собой через наземные станции.

Используя многозвенную спутниковую ретрансляцию, мы можем охватить пользователей, которые обычно находятся за пределами досягаемости архитектуры прямой ретрансляции (например, пользователей на кораблях или в самолетах), а также наземных пользователей. Это не только позволяет традиционным сервисам доставлять свои услуги почти всем пользователям во всем мире, но также полезно для таких сервисов, как глобальное уведомление о чрезвычайных ситуациях и реагирование на стихийные бедствия.

Для достижения глобального покрытия и низкой задержки в рамках группировки потребуются сотни или даже тысячи спутников – в зависимости от высоты орбиты спутника, – включая более сложные многооболочечные сети или сети космической ретрансляции. Следовательно, ожидается, что 6G будет развивать эту архитектуру с помощью новых критически важных методов маршрутизации и адресации. Это, в свою очередь, позволит спутниковым группировкам функционировать в качестве общей платформы маршрутизации, которую можно использовать как базовую технологию, в буквальном смысле возносящую систему мобильной связи в небеса.

32.2.4. Мультисервисное обслуживание

Спутниковая технология расширит зону покрытия до районов, которые недостаточно обслуживаются существующими беспроводными сетями. Спутник LEO может покрывать область с радиусом около 300 км. Но с помощью ISL или наземных ретрансляционных станций область покрытия можно расширить и в конечном итоге охватить большую часть земного шара. В результате операторы смогут охватить гораздо большую зону покрытия и реализовать новые возможности обслуживания. Ожидается, что спутниковые сети будут не только выступать в качестве простых небесных маршрутизаторов или ретрансляторов, но и станут предоставлять различные типы услуг. Такие сети, хотя в первую очередь и нацелены на расширение систем мобильной связи, должны позволять предоставлять несколько услуг через общее, но повсеместное ядро. Для этого перед туннелированием через спутниковую сеть услуги должны быть отображены на общем уровне – обычно на уровне IP.

32.3. Особенности архитектуры

Интегрированная спутниковая сеть в системе 6G будет отличаться от наземных сетей, во многом из-за непрерывного движения спутников. По этой причине важно хорошо понимать ключевые особенности архитектуры, чтобы реализовать полное трехмерное покрытие. В этом смысле наземная система мобильной связи может быть действительно расширена вплоть до космоса, особенно когда определенная ее часть размещается в спутниковых системах.

32.3.1. Задержка

Использование спутников LEO позволяет достичь меньшей задержки, чем та, которая обеспечивается существующими оптоволоконными сетями. В радиочастотной передаче, такой как микроволновая, задержка в системе передачи приблизительно равна скорости света в вакуу-

ме. В волоконно-оптических системах наличие преломления снижает скорость передачи примерно до 68 % от скорости света в вакууме. На больших расстояниях это снижение может быть значительным по сравнению с передачей по радиочастотному каналу, и, несмотря на большое расстояние до спутника LEO, спутниковый маршрут может быть быстрее. Микроволновые каналы типа «точка–точка» обычно обеспечивают низкую задержку, но во многих случаях наземное развертывание оборудования затруднено (например, над горами) или невозможно (например, над океанами). В этих случаях для обеспечения оптимального качества обслуживания с точки зрения малой задержки могут использоваться спутники.

В частности, нам необходимо рассмотреть два случая. Первый – это случай с несколькими переходами, который включает передачу трафика через несколько спутников и на большие расстояния (например, вокруг Земли). В этом случае объединяются несколько ISL. Второй вариант – прямая ретрансляция, которая предполагает использование одного спутника для ретрансляции наземного трафика. Иногда такую ретрансляцию называют «коленчатым каналом», и это понятие также применимо к конфигурации HAPS. Следующие рассуждения основаны на задержке распространения без учета задержки оборудования и на предположении, что оптоволоконный тракт напрямую соединяет две наземные точки без боковых отклонений. В действительности всегда будет некоторое отклонение, в результате чего длина оптоволоконного тракта будет больше, чем длина прямой линии связи, что, в свою очередь, приводит к более сложным рассуждениям.

Случай с несколькими переходами

Общая геометрия спутниковой трассы по сравнению с наземной трассой показана на рис. 32.3а. Здесь путь включает соединение восходящей и нисходящей линий связи, а также множественные соединения между спутниками. Для простоты первый и последний спутники показаны непосредственно над наземными пользователями, а спутниковый тракт представлен в виде дуги (пунктирная линия). В действительности спутник может не находиться непосредственно над наземными пользователями, и показанный путь передачи будет состоять из множества отрезков прямых линий между несколькими спутниками. Расстояние между двумя точками a и b на земле обычно рассчитывается по формуле *заверсинуса* и обозначается на рисунке как S_{ab} . Расстояние пути по спутниковому каналу увеличивается на коэффициент, связанный с высотой, как показано на рис. 32.3а.

Основными факторами, влияющими на разницу в распространении между наземным волокном и космической спутниковой линией

связи, являются снижение скорости света в волокне из-за показателя преломления материала волокна и увеличенная длина спутникового тракта, который теперь также включает два перехода «земля–спутник». Для радиочастотной линии связи «земля–космос» и для ISL задержка распространения сигнала определяется скоростью света в вакууме. Мы предполагаем, что показатель преломления в волокне равен 1,47; следовательно, скорость распространения сигнала составляет примерно 68 % скорости света в вакууме. Длины пути наземного сигнала между двумя точками a и b обычно рассчитываются как путь по дуге большого круга (ортодрому) между этими двумя точками на земной поверхности, показанный как S_{ab} на рис. 32.3а. Спутниковый тракт представляет собой серию оптических каналов «точка–точка» в свободном пространстве. При наличии достаточно высокой плотности спутников промежуточные звенья тракта аппроксимируются дугой на сфере с радиусом, равным радиусу Земли плюс высота спутника. Общая задержка распространения спутникового тракта может быть рассчитана, исходя из этого расстояния плюс длина двух каналов доступа «земля–космос». Следовательно, если взять любые две точки на земле, a и b , и знать высоту орбиты спутника, то можно сравнить задержку распространения сигнала по спутниковому тракту с наземным оптоволоконным трактом. Это может, например, помочь нам определить, выгодно ли использование спутников. Рисунок 32.4 иллюстрирует это на примере высокочастотного финансового трейдинга, где уменьшение задержки имеет критически важное значение.

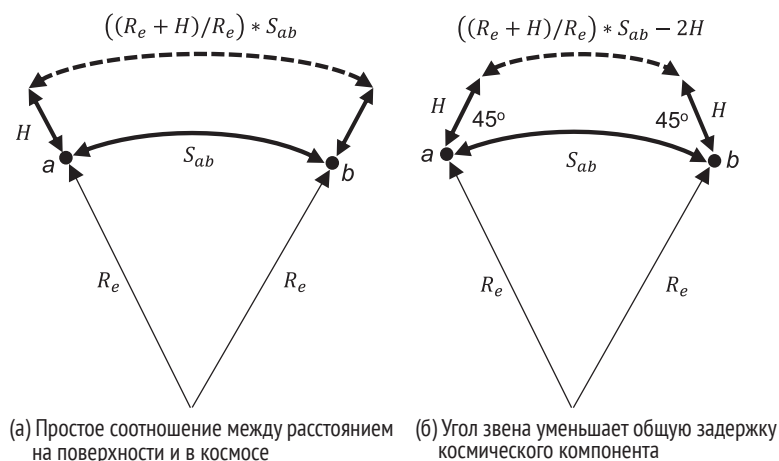


Рис. 32.3. Геометрия спутниковой связи

Мы можем видеть, что, например, чтобы быть полезным для финансового трейдинга между Лос-Анджелесом и Нью-Йорком, максимальная орбита спутника LEO должна составлять примерно 700 км при условии

нулевой задержки переадресации между спутниками, относительно которых также предполагается, что они находятся на идеальной дуге между Лос-Анджелесом и Нью-Йорком. Мы также можем видеть, что верхняя граница возникает там, где спутник находится на орбите примерно 2000 км, и расстояние, которое необходимо преодолеть, чтобы оправдать такой высокий спутник, должно составлять почти половину окружности Земли. Подразумевается, что для достижения более низкой задержки, чем в оптоволоконных каналах, потребуется использование LEO на высоте 600 км и ниже, предполагая, что все пути относительно прямые (как на земле, так и в космосе).

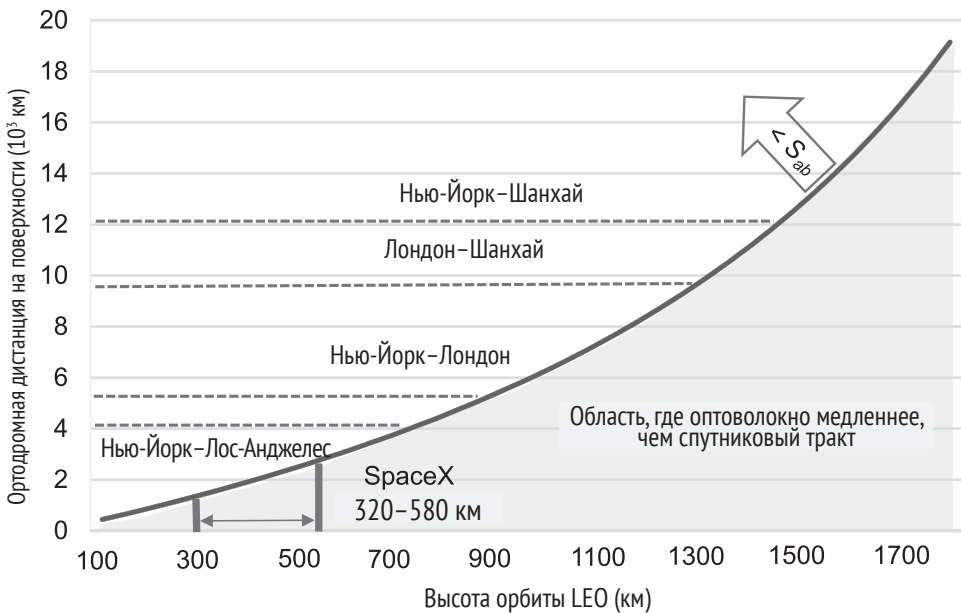


Рис. 32.4. Эффективная область, где LEO обеспечивает более низкую задержку распространения, при условии использования прямых трактов и отсутствия задержек оборудования на каждом шаге

В предыдущих рассуждениях мы рассмотрели сценарий, в котором спутник находится непосредственно над наземной станцией или оконечным устройством. Стоит отметить, что на самом деле общий путь может быть короче, чем показано на рисунке. В лучшем случае спутниковый тракт проходит по дуге большого круга между наземными станциями или оконечными устройствами, которые наблюдают оба спутника с умеренным наклоном, как показано на рис. 32.3б. Однако это зависит от относительного расположения наземных станций или оконечного оборудования и плотности группировки. Из-за движения спутников относительно наземных станций или терминалов эта минимальная задержка будет временным явлением. Пользователь, ве-

роятно, столкнется с изменением задержки распространения по мере движения спутников. Тем не менее этот пример дает хорошую оценку нижней границы задержки.

Случай с одним переходом

В предыдущем случае мы предположили, что спутниковая сеть ретранслирует трафик через несколько спутников. Однако в определенных сценариях пользовательский трафик может быть напрямую ретранслирован на другую наземную станцию или оконечное устройство через один переход, в отличие от ретрансляции через несколько спутников; напомним, что такую ретрансляцию иногда называют «коленчатым каналом». В случае созвездия LEO использование прямой ретрансляции может играть большую роль в организации передачи трафика на границе сети. Однако стоит отметить, что использование услуг через спутник GEO для такой конфигурации приведет к чрезмерной задержке.

В случае однопереходной передачи через спутник мы сравниваем задержку передачи сигнала между двумя пользователями в оптоволокне, уложенном на земле, и через две линии связи земля–космос. Для простоты рассуждений мы пренебрегаем влиянием кривизны Земли и предполагаем, что наземное оптоволокно напрямую соединяет двух пользователей.

Задержка распространения по космическому тракту будет непрерывно изменяться из-за движения спутника. Минимальная задержка будет иметь место, когда спутник находится на одинаковом удалении от пользователей. Мы предполагаем, что спутники будут переключаться, чтобы обеспечить оптимальную пропускную способность сети, и фактически переключение чаще всего происходит, когда спутник находится непосредственно над одним из пользователей. Заметим также, что как максимальная, так и минимальная задержки распространения также будут изменяться в зависимости от высоты спутника.

На рис. 32.5а показан пример как минимальной, так и максимальной задержки распространения для спутника LEO, работающего на высоте 395 км. Мы видим, что на небольших расстояниях наземный оптоволоконный тракт обеспечивает меньшую задержку. Обратите внимание, что задержка оптоволоконна растет быстрее, чем задержка космического тракта, из-за снижения скорости света в оптоволокне. Если говорить о минимальной задержке спутникового тракта, то мы видим, что на расстоянии «точка–точка» приблизительно 750 км задержка в спутниковом тракте лучше, чем в оптоволокне. На расстоянии 750 км все еще будут случаи, когда задержка спутника превышает задержку наземного волокна из-за движения спутника. Тем не менее мы можем наблюдать, что максимальная задержка спутникового тракта совпадает с задерж-

кой наземного оптоволокну, когда длина оптоволокну составляет приблизительно 1000 км.

Аналогичное соотношение можно увидеть для HAPS, которые работают на меньшей высоте. Например, как показано на рис. 32.5б, для HAPS, работающего на высоте 20 км, задержка распространения от точки до точки превышает задержку, возникающую в оптоволокну на расстояниях примерно 50 км. Но мы предполагаем, что длина оптоволокну – это расстояние прямой видимости. В действительности длина волокна будет больше, что, в свою очередь, увеличивает задержку в нем, в то время как спутник обеспечивает лучшее быстродействие по сравнению с оптоволокну.

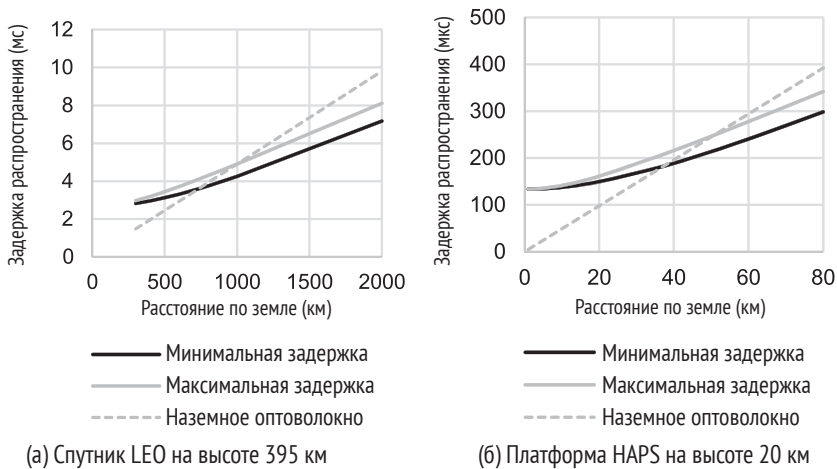


Рис. 32.5. Задержка распространения сигнала для спутников LEO и HAPS

Как для HAPS, так и для спутников соотношение, при котором минимальная и максимальная задержки превышают задержку в оптоволокну, является фиксированным. Это можно увидеть на рис. 32.6, который показывает области, используемые как для платформ HAPS, так и для спутников LEO. Область, обозначенная пунктирной линией, представляет собой переходную зону между обычным и орбитальным полетами, а HAPS обычно работает на высоте от 20 до 25 км. С другой стороны, хотя конкретная граница не определена, спутники LEO обычно работают в районе более 300 км.

Даже когда задержки спутника или HAPS превышают задержки в оптоволокну, преимущества HAPS и спутников заключаются в их повсеместном покрытии. Когда не требуется укладка оптоволоконного кабеля, услуги можно развертывать быстрее (и зачастую дешевле).

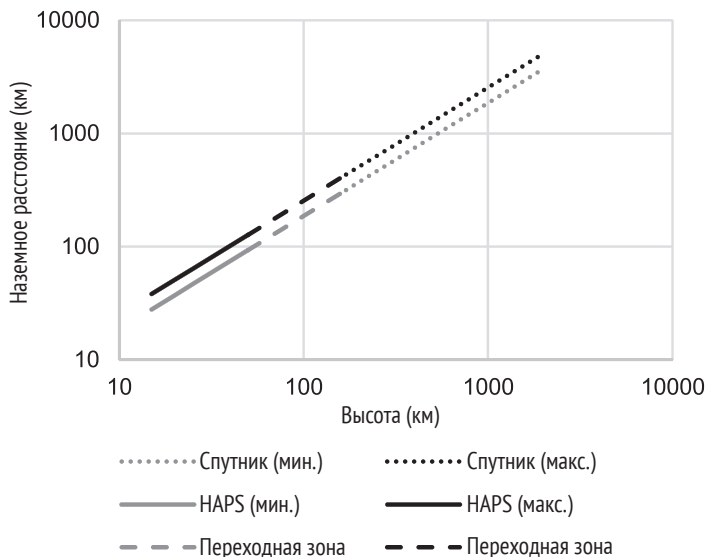


Рис. 32.6. Переходная зона, где задержка HAPS или спутника эквивалентна наземному оптоволоконному соединению «точка–точка»

Промежуточный итог

Анализ задержки для HAPS и спутников, включая случаи как с одним переходом, так и с несколькими, показывает, что передача сигнала через неназемную часть сети может обеспечить меньшую задержку, чем та, которая достигается с помощью нынешнего наземного оптоволокну. Во всех случаях достигаемое преимущество зависит от высоты платформы неназемной сети (non-terrestrial network, NTN) и расстояния между двумя точками на земле. В случае использования спутников будут некоторые различия в быстродействии тракта из-за движения спутника, что увеличит задержку в наихудшем случае, с которым сталкивается пользователь.

Что касается случая одиночного перехода, для типичных высот на низкоорбитальной орбите, составляющих приблизительно 400 км, характеристики спутникового тракта начинают равняться характеристикам наземного оптоволокну, когда расстояние между пользователями составляет приблизительно 750 км. На расстояниях около 1000 км спутник начинает работать лучше, чем прямая оптоволоконная линия между двумя пользователями. Это преимущество продолжает расти с увеличением расстояния между пользователями и будет ограничиваться только зоной радиопокрытия, которая зависит от высоты. Важно учитывать тот факт, что наземные пользователи редко имеют прямые оптоволоконные линии связи, и в большинстве случаев длина оптово-

локонного тракта превышает географическое расстояние между пользователями. Предполагается, что наземные соединения также будут включать в себя несколько узлов обслуживания между географически удаленными пользователями. Речь идет о широкополосных услугах, в которых пользователи могут столкнуться с несколькими элементами маршрутизации, каждый из которых увеличивает задержку. Их наличие зависит от сети, но обычно сокращает расстояние, на котором NTN обеспечивает лучшую задержку по сравнению с оптоволоконном.

В случае многозвенного спутникового тракта прирост быстродействия аналогичен предыдущим наблюдениям. Основное отличие заключается в том, что наземная часть сети ограничена только структурой группировки, а не зоной покрытия одного спутника. При условии правильной конструкции сетевого элемента космическая часть сети может быть спроектирована так, чтобы минимизировать задержку для каждого узла. Что касается индивидуальных сценариев пользователя, движение спутника приводит к некоторому изменению задержки. В случае с одним переходом изменение происходит из-за движения одного спутника относительно пользователей. Однако в случае со множеством переходов каждый пользователь получает доступ к разным спутникам, и каждый спутник может вносить вклад в дополнительную задержку на тракте. В конечном итоге это может привести к более широкому диапазону изменения задержки, но, поскольку используется несколько спутников, эквивалентная задержка наземного тракта будет значительной, и с помощью многозвенного спутникового тракта все же можно добиться снижения общей задержки.

Прирост быстродействия, обеспечиваемый NTN, обусловлен особенностями современной оптоволоконной технологии. Стоит отметить, что в отрасли проводятся исследования по разработке полых волокон, которые уменьшат задержку в наземном тракте. Даже несмотря на то, что мы все еще находимся на ранней стадии исследований, если будут разработаны оптоволоконные системы с использованием полых волокон, итоговый выигрыш в задержке, обеспечиваемый NTN, значительно уменьшится. Тем не менее это не рассматривается как препятствие для развертывания NTN, которая в основном направлена на достижение повсеместного покрытия. Еще стоит отметить, что наш анализ касается только задержки.

Хотя HAPS также демонстрируют уменьшение задержки и такие же характеристики, как и спутники LEO, они располагаются на гораздо меньшей высоте и обеспечивают лучшие характеристики в более узком диапазоне расстояний. Поскольку HAPS являются условно стационарными, они обеспечивают быстродействие, эквивалентное тому, которое обеспечивают оптоволоконные кабели для пользователей, расположен-

ных на расстоянии от 35 до 50 км друг от друга. Тот факт, что HAPS являются стационарными (даже с учетом незначительного перемещения), также означает фиксированную длину пути между двумя подключенными пользователями, которые не наблюдают значительных различий в задержке. Эта характеристика основана на предположении, что HAPS обеспечивает простую связь между двумя пользователями. Как и в описанных выше случаях со спутниками, мы не учли задержку обработки сигнала в оборудовании. Следовательно, если процесс оказания услуги может быть обеспечен непосредственно на платформе HAPS, без транзита данных в другое наземное местоположение, можно добиться дополнительного сокращения задержки.

Хотя и HAPS, и спутники могут обеспечить лучшую задержку в тракте, нам также необходимо учитывать другие факторы, которые влияют на задержку для каждой сети. К ним относятся угол места, плотность группировки, тип группировки и механизмы маршрутизации. Показатель задержки можно улучшить, если наземная станция (или пользователь) может видеть несколько спутников и впоследствии выбирать тот, который минимизирует задержку. Возможность выбора может зависеть и от типа используемого интерфейса.

32.3.2. Модели подключения

Ожидается, что услуги, предоставляемые спутниковыми сетями в системе 6G, обеспечат повсеместную возможность подключения для наземных пользователей. Безупречная интеграция с наземными сетями для предоставления услуг повсеместного доступа имеет первостепенное значение. Очевидно, в перечень услуг будут входить как прямые услуги (например, прямая широкополосная связь через выделенные спутниковые приемники), так и косвенные услуги, такие как транспортная сеть мобильной связи или другие услуги ретрансляции (например, услуги передачи данных для коммерческих авиарейсов). В сети 6G также потребуется возможность прямой связи пользователь–пользователь и машина–машина, в том числе между транспортными средствами (т. е. V2X). Для реализации разнообразных вариантов использования, упомянутых в главе 6, могут быть предусмотрены различные типы поддержки со стороны архитектуры.

Для обеспечения трехмерного покрытия наземным станциям 6G могут потребоваться новые способы интеграции со спутниковыми сетями. Например, спутниковые сети могут действовать как транзитные сети для расширения наземных беспроводных услуг, как показано на рис. 32.7. Спутниковые сети также могут обслуживать прямые соединения сети 6G, как видно на рис. 32.8. Это касается услуг как фиксированной, так и мобильной связи. На рис. 32.8 спутниковая сеть имеет прямое

соединение с пользовательским оборудованием. Спутниковая сеть передает как пользовательские данные, так и сигналы управления в одном туннеле. Не исключено, что обработка пользовательских данных может производиться напрямую в спутниковой сети как разновидность полезной нагрузки. В этом случае спутниковая сеть может предоставлять отдельные туннели.

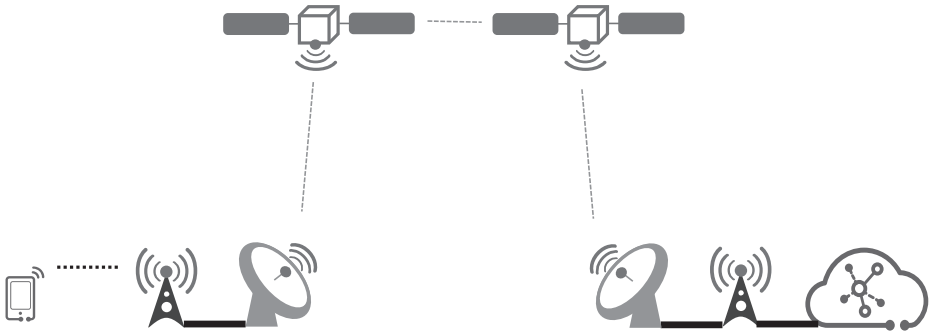


Рис. 32.7. Спутник, предоставляющий мобильную услугу (расширение охвата услуги)

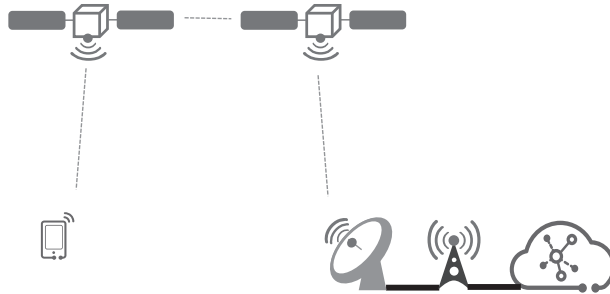


Рис. 32.8. Спутник, обслуживающий подвижную связь (прямое подключение)

Сценарий мобильной связи также может быть расширен для объединения нескольких пользователей с помощью устройства с одним спутниковым каналом. Поскольку каждый пользователь может передавать данные в разные места, объединяющее устройство должно участвовать в системе спутниковой маршрутизации для соответствующей обработки трафика (рис. 32.9).

Для простоты архитектура на рис. 32.9 показывает маршрутизацию трафика на земле. В некоторых случаях маршрутизация может быть выполнена стационарным летательным аппаратом (например, HAPS) с транзитом данных через спутниковую сеть. С точки зрения архитектуры спутниковой сети HAPS и наземные устройства агрегации пользовательского трафика идентичны.

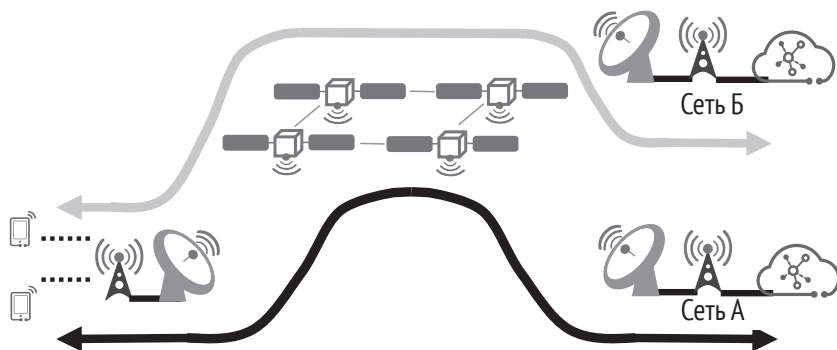


Рис. 32.9. Агрегированные услуги мобильной связи на земле

Трехмерное покрытие особенно важно для технологии IoT, когда она применяется в удаленных районах с ограниченным развертыванием наземных базовых станций или без них (например, датчики, используемые для мониторинга лесных пожаров и изменений климата). Спутниковые сети могут поддерживать подключение к интернету вещей или сети датчиков, используя модель, показанную на рис. 32.7. Поскольку сенсорные устройства IoT потребляют небольшое количество энергии и часто находятся в ждущем режиме, они могут потребовать специальной поддержки со стороны спутниковой сети, например специальных команд управления для включения устройств.

23.3.3. Маршрутизация в космосе

Хотя протоколы маршрутизации в космосе могут казаться похожими на протоколы, используемые в наземных сетях, различия являются критическими, и поэтому спутниковую сеть следует рассматривать как отдельный уровень маршрутизации в составе системы 6G. Аналогично иерархии внутри обычной многоуровневой сети, взаимоотношения между сетями тоже имеют характер «клиент–сервер». В этом случае спутник выступает как серверный уровень для данных, предоставляемых клиентом, наземной сетью или прямым пользователем. Способы передачи данных по сети различаются в зависимости от конкретной технологии, используемой в клиентской сети.

Спутниковая сеть призвана обеспечить возможность передачи данных в пакетной форме. Поскольку эта сеть будет в основном подключаться к сетям на основе IP, в архитектуре спутниковой сети клиентские сигналы будут передаваться с использованием модели туннеля, аналогичной той, которая применяется в наземных сетях передачи данных. Таким образом, пользовательские соединения будут представлять собой туннели между точками доступа пользователей к спутниковым сетям (или внутри самих спутников) и наземными шлюзами или другими

пользователями, в зависимости от ситуации. Поскольку при передаче через спутниковый канал невозможно напрямую реализовать маршрутизацию интернета или даже маршрутизацию поставщика услуг, требуется модель туннеля, которая инкапсулирует и передает пользовательские данные с адресами, подходящими для спутниковой сети.

В наземных сетях система маршрутизации направляет пакет в конечный пункт назначения, который обычно находится рядом с конкретным шлюзом. Маршрутизация к этому шлюзу может считаться оптимальной, если она основана на совпадении самого длинного префикса. Однако в определенных сценариях работы спутниковой сети (например, из-за проблем с питанием) может быть более целесообразным выполнить маршрутизацию к альтернативному промежуточному шлюзу, а затем окончательную маршрутизацию к месту назначения по наземной сети, как показано на рис. 32.10.

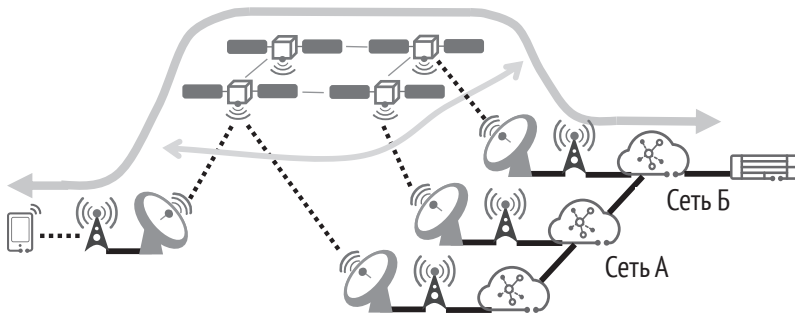


Рис. 32.10. Доставка данных на ближайшую наземную станцию (когда могут существовать несколько вариантов пути)

Система маршрутизации должна будет поддерживать вышеупомянутую функцию, и поэтому механизмы маршрутизации в космосе могут нуждаться в дальнейшем улучшении для интеграции с 6G. Например, в наземных сетях и существующих протоколах маршрутизации адреса источника и получателя основаны на подсетях и не включают в себя прямую информацию о географическом расположении устройств. Поскольку при этом не учитывается мобильность сетевых элементов, при разработке новой архитектуры следует учитывать поддержку мобильности с помощью ряда механизмов. Поэтому нам необходимо провести исследование того, как идентификатор и местоположение каждой конечной станции в сочетании с географической маршрутизацией могут решить проблему мобильности в этом контексте.

В спутниковых сетях 6G схема адресации может быть основана на географических адресах, т. е. адресами источника и получателя считаются соответствующие географические положения, а не IP-адреса. Если бы маршрут можно было определить по географическому расстоянию

между пунктом назначения и многими известными промежуточными спутниками, стало бы концептуально легче определить затраты на маршрутизацию. Кратчайший путь между двумя точками на земном шаре называется маршрутом большого круга, или ортодромией, как показано на рис. 32.11. Поэтому маршрутизация, основанная на минимизации этого расстояния, называется *ортодромной маршрутизацией*, и дальнейшие исследования по этому вопросу по-прежнему представляют интерес.

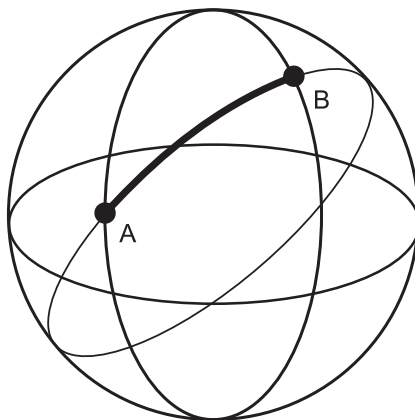


Рис. 32.11. Ортодромия относительно большого круга

32.3.4. Эксплуатация, администрирование и обслуживание

В настоящее время в наземных сетях не существует операционная модель, которая учитывает движение спутников над различными административными регионами или странами, поэтому может потребоваться ее адаптация с помощью различных систем ОА&М и маршрутизации.

В свете этого следует рассмотреть возможность создания системы связи для ОА&М интегрированной спутниковой сети 6G, что, в свою очередь, позволит наземным центрам управления управлять любым спутником в созвездии, в то время как эти спутники посылают обновления состояния и ответы. Иными словами, требования ОА&М не будут выполняться только в том случае, если спутник неисправен или находится вне пределов досягаемости наземной станции.

Это означает, что пакетная сеть также должна поддерживать пересылку пакетов данных ОА&М и их отправку в соответствии с бортовым программным обеспечением спутников. Мы можем достичь этого с помощью самых разных методов, среди которых наиболее распространенный метод для наземных пакетных сетей включает перенос пакетов

как уровня управления, так и уровня данных по общим каналам связи между маршрутизаторами. Для ОА&М в спутниковых сетях наиболее экономичный метод, вероятно, будет включать общий уровень управления маршрутизацией, но отдельные уровни данных и, возможно, упреждающий контролируемый подканал QoS на ISL. По причинам, обсуждавшимся ранее, невозможно использовать существующие уровни управления ОА&М, такие как OSPF/ISIS, или существующую адресацию ОА&М.

Более того, также важно определить интерфейс между спутниковой и наземной системами ОА&М, если они принадлежат разным бизнес-объектам, и это крайне важно для обеспечения целостной эксплуатации и управления услугами. Поскольку это может быть похоже на сценарии многопользовательской экосистемы, вы можете вернуться к главе 31, чтобы оценить соответствующий объем исследований.

Из-за сложной природы спутниковых группировок в эксплуатации и администрировании спутниковых сетей значимую положительную роль может сыграть машинное обучение.

32.4. Источники

- [1] *Space exploration holdings, LLC, SpaceX non-geostationary satellite system, attachment A, technical information to supplement schedule S.* <https://fcc.report/IBFS/SAT-MOD-20181108-00083/1569860>.
- [2] *Space exploration holdings, LLC, SpaceX non-geostationary satellite system, attachment A, technical information to supplement schedule S.* <https://fcc.report/IBFS/SAT-LOA-20200526-00055/2378671>.
- [3] *Kuiper systems LLC, technical appendix, application of Kuiper systems LLC for authority to launch and operate a non-geostationary satellite orbit system in Ka-band frequencies.* <https://fcc.report/IBFS/SAT-LOA-20190704-00057/1773885>.
- [4] 3GPP, *Study on new radio (nr) to support non-terrestrial networks*, 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.811, July 2020, version 15.3.0. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3234>.
- [5] 3GPP, *Solutions for NR to support non-terrestrial networks (ntn)*, 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.821, Dec. 2019, version 16.0.0. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3525>.

Резюме части VI

В части VI мы рассмотрели шесть ключевых аспектов архитектуры, которые вместе определяют структуру сети 6G. Архитектура 6G в полной мере охватит и интегрирует связь, вычисления и хранение. Система будущей мобильной связи не станет фокусироваться только на отправке данных в нужное место для обработки. Напротив, сама система мобильной связи станет оптимальным местом для обработки соответствующих данных с учетом ее повсеместного охвата и граничных вычислительных ресурсов. Однако для этого требуется глубокая конвергенция всех доступных ресурсов от всех заинтересованных сторон для создания управляемой открытой среды. Без этого не обойтись, если мы хотим создать универсальную и сверхскоростную инфраструктуру ICT.

По мере того как структурированная и многоуровневая сетевая архитектура становится более распределенной, в конечном итоге она трансформируется в новый тип архитектуры, которая использует периферийный интеллект. Он будет автономным и саморазвивающимся, но применимым к крупномасштабным сетям. Обеспечивая сверхвысокую скорость передачи данным, технология 6G будет не просто беспроводным соединением. В ее состав будут входить базовые станции и оконечные устройства, способные выполнять сканирование окружающей среды. Эти функции заложат прочную основу для подключения всех интеллектуальных устройств через интеллектуальную и надежную платформу. Таким образом, 6G станет основой вертикальной инфраструктуры, переопределив функциональные границы между оконечным устройством, каналом связи и облаком.

Кроме того, важно с самого начала проектирования 6G работать вместе с вертикальными отраслевыми заказчиками. Наш опыт развития отраслевых применений 5G показывает, что консорциумы 5GAA (автомобильная промышленность) и 5G-ACIA (обрабатывающая промышленность) играют важную роль в продвижении 5G и выходе на новые потенциальные рынки. Ранние отзывы потенциальных клиентов – не только с точки зрения бизнеса, но и с точки зрения технологий – имеют решающее значение для корректировки направления развития 5G. Уроки, извлеченные из 5G, пригодятся для развития отраслевых применений 6G. Секрет успеха в том, чтобы подходящие отраслевые платформы были сформированы в нужное время, с правильными партнерами и для правильных целей.

Часть VII



Подведение итогов и планы на будущее

Глава 33

Экосистема 6G и дорожная карта развития

33.1. Инициативы и экосистема 6G

В 2019 году мы стали свидетелями того, как связь 5G стала реальностью, когда началось глобальное развертывание системы. Хотя основные усилия направлены на достижение более высокого качества обслуживания и более зрелые применения в вертикальных отраслях за счет использования стандартизированной структуры NR, на горизонте восходит система 6G, стремящаяся решить гораздо больше проблем, с которыми сталкивается общество, и сулящая множество новых возможностей. Мы ожидаем, что в ближайшее десятилетие отрасль продолжит технологические исследования, согласование экосистем и глобальную стандартизацию в отношении 6G.

С 2018 года был запущен ряд инициатив в области исследований 6G. Например, промышленные и научные сообщества в Европе, Китае, Японии, Южной Корее и США занимаются определением типичных сценариев применения, ключевых возможностей и потенциальных технологий для беспроводной сети следующего поколения, которая, как ожидается, будет коммерциализирована в 2030-е гг.

33.1.1. Инициативы ITU-R

В течение последних 30 лет подразделение радиосвязи Международного союза электросвязи, известное как ITU-R, координировало усилия правительства и отрасли по разработке глобальной широкополосной системы IMT. Оно успешно руководило разработкой IMT-2000 (3G) [1], IMT-Advanced (4G) [2] и IMT-2020 (5G) [3], и теперь начинает новый цикл до 2030 г. и далее.

В феврале 2020 г. рабочая группа ITU-R WP 5D согласовала план работы по предварительному проекту Отчета ITU-R M.[IMT.FUTURE

TECHNOLOGY TRENDS] с упором на радиочастотные технологии. Ожидается, что этот документ будет завершен в июне 2022 года и обеспечит широкий технический обзор наземных систем IMT, которые появятся в 2030 году и в последующий период. Ожидается, что ITU-R также запустит исследование видения технологий IMT следующего поколения примерно в июне³⁰ 2021 г. и завершит его до Всемирной конференции радиосвязи в 2023 году (ВКР-23).

33.1.2. Региональные мероприятия

Европа: рамочная программа Horizon 2020

Рамочные программы (framework program, FP) – это проекты финансирования, созданные Европейским союзом (ЕС) и Европейской комиссией для поддержки и продвижения исследований; с 1984 года они предоставляют финансовую поддержку для реализации европейской политики в области исследований и инноваций. Например, в проекте WINNER [4] в FP6 и его последующих мероприятиях были определены основы для технологических инноваций 4G LTE, а проект METIS [5] в FP7 и его последующие мероприятия заложили прочную основу для успеха технологий 5G. Horizon 2020 – это восьмой рамочный проект на период с 2014 по 2020 г., сфокусированный на инновационной науке, промышленном лидерстве и решении социальных проблем. Что касается исследований, задача Horizon 2020 – стать новым исследовательским и инновационным проектом, определяющим новое партнерство между административными органами ЕС, промышленностью, малыми и средними предприятиями и исследовательскими институтами. (В 2021 г. стартовала новая программа под названием Horizon Europe, рассчитанная на период до 2027 г., с бюджетом 95 млрд евро. – *Прим. перев.*)

В мае 2020 г. NetWorld2020 – Европейская технологическая платформа коммуникационных сетей и услуг – опубликовала официальный документ [6] о Программе стратегических исследований и инноваций Европейского союза (SRIA) на 2021–2027 годы. Еврокомиссия использовала этот документ при подготовке рамочной программы Horizon Europe. Программа, созданная более чем 130 авторами из промышленных и научных организаций по всему миру, предлагает общие рекомендации и подробно обсуждает различные аспекты коммуникаций к 2030 г. Она охватывает следующие вопросы: глобальные мегатенденции; основы политики и KPI до 2030 года; дискуссии о человекоориентированных и вертикальных услугах; технологические аспекты системной архитектуры, периферийных вычислений, радиотехнологий, оптических сетей, безопасности, спутников и устройств; новые тенденции.

³⁰ На момент подготовки перевода была опубликована обзорная статья о работе ITU-R WP 5D в этом направлении: <https://techblog.comsoc.org/2021/06/15/development-of-imt-vision-for-2030-and-beyond-from-itu-r-wp-5d/>.

Китай: инициативы 6G

В 2019 г. была создана Китайская группа развития передовых технологий IMT-2030 [7]. Было сформировано несколько рабочих групп в соответствии с различными направлениями исследований 6G для мобильной связи до 2030 г. и далее, охватывающих требования и видение 6G, спектр, потенциальные технологии радиодоступа, сетевые технологии и т. д. Группа развития IMT-2030 объединяет всех, кто занимается изучением 6G, включая университеты, исследовательские институты, операторов, поставщиков и отраслевые вертикали.

В ноябре 2019 г. Министерство науки и технологий Китая совместно с пятью другими министерствами провело стартовое совещание, чтобы официально объявить о начале исследований и разработок в области 6G [8]. На встрече были созданы национальная рабочая группа по исследованиям и разработкам в области технологий 6G и общая группа экспертов – эти группы будут заниматься вместе с министерствами определением стратегии исследований и разработок в области 6G. Китайская ассоциация стандартов связи (CCSA) TC5 WG6 [9] отвечает за изучение передовых стандартов беспроводных технологий. В настоящее время она работает над отчетом «Исследование перспектив и требований к системам мобильной связи поколения B5G».

Япония: стратегия развития после 5G

В январе 2020 года Министерство внутренних дел и коммуникаций Японии (MIC) провело «Круглый стол по стратегии развития после 5G» [10]. Целью круглого стола является выработка комплексной стратегии удовлетворения потребностей и технического прогресса, ожидаемого после выхода за рамки 5G и перехода к следующему поколению связи.

MIC также объявил об общих чертах своей стратегии развития после 5G [11] в апреле 2020 года и приступило к общественному обсуждению. Двумя месяцами позже, в июне 2020 года, MIC опубликовало «Стратегию развития 5G – дорожную карту по направлению к 6G» [12].

Южная Корея: стратегия исследований и разработок для мобильной связи будущего

В августе 2020 года Министерство науки и ICT Южной Кореи (MSIT) опубликовало отчет под названием «Стратегия исследований и разработок для будущей мобильной связи 6G» [13]. В отчете говорится, что MSIT инвестирует 200 млрд вон (денежная единица Южной Кореи) в течение пяти лет, начиная с 2021 года, для поощрения технологических инноваций в 6G, а также для развития международного сотрудничества, укрепления отраслевой экосистемы и лидерства на мировом рынке 6G.

США: продвижение лидерства в сфере 6G

В марте 2019 года Федеральная комиссия по связи (FCC) разрешила использование диапазона частот от 95 ГГц до 3 ТГц для экспериментальных целей [14], открыв путь для исследований и испытаний беспроводных технологий, выходящих за рамки 5G. Диапазон спектра, разрешенный FCC, имеет значительный потенциал для создания будущей беспроводной связи с чрезвычайно широкой полосой пропускания.

В мае 2020 года Альянс решений для телекоммуникационной отрасли (ATIS) опубликовал призыв «Содействовать лидерству США на пути к 6G» [15]. ATIS заявил, что разработка 6G уже началась, и поделился своим видением сотрудничества между правительством, академическими кругами и отраслями промышленности, нацеленного на то, чтобы США оставались в авангарде технологического лидерства в течение следующего десятилетия.

33.1.3. Мнения промышленности и научных кругов

Заглядывая в будущее на следующие 10 лет, промышленные и научные сообщества активно обмениваются своими взглядами на конференциях IEEE и в многочисленных официальных документах.

В ноябре 2019 г. China Mobile опубликовала бюллетень «Представления и ожидания на период до 2030 и далее». В документе представлены начальные KPI для 6G и описаны пять основных характеристик сетей 6G, а именно «обслуживание по запросу, чрезвычайная простота, программируемая сеть, эндогенный интеллект и эндогенная безопасность» [16].

В январе 2020 г. NTT DOCOMO опубликовала бюллетень «Эволюция 5G и 6G». В документе обсуждается, как технологии мобильной связи будут развиваться в направлении 6G, а также описаны соответствующие требования, варианты использования и потенциальные технологии. Было высказано мнение, что 6G расширит программу «Третья волна: решение социальных проблем и создание ценностей, ориентированных на человека», предложенную консорциумом 5G в рамках 20-летнего цикла изменения услуг мобильной связи [17].

Компания Samsung опубликовала официальный доклад по 6G в июле 2020 г. Ее видение 6G – «привнести новые возможности гиперподключения во все уголки жизни». В докладе подробно обсуждаются соответствующие мегатенденции, услуги, требования и технологии-кандидаты, а также прогнозируются сроки завершения стандартизации и коммерциализации 6G [18].

Начиная с 2019 г. 6G – это горячая тема для дискуссий и обширных исследований в научном сообществе. Она занимает видное место в повестке дня почти всех соответствующих конференций, семинаров и практикумов, включая Международную конференцию по акусти-

ке, речи и обработке сигналов (ICASSP) 2020, Конференцию IEEE по беспроводной связи и сетевым технологиям (WCNC) 2020, Европейскую конференцию по сетям и связи (EuCNC) 2020, Международную конференцию IEEE по связи (ICC) 2020, Международный симпозиум IEEE по персональной, стационарной и мобильной радиосвязи (PIMRC) 2020, Конференцию IEEE по автомобильным технологиям (VTC-Fall) 2020 и Конференцию IEEE по глобальным коммуникациям (GLOBECOM) 2020. Например, на ICASSP в мае 2020 года тема 6G стала предметом широкого обсуждения в основных докладах [19], а также отраслевой панели «Проблема 5G/6G – гонка эффективных технологий и принципов проектирования» [20]. Кроме того, было опубликовано множество исследовательских работ по различным предметным областям, охватывающим прогнозы и требования к вариантам использования и потенциальным технологиям.

6Genesis – одна из флагманских инициатив национальной программы финансирования исследований Академии Финляндии, проводимой Университетом Оулу. Это восьмилетняя исследовательская инициатива по разработке, внедрению и тестированию ключевых технологий для 6G [21], начатая в 2018 г. Программа 6Genesis Flagship спонсирует саммит 6G Wireless Summit, ежегодную конференцию, проводимую Университетом Оулу. Первый саммит прошел 24–26 марта 2019 г. в Леви в финской Лапландии, где был выпущен первый официальный бюллетень по 6G, озаглавленный «Ключевые движущие силы и исследовательские задачи развития повсеместного беспроводного интеллекта 6G» [22]. Второй саммит прошел в формате виртуального мероприятия в феврале 2020 г. в связи со вспышкой COVID-19. Всего в рамках программы 6G Flagship опубликовано 11 новых официальных документов по 6G [23].

Еще один заслуживающий внимания документ – «Коммуникации в эпоху 6G», опубликованный исследователями из Nokia Bell Laboratories в марте 2020 г. [24]. В документе обсуждаются различные аспекты 6G, включая новые варианты использования, ключевые требования и показатели производительности, а также фундаментальные технологические преобразования, которые, скорее всего, будут определены для систем 6G. Далее мы приводим перечень других заслуживающих внимания обсуждений.

- [25] определяет ключевые движущие силы 6G, включая мультисенсорные приложения XR, подключенную робототехнику и автономные системы, беспроводное взаимодействие мозга и компьютера, а также технологии блокчейна и распределенного реестра. Здесь также представлены пять рекомендаций с точки зрения тенденций, проблем и связанных с ними исследований в 6G.

- [26] описывает 6G как многомерную и автономную сетевую архитектуру, объединяющую космические, воздушные, наземные и подводные сети для обеспечения полного покрытия и неограниченной беспроводной связи. Эта работа определяет несколько многообещающих технологий для экосистемы 6G, включая терагерцовую связь, сверхмассивный ММО, большие интеллектуальные поверхности и голографическое формирование луча, мультиплексирование на основе орбитального углового момента, связь на основе лазерного излучения и видимого света, совместное использование спектра на основе блокчейна, квантовую связь и вычисления, молекулярную связь и интернет нановещей.
- [27] представляет ряд базовых принципов для будущего развития с диаграммой, отображающей время, частотные диапазоны и использование космических ресурсов 6G. Авторы проанализировали различные результаты исследований, связанных с проектированием 6G, в том числе многополосные сверхбыстрые методы передачи, сверхгибкие интегрированные сетевые структуры и многорежимную многодоменную совместную передачу, а также машинное обучение и интеллектуальные технологии с использованием больших данных. Авторы также исследовали три открытых вопроса – электропитание, безопасность сети и проектирование оборудования – и предложили возможные решения.
- [28] суммирует основные проблемы и потенциал каждой входящей в состав 6G перспективной технологии и связывает каждую из них с конкретными вариантами использования. В нем определяется ряд вариантов использования 6G, включая VR, голографическое телеприсутствие (виртуальную телепортацию), электронное здравоохранение, повсеместную связь, индустрию 4.0 и робототехнику, а также беспилотные транспортные средства и соответствующие вспомогательные технологии. Анализ показывает, что для удовлетворения требований потребуются новые прорывные коммуникационные технологии, инновационные сетевые архитектуры и интеграция сетей и искусственного интеллекта.
- [29] предлагает понимание связи 6G с точки зрения человека. Авторы считают, что в дополнение к расширению 5G сеть 6G должна быть ориентирована на человека с точки зрения более высокой пропускной способности и меньшей задержки.

Есть также много содержательных статей, которые сосредоточены на сценариях применений 6G, таких как тактильный интернет, спектр выше 100 ГГц, связь с использованием искусственного интеллекта и свя-

занные с ними потенциальные технологии. Особого внимания заслуживают упомянутые ниже работы.

- [30] предполагает, что технологии субтерагерцового диапазона и VLC являются мощными средствами развития 6G.
- [31] описывает многообещающие исследования, которые включают технологию массивного MIMO, цифровое формирование диаграммы направленности и антенные решетки.
- [32] вводит концепцию уровня семантической эффективности (semantic-effectiveness, SE) как основной части будущих коммуникационных архитектур.
- [33] представляет собой всеобъемлющий обзор типов межмашинной связи в 6G, включая соответствующие показатели производительности и ключевые вспомогательные технологии.
- [34] предлагает анализ фундаментальных возможностей, проблем и подходов, связанных с созданием будущих систем беспроводной связи, сканирования и определения местоположения для диапазонов частот выше 100 ГГц.

Использование искусственного обучения и машинного интеллекта способно совершить революцию в системах беспроводной связи, поэтому является привлекательной темой для исследований в области 6G. В статье [35] представлено подробное руководство по использованию машинного обучения с особым акцентом на добавление разнообразных применений в будущие беспроводные сети. В ней обсуждается ряд новых применений, таких как беспилотные летательные аппараты (БПЛА), беспроводная виртуальная реальность, кеширование и вычисления на границе мобильной сети, интернет вещей и управление использованием спектра. Также обсуждаются перспективы использования машинного обучения для решения различных проблем беспроводной связи. В [36] авторы утверждают, что 6G выйдет за рамки мобильного интернета – 6G потребуются для поддержки всеохватывающих сервисов ИИ, от базовой сети до подключенных пользовательских устройств. Кроме того, они предсказывают, что ИИ будет играть решающую роль в проектировании и оптимизации архитектур, протоколов и операций 6G.

33.2. Дорожная карта до 2030 года и далее

В то время как 5G постоянно развивается и появляются новые версии, рабочая группа ITU-R WP 5D запустила исследовательский проект по изучению технологических тенденций систем IMT до 2030 г. и в последующий период. Предварительный график исследования предполагает, что оно будет завершено в июне 2022 г., и указывает, что еще

один проект, посвященный перспективам развития систем ИМТ, стартует в середине³¹ 2021 г. с ожидаемой датой завершения в середине 2023 года, до проведения WRC-23. В этом концептуальном докладе будут представлены рекомендации по вариантам использования, КРІ и полосам спектра 6G.

Ожидается, что после публикации данного доклада исследования технологий 6G будут продолжены, потому что углубленные исследования коммерческих перспектив и моделей, подчеркивающих уникальные конкурентные преимущества 6G, так же важны для всей экосистемы, как и дискуссии о базовых технологиях. Изучая опыт, накопленный в 5G, мы можем видеть, что разные вертикальные отрасли имеют очень разные требования, и даже одна вертикальная отрасль, охватывающая несколько регионов, предъявляет к экосистеме связи чрезвычайно разные требования. Понадобится время, чтобы полностью проанализировать, как эти требования повлияют на проектирование систем беспроводной связи, поэтому 3GPP может начать общее исследование новых радиоинтерфейсов и базовых сетей примерно в конце 2025 г. (R20), а в конце 2027 г. может приступить к формированию спецификаций. В настоящее время мы ожидаем, что первая версия стандарта 6G будет выпущена примерно в 2030 году, как показано на рис. 33.1.

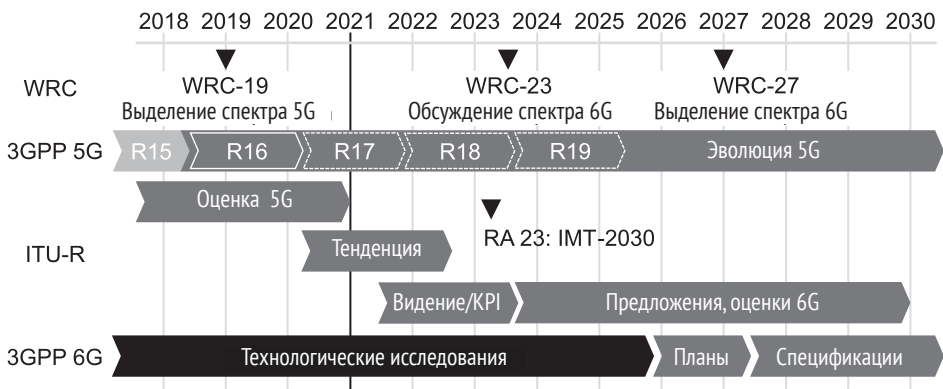


Рис. 33.1. Предполагаемый график исследований и стандартизации 6G в ITU-R и 3GPP

В течение следующего десятилетия мы вступим в эпоху, когда люди, вещи и искусственный интеллект будут связаны между собой. Система связи будет воспринимать физический мир путем анализа распространения собственных сигналов и формировать цифровое видение окружающей среды для использования искусственным интеллект. Предлагая расширенные услуги связи и возможности нативного искус-

³¹ Фактически стартовал в июле 2021 г. – Прим. перев.

ственного интеллекта, новое поколение беспроводной связи позволит обществу сделать следующий шаг вперед в решении проблем, с которыми мы сталкиваемся в настоящее время.

Чтобы сформировать представление о 6G и определить технологии, которые сделают 6G реальностью, необходимо, чтобы отраслевые, академические и системные участники работали вместе, как делали до этого в 4G и 5G. Стандартизация 6G во всем мире, несомненно, заложит основу для успешного развития на десятилетия вперед.

33.3. Источники

- [1] *Framework for the radio interface(s) and radio sub-system functionality for international mobile telecommunications-2000 (IMT-2000)*, Recommendation ITU-R M.1035, Mar. 1994.
- [2] *Framework and overall objectives of the future development of IMT 2000 and systems beyond IMT-2000*, Recommendation ITU-R M.1645, June 2003.
- [3] *IMT Vision – framework and overall objectives of the future development of IMT for 2020 and beyond*, Recommendation ITU-R M.2083-0, Sept. 2015.
- [4] WINNER, May. 2020. <http://www.ist-winner.org/>.
- [5] METIS, Sept. 2020. <https://metis2020.com/>.
- [6] *Smart networks in the context of NGI*, Strategic Research and Innovation Agenda 2021–27, Sept. 2020. [https://bscw.5g-ppp.eu/pub/bscw.cgi/d367342/Networld2020 %20SRIA %202020 %20Final %20Version %202.2 %20.pdf](https://bscw.5g-ppp.eu/pub/bscw.cgi/d367342/Networld2020%20SRIA%202020%20Final%20Version%202.2%20.pdf).
- [7] *IMT-2030*, Sept. 2020. http://www.cbdio.com/BigData/2020-01/20/content_6154253.htm.
- [8] *MST 6G*, Sept. 2020. <http://www.chinanews.com/sh/2019/11-07/9001283.shtml>.
- [9] *CCSA TC5 WG6*, Sept. 2020. [http://www.ccsa.org.cn/navMore?title=TC5 %3A %20 %E6 %97 %A0 %E7 %BA %BF %E9 %80 %9A %E4 %BF %A1&index=0&menuIndex=3](http://www.ccsa.org.cn/navMore?title=TC5%3A%20%E6%97%A0%E7%BA%BF%E9%80%9A%E4%BF%A1&index=0&menuIndex=3).
- [10] *Appeal for opinions on strategy outline of beyond 5G promotion*, Apr. 2020. https://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/pressrelease/2020/4/14_1.html.
- [11] *Beyond 5G promoting strategy (overview) (ver.1.0)*, Sept. 2020. https://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/presentation/pdf/200414_B5G_ENG_v01.pdf.
- [12] *Beyond 5G promotion strategy*, Sept. 2020. https://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/presentation/pdf/Beyond_5G_Promotion_Strategy-Roadmap_towards_6G.pdf.
- [13] *The research and development strategy for future mobile communications of 6G*, Sept. 2020. https://www.msit.go.kr/web/msipContents/contentsView.do?catelD=_policycom2&artId=3015098.

- [14] *FCC opens spectrum Horizons for new services & technologies*, Sept. 2020. <https://www.fcc.gov/document/fcc-opens-spectrum-horizons-news-services-technologies>.
- [15] *Promoting U.S. leadership on the path to 6G*, July 2020. <https://www.atis.org/wp-content/uploads/2020/07/Promoting-US-Leadership-on-Path-to-6G.pdf>.
- [16] *The vision and requirements for 2030+*, Nov. 2019. <https://www.shangyexinzhi.com/article/372578.html>.
- [17] NTT DOCOMO, inc. *White Paper 5G evolution and 6G*, Jan. 2020. https://www.nttdocomo.co.jp/english/binary/pdf/corporate/technology/whitepaper_6g/DOCOMO_6G_White_PaperEN_20200124.pdf.
- [18] Samsung, *6G vision: The next hyper-connected experience for ALL*, July [https://cdn.codeground.org/nsr/downloads/researchareas/6G %20Vision.pdf](https://cdn.codeground.org/nsr/downloads/researchareas/6G%20Vision.pdf).
- [19] *6G: From connected everything to connected intelligence*, May 2020. <https://2020.ieeeicassp.org/industry-program/keynotes/6g-from-connectedeverything-to-connected-intelligence/>.
- [20] *The B5G/6G Challenge – a race of enabling technologies and design principles*, May 2020. <https://2020.ieeeicassp.org/industry-program/panels/the-b5g-6g-challenge-a-race-of-enabling-technologies-and-design-principles/>.
- [21] M. Katz, M. Matinmikko-Blue, and M. Latva-Aho, *6Genesis flagship program: Building the bridges towards 6G-enabled wireless smart society and ecosystem*, in Proc. 2018 IEEE 10th Latin-American Conference on Communications (LATINCOM). IEEE, 2018, pp. 1–9.
- [22] M. Latva-Aho and K. Leppänen, *Key drivers and research challenges for 6G ubiquitous wireless intelligence (white paper)*, 6G Flagship, Oulu, Finland, 2019.
- [23] *11 new 6G research white papers published*, Sept. 2020. <https://www.6gchannel.com/>.
- [24] H. Viswanathan and P. E. Mogensen, *Communications in the 6G era*, IEEE Access, vol. 8, pp. 57 063–57 074, 2020.
- [25] W. Saad, M. Bennis, and M. Chen, *A vision of 6G wireless systems: Applications, trends, technologies, and open research problems*, IEEE Network, vol. 34, no. 3, pp. 134–142, 2019.
- [26] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, and P. Fan, *6G wireless networks: Vision, requirements, architecture, and key technologies*, IEEE Vehicular Technology Magazine, vol. 14, no. 3, pp. 28–41, 2019.
- [27] P. Yang, Y. Xiao, M. Xiao, and S. Li, *6G wireless communications: Vision and potential techniques*, IEEE Network, vol. 33, no. 4, pp. 70–75, 2019.
- [28] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, *Toward 6G networks: Use cases and technologies*, IEEE Communications Magazine, vol. 58, no. 3, pp. 55–61, 2020.

- [29] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, *From a human-centric perspective: What might 6G be?* arXiv preprint arXiv:1906.00741, 2019.
- [30] E. C. Strinati, S. Barbarossa, J. L. Gonzalez-Jimenez, D. Ktenas, N. Cassiau, L. Maret, and C. Dehos, *6G, the next frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication*, IEEE Vehicular Technology Magazine, vol. 14, no. 3, pp. 42–50, 2019.
- [31] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, *Massive MIMO is a reality; what is next?: Five promising research directions for antenna arrays*, Digital Signal Processing, vol. 94, pp. 3–20, 2019.
- [32] P. Popovski, O. Simeone, F. Boccardi, D. Gündüz, and O. Sahin, *Semantic-effectiveness filtering and control for post-5G wireless connectivity*, Journal of the Indian Institute of Science, vol. 100, no. 2, pp. 435–443, 2020.
- [33] N. H. Mahmood, H. Alves, O. A. López, M. Shehab, D. P. M. Osorio, and M. Latvaaho, *Six key enablers for machine type communication in 6G*, arXiv preprint arXiv:1903.05406, 2019.
- [34] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, *Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond*, IEEE Access, vol. 7, pp. 78 729–78 757, 2019.
- [35] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, *Artificial neural networks-based machine learning for wireless networks: A tutorial*, IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3039–3071, 2019.
- [36] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, *The roadmap to 6G: AI empowered wireless networks*, IEEE Communications Magazine, vol. 57, no. 8, pp. 84–90, 2019.

Предметный указатель

А

Автоэнкодер [146](#), [150](#)
 эволюционный [150](#)

Алгоритм
 динамического программирования
 Витерби [159](#)
 максимизации ожидания [159](#)

Антенна
 четырёхзаходная спиральная [169](#)

Апикальный импульс [100](#)

Аппроксиматор функций [143](#)

Б

Байесовская сеть [157](#)
 динамическая [158](#)

Благонадежность [541](#)

Блокчейн [552](#)

Бюджет канала [63](#)

В

Вертикальная отрасль [107](#)

Взаимная информация [148](#)

Виртуальный город [96](#)

Волны Ценнека [168](#)

Высотная платформа [122](#)

Вычисления
 граничные [293](#)
 квантовые [284](#)
 нейроморфные [283](#)
Вычислительный электромагнетизм [170](#)

Г

Глубокая нейросеть [130](#)
Глубокая периферийная сеть [45](#)
Голографический MIMO [469](#)
Голография
 метаповерхностная [250](#)

Граничные вычисления [45](#)

Граничный узел [518](#)

Графен [238](#)

Д

Динамический байесовский вывод [156](#)

Дисплей
 голографический [84](#)
 светового поля [84](#)

Добросовестность [64](#)

Долговременная краткосрочная память [144](#)

Ж

Жидкие кристаллы [238](#)

З

Зеленая радиосеть [56](#)

Зрение
 бинокулярное [77](#)
 монокулярное [77](#)

И

Извлечение признаков [102](#)
Излучательное ближнее поле [463](#)

Индустрия 4.0 [106](#)

Интеллект
 всеобщий [17](#)
 подключенный [17](#)
Интеллектуальная
 поверхность [140](#), [224](#)
 среда [167](#)

Информационная метаповерхность [168](#)

Информационная структура [187](#)

Искусственная нейронная сеть [142](#)
 генеративно-состязательная [151](#)
 глубокая [143](#)

мелкая 142
 рекуррентная 144
 сверточная 144
 скрытый слой 144

Истинный цвет 77

К

Канал

боковой. См. Канал одноранговый прямой
 одноранговый прямой 489

Канальное кодирование 399

Квантовый бит (кубит) 284

Киборг 113

Кобот 112

Л

Лаг движение–фотон 73

Локализация

активная 365
 относительная 91
 пассивная 365
 семантическая 92

Локальный граничный узел 135

М

Марковский процесс принятия решений 147

Материалы с фазовым переходом 238

Машинное обучение 130

Мемристор 283

Метаповерхность 249

Многопользовательское взаимодействие 571

Многостороннее участие 571

Множественный доступ 427

неортогональный 429
 ортогональный 429

Модель многостороннего доверия 550

Молекулярное поглощение 228

Н

Новое радио (NR) 312

О

Обработка естественного языка 144

Обучение

без учителя 146
 многоагентное с подкреплением 155
 с подкреплением 146
 с учителем 145
 федеративное 152

Орбитальный угловой момент 170

Ориентированная на пользователя сеть 524

Ортодромная маршрутизация 606

Осведомленность о ситуации 472

П

Плазмонный материал 240

Плотность подключения в области 62

Поверхность

интеллектуальная большая 168
 интеллектуальная отражающая 167
 интеллектуальная реконфигурируемая 167

Поляризационное мультиплексирование 346

Полярный код 404

Приближенная передача сообщений 441

Принцип композиционности 184

Пропускная способность области 62

Пространственный нестационарный
 канал 224

Просьюмер 47

Прямой детектор 262

Прямой суперканал 490

Р

Распределение ресурсов 317

Распределенный консенсус 575

Распределенный реестр 552

Расстояние Рэлея 467

Расщепление луча 270

Резонатор с разъемным кольцом 252

Реконфигурируемая интеллектуальная
 поверхность 461

С

Сверхбольшая апертурная решетка 169

Сверхмассивное подключение 445

Связь

в видимом свете 49

дуплекс с временным разделением 208

дуплекс с частотным разделением 209

на поверхностных волнах 168

с обратным рассеянием 364

терагерцовая (ТГц) 49

Сетевое кодирование

аналоговое 416

линейное 416

Сигнатура NOMA 436

Система с множеством входов и множеством выходов 66

Сканирование 351

в ближнем поле 252

визуальное 351

радиочастотное 351

Скрытая марковская модель 158

Смарт-контракт 575

Совместное кодирование источника и канала 410

Спектральное распознавание 100

Спектроскопия

во временной области 269

непрерывной волны 269

Спутниковая группировка

дельта-паттерн Уокера 591

орбитальный шов 592

полярная 591

смежность 592

Станция на высотной платформе 44

Степени свободы 165

Сфера Блоха 284

Т

Теория

информационных узких мест 148

разреженной выборки 191

сверхвысокого разрешения 188

семантической коммуникации 183

электромагнитной информации 165

Технология распределенного реестра 573

Трансмиссия 140

У

Узел радиовычислительный 519

Уровень

пользователя 526

управления 526

Ф

Федеративное обучение 52

Формула гаверсинуса 595

Фотонный кристалл 239

Ц

Цифровой двойник 41, 536

Ч

Частота дискретизации Найквиста 189

Э

Энергия

передачи 194

статическая 194

Эффект цветовых полос 77

Я

Якорное устройство 296

Книги издательства «ДМК ПРЕСС» можно купить оптом и в розницу в книготорговой компании «Галактика» (представляет интересы издательств «ДМК ПРЕСС», «СОЛОН ПРЕСС», «КТК Галактика»).

Адрес: г. Москва, пр. Андропова, 38;

Тел.: +7(499) 782-38-89. Электронная почта: **books@alians-kniga.ru**.

При оформлении заказа следует указать адрес (полностью), по которому должны быть высланы книги; фамилию, имя и отчество получателя.

Желательно также указать свой телефон и электронный адрес.

Эти книги вы можете заказать и в интернет-магазине: **www.a-planeta.ru**.

Вэнь Тонг, Пейин Чжу
(редакторы)

Сети 6G. Путь от 5G к 6G глазами разработчиков

От подключенных людей и вещей
к подключенному интеллекту

Главный редактор *Мовчан Д. А.*
dmkpress@gmail.com

Перевод с английского *Яценков В. С.*

Корректор *Синяева Г. И.*

Верстка *Паранская Н. В.*

Дизайн обложки *Мовчан А. Г.*

Формат 70×100¹/₁₆.

Усл. печ. л. 50.7. Печать цифровая.

Тираж 200 экз.

Веб-сайт издательства: **www.dmkpress.com**



От коммуникаций между людьми и интернета вещей к повсеместно подключенному интеллекту

Первая книга о беспроводной связи 6G дает развернутое представление о будущем 6G – эпохе всеобщего интеллекта, которая будет определять облик нового мира.

Среди грядущих инноваций:

- ориентированная на человека связь с высокой реалистичностью общения;
- сканирование, локализация и визуализация;
- подключенное машинное обучение и сетевой искусственный интеллект (ИИ);
- индустрия 4.0 и выше, с подключенным интеллектом, умными городами и технологиями;
- спутниковая мегагруппировка, обеспечивающая трехмерное покрытие беспроводной сетью всей земной поверхности.

В книге рассматриваются новые радиоинтерфейсы и сетевые технологии, интеграция услуг сканирования и связи, а также наземных и неназемных сетей. Также обсуждаются новые сетевые архитектуры, обеспечивающие функционирование сетевого ИИ, ориентированные на пользователя сети, и нативная защита данных.

Издание адресовано исследователям, инженерам, разработчикам, менеджерам, работающим в области беспроводной связи, и всем, кто интересуется технологиями будущего.

Вэнь Тонг – технический директор Huawei Wireless и главный научный сотрудник Huawei 5G. Доктор Тонг также является научным сотрудником IEEE и членом Канадской инженерной академии. Он был удостоен награды IEEE Communications Society «Выдающийся лидер отрасли» и медали Р. А. Фессендена.

Пейин Чжу – старший вице-президент по исследованиям беспроводной связи в Huawei и научный сотрудник Huawei. Она также является научным сотрудником IEEE и членом Канадской инженерной академии.

Интернет-магазин:
www.dmkpress.com

Оптовая продажа:
КТК «Галактика»
books@aliens-kniga.ru

 **CAMBRIDGE**
UNIVERSITY PRESS


ИЗДАТЕЛЬСТВО
www.dmk.rf

ISBN 978-5-97060-995-8



9 785970 609958 >