面向大规模物联网的随机接入：现状、挑战与机遇

范平志，李里，陈欢，程高峰，杨林杰，汤小波
（西南交通大学信息科学与技术学院，四川 成都 611756）

摘 要：在传统通信系统中，随机接入是终端与网络之间建立无线链路的必经过程，只有在随机接入完成之后，终端与网络之间才能正常进行数据传输。聚焦大规模节点物联网，首先阐述了大规模物联网随机接入的基本特征，分析了传统授权随机接入（RA）方法对于大规模物联网应用的局限性；然后从系统模型、协议框架、算法流程、系统性能等多个角度，讨论免授权 RA、半免授权 RA、无用户标识 RA 三大类新型大规模随机接入方法，分析各类方法优缺点，并指出面临的挑战与潜在的机遇；最后对大规模物联网随机接入涉及的关键技术与共性问题进行了总结。

关键词：大规模物联网；随机接入；免授权；无用户标识 RA；非正交多址
中图分类号：TN929.5
文献标识码：A
DOI: 10.11959/j.issn.1000−436x.2021098

Random access for massive Internet of things: current status, challenges and opportunities

FAN Pingzhi, LI Li, CHEN Huan, CHENG Gaofeng, YANG Linjie, TANG Xiaobo
School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China

Abstract: In conventional communication systems, the random access (RA) is a necessary procedure for a user equipment (UE) to establish wireless connection to the network before actual data transmission. Focusing on the massive IoT with large number of nodes, the fundamental characteristics and limitations of conventional grant-based RA were firstly discussed. Then, the grant-free RA, semi-grant-free RA and unsourced RA for massive IoT were addressed in details from multiple perspectives, including their system models, protocol frameworks, algorithm procedures, system performance, related pros and cons, as well as critical challenges and potential opportunities. Finally, some highlights on the related key techniques and common issues in random access for massive IoT were concluded.

Keywords: massive Internet of things, random access, grant-free, unsourced random access, non-orthogonal multiple access

1 引言

“万物互联”已成为通信行业的重要发展方向，未来大部分终端设备都将连接网络，逐渐聚合为规模庞大、高度智能的全球物联网。2019 年 12 月 16 日，全球移动通信系统协会预测：全球物联网设备将从 2018 年的 103 亿个增长到 2025 年的 250 亿个，其中企业物联网连接数将在 2024 年超过消费者物联网连接数，并在 2025 年达到 133 亿个[1]。因此，物联网不仅将走进人们生活的方方面面，也将深度融入各个垂直行业。人们有理由相信，在不远的将来，物联网规模将远超移动互联网，成为下一个万亿元

收稿日期：2021−01−18；**修回日期**：2021−04−08
基金项目：国家重点研发计划基金资助项目（No.2018YFB1801104）；国家自然科学基金资助项目（No.62071394）；四川省国际科技创新合作基金资助项目（No.2020YFH0011）
Foundation Items: The National Key Research and Development Program of China (No.2018YFB1801104), The National Natural Science Foundation of China (No.62071394), Sichuan Science and Technology Program (No.2020YFH0011)

规模的产业。5G 时代的来临成为物联网发展的催化剂[2]，而已开启研究的 6G 系统[3]更有可能成为全球大规模智能物联网的承载者。为了支撑未来大规模物联网应用，国际标准化组织 3GPP（3rd Generation Partnership Project）早在 Release-13 版本就将大规模机器类通信（mMTC, massive machine type communication）作为其主要应用场景之一。而随着 2020 年 5G 商用元年的到来，国际电信联盟（IMT, International Telecommunication Union）组织成员华为海思、诺基亚上海贝尔实验室、高通、爱立信、中兴、联想等在 Release-15 及后续版本制定中，针对 mMTC 场景进行了大量仿真实测[4-6]，对 mMTC 新型空口接入方案的可支持网络密度、覆盖能力、时延属性、设备电池自持寿命、吞吐率等关键指标进行了大量分析论证。总体认为，无论是业界已经商用化布置的窄带物联网（NB-IoT, narrow band Internet of things）、LoRa（long range radio）、增强机器类通信（eMTC, enhanced MTC）技术[7]，还是新近发布的 Release-15 版本都不足以支撑未来大规模物联网业务。

大规模物联网应用主要面向智慧城市、智慧交通、智能电网、现代物流、现代农业和环境检测等以传感器和数据采集为目标的场景，终端设备包括智能水电气表、智能垃圾桶、物流跟踪器、自动贩卖机、集装箱标签、农业/环境/工业中的各种传感设备等，其业务具有以下显著区别于传统蜂窝网络人与人通信的基本特征。

1) 终端设备数量庞大且覆盖广泛。根据 3GPP 规划，mMTC 场景应满足每平方千米 100 万的连接密度[8]，未来甚至要求满足每平方千米 1 000 万的连接密度[9]。

2) 数据分组短。国际电信联盟认为 mMTC 场景下数据分组通常较短，一般不超过 1 000 个字节。

3) 未来物联网通信业务呈现较强偶发性和稀疏性[10]。

4) 时延容忍度高。

5) 终端设备小型化、低成本、低功耗。3GPP 要求终端设备的电池寿命应达到 10 年以上[10]。

然而，目前尚不存在一种通信技术能够解决上述大规模物联网的各类新需求、新挑战。因此，有必要从网络协议、信号处理方法、硬件设备等多个层次出发，综合应用网络切片、大规模多天线、智能反射面等大量新兴通信前沿技术，以全面支撑未来车联网、智能家居、数字农业等各类物联网场景，如图 1 所示。

特别地，对于大规模物联网高达千万的空闲用户、上万的同时活跃用户、短分组数据发送、低能耗限制的业务特征，传统随机多址接入和编码技术面临巨大的挑战。在传统通信系统中，随机接入是终端与网络之间建立无线链路的必经过程，只有在随机接入完成之后，终端与网络之间才能正常进行数据传输。然而，如文献[11]所述，当小区内用户数超过 3 万个时，随机接入性能开始明显退化。即使采取接入类禁止（ACB，access class barring）[12]、自适应接入类禁止（A-ACB，adaptive ACB）[13]、扩展接入禁止（EAB，extended access barring）[12-14]等过载控制技术，随着小区内支持用户个数的持续增加，整个系统的接入成功概率和前导序列碰撞概率也会大幅恶化。其根本原因是已有基于极低控制开销、长码，以及与时延无关的信息理论极限和编码技术无法适用于大规模物联网应用。经典的 K 用户多址接入（K-user multiple access）模型无法匹配大规模物联网场景下的随机接入特性，“碰撞概率波动极大，忙时忙死，闲时闲死；有效负载太低，

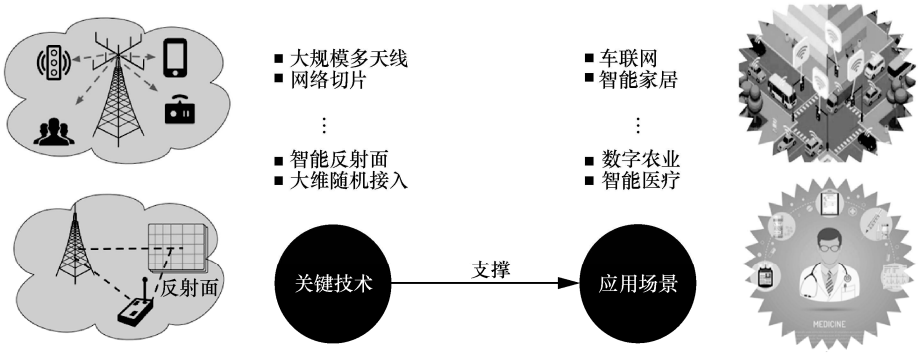

图 1　大规模物联网关键技术和典型应用场景

随机接入开销极大”，已有理论碰撞模型难以刻画实际应用场景中的信号混叠特征。因此，亟须突破传统长期演进技术（LTE, long term evolution）框架下基于授权的随机接入（GB-RA, grant-based random access）模式，设计新型大维随机接入方法。

首先，改良基于授权的随机接入方法仍在持续进行，相继出现了自适应 EAB[14]、多功能 ACB[15]、基于机器学习的 ACB[16]等新型过载控制方法。另一方面，基于压缩感知技术的免授权随机接入（GF-RA, grant-free random access）方法[17-18]实现了对传统随机接入机制的一次突破，使网络用户容量、接入能效的数量级提升成为可能。同时，非正交多址（NOMA, non-orthogonal multiple access）技术迅速发展，出现了功率域 NOMA[19]、码域 NOMA[20-21]、角度域 NOMA[22]等多种技术。如何利用非正交技术支撑大规模随机多址接入也是该领域的重要研究方向，如基于功率域 NOMA[23]、SCMA（sparse code multiple access）、PDMA（pattern devision multiple access）[24]、MUSA（multi-user shared access）[25]等技术的一系列新型随机多址接入方法。这些随机接入方法虽不需要基站授权，但相较于免授权随机接入又有一定的额外控制开销，因此，可将它们称为半免授权随机接入方法。最后，基于不区别传输内容是用户身份还是有效负载、全网用户使用统一码本等核心思想，文献[26]提出了随机多址码概念，分析了大规模随机多址码能效与谱效间的理论约束关系，从信息论角度揭示了大规模随机接入系统性能上界（即 Polyanskiy 界），并清晰表明无论是网络理论体系的 ALOHA 协议、时隙化 ALOHA（slotted ALOHA）协议等技术，还是传统信息论与编码体系的码分多址（CDMA, code division multiple access）、NOMA 等技术，与 Polyanskiy 界都存在相当距离，还有明显提升空间，为大规模随机接入指出了新的研究方向。

针对上述大规模物联网随机接入应用需求，结合笔者部分研究工作，本文将主要分析讨论如下内容。

1) 阐明大规模物联网随机接入巨量用户共存、活跃用户稀疏化、小分组传输的典型特征，分析传统基于授权的随机接入机制无法满足大规模物联网应用需求的基本事实和根本原因，揭示设计新型随机接入方法的必要性。

2) 从研究动机、系统模型、协议框架、算法流程、系统性能等多个角度分析免授权 RA、半免授权 RA、无用户标识 RA 三大类新型大规模随机接入技术，分析各类方法优缺点，并指出这些方法面临的挑战与潜在的机遇。

3) 对大规模物联网随机接入涉及的关键技术与共性问题进行总结。揭示各主要技术方向上亟须解决的开放性问题，如算法复杂度过高难题，信号同步难题，缺乏理论界难题，共存用户个数、活跃用户检测精度、接入开销三者间互相制约难题等。

2　基于授权的随机接入

LTE 标准中 GB-RA 方法主要面向人与人（H2H, human to human）通信场景。因此，GB-RA 方法适用于活跃用户数较少、传输数据量较大、传输速率要求较高的业务。但在未来大规模物联网随机接入场景中，由于前导序列资源池规模有限，如 LTE 中用于随机接入的正交前导序列仅有 64 个[27]，势必造成严重的接入用户碰撞问题，进而使各类接入指标严重退化，用户体验急剧下降。

本节首先讨论传统 GB-RA 过程，并具体分析其在大规模物联网应用中的局限性。然后，对几种 GB-RA 重要改进方案进行介绍。最后，根据其性能表现，指出设计全新随机接入方法的必要性。

2.1　GB-RA 流程

在传统 LTE 标准中，GB-RA 流程由用户（UE）与基站间的四步握手过程组成，用户与基站间的信令交互过程如图 2 所示[27]。

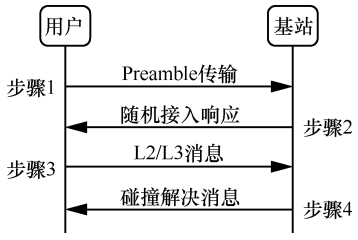

图 2　用户与基站间的信令交互过程

步骤 1　各活跃用户从 ZC（Zadoff-Chu）序列循环移位生成的序列资源池中随机选择一个前导序列向基站发送。基站检测接收到的前导序列，并通过该检测过程估计各活跃用户时间提前量（TA, timing advance），随后为每个检测到的前导序列分配上行资源授权（UG, uplink grant）和临时身份标识。

步骤 2　基站将每个前导序列对应的 TA 值和

UG 信息打包在随机接入响应消息中，并向所有用户广播。

步骤 3 用户对收到的随机接入响应消息进行解析得到自身上行资源授权信息，随后在基站指定的上行资源上发送 L2/L3 消息。其中，L2/L3 消息由用户连接请求消息和基站分配的临时身份标识等构成。

步骤 4 若基站正确解调某用户的 L2/L3 消息，则向该用户反馈碰撞解决确认信息。反之，基站将不会向该用户反馈确认消息，用户由此判断此次接入失败。

在步骤 1 中，若有 2 个及以上的用户选中同一种前导序列进行发送，则会出现用户碰撞事件，最终导致接入失败。因此，有限的前导序列资源池规模严重制约了传统 LTE 标准中 GB-RA 方法的接入能力。

此外，物理层上行共享信道（PUSCH, physical uplink shared channel）作为基站侧可为用户上行传输分配的资源，其有限的数量同样制约了 GB-RA 方法的应用前景。在大规模物联网随机接入场景中，无碰撞的用户数大于系统可分配 PUSCH 资源数时，某些用户会由于无法获得上行授权而导致接入失败。这一现象不仅浪费了物联网用户接入能量，也会降低前导序列资源池的利用效率。这是 GB-RA 方法不适于大规模物联网的另一重要原因。

2.2 GB-RA 的各类改进方案

基于上述分析，前导序列资源池规模过小和 PUSCH 资源数有限是导致 GB-RA 方法不适用于大规模物联网随机接入的 2 个重要因素。针对这 2 种资源的稀缺性问题，业界提出了许多针对 GB-RA 的增强方案，主要改进思路可分为过载控制设计（节流）、前导序列资源池扩展（开源）和上行授权资源碰撞解决（善用）3 个技术方向。

1) 过载控制设计

ACB 机制[12]是最经典的用户过载控制方法。基站广播 ACB 因子，活跃用户将本地产生的随机数与侦听到的 ACB 因子进行比较，根据自身设备类型和比较结果决定是否发起随机接入过程，若不能发起随机接入，用户进行随机回退。ACB 机制提出后，出现了大量基于 ACB 机制的改良，如自适应/动态 ACB（A/D-ACB, adaptive/dynamic ACB）[13, 28-29]、EAB[14, 30]、基于服务质量（QoS, quality of service）需求的 ACB[31-32]、结合机器学习的 ACB[16, 33-35]、上行授权资源受限下的 ACB[36]、基于时间提前量的 ACB[37-38]等方案。图 3 总结了基于 ACB 机制的各类代表性过载控制方案，相关仿真结果可见 2.3 节。

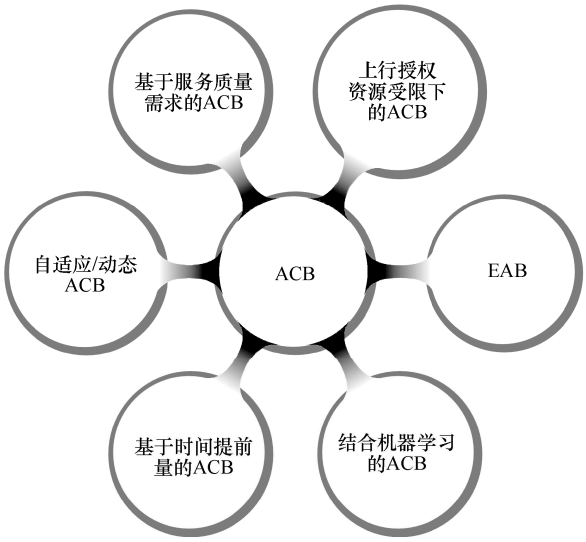

图 3 基于 ACB 机制的过载控制方案

2) 前导序列资源池扩展

文献[39]将 m 序列进行循环多普勒时延移位，以在序列长度受限下增加序列个数。文献[40-41]提出了 2 种后向兼容的前导序列资源池扩展方法，将正交 ZC 序列分别与 m 序列或 all-top 序列[42]进行点乘，从而构造出大量具有低相关区属性的前导序列。表 1 比较了文献[41]中所提序列集与传统 ZC 序列集的各项指标，其中 L 为序列长度，N_{CS} 为 ZC 序列循环移位值，ZCm 表示 ZC 序列与 m 序列按位相乘后得到的低相关序列，ZCa 表示 ZC 序列与 all-top 序列按位相乘后得到的低相关序列。对比表 1 的最后一列可知，通过将原 ZC 序列间的零相关区弱化为低相关区，可显著增大前导序列个数。前导序列资源池扩展后，低相关区序列集可提供的物理随机接入信道（PRACH, physical random access channel）数量比原 ZC 序列集提升了 $L+1$ 倍。

改变活跃用户前导序列发送结构同样能实现前导序列资源池的扩展。3GPP 在文献[43-44]等技术规范中建议使用多阶段前导序列传输策略来克服其数量不足。

图 4 对比了单阶段前导序列传输方式与两阶段前导序列传输方式，其中 CP（cyclic prefix）表示循环前缀，GT（guard time）表示保护时间。在单阶段方式中，前导序列需为一个较长的 ZC 序列；在多阶段方式中，前导序列长度、CP 以及 GT 的选

表 1　　**ZC 序列集与低相关区序列集特性比较**

序列类型	每个子集内序列个数	低相关区内相关值大小	子集个数	可达 PRACH 数量
ZC	$\left\lfloor \frac{L}{N_{\mathrm{CS}}} \right\rfloor$	0	$L-1$	$(L-1)\left\lfloor \frac{L}{N_{\mathrm{CS}}} \right\rfloor$
	$(L-1)\left\lfloor \frac{L}{N_{\mathrm{CS}}} \right\rfloor$	$\sqrt{L}$	1	
ZCm	$(L+1)\left\lfloor \frac{L}{N_{\mathrm{CS}}} \right\rfloor$	$\sqrt{L+1}$	$L-1$	$(L^2-1)\left\lfloor \frac{L}{N_{\mathrm{CS}}} \right\rfloor$
ZCa	$\approx\left\lfloor \frac{L}{N_{\mathrm{CS}}} \right\rfloor$	$\sqrt{L}$	L	$(L^2-1)\left\lfloor \frac{L}{N_{\mathrm{CS}}} \right\rfloor$

择都更加灵活。这种将多个较短前导序列进行组合拼接为某个用户对应前导序列的方法能显著增大前导序列资源池规模。以文献[43]中单阶段采用长 139 的 ZC 序列，两阶段采用 2 个长 71 的 ZC 序列进行拼接为例，两阶段方式的可用序列总数是单阶段的 81 倍。此外，文献[45]在每个单阶段使用由不同 ZC 根序列生成的子前导序列，进一步增加了可用前导序列的总体数量。

(a) 单阶段前导序列传输

(b) 两阶段前导序列传输

图 4　单阶段前导传输与两阶段前导传输

3) 上行授权资源碰撞解决

如前文所述，若某个前导序列被多个用户同时选中（即发生前导序列碰撞），则这些用户随后将在相同上行授权信道资源上发送各自 L2/L3 连接请求消息，导致多个 L2/L3 连接请求消息互相干扰，无法正确解调，也造成上行授权资源浪费。因此，如何在发生前导序列碰撞事件时仍能正确地解调多个互相干扰的 L2/L3 连接请求消息，对提升 GB-RA 性能有重要意义。表 2 梳理了现有文献中关于上行授权资源碰撞的解决策略。

2.3　挑战与机遇：引入新型随机接入方法的必要性

显然，传统 GB-RA 方法严重受限于其前导序列资源池规模和上行授权资源数量。虽然可采用过载控制、碰撞解决等新机制进行改良，但由表 2 可知，这些增强型 GB-RA 方法往往引入了复杂的信令交互过程，仍不适用于大规模随机接入场景。具体而言，图 5[11]和图 6[11]展示了各类基于过载控制的 GB-RA 方法的接入成功概率和前导序列碰撞概率仿真结果。由图 5 和图 6 可知，当小区内的 UE 总数为 30 000 时，大部分随机接入机制中 UE 的接入成功率不足 40%，且前导碰撞概率上升明显，显然，这与未来大规模物联网的每平方千米 100 万连接密度的要求还有较大差距。因此，为满足未来大规

表 2　　**上行授权资源碰撞解决策略对比**

策略	技术特点	文献	主要方法	特性分析
碰撞控制	随机选择 PUSCH 资源	文献[46]	基站为一个前导序列分配多个上行 PUSCH 资源，用户随机选择一个 PUSCH 资源来传输 L2/L3 连接请求消息	PUSCH 资源随机选择策略能在一定程度上降低上行授权资源碰撞概率
	限制连接请求消息传输	文献[47]	基站根据当前可分配 PUSCH 资源数，广播消息传输限制因子，以此控制用户发送连接请求消息的数量	未知确切前导序列碰撞程度时，设定限制因子会增大基站侧控制开销，易造成 PUSCH 资源浪费
资源复用	时域非正交技术	文献[48]	根据用户上行到达时间的不同将其划分为不同群组，每个群组复用同一个 PUSCH 资源	该方法需要估计用户上行到达时间，且每个资源复用组的生成过程增大了接入能耗
	空域非正交技术	文献[49]	根据空间位置关系识别发生前导序列碰撞的多个用户，并将它们视为一个 NOMA 组，复用同一个 PUSCH 资源	对碰撞用户的识别能力有限，限制了其在大规模随机接入场景中的应用
碰撞解调	分布式导频正交化	文献[50]	将 L2/L3 连接请求消息中包含的参考导频信号正交化，并采用多天线技术对多个 L2/L3 连接请求消息进行解调	正交参考导频信号个数有限，不足以支撑大量用户碰撞情况
	子帧间串行干扰消除（SIC, successive interference cancellation）技术	文献[51]	在基础侧部署缓存器，用于存储发生碰撞的 L2/L3 连接请求消息，再利用子帧间 SIC 技术进行解调	所提子帧间迭代式 SIC 技术仅适用于少量用户碰撞的情况

模小型化、低成本、低功耗的物联网随机接入需求，设计新型随机接入方法非常必要。

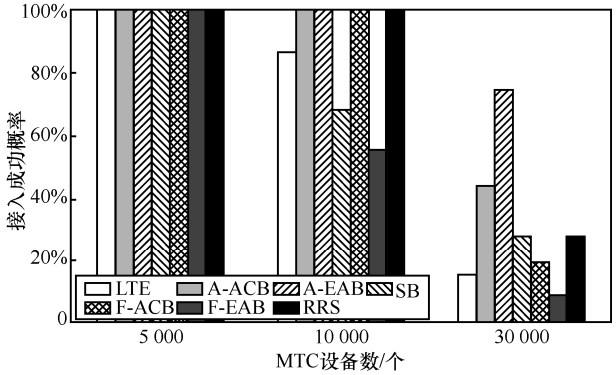

图 5　基于过载控制的 GB-RA 方法的接入成功率

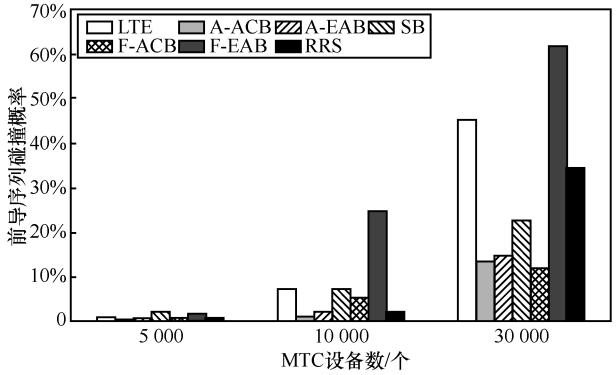

图 6　基于过载控制的 GB-RA 方法的前导序列碰撞概率

3　免授权随机接入方法

由于当前基于授权的随机接入方法难以满足未来大规模物联网新场景，必须引入全新机制。近年来迅速发展的 NOMA 技术可在时频资源上实现多用户高过载传输，为突破随机接入传统方法提供了信号处理层面的基石。在此背景下，3GPP 于 2016 年正式提出 GF-RA 概念[52]，相继出现了功率域−免授权非正交多址[18]、基于码的免授权非正交多址[24-25, 53]、基于交织的免授权非正交多址[54]等一系列免授权传输技术。

上述各类免授权传输技术都做了以下重要共性假设[55]，即有效数据传输前，基站已完成了理想的活跃用户身份检测、信道估计和信号同步。如何以免授权方式完成上述工作，成为免授权随机接入需要解决的核心难题之一。

为解决上述挑战，研究人员利用大规模物联网通信的稀疏特性，采用压缩感知技术[56]来实现 GF-RA。GF-RA 可同时完成活跃用户身份检测和信道状态估计，显著降低接入开销，成为实现大规模物联网免授权传输的重要研究方向。本节针对免授权传输中的重要环节 GF-RA 进行讨论，重点围绕核心压缩感知技术，从理论框架、GF-RA 和性能界 3 个方面展开讨论，最后给出所面临的一系列挑战和机遇。

3.1　压缩感知技术

压缩感知是由斯坦福大学 Donoho 教授及其学生 Candes 在信号稀疏分解和逼近理论基础上进一步发展而来的信号处理方法[57-58]，在数字图像处理[59-60]、通信信道估计[61]等领域被广泛应用。压缩感知数学模型如图 7 所示，其中，$\boldsymbol{z}$ 为观测信号 $\boldsymbol{y}$ 在保存或传输过程中遇到的加性噪声。

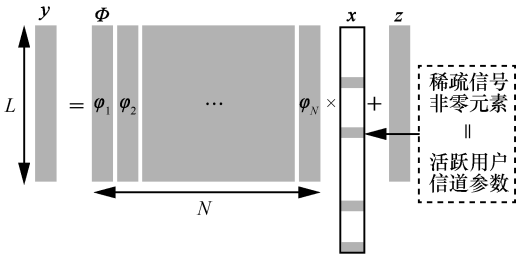

图 7　压缩感知数学模型

若信号 $\boldsymbol{x}\in\mathbb{C}^{N\times 1}$ 为只含 $K(K\ll N)$ 个非零元素的稀疏信号，则利用测量矩阵 $\boldsymbol{\Phi}$ 对 $\boldsymbol{x}$ 进行观测，可得长度为 $L(K<L<N)$ 的测量信号为

$$\boldsymbol{y}=\boldsymbol{\Phi}\boldsymbol{x}+\boldsymbol{z}$$

若测量矩阵 $\boldsymbol{\Phi}$ 满足有限等距性质（RIP, restricted isometry property）准则，接收端可以基于信号 $\boldsymbol{y}$ 实现稀疏信号 $\boldsymbol{x}$ 的精确重构。如果实际应用中原始信号并不稀疏，可对原始信号进行正交变换得到其稀疏表示 $\boldsymbol{x}$[62]。

1) 随机接入借助压缩感知的可行性

考虑小区用户数为 N 的蜂窝上行链路，以 $x_n\in\{0,1\}$ 表征任意用户 n 的活跃状态。由于在大规模物联网随机接入过程中大部分设备处于睡眠模式，属于非活跃状态，则由 $x_n, 1\leqslant n\leqslant N$ 构成的信号向量 $\boldsymbol{x}=[x_1,x_2,\cdots,x_N]$ 天然具备稀疏特性。这使基于压缩感知的活跃用户检测成为可能。

具体而言，基站预先给每个用户分配一条长度为 L 的前导序列 $\boldsymbol{\varphi}_n$。假设所有用户已完成上行帧同步，活跃用户发送前导序列。前导序列传输接收模型如图 7 所示。当以各用户前导序列为列构成的测量矩阵 $\boldsymbol{\Phi}=[\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_2,\cdots,\boldsymbol{\varphi}_N]$ 满足 RIP 准则时，基站

可使用压缩感知重构算法，基于接收的信号 $\boldsymbol{y}$ 对活跃用户身份进行精准估计，即对图 7 中 $\boldsymbol{x}$ 的非零元素索引值进行精准重构。

当前压缩感知研究主要关注信号的稀疏表示、测量矩阵构造方法，以及重构算法设计。由于大规模物联网随机接入过程中，待恢复信号 $\boldsymbol{x}$ 已天然具备稀疏特性，因此本节将聚焦后 2 个研究问题。

2) 测量矩阵的评估准则和构造方法

文献[62]提出，为精准重构稀疏信号 $\boldsymbol{x}$，测量矩阵 $\boldsymbol{\Phi}$ 应满足 K 阶 RIP 准则，即存在有限等距常数（RIC, restricted isometry constant）$\delta_K \in (0,1)$，使式(1)成立。

$$(1-\delta_K)\|\boldsymbol{x}\|_2^2 \leqslant \|\boldsymbol{\Phi x}\|_2^2 \leqslant (1+\delta_K)\|\boldsymbol{x}\|_2^2 \qquad (1)$$

文献[62]还进一步指出独立同分布的高斯随机矩阵满足上述准则，可用于构造普适性压缩感知测量矩阵。

但对于其他矩阵类型，要证明其是否满足 RIP 准则往往是 NP 难题。因此，相继出现了 Spark 理论[63]以及统计有限等距性质（STRIP, statistic restricted isometry property）准则[64]等更简单、定性的构造准则，为设计测量矩阵提供了其他理论依据。表 3 列举了几种代表性序列构造方法，并比较了基于它们所构造测量矩阵的重构精度、存储开销等性能。

表 3　几种代表性序列及其测量矩阵性能比较

所用序列类型	所满足评估准则	所达重构精度	所需硬件条件
Gaussian 随机序列	RIP 准则	较高，设为 P_{correct}	高
Bernoulli 随机序列	RIP 准则	$\approx P_{\text{correct}}$	低
Kasami 序列[65]	STRIP 准则	$> P_{\text{correct}}$	低
Udaya-Siddiq 序列[66]	相干性准则	$> P_{\text{correct}}$	低
Rademacher 序列	Spark 下界	$> P_{\text{correct}}$	低

3) 压缩感知稀疏信号重构算法

文献[62]指出，当测量矩阵 $\boldsymbol{\Phi}$ 满足 RIP 准则时，可将重构视为线性规划问题。基于此，一系列基于凸优化的稀疏信号重构算法被相继提出，如基追踪（BP, basis pursuit）[67]、迭代阈值（IT, iterative thresholding）[68]以及 LASSO（least absolute shrinkage and selection operator）[69]等。凸优化方法能依靠较少的测量样本实现较高的重构精度，但极高的算法复杂度降低了其在通信系统中的实用性。为增强算法实用性，相继出现了一系列贪婪算法，如匹配追踪（MP, matching pursuit）[70]、正交匹配追踪（OMP, orthogonal matching pursuit）[71]以及基于最大后验概率的压缩采样匹配追踪（MAP-CoSaMP, maximum a posteriori-compressive sampling matching pursuit）[72]等。

相较凸优化算法，贪婪算法实现了更低的运算复杂度，但在重构精度上有所损失。文献[73]提出的近似消息传递（AMP, approximated message passing）在信号重构精度和算法复杂度两方面实现了更优折中，成为近年来基于压缩感知进行活跃用户检测的主流算法。不同重构算法的计算复杂度的比较如表 4 所示。

表 4　不同重构算法的计算复杂度的比较

重构算法	计算复杂度
BP	$O(N^3)$
IT	$O(N^2)$
LASSO	$O(NL\log N)$
MP	$O(N\log N)$
OMP	$O(KNL)$
MAP-CoSaMP	$O(NL)$
AMP	$O(NL)$

3.2　基于压缩感知的用户身份与信道状态联合检测

文献[17]针对 CDMA 系统中的多用户检测问题提出采用压缩感知技术。此后，文献[74]针对大规模随机接入场景，提出基于压缩感知的活跃用户和信道状态联合检测方法，将整个上行传输过程划分为 2 个部分，即活跃用户身份与信道状态检测阶段、活跃用户数据传输阶段。

1) 基于最小均方误差（MMSE, minimum mean-square error）降噪的 AMP

文献[74]提出基于 MMSE 进行降噪的算法，即 MMSE-AMP，来完成活跃用户检测。该算法最核心的改进是将原 AMP 中软阈值降噪函数重新设计为

$$\eta(\tilde{x}_n, g_n, \sigma_t) = \frac{g_n^2\left(g_n^2+\sigma_t^2\right)^{-1}\tilde{x}_n}{1+\dfrac{1-\lambda}{\lambda}\dfrac{g_n^2+\sigma_t^2}{\sigma_t^2}\exp\left(-\Delta\left|\tilde{x}_n\right|^2\right)}$$

其中，$\Delta=\sigma_t^{-2}-(g_n^2+\sigma_t^2)^{-1}$，$\lambda=K/N$ 表示稀疏度，σ_t^2 表示第 t 次迭代的均方误差，g_n 表示用户 n 的大尺度衰落系数且假设基站已知。MMSE-AMP

与原 AMP 性能比较如图 8 所示（图 8 中部分结果源自文献[74]），其中 P_k 为发射功率。从图 8 可以看出，与传统 AMP 相比，在相同序列长度下 MMSE-AMP 有更佳检测性能。此后，文献[75]将多天线系统中的压缩感知求解问题等效为多测量向量（MMV, multiple measurement vector）的数据重构问题，从而将 MMSE-AMP 应用范围扩展至多天线系统。

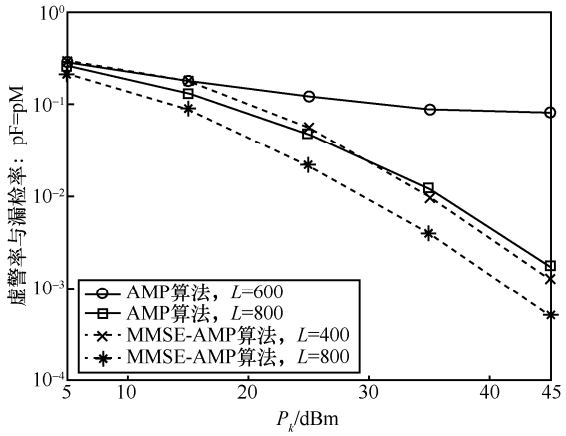

图 8　MMSE-AMP 与原 AMP 性能比较

文献[76]针对运算能力受限的低功耗物联网设备，分析了 MMSE-AMP 在大规模多天线系统中的性能。分析结果显示，增加接收端天线可有效提升 MMSE-AMP 检测性能。但同时，大规模多天线系统存在饱和效应，当天线数量增大到一定程度后所得增益不再明显。

文献[77]针对实际基站往往无法预知整个服务小区内全体用户活跃稀疏度的问题，提出改进的交叉校验 MMSE-AMP（ICV-AMP, improved cross validation aided MMSE-AMP）。该算法将用户前导序列分为观测部分和校验部分，一方面使用观测部分同时估计活跃用户身份和全体用户稀疏度，另一方面使用校验部分对所估计稀疏度准确性进行校验，通过多次迭代上述过程得到与假设稀疏度理想已知时 MMSE-AMP 相近似的检测性能，而所提算法复杂度并没有明显增加。

2) 嵌入有效负载比特的随机接入方法

文献[76]的另一重要贡献在于提出了嵌入信息位（EIB, embedded information bit）的随机接入方法。该方法允许用户在接入过程中直接使用前导序列传递少量有效负载比特，从而实现随机接入和数据传输的一步式完成。该方法充分匹配了本文引言中所述大规模物联网小分组传输的基本特征。

以某用户需要传输 τ bit 有效负载为例。基站预先给该用户分配 2^τ 条合法前导序列。该用户根据 τ bit 有效负载实际取值，从 2^τ 条合法序列中选择一条作为本次随机接入的前导序列进行发送。这等价于将该用户扩展为 2^τ 个虚拟用户，然后让 τ bit 有效负载实际取值映射出的某个虚拟用户进入活跃状态，最后基站利用压缩感知算法检测相关虚拟活跃用户。显然，一旦虚拟活跃用户被正确检测，则真实活跃用户和 τ bit 有效负载都被成功检测。

该方法的难点在于，小区用户总数 N 如果以虚拟用户个数统计会被扩大若干倍，如果直接使用现有重构算法会导致检测性能严重恶化。因此 Larsson 等[76]对 MMS-AMP 再次改进，提出了 M-AMP（modified-AMP），有效提升了 EIB 方法传输性能。但相较于采用 AMP 进行活跃用户检测后再进行数据传输的方法，其检测精度存在一定差距，等价于用检测精度的下降置换接入加传输一步式完成能力。

3.3　基于压缩感知的随机接入性能界

基于压缩感知的活跃用户检测的性能受序列长度 L、总用户数 N，以及活跃用户数 K 等诸多因素综合影响。定量刻画活跃用户检测精度与各制约因素之间的关系有重要意义。本节旨在回顾相关研究现状并给出笔者最新的一些研究结果。

1) 基于信息论的序列长度下界

文献[78]从信息论角度出发，推导出在加性白高斯噪声（AWGN, additive white Gaussian noise）信道中实现大规模随机接入无误检测所需前导序列长度下界为

$$L=\frac{Nf\left(\frac{K}{N}\right)}{\frac{1}{2}\log\left(1+KP\right)}$$

$$f\left(x\right)=-x\log x-\left(1-x\right)\log\left(1-x\right)$$

上述理论成果的局限在于，其分析并未针对任何具体检测算法，所得下界较松弛，与 AMP 实测性能存在较大差距，理论预测效果不佳。

2) 基于压缩感知理论的序列长度下界研究

文献[62]证明，如果测量矩阵满足 $\delta_{2K}\leqslant\sqrt{2}-1$，可通过求解最小化 l_1 范数来得到唯一稀疏解。在此基础上，文献[79]指出，若由高斯随机序列或伯努利随机序列来构成测量矩阵，则序

列长度 L 必须满足

$$L \geqslant CK \log\left(\frac{N}{L}\right)$$

其中，C 为常数。上式表明序列长度下界与稀疏度及总用户数有关，但文献[79]并未给出常数 C 的具体取值或求解方法，导致在实际应用中仍无法指导用户前导序列的最短化选择。

3) 基于 AMP 状态演进的序列长度下界研究

针对 AMP，文献[80]讨论了信号重构成功概率与 $\rho = K/L$ 和 $\delta = L/N$ 间的约束关系。但是，在其分析中并未考虑衰落、噪声等信道因素对 AMP 检测性能的影响。文献[74]通过分析 MMSE-AMP 状态演化（SE, state evolution）过程，给出了经历衰落信道时，MMSE-AMP 重构成功概率的分析方法。基于文献[74]，文献[81]进一步提出采用 MMSE-AMP 时，为保证较高检测精度，所需序列长度应满足

$$L \geqslant \frac{N\,\mathrm{MSE}(\sigma_{\mathrm{a}}) + \dfrac{1}{P}}{{\sigma_{\mathrm{a}}}^2}$$

其中，P 为信噪比，$\mathrm{MSE}(\sigma_{\mathrm{a}})$ 为 MMSE-AMP 的状态演进函数[74]。图 9 展示了文献[81]所推导的 MMSE-AMP 序列长度理论下界与测试值的对比，其中，目标检测精度为虚警和漏报概率均小于 0.05。从图 9 可以看出，在不同的稀疏度和发射功率下，该理论下界与测试值均较接近。

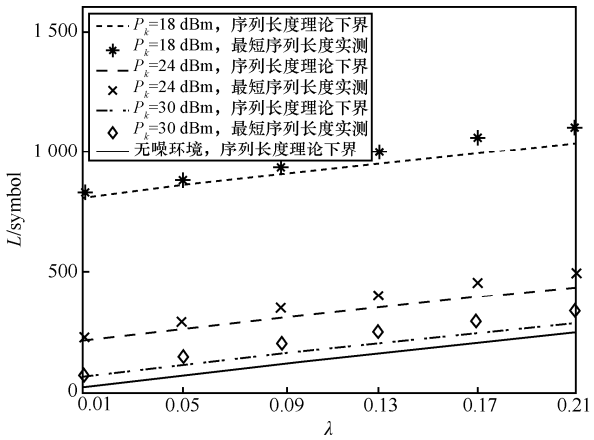

图 9　MMSE-AMP 序列长度理论下界与测试值对比

3.4　免授权随机接入面临的挑战与机遇

压缩感知技术及其在大规模免授权随机接入中的应用潜力巨大，但同时存在很多亟待解决的问题，具体如下。

1) 低复杂度、高精度压缩感知重构算法设计

由表 4 可知，即使是目前较先进的压缩重构算法 AMP 也需消耗相当计算资源。而多天线系统是支撑未来大规模物联网应用的另一重要技术，多天线信号处理加压缩感知是重要研究方向[75-76]，必然给整个通信系统带来极高运算复杂度，影响算法实用性。因此，成数量级减少压缩感知算法复杂度有重要意义。另一方面，压缩感知算法的重构精度直接决定大规模随机接入的检测成功概率，也是关键性能指标。然而，根据 3.2 节理论分析和各类约束关系可知，压缩感知重构算法复杂度和检测精度相互制约。例如，文献[82]基于二阶 Reed-Muller 序列提出的逐层检测算法，虽然以较短序列长度支撑了庞大用户群体，节省了运算复杂度，但其检测能力非常有限。因此，如何实现复杂度更低、检测精度更高的压缩感知重构算法，有效降低基站和用户端序列存储开销和运算开销极具挑战。

2) 前导序列长度理论下界分析

压缩感知框架下，用户所需发送的前导序列长度理论下界值可指导序列长度的自适应选择，减少序列存储、传输开销，具有重要研究意义。然而正如 3.3 节所述，当前已有的各类理论结果存在应用场景受限、过于松弛、存在未具化常参数等问题。具体而言，Donoho 等[80]对 AMP 重构精度的约束关系进行了分析，但未考虑噪声带来的影响，前导序列和重构算法抗噪性能分析仍有大量空白。从图 9 可以看出，文献[81]所得理论下界在稀疏度数值增大时，与测试值间差距逐渐变大。因此，压缩感知框架下的前导序列长度理论下界分析仍然任重而道远。

3) 异步场景下的随机接入难题

已有模型大都假设用户实现了理想信号同步。但在实际免授权随机接入过程中，不同用户存在不同定时偏差，这导致基站侧接收到的多个前导序列无法完全同步。前导序列异步现象会导致基于压缩感知的活跃用户检测算法性能急剧恶化，甚至完全失效，这也是压缩感知算法在大规模随机接入实用化中面临的一大挑战，但目前对该问题的研究较少。

异步场景下，现有活跃用户压缩感知检测方法均假设由定时偏差导致的前导序列最大位移量 γ 非常有限。在此假设下，可借鉴 3.2 节所述 EIB 方法来处理异步问题，即将每种序列位移可能情况映射为一个虚拟用户，这样等价于将测量矩阵列维度

和用户总数扩大了 γ 倍，基站仍可利用压缩感知重构算法检测活跃用户及其定时偏差[83-84]，但此方法也会极大地牺牲检测精度。文献[84]调整了异步场景下 MMSE-AMP 的降噪函数，并给出不同用户的差异化判决门限，实现了对活跃用户身份、信道状态以及定时偏差的联合估计。但其检测性能依赖于足够长的前导序列。综上，异步场景下的大规模随机接入问题仍处于探索阶段，有待深入研究。

4) 一步式随机接入和数据传输方法

前述免授权随机接入方法大多关注活跃用户检测问题和信道状态估计问题，少有关注有效负载的传输方案。EIB 方法虽然可实现一步式随机接入和数据传输，但需占用大量前导序列，且形成的数据负载能力也非常有限。然而，一步式随机接入和数据传输方法对于减少大规模物联网业务平均通信时延和控制开销具有重要意义，值得研究。

4 半免授权大规模随机接入方法

免授权随机接入使蜂窝网络支撑海量用户共存成为可能。但网络的随机特性决定了免授权随机接入只能在宏观统计意义上达成全体用户的平均接入成功概率和长期通信服务质量。对于高负载时隙的特定用户，免授权随机接入仍无法保证其接入性能与后续通信质量。为更好地平衡个体与整体的用户体验，半免授权随机接入方法成为另一种解决途径。

半免授权随机接入的技术特点如下。1) 部分用户在发送有效负载前需消耗一定控制开销，但系统整体接入开销仍远低于基于授权的传统随机接入过程；2) 完成授权不再成为用户可传输有效负载的必备前提条件；3) 形式上往往呈现基于授权的传统随机接入与前述免授权随机接入相融合的特征；4) 以非正交多址、功率控制、资源调配、压缩感知等为关键技术。

文献[23]提出半免授权随机接入概念，并基于功率域非正交多址技术[18, 85]与认知无线电机理[86]设计了其具体实现方案。

对小区用户分组或分层是实现半免授权随机接入的另一设计思路，由 3GPP 提案[87]提出。

在此方向上，文献[81]开展了相关研究，提出了一种两级式群组化网络拓扑结构和与之配合的两步式自适应随机接入方法，部分研究结果将在 4.2 节进行描述。

4.1 功率域非正交多址与认知无线电机理半免授权随机接入方法

文献[23]所提半免授权功率域非正交多址（SGF-NOMA, semi-grant-free power domain non-orthogonal multiple access）方案的核心思想和主要步骤如下。1) 将小区用户划分为高优先级的授权主用户和低优先级的免授权次用户；2) 根据认知无线电技术中干扰可消除原理提出两类用户间的配对策略；3) 形成配对的用户组基于功率域 NOMA 技术复用系统信道资源；4) 配对组中主用户沿用基于授权的随机接入方法，而次用户采用新型免授权非正交随机接入方法。

基于上述思路，文献[23]假设了 2 种典型场景，初步提出了 2 种配对方案，并对不同场景下使用不同配对方案的 4 种组合进行了理论性能分析。2 种典型场景具体如下：场景 1，授权用户靠近基站，免授权用户位于小区边缘，基站侧 SIC 过程中首先检测授权用户消息；场景 2，免授权用户靠近基站，授权用户位于小区边缘，基站侧 SIC 过程中首先检测免授权用户消息。2 种配对方案分别为开环方案（OL, open-loop）[88]和分布式竞争方案（DC, distributed-contention）[89]。场景 1 采用开环方案，如图 10 所示；场景 2 采用分布式竞争方案，如图 11 所示。2 种配对方案的主要步骤与特性对比如表 5 所示。

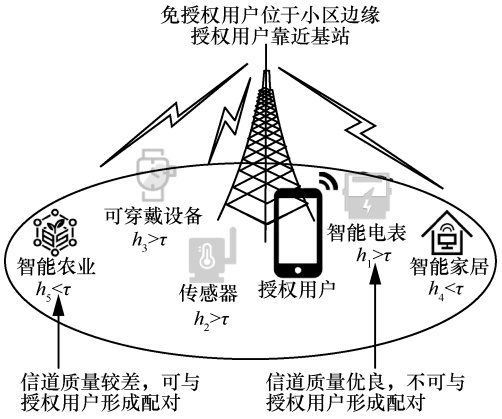

图 10 场景 1 采用开环方案

4.2 基于分组的半免授权随机接入方法

利用分组或分层技术支撑大规模随机接入一直是标准提案和学术文献探讨的重要方向[87, 90]。文献[81]聚焦低功耗、时延不敏感型 mMTC 业务，提出了群组化两步式半免授权随机接入方法。其核心思想和主要步骤如下。1) 根据地理位置将小区用户分为若干大群，再在每个大群内将互相邻近、活跃

度相似的多个用户分为一组，每个用户组选择一个头节点用户代表该组全体成员完成随机接入和数据传输，从而形成图 12 所示的两级式群组化网络拓扑结构。2) 与群组化网络架构相配合，将用户随机接入过程划分为 2 个时隙，在第一时隙，根据负载估计结果，基站调度各群接入优先级和各群内头节点用户的前导序列长度；在第二时隙，各群内头节点用户基于压缩感知技术完成免授权随机接入。其关键技术在于：1) 群组化网络拓扑结构的低开销建立与维护方法；2) 基站侧负载估计算法与用户压缩感知算法；3) 头节点用户前导序列长度自适应选择算法；4) 半免授权随机接入方法整体性能优化。

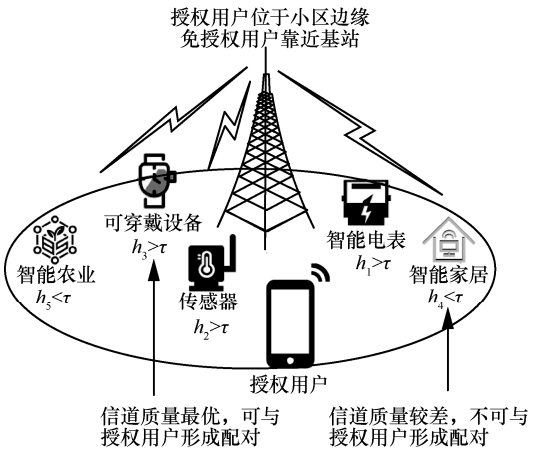

图 11　场景 2 采用分布式竞争方案

1) 群组化网络拓扑架构的建立与维护

蜂窝小区内物联网用户两级式群组化网络主要通过用户注册、用户组初始化、周期性维护三大操作形成。

①用户注册：用户开机或第一次切换入小区时，向基站发送注册消息；基站根据该消息确定用户的群归属，决策该用户是否初始化为头节点用户，然后将相关决策反馈给用户。

②用户组初始化：头节点用户广播建组消息；邻近的非头节点用户侦听并向头节点用户发送入组申请；用户组初步建立后通过最小能耗算法重选最佳头节点用户；新选头节点用户向基站上报用户组构成情况。

③周期性维护：基站周期性向所有用户广播群组更新消息；侦听到该消息的头节点用户重新发起建组流程；非头节点用户基于 k-means 聚类算法[91]执行用户组成员优化过程；优化后，由头节点用户再次将用户组构成情况上报给基站。

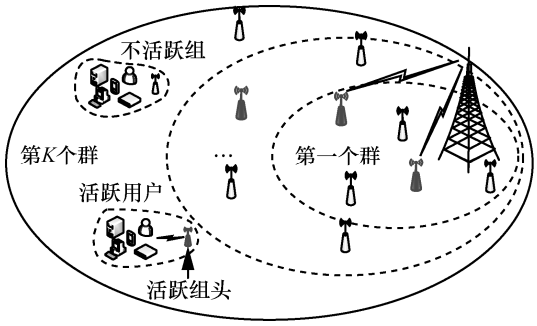

图 12　蜂窝小区内物联网用户两级式群组化网络拓扑结构

2) 两步式自适应随机接入方法

在第一时隙，接收到基站广播的随机接入机会（RAO, random access opportunity）消息后，活跃用户组头节点首先发送群级前导序列，该前导序列属于传统正交序列类型，如 m 序列；然后，基站对各群负载情况进行估计，决定各群接入优先级以及头节点用户免授权随机接入时所用非正交前导序列长度，并广播这些决策结果。

在第二时隙，头节点用户侦听基站决策结果以调整本地前导序列长度，并在基站指示的群接入时隙内执行第 3 节所述基于压缩感知的免授权随机接入过程；基站采用 3.2 节所述 MMSE-AMP[74]联合检测头节点用户身份和相关信道状态；最后，基站向头节点用户反馈接入结果，若成功，基站指示用于数据传输的信道资源块。

表 5　2 种配对方案的主要步骤与特性对比

方案名称	配对过程	优缺点
开环方案	1) 基站广播信道增益门限值 τ，免授权用户将自身上行信道增益与门限值比较 2) 场景 1 下，信道增益低于 τ 的任意免授权用户可与授权用户构成 NOMA 组；场景 2 下，信道增益高于 τ 的任意免授权用户可参与构成 NOMA 组	优点：控制开销低，允许大量免授权用户复用信道资源 缺点：复用系统资源的免授权用户数随机波动，数量大时导致多用户检测复杂度极高，传输性能下降
分布式竞争方案	1) 基于开环方案确定可配对的免授权用户范围 2) 开启竞争时间窗口，免授权用户根据自身上行信道增益获取回退值 τ_m 3) 场景 1 下，自身信道增益最差的免授权用户可与授权用户构成 NOMA 组；场景 2 下，自身信道增益最好的免授权用户可参与构成 NOMA 组	优点：复用信道资源的免授权用户数量可控，降低多用户检测算法复杂度，保证传输性能 缺点：控制复杂，网络中共存用户总数下降

群组化两步式半免授权随机接入方法性能仿真分析如下。群组化小区与未群组化小区活跃用户身份检测性能对比如图 13 所示，其中，小区总用户数 $N=100\,000$，小区活跃用户稀疏度 $\lambda=0.05$，前导序列长度固定时为 $L=512$，K 表示群个数，M 表示群内组个数，pM 和 pF 分别表示活跃用户漏检率和虚警率。

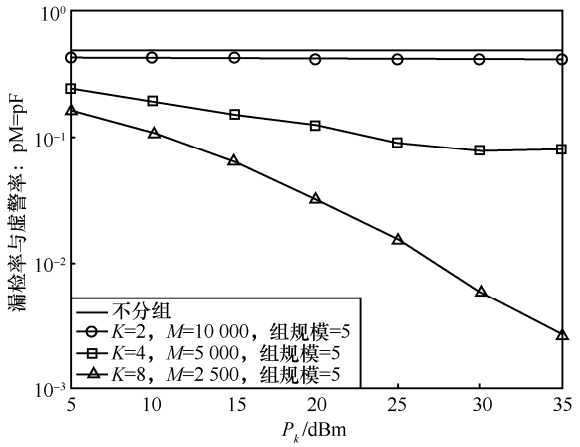

图 13　群组化小区与未群组化小区活跃用户身份检测性能对比

根据图 13 可知，群组化后活跃用户检测精度明显高于未群组化的传统小区，并且群个数越多，增益越大。但群个数增加也会带来接入时延增高问题。另一方面，根据 3.3 节所述压缩感知理论可知，不同稀疏度下，完成精准重构所需序列长度也不相同。通过改变用户平均活跃度，在保证活跃用户身份检测精度达到 pF=pM < 0.05 的前提下，利用自适应序列长度选择算法确定活跃用户所发前导序列长度。自适应序列长度随稀疏度变化情况如图 14 所示。从图 14 可以看出，相对于 $L=512$ 的固定序列长度方法，自适应序列长度选择算法显著缩短了前导序列长度，减少了接入开销和能耗。

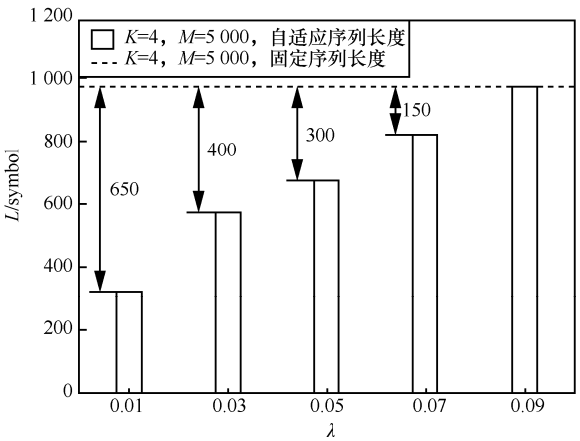

图 14　自适应序列长度随稀疏度变化情况

相较于第 3 节所述免授权随机接入方法，基于分组的半免授权随机接入方法具有以下优势。1) 以形成群组化网络拓扑结构所需的少量调度、控制开销为代价，显著降低大规模随机接入用户碰撞概率；2) 基于群负载估计的自适应序列长度选择算法，可灵活匹配各群在用户活跃度、群内用户总数等属性上的差异性。

4.3　半免授权随机接入面临的挑战与机遇

半免授权随机接入方法在功率分配、资源调度、信道估计、干扰控制、接入与传输整体方案、理论界等许多问题上都有待进一步研究。

1) 用户划分策略

用户划分策略是半免授权随机接入方法中的重要设计环节。3GPP 提案[92]根据用户所需传输分组长度完成分类，将传输长分组的用户划分为授权用户，将传输短分组的用户划分为免授权用户。前述 SGF-NOMA 系统根据用户 QoS 需求完成分类，将高 QoS 需求用户划分为授权用户，将低 QoS 需求用户划分为免授权用户。文献[81]所提群组化两步式随机接入系统根据用户地理位置完成群划分，并在此基础上进行组划分。如何根据不同时延、带宽、可靠性要求划分用户类型或形成群组以提高系统整体性能，以及如何低开销地获取完成用户划分所需的信息，都值得进一步探讨。

2) 用户成组方案

前述 SGF-NOMA 系统提出了开环和分布式竞争 2 种用户配对成组方案，本质上遵循了功率域 NOMA 技术依据接收功率大小形成用户组的原理。文献[93]在此基础上，进一步考虑上行信道状态的影响，提出自适应功率分配方案。对于两级式群组化随机多址接入系统，依据用户地理位置和通信行为，让地理位置毗邻、活跃度相近的用户构成用户组。这里仅给出了成组准则，但实际系统中如何进行分布式成组决策与控制，还有待深入研究。

3) 多用户干扰控制与分析

半免授权随机接入与数据传输方法使多个用户复用相同信道资源，提高了系统频谱效率，但也引入了多用户干扰。因此，需要设计多用户干扰控制机制并开展相关性能分析。SGF-NOMA 系统通过信道增益阈值和竞争时间窗口来调节复用授权用户信道资源的免授权用户数量。但信道增益阈值、竞争时间窗口等控制因素对系统干扰分布、中断概率等性能的影响尚无分析。基于分组或分层技

术来实现半免授权随机接入时，组内用户间通信采用 D2D（device-to-device）通信[94-95]，组头节点与基站则采用小区蜂窝通信资源，可以采用文献[96]所提 δ_D-ILA 区域限制干扰控制算法，但该算法有待进一步改进。

4) 系统理论性能分析

半免授随机接入方法尚处于初步研究阶段，主要关注切实可行的算法、方案，但对所提算法、方案能达到的理论性能极限研究不多。例如，文献[23]对用户中断概率进行了分析，笔者对用户所需前导序列长度下界进行了分析，但对于接入成功概率、平均接入时延、系统用户容量、用户检测精度等，尚需进行相关理论研究。此外，对于信道状态估计误差影响、控制开销大小等实际因素也需要给予关注。

5　无用户标识随机接入方法

前文已指出，由于大规模物联网中海量用户共存，沿用基于授权的随机接入方法将导致极端碰撞概率和高额信令开销。为解决海量低活跃度用户在相同时频资源上的多址难题，麻省理工学院的 Polyanskiy[26]提出了无用户标识随机接入（URA, unsourced random access）概念，也被称为随机接入码（RAC, random access code）技术。由于所有用户共享相同的码本，用户传输前不需要完成身份标识序列（SS, signature sequence）的分配，因此 URA 也被视为一种无协调免授权大规模随机接入（UGFMA, uncoordinated grant-free massive random access），以便与基于协调的免授权大规模随机接入（CGFMA, coordinated grant-free massive random access），如基于压缩感知的免授权随机接入相区别[97]。协调免授权与无协调免授权随机接入在应用场景、典型接入性能、方法特性上的对比如表 6 所示。

URA 的核心特点如下。1) 接收端不需要识别活跃用户的身份标识，需要传递自身标识的活跃用户可以将标识信息内嵌入数据负载中。该机制使所有用户可以共享相同的码本进行传输，避免了在海量用户中进行复杂的用户身份标识认证与分配过程。2) 使用所有活跃用户消息译码的平均错误概率衡量系统整体接入与传输性能。3) 针对传统信息论不适用于大规模短分组通信场景的现状，提出有限码长理论作为随机接入码分析工具。

文献[26]不仅提出了全新的随机接入框架，还推导了该框架的性能边界。其理论界分析表明，ALOHA 协议、slotted ALOHA 协议、时分多址等多种经典的接入或多址方法都远远未达理论上界，还有明显的性能提升空间。因此，近年来 URA 技术在业界内引起了广泛关注，一系列 URA 新框架和 RAC 新编码方案被相继提出，成为当前研究热点。

从 URA 框架下的用户行为特征来看，URA 也可视为免授权随机接入的一种。但由于其概念的新颖性和方法的基础性，URA 值得特别关注。下面将基于 URA 技术的实现细节，重点讨论三大类研究方向：基于 T-fold ALOHA 协议的 URA 方法、基于编译码增强的 URA 方法、基于用户位置与发送内容相关性的 URA 方法。

5.1　基于 T-fold ALOHA 协议的 URA 方法

Polyanskiy 揭示了 URA 框架的普适性，即已有大部分随机接入或多址方案，如 ALOHA 协议、slotted ALOHA 协议、时分多址、码分多址、视干扰为噪声（TIN, treat interference as noise）技术等都可看作 URA 的特例。分析结果表明，这些传统接入或多址技术并不适用于巨连接场景，与 URA 理论界存在巨大差距。

Ordentlich 和 Polyanskiy 提出了第一个适用于大规模巨连接场景的级联编码结构 URA 策略，即 T-fold ALOHA 协议[98]。T-fold ALOHA 协议原理如图 15 所示，主要步骤如下。1) 用户利用重叠码（SC,

表 6　　无协调免授权与协调免授权（基于压缩感知的免授权）对比

方案名称	应用场景	典型接入性能	优缺点
无协调免授权	完成大规模随机接入场景中活跃用户的数据传输	使用信道资源个数为 30 000，信噪比为 3 dB，活跃用户个数为 200 时，误码率可达 0.05	优点：支持总用户数量大，不需要显式的活跃用户身份检测环节，适用于低时延传输场景 缺点：无前导序列，信道估计困难，译码器复杂度较高
基于压缩感知的免授权[74-77]	完成大规模随机接入场景中活跃用户身份检测与信道估计	总用户个数为 4 000，前导序列长度为 800，信噪比为 3 dB,活跃用户个数为 200 时，误码率可达 0.012	优点：联合完成活跃用户身份检测与信道估计，天线个数充分多时，检测误差可迫近零 缺点：前导序列开销较大，支持总用户数受前导序列长度制约

superposed code）作为外码（OC, outer code）对用户原始数据进行第一级编码，该级编码使接收端可从接收信号中分离不同用户消息。2) 用线性二进制码（LBC, linear binary code）作为内码（IC, inner code）对第一级输出码字进行第二级编码，该级编码使接收端可对抗信道衰落影响；第二级编码完成后，数据分组的发送过程与传统 ALOHA 协议类似，且均假设经历以时隙为单位的时分信道。3) 用户随机选择若干时隙，在各选定时隙上重复传输同一数据分组。在 ALOHA 协议中，如果某个时隙被多个用户选中，则其会承载多个用户数据分组，导致多用户数据间互相干扰，造成译码失败。因此，相对 ALOHA 协议而言，T-fold ALOHA 协议的最大改进在于引入了重叠码，使每个时隙上最多可同时承载 T 个用户数据分组，而 T 的大小由重叠码类型和重叠码参数决定。相关仿真表明，在相同系统参数配置下，达到相同误码率性能时，T-fold ALOHA 协议所需发送功率远低于原 ALOHA 协议和 TIN 技术[98]。

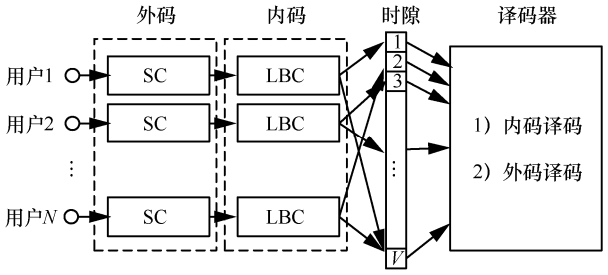

图 15 T-fold ALOHA 协议原理

文献[99]对文献[98]中的 T-fold ALOHA 协议进行了改进，重新设计了用户传输时隙选择方法，提出了 T-fold IRSA（irregular repetition slotted ALOHA）协议。该文推导了 T-fold IRSA 方案可达理论界，并利用密度进化法优化了活跃用户数固定和活跃用户数服从泊松随机分布 2 种场景下的系统参数，有效提高了 T-fold ALOHA 协议的系统性能。文献[100]研究了 T-fold ALOHA 协议在瑞利衰落信道下的系统性能。文献[101]在文献[100]的基础上，进一步考虑了时间异步问题对 T-fold ALOHA 协议性能产生的影响。这些面向实际工程难题的研究都进一步提高了 T-fold ALOHA 协议的可行性。

5.2 基于编译码增强的 URA 方法

在原 T-fold ALOHA 协议[98]中，为降低计算复杂度，采用线性二进制码作为内码，采用 BCH 码校验阵构造外码。仿真结果表明，相较传统随机接入方法，T-fold ALOHA 协议实现了性能增益。但它与随机接入码 Polyanskiy 界[26]之间仍存在明显差距。

为此，研究人员尝试采用纠错能力更强的码来充当内码与外码，并在接收端摒弃了内码与外码各自独立译码的架构，转而设计新型联合译码算法。此类方法按照采用的编码类型可分为基于低密度奇偶校验码（LDPC, low-density parity-check code）的 T-fold ALOHA 协议和基于极化码（PC, polar code）的 T-fold ALOHA 协议。上述 2 类方法的一般化原理流程如图 16 所示。

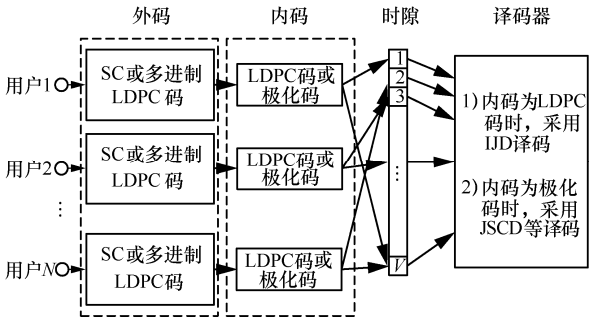

图 16 基于编译码增强的 T-fold ALOHA 原理流程

1) 基于 LDPC 码的 T-fold ALOHA 协议增强

文献[102]采用多进制 LDPC 码作为 T-fold ALOHA 协议中的外码，而内码采用 LBC。相应地，在接收端采用 q 阶和积算法（QSPA, q-ary sum-product algorithm）作为外码模块译码算法，采用最大似然算法作为内码模块译码算法。此外，外码译码器与内码译码器根据因子图进行信息交互，整体形成迭代式联合译码器（IJD, iterative joint decoder）。

文献[103]则将用户数据分为 2 个部分。第一部分为扩频序列，用于标识用户身份并通过压缩感知算法在接收端进行感知；第二部分为有效负载，使用 LDPC 码进行编码保护。在接收端，文献[103]针对 T 用户二进制输入高斯多址信道设计了迭代联合译码算法，还利用互信息转换图（EXIT Charts, extrinsic information transfer charts）工具优化所用 LDPC 码内码因子图，提高性能增益。

文献[104]同样将用户数据分为 2 个部分，即前导序列和有效负载，但区别在于：1) 只对所发前导序列在前导序列集中的索引号以及所发消息在消息集中的索引号进行编码，而不对前导序列本身和消息内容进行编码；2) 根据前导序列索引号来决定编码模块使用的交织器，本质上等价于引入了稀疏

交织多址（sparse IDMA, sparse version of interleave-division multiple access）技术。其译码端沿用了文献[103]中提出的迭代联合译码器。

总体来看，上述方法通过引入纠错能力强大的 LDPC 码，有效提高了 T-fold ALOHA 协议的系统性能，缩小了与随机接入码理论界的差距。

2) 基于极化码的 T-fold ALOHA 协议增强

文献[101]指出相较于文献[102-103]提出的联合迭代译码器，将 TIN 技术与 SIC 技术相结合，可实现更低复杂度、更强纠错能力的译码方法。受此启发，文献[105]也在接收端同时采用了 TIN 技术和 SIC 技术。文献[105]与文献[101]的主要区别在于用极化码取代 LDPC 码作为 T-fold ALOHA 协议内码。此外，文献[105]还基于随机接入码理论推导了 T-fold ALOHA 协议以极化码为内码，以 TIN 加 SIC 技术为译码方法时的可达性能界。仿真结果显示，文献[105]所提方法相较文献[101]有进一步性能提升。

文献[106]研究了以极化码为内码的 T-fold IRSA 协议，在接收端设计了 2 种译码算法，即迭代式联合译码算法和联合串行消除译码（JSCD, joint successive cancellation decoding）算法，此外还优化了极化码冻结比特挑选方法。仿真结果表明，在 T-fold IRSA 协议框架下，相比于以 LDPC 码为内码，采用极化码为内码能带来更高性能增益。

文献[107]在发射端将一次待发送消息分为 2 个部分。第一部分消息用于映射扩频序列，第二部分消息先进行极化码编码，再使用第一部分消息所选扩频序列进行扩频。在接收端，先依据序列相关性检测用户扩频序列，从而确定第一部分消息，再使用基于最小均方误差准则的对数似然比估计器和极化码串行消除译码（SCD, successive cancellation decoding）算法来检测第二部分消息。

文献[108]令每个用户按照某种概率分布随机选择极化码码长和传输功率来完成编码和射频发射。每个用户选定的码长、传输功率等参数利用压缩感知技术进行感知。接收端基于感知到的用户编码和传输参数，使用多用户 SIC 和极化码串行消除算法进行信号检测。

3) 编码增强型 URA 方法性能对比分析

为比较前述各代表性 URA 方法的性能，本文分析了译码错误概率 $\varepsilon<0.05$ 时所需的最低信噪比，如图 17 所示。显然，相较原始 T-fold ALOHA 协议[98]，通过设计用户传输时隙挑选方法，T-fold IRSA 协议[99]大幅提升了系统性能。而无论是以 LDPC 码[100]为内码还是以极化码[105]为内码都可进一步增强 URA 方法性能。更具体而言，随着用户规模不断增大，相较于以 LDPC 码为内码以极化码为内码的方法能够获得 1 dB 以上的性能增益，距离随机码理论界仅相差约 1 dB[109]。最后，在上述所有 URA 方法中，极化码结合扩频序列随机映射的方法[107]在活跃用户数少于 225 时取得最佳系统性能；当活跃用户数大于 250 时，极化码结合随机选择码长、发射功率的方法[108]实现了最佳系统性能。

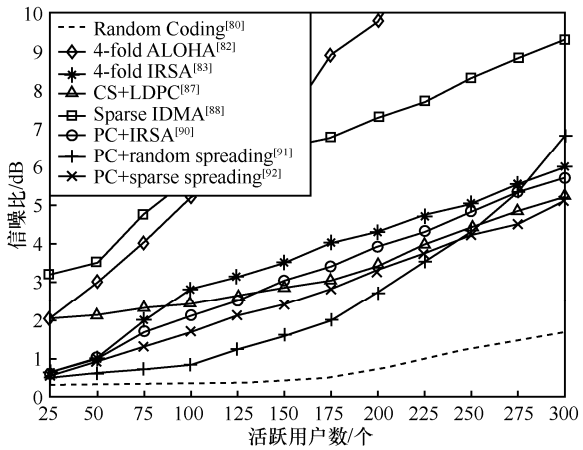

图 17　各类代表性 URA 方法系统性能比较

5.3　基于用户位置与发送内容相关性的 URA 方法

URA 概念和 T-fold ALOHA 协议[26, 98]这一具体实现方法被提出时，并没有考虑各用户活跃状态间的相关性，尤其未考虑活跃状态与待发消息内容间的相关性，而是假设各用户活跃状态完全随机且相互独立。但在很多物联网应用中，由于类型相同、功能相似、所处地理位置毗邻等原因，多个设备发送消息内容可能是对同一事件的观测结果。由于待发消息内容的一致性，导致各用户活跃状态呈现明显相关性。基于上述事实，区别于各用户活跃状态随机且独立模型，文献[109]引入各用户活跃状态相关性，提出了一种多用户群体活跃或群体静默的新模型，丰富了 URA 体系下接入模型。基于事件驱动的随机接入模型如图 18 所示[109]，其主要特点为特定物理事件（如 ξ_1）会导致特定集合内（如 Λ_1）全部用户（如 $X_k \in \Lambda_1$）进入活跃状态。

文献[109]允许用户发送 2 种消息，即标准消息和报警消息。当各用户活跃行为遵循传统独立活跃模型时，用户发送标准消息。当各用户活跃行为遵

循新型相关性活跃模型时，由特定物理事件激活与之对应的特定用户集合。激活后，该集合中所有用户都发送相同报警消息，以提高报警消息检测可靠性。接收端译码器既能检测标准消息又能检测报警消息。文献[109]更新了传统 URA 模型对错误激活概率的定义，其将报警消息对应用户集合外的其他用户处于活跃状态的概率视为错误激活概率。上述随机接入方式被称为带报警的随机接入码（ARAC, alarm random access code）技术。文献[109]根据随机接入码[26]理论，推导了该方案的频谱效率与检测性能界。理论与仿真结果显示，该方案利用设备间活跃状态与发送内容的相关性，以牺牲一定频谱效率为代价，大大提高了接收端检测可靠性。

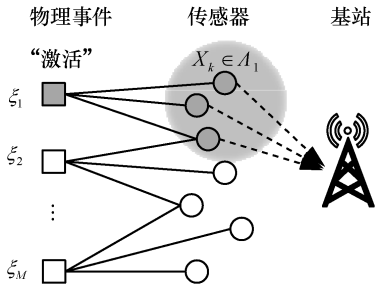

图 18 基于事件驱动的随机接入模型

文献[110]提出一种基于用户类型的多址（TBMA, type-based multiple access）方法。其核心思想如下。1) 让监测同一特定事件的所有传感器使用相同前导序列和码本；2) 不同码本之间并不完全正交，以增大码本个数；3) 针对此种接入方式，为接收端设计相应的新型消息传递算法；4) 基站直接通过检测接收到的码本即可判断特定事件是否发生，避免对活跃传感器逐个译码。仿真结果表明，在接入用户间引入相关性，有利于提高系统性能。

文献[111]在基于 LDPC 码的 T-fold ALOHA 协议中进一步考虑了用户发送内容的相关性。发送相同内容的活跃用户被映射到同一时隙进行接入与传输。仿真结果显示，此类根据内容决定传输模式的方式也有利于提高系统性能。

5.4 URA 方法的机遇与挑战

面向未来“万物互联”新时代，URA 方法具备特别优势，主要体现在以下几个方面。

1) 更低接入能耗

低功耗、低成本、低复杂度是未来 mMTC 业务对物联网终端设备的核心要求，而降低接入能耗则是提出 URA 方法的原始动机之一。URA 方法简化了随机接入步骤，避免了在随机接入过程中收发大量控制信令；收发端均采用先进编译码技术，降低了达到误码性能指标时所需接入能耗。研究表明，达到相同误码性能时，URA 方法所需发送功率远低于传统随机接入方法，并且还有较大提升空间。

2) 兼容时延不敏感场景

在 T-fold ALOHA 协议等代表性 URA 方法中，活跃用户随机选择若干时隙重复发送自身数据分组。若某个时隙上承载的用户数据分组个数小于阈值 T，则该时隙上可实现高译码成功概率。因此，如果单个活跃用户的有效负载可在更大时隙范围内随机上传，则有利于降低用户数据分组间的相互干扰。这等价于以更大的传输时延置换更高的译码性能。而在未来大规模物联网应用中确实存在如智能农业、智能读表等时延不敏感类业务，这为发挥 URA 方法特色提供了现实场景。

3) 支撑随机接入与数据传输一步式解决方案

免授权随机接入的核心目标就是提供随机接入和数据传输的一步式解决方案。但基于压缩感知技术构建的免授权方法，其数据传输能力非常有限，每次接入往往只能负载几到几十比特[76]。而 URA 方法不再区分用户身份标识和有效负载，一般性地将用户身份识别和有效负载传输统一为二进制比特的编译码问题，提高了一步式解决方案的数据承载能力，极有可能成为实现免授权随机接入的优选技术方向。

4) 开放性框架

URA 方法重在提供一种随机接入的新型框架。该框架具有充分开放性，一方面可以通过选取、组合、改造各种现有编译码技术来进一步逼近理论性能界；另一方面，URA 方法也易于与其他前沿技术相结合。例如，虽然目前 URA 方法大多假设单天线场景，但可方便地与大规模多天线技术相融合，从而大幅提升网络可支持的总用户数和活跃用户数。

虽然 URA 方法优势明显，但仍面临一些技术挑战，主要有如下几个方面。

1) 新型重叠编码设计

T-fold ALOHA 协议以重叠码为外码，来实现每个时隙最多可承载 T 个用户数据分组。T 的最大取值成为决定其用户承载能力的关键。现有重叠码设计可构造出足够大的码本空间，但码字重叠后仍可

区分的用户数T却仍十分有限，所需码字长度随T呈指数增长。具体而言，现有 T-fold ALOHA 协议普遍只做到$T\leqslant 5$。因此，设计新型重叠码，使相同码字长度条件下可区分重叠用户数T明显提高，达到$T\geqslant 10$水平，对提升 T-fold ALOHA 协议实用化具有重要意义。

2) 非理想化活跃用户检测

现有基于编码增强的 URA 方法往往将活跃用户身份检测与有效负载译码解耦，即先采用压缩感知技术进行活跃户身份检测，再根据感知到的用户扩频序列或交织规则进行有效负载部分译码。显然，前端活跃用户身份检测（即扩频序列或交织规则等的感知）的结果会严重影响后端有效负载的译码性能。现有文献均假设理想活跃用户身份检测，间接回避了上述问题。因此，研究活跃用户身份检测误差对系统整体性能的影响，提高系统抗感知误差能力具有重要实用价值。

3) 活跃用户相关性影响与利用

虽然已有研究开始利用用户活跃状态相关性来提高系统检测性能，但尚处于研究初期阶段[92-94]，并未量化用户间行为、分析发送内容相关性等深入问题。例如，相同物理事件可在不同地理位置同时发生，而处于不同地理位置的传感器将经历差异化通信信道，因此直接依赖接收信号数值判断特定事件发生的判定方法[109]容易出现检测错误。此外，高效的小区用户快速成群算法也对提高该类方法性能有明显作用。报警消息对应特定用户集的筛选方法，以及用户如何支持 2 种及以上报警消息也值得进一步研究。

6　结束语

综上所述，在大规模物联网随机接入研究领域，已涌现出丰富研究成果。本文从系统模型、协议框架、算法流程、系统性能、重要结论等多个角度对其进行了讨论与梳理。由于已有技术还存在很多理想假设和开放性问题，因此本文第 2～5 节给出了各类随机接入方法面临的挑战与机遇。

具体而言，由于基于授权的传统随机接入方法与大规模物联网通信的新型特征出现失配，必然遭遇性能回退乃至完全失效的困境。针对传统授权式随机接入的弊端，业界已提出基于控制负载、基于序列资源池扩展、基于碰撞解决等多种改进方案。然而，这些改进方案仍难以从根本上实现突破。因此必须引入新型随机接入方法。基于这一背景，可以认为免授权随机接入、半免授权随机接入以及基于编码的无用户标识随机接入都针对大规模物联网核心需求和本质特征进行了重新设计，并与大规模物联网通信新型特征形成了良好的匹配，具有很大的扩展空间。如图 19 所示，压缩感知技术、负载估计算法、非正交多址技术、群组化通信策略、D2D 通信、有限长编码理论等大量前沿技术是支撑这 3 类随机接入方法，构建新型大规模随机接入系统的核心。从这些信号处理技术入手，也是理解、掌握乃至改进新型随机接入方法的重要途径。

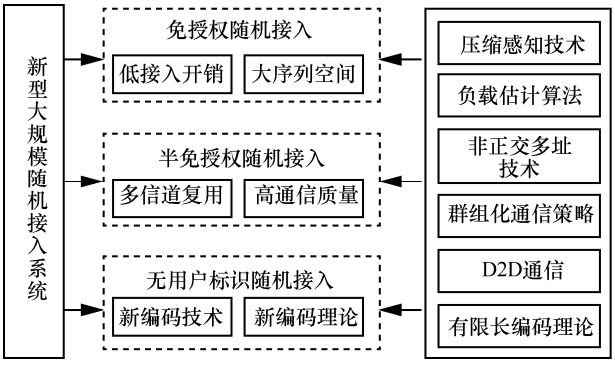

图 19　新型大规模随机接入系统关键技术

从 3 类方法研究现状来看，它们都旨在实现接入开销、接入成功概率、接入时延、接入用户数量等核心性能指标间的更佳折中。其中，免授权随机接入和无用户标识随机接入更追求“零开销”特性，而半免授权随机接入允许投入一定的必要开销或容忍一定的必要时延，以期明显提高接入质量。

大规模物联网随机接入面临的挑战与机遇如图 20 所示。从 3 类方法面临的挑战来看，它们都有如下共性问题。1) 电磁环境复杂多变，如何高效实现频谱感知？2) 用户检测运算代价高昂，如何降低算法复杂度？3) 物联网业务类型多样，如何匹配各业务特性需求？4) 各类新方法新框架性能边界未知，有何理论界指导？更具体而言，可得以下结论。1) 低复杂度高精度压缩感知重构算法设计、用户前导序列设计、异步难题解决方法、压缩理论框架下序列长度理论下界分析等都是免授权随机接入技术亟须开展的研究课题。2) 用户划分策略、用户成组方案、多用户干扰控制算法、多用户功率分配方法、系统理论界推导等是提高半免授权随机接入性能的重要研究方向。3) 增强重叠码用户区分能力、

降低活跃用户检测误差影响、活跃用户相关性系统化研究则是无用户标识随机接入需要克服的挑战。

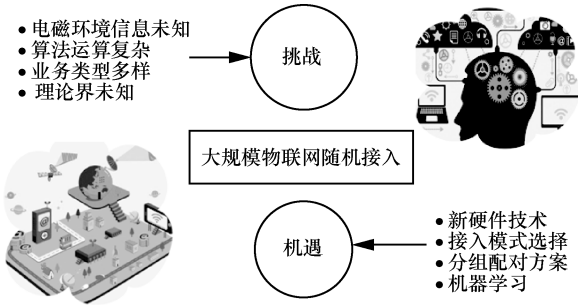

图 20　大规模物联网随机接入面临的挑战与机遇

3 类方法未来潜在的机遇如下。1) Polyanskiy 界等理论结果表明现有方法距理论界还有相当距离，新型随机接入方法还有明显性能提升空间。2) 反向散射、智能反射面[112]等新型材料和硬件射频技术的出现，为提升大规模物联网各项性能提供了新的物理基础。3) 基于人工智能的生成对抗网络、联邦学习等新型算法大量出现，智能与通信相结合已成为重要研究方向，对于提升大规模随机接入性能极具潜力。

参考文献：

[1] ALEX G, CAMEN L, SYLWIA K. IoT connections forecast: the rise of enterprise[R]. (2019-12-16)[2021-01-18].

[2] 张平, 陶运铮, 张治. 5G 若干关键技术评述[J]. 通信学报, 2016, 37(7): 15-29.
ZHANG P, TAO Y Z, ZHANG Z. Survey of several key technologies for 5G[J]. Journal on Communications, 2016, 37(7): 15-29.

[3] 尤肖虎, 尹浩, 邬贺铨. 6G 与广域物联网[J]. 物联网学报, 2020, 4(1): 3-11.
YOU X H, YIN H, WU H Q. On 6G and wide-area IoT[J]. Chinese Journal on Internet of Things, 2020, 4(1): 3-11.

[4] 3GPP. Considerations and evaluation results for IMT-2020 for mMTC connection density: R16-1808080[S]. (2018-08) [2021-01-18].

[5] 3GPP. IMT-2020 connection density evaluation results (mMTC): R16-1808866[S]. (2018-08) [2021-01-18].

[6] 3GPP. IMT 2020 self-evaluation: mMTC coverage, data rate, latency & battery life: R17-1905187[S].(2019-04) [2021-01-18].

[7] PALATTELLA M R, DOHLER M, GRIECO A, et al. Internet of things in the 5G era: enablers, architecture and business models[J]. IEEE Journal on Selected Areas in Communications, 2016, 34(3): 510-527.

[8] 3GPP. Considerations and evaluation results for IMT-2020 for mMTC connection density: R1-1903968[S].(2019-04) [2021-01-18].

[9] ZHANG Z Q, XIAO Y, MA Z, et al. 6G wireless networks: vision, requirements, architecture, and key technologies[J]. IEEE Vehicular Technology Magazine, 2019, 14(3): 28-41.

[10] 3GPP. IMT 2020 self evaluation: mMTC coverage, data rate, latency & battery life: R1-1905187[S]. (2019-04) [2021-01-18].

[11] DE ANDRADE T P C, ASTUDILLO C A, SEKIJIMA L R, et al. The random access procedure in long term evolution networks for the Internet of things[J]. IEEE Communications Magazine, 2017, 55(3): 124-131.

[12] 3GPP. Study on RAN Improvements for machine-type communications: TR 37.868[S]. (2011) [2021-01-18].

[13] DUAN S Y, SHAH-MANSOURI V, WONG V W S. Dynamic access class barring for M2M communications in LTE networks[C]//2013 IEEE Global Communications Conference. Piscataway: IEEE Press, 2013: 4747-4752.

[14] CHENG R G, CHEN J, CHEN D W, et al. Modeling and analysis of an extended access barring algorithm for machine-type communications in LTE-A networks[J]. IEEE Transactions on Wireless Communications, 2015, 14(6): 2956-2968.

[15] JANG H S, JIN H, JUNG B C, et al. Versatile access control for massive IoT: throughput, latency, and energy efficiency[J]. IEEE Transactions on Mobile Computing, 2020, 19(8): 1984-1997.

[16] DI C, ZHANG B, LIANG Q L, et al. Learning automata-based access class barring scheme for massive random access in machine-to-machine communications[J]. IEEE Internet of Things Journal, 2019, 6(4): 6007-6017.

[17] SCHEPKER H F, DEKORSY A. Compressive sensing multi-user detection with block-wise orthogonal least squares[C]//Vehicular Technology Conference. Piscataway: IEEE Press, 2013: 1-5.

[18] 田飞燕, 陈晓明, 钟财军, 等. 6G 蜂窝物联网的大规模接入技术[J]. 物联网学报, 2020, 4(1): 92-103.
TIAN F Y, CHEN X M, ZHONG C J, et al. Massive access technology in 6G cellular Internet of things network[J]. Chinese Journal on Internet of Things, 2020, 4(1): 92-103.

[19] DING Z G, YANG Z, FAN P Z, et al. On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users[J]. IEEE Signal Processing Letters, 2014, 21(12): 1501-1505.

[20] NIKOPOUR H, BALIGH H. Sparse code multiple access[C]//IEEE International Symposium on Personal Indoor and Mobile Radio Communications. Piscataway: IEEE Press, 2013: 332-336.

[21] REN B, WANG Y M, DAI X M, et al. Pattern matrix design of PDMA for 5G UL applications[J]. China Communications, 2016, 13(Supplement2): 159-173.

[22] HU X L, ZHONG C J, CHEN X M, et al. Cluster grouping and power control for angle-domain mmwave MIMO NOMA systems[J]. IEEE Journal of Selected Topics in Signal Processing, 2019, 13(5): 1167-1180.

[23] DING Z G, SCHOBER R, FAN P Z, et al. Simple semi-grant-free transmission strategies assisted by non-orthogonal multiple access[J]. IEEE Transactions on Communications, 2019, 67(6): 4464-4478.

[24] TANG W W, KANG S L, REN B, et al. Uplink grant-free pattern division multiple access (GF-PDMA) for 5G radio access[J]. China Communications, 2018, 15(4): 153-163.

[25] YUAN Z F, YAN C L, YUAN Y F, et al. Blind multiple user detection for grant-free MUSA without reference signal[C]//2017 IEEE 86th Vehicular Technology Conference. Piscataway: IEEE Press, 2017: 1-5.

[26] POLYANSKIY Y. A perspective on massive random-access[C]//2017 IEEE International Symposium on Information Theory. Piscataway:

IEEE Press, 2017: 2523-2527.

[27] SESIA S, TOUFIK I, BAKER M. LTE——the UMTS long term evolution from theory to practice[M]. New York: Wiley, 2009.

[28] WIRIAATMADJA D T, CHOI K W. Hybrid random access and data transmission protocol for machine-to-machine communications in cellular networks[J]. IEEE Transactions on Wireless Communications, 2015, 14(1): 33-46.

[29] DUAN S Y, SHAH-MANSOURI V, WANG Z H, et al. D-ACB: adaptive congestion control algorithm for bursty M2M traffic in LTE networks[J]. IEEE Transactions on Vehicular Technology, 2016, 65(12): 9847-9861.

[30] WALI P K, DAS D. Optimization of barring factor enabled extended access barring for energy efficiency in LTE-advanced base station[J]. IEEE Transactions on Green Communications and Networking, 2018, 2(3): 830-843.

[31] LIN T M, LEE C H, CHENG J P, et al. PRADA: prioritized random access with dynamic access barring for MTC in 3GPP LTE-A networks[J]. IEEE Transactions on Vehicular Technology, 2014, 63(5): 2467-2472.

[32] SIM Y, CHO D H. Performance analysis of priority-based access class barring scheme for massive MTC random access[J]. IEEE Systems Journal, 2020, 14(4): 5245-5252.

[33] TELLO-OQUENDO L, PACHECO-PARAMO D, PLA V, et al. Reinforcement learning-based ACB in LTE-A networks for handling massive M2M and H2H communications[C]//2018 IEEE International Conference on Communications. Piscataway: IEEE Press, 2018: 1-7.

[34] CHEN Z Q, SMITH D B. Heterogeneous machine-type communications in cellular networks: random access optimization by deep reinforcement learning[C]//2018 IEEE International Conference on Communications. Piscataway: IEEE Press, 2018: 1-6.

[35] JIANG N, DENG Y S, NALLANATHAN A. Deep reinforcement learning for discrete and continuous massive access control optimization[C]//2020 IEEE International Conference on Communications. Piscataway: IEEE Press, 2020: 1-7.

[36] JANG H S, JIN H, JUNG B C, et al. Recursive access class barring for machine type communications with PUSCH resource constraints[C]//2019 IEEE International Conference on Communications. Piscataway: IEEE Press, 2019: 1-6.

[37] WANG Z H, WONG V W S. Joint access class barring and timing advance model for machine-type communications[C]//2014 IEEE International Conference on Communications. Piscataway: IEEE Press, 2014: 2357-2362.

[38] WANG Z H, WONG V W S. Optimal access class barring for stationary machine type communication devices with timing advance information[J]. IEEE Transactions on Wireless Communications, 2015, 14(10): 5374-5387.

[39] SCHREIBER G, TAVARES M. 5G new radio physical random access preamble design[C]//2018 IEEE 5G World Forum. Piscataway: IEEE Press, 2018: 215-220.

[40] PITAVAL R A, POPOVIC B M, BERGGREN F, et al. Overcoming 5G PRACH capacity shortfall by combining Zadoff-Chu and M-sequences[C]//2018 IEEE International Conference on Communications. Piscataway: IEEE Press, 2018: 1-6.

[41] PITAVAL R A, POPOVIC B M, WANG P, et al. Overcoming 5G PRACH capacity shortfall: supersets of Zadoff-Chu sequences with low-correlation zone[J]. IEEE Transactions on Communications, 2020, 68(9): 5673-5688.

[42] HEATH R W, STROHMER T, PAULRAJ A J. On quasi-orthogonal signatures for CDMA systems[J]. IEEE Transactions on Information Theory, 2006, 52(3): 1217-1226.

[43] 3GPP. On NR RACH preamble design: R1-1702065[S]. (2017-02) [2021-01-18].

[44] 郭伯仁, 张欣, 杨鸿文. 基于子前导序列的IoT系统随机接入方案[J]. 北京邮电大学学报, 2018, 41(5): 110-114.
GUO B R, ZHANG X, YANG H W. A random access scheme based on sub-preambles for IoT systems[J]. Journal of Beijing University of Posts and Telecommunications, 2018, 41(5): 110-114.

[45] 3GPP. On NR RACH preamble design: R1-1704540[S]. (2017-04) [2021-01-18].

[46] KIM J S, MUNIR D, HASAN S F, et al. Enhancement of LTE RACH through extended random access process[J]. Electronics Letters, 2014, 50(19): 1399-1400.

[47] SARI R F, HARWAHYU R, CHENG R G. Load estimation and connection request barring for random access in massive C-IoT[J]. IEEE Internet of Things Journal, 2020, 7(7): 6539-6549.

[48] JANG H S, PARK H S, SUNG D K. A non-orthogonal resource allocation scheme in spatial group based random access for cellular M2M communications[J]. IEEE Transactions on Vehicular Technology, 2017, 66(5): 4496-4500.

[49] LIANG Y N, LI X, ZHANG J Y, et al. Non-orthogonal random access for 5G networks[J]. IEEE Transactions on Wireless Communications, 2017, 16(7): 4817-4831.

[50] KIM T, JUNG B C, SUNG D K. An enhanced random access with distributed pilot orthogonalization for cellular IoT networks[J]. IEEE Transactions on Vehicular Technology, 2020, 69(1): 1152-1156.

[51] KIM T, JUNG B C. An enhanced random access with inter-frame successive interference cancellation for stationary cellular IoT networks[J]. IEEE Wireless Communications Letters, 2020, 9(5): 606-610.

[52] 3GPP. Uplink multiple access schemes for NR: R1-165174[S]. (2016-05) [2021-01-18].

[53] BAYESTEH A, YI E, NIKOPOUR H, et al. Blind detection of SCMA for uplink grant-free multiple-access[C]//2014 11th International Symposium on Wireless Communications Systems. Piscataway: IEEE Press, 2014: 853-857.

[54] 3GPP. Contention-based non-orthogonal multiple access for UL mMTC: R1-164269[S]. (2016-05) [2021-01-18].

[55] SHAHAB M B, ABBAS R, SHIRVANIMOGHADDAM M, et al. Grant-free non-orthogonal multiple access for IoT: a survey[J] IEEE Communications Surveys & Tutorials, 2020, 22(3): 1805-1838.

[56] WANG B C, DAI L L, YUAN Y F, et al. Compressive sensing based multi-user detection for uplink grant-free non-orthogonal multiple access[C]//2015 IEEE 82nd Vehicular Technology Conference. Piscataway: IEEE Press, 2015: 1-5.

[57] DONOHO D L. Compressed sensing[J]. IEEE Transactions on Information Theory, 2006, 52(4): 1289-1306.

[58] CANDÈS E J. Compressive sampling[J]. Marta Sanz Solé, 2006, 17(2): 1433-1452.

[59] 李佳, 高志荣, 熊承义, 等. 加权结构组稀疏表示的图像压缩感知重构[J]. 通信学报, 2017, 38(2): 196-202.

LI J, GAO Z R, XIONG C Y, et al. Image compressive sensing recovery based on weighted structure group sparse representation[J]. Journal on Communications, 2017, 38(2): 196-202.

[60] 王金铭, 叶时平, 尉理哲, 等. 半张量积压缩感知模型的快速重构方法[J]. 通信学报, 2018, 39(7): 26-38.
WANG J M, YE S P, YU L Z, et al. Fast reconstruction method for compressed sensing model with semi-tensor product[J]. Journal on Communications, 2018, 39(7): 26-38.

[61] 袁伟娜, 严秋. 基于压缩感知的FBMC/OQAM系统信道估计方法[J]. 通信学报, 2019, 40(12): 98-104.
YUAN W N, YAN Q. Channel estimation method based on compressive sensing for FBMC/OQAM system[J]. Journal on Communications, 2019, 40(12): 98-104.

[62] CANDÈS E J. The restricted isometry property and its implications for compressed sensing[J]. Comptes Rendus Mathematique, 2008, 346(9-10): 589-592.

[63] DONOHO D L, ELAD M. Optimally sparse representation in general (nonorthogonal) dictionaries via l minimization[J]. Proceedings of the National Academy of Sciences of the United States of America, 2003, 100(5): 2197-2202.

[64] CALDERBANK R, HOWARD S, JAFARPOUR S. Construction of a large class of deterministic sensing matrices that satisfy a statistical isometry property[J]. IEEE Journal of Selected Topics in Signal Processing, 2010, 4(2): 358-374.

[65] NOUASRIA H, ET-TOLBA M. Sensing matrix based on Kasami codes for compressive sensing[J]. IET Signal Processing, 2018, 12(8): 1064-1072.

[66] GU Z, ZHOU Z. C., YANG Y, et al. Deterministic compressed sensing matrices from sequences with optimal correlation[J]. IEEE Access, 2019, 7: 16704-16710.

[67] CHEN S S, SAUNDERS D M A. Atomic decomposition by basis pursuit[J]. Siam Review, 2001, 43(1): 129-159.

[68] DAUBECHIES I, DEFRISE M, DE MOL C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint[J]. Communications on Pure and Applied Mathematics, 2004, 57(11): 1413-1457.

[69] APPLEBAUM L, BAJWA W U, DUARTE M F, et al. Asynchronous code-division random access using convex optimization[J]. Physical Communication, 2012, 5(2): 129-147.

[70] MALLAT S G, ZHANG Z F. Matching pursuits with time-frequency dictionaries[J]. IEEE Transactions on Signal Processing, 1993, 41(12): 3397-3415.

[71] TROPP J A, GILBERT A C. Signal recovery from random measurements via orthogonal matching pursuit[J]. IEEE Transactions on Information Theory, 2007, 53(12): 4655-4666.

[72] LEE N. MAP support detection for greedy sparse signal recovery algorithms in compressive sensing[J]. IEEE Transactions on Signal Processing, 2016, 64(19): 4987-4999.

[73] DONOHO D L, MALEKI A, MONTANARI A. Message passing algorithms for compressed sensing: I. Motivation and construction[C]//Information Theory Workshop. Piscataway: IEEE Press, 2010: 1-5.

[74] CHEN Z L, SOHRABI F, YU W. Sparse activity detection for massive connectivity[J]. IEEE Transactions on Signal Processing, 2018, 66(7): 1890-1904.

[75] LIU L, YU W. Massive connectivity with massive MIMO-part I: device activity detection and channel estimation[J]. IEEE Transactions on Signal Processing, 2018, 66(11): 2933-2946.

[76] SENEL K, LARSSON E G. Grant-free massive MTC-enabled massive MIMO: a compressive sensing approach[J]. IEEE Transactions on Communications, 2018, 66(12): 6164-6175.

[77] YANG L J, FAN P Z, LI L, et al. Cross validation aided approximated message passing algorithm for user identification in mMTC[J]. IEEE Communication Letters, submitted, 2021.

[78] CHEN X, CHEN T Y, GUO D N. Capacity of Gaussian many-access channels[J]. IEEE Transactions on Information Theory, 2017, 63(6): 3516-3539.

[79] BARANIUK R G. A lecture on compressive sensing[J]. IEEE Signal Processing Magazine, 2007, 24(4): 1-9.

[80] DONOHO D L, MALEKI A, MONTANARI A. Message-passing algorithms for compressed sensing[J]. Proceedings of the National Academy of Sciences of the United States of America, 2009, 106(45): 18914-18919.

[81] 张云昊, 徐修强, 范平志, 等. 一种通信的方法及装置: PCT/CN2020/109865[P]. (2020-08-18) [2021-01-18].
ZHANG Y H, XU X Q, FAN P Z, et al. A communication scheme and its installation: PCT/CN2020/109865[P]. (2020-08-18) [2021-01-18].

[82] ZHANG C, XU Y, GAO N, et al. Low complexity decoding scheme of raptor-like LDPC code in sufficient SNR scenarios[C]//2019 IEEE 90th Vehicular Technology Conference. Piscataway: IEEE Press, 2019: 1-5.

[83] SHAH-MANSOURI V, DUAN S, CHANG L H, et al. Compressive sensing based asynchronous random access for wireless networks[C]//Wireless Communications & Networking Conference. Piscataway: IEEE Press, 2013: 884-888.

[84] LIU C T, SU H J, TAKANO Y. Sparse activity, timing detection and channel estimation for grant-free uplink communications[C]//2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications. Piscataway: IEEE Press, 2020: 1-7.

[85] SAITO Y, KISHIYAMA Y, BENJEBBOUR A, et al. Non-orthogonal multiple access (NOMA) for cellular future radio access[C]//2013 IEEE 77th Vehicular Technology Conference. Piscataway: IEEE Press, 2013: 1-5.

[86] DING Z G, FAN P Z, POOR H V. Impact of user pairing on 5G non-orthogonal multiple access[J]. IEEE Transactions on Vehicular Technology, 2014, 65(8): 6010-6023.

[87] 3GPP. Pull based RAN overload control: R2-104870[S]. (2010-08) [2021-01-18].

[88] ZHAO Q, TONG L. Opportunistic carrier sensing for energy-efficient information retrieval in sensor networks[J]. EURASIP Journal on Wireless Communications and Networking, 2005, 2005(2): 1-11.

[89] BLETSAS A, KHISTI A, REED D P, et al. A simple cooperative diversity method based on network path selection[J]. IEEE Journal on Selected Areas in Communications, 2006, 24(3): 659-672.

[90] AROUK O, KSENTINI A, TALEB T. Group paging-based energy saving for massive MTC accesses in LTE and beyond networks[J]. IEEE Journal on Selected Areas in Communications, 2016, 34(5): 1086-1102.

[91] TU C Y, HO C Y, HUANG C Y. Energy-efficient algorithms and evaluations for massive access management in cellular based machine to

machine communications[C]//Vehicular Technology Conference. Piscataway: IEEE Press, 2011: 1-5.

[92] 3GPP. Intel corporation, grant-less and non-orthogonal UL transmissions in NR: R1-167698[S]. (2016-08) [2021-01-18].

[93] YANG Z, XU P, HUSSEIN J A, et al. Adaptive power allocation for uplink non-orthogonal multiple access with semi-grant-free transmission[J]. IEEE Wireless Communications Letters, 2020, 9(10): 1725-1729.

[94] LIU C, NATARAJAN B. Ergodic capacity in D2D underlay networks with varying user distribution[C]//2016 13th IEEE Annual Consumer Communications & Networking Conference. Piscataway: IEEE Press, 2016: 325-329.

[95] CHEN R, LIAO X W, ZHU S H, et al. Capacity analysis of device-to-device resource reusing modes for cellular networks[C]//IEEE International Conference on Communication. Piscataway: IEEE Press, 2012: 64-68.

[96] MIN H, LEE J, PARK S, et al. Capacity enhancement using an interference limited area for device-to-device uplink underlaying cellular networks[J]. IEEE Transactions on Wireless Communications, 2011, 10(12): 3995-4000.

[97] WU Y P, GAO X Q, ZHOU S D, et al. Massive access for future wireless communication systems[J]. IEEE Wireless Communications, 2020, 27(4): 148-156.

[98] ORDENTLICH O, POLYANSKIY Y. Low complexity schemes for the random access Gaussian channel[C]//2017 IEEE International Symposium on Information Theory. Piscataway: IEEE Press, 2017: 2528-2532.

[99] GLEBOV A, MATVEEV N, ANDREEV K, et al. Achievability bounds for T-fold irregular repetition slotted ALOHA scheme in the Gaussian MAC[C]//2019 IEEE Wireless Communications and Networking Conference. Piscataway: IEEE Press, 2019: 1-6.

[100] KOWSHIK S S, ANDREEV K, FROLOV A, et al. Energy efficient random access for the quasi-static fading MAC[C]//2019 IEEE International Symposium on Information Theory. Piscataway: IEEE Press, 2019: 2768-2772.

[101] ANDREEV K, KOWSHIK S S, FROLOV A, et al. Low complexity energy efficient random access scheme for the asynchronous fading MAC[C]//2019 IEEE 90th Vehicular Technology Conference. Piscataway: IEEE Press, 2019: 1-5.

[102] USTINOVA D, GLEBOV A, RYBIN P, et al. Efficient concatenated same codebook construction for the random access Gaussian MAC[C]//2019 IEEE 90th Vehicular Technology Conference. Piscataway: IEEE Press, 2019: 1-5.

[103] GLEBOV A, MEDOVA L, RYBIN P, et al. On LDPC code based massive random-access scheme for the Gaussian multiple access channel[C]//Internet of Things, Smart Spaces, and Next Generation Networks and Systems. Berlin:Springer, 2018, 111(18): 162-171.

[104] PRADHAN A, AMALLADINNE V, VEM A, et al. A joint graph based coding scheme for the unsourced random access Gaussian channel[C]//2019 IEEE Global Communications Conference. Piscataway: IEEE Press, 2019: 1-6.

[105] ANDREEV K, MARSHAKOV E, FROLOV A. A polar code based TIN-SIC scheme for the unsourced random access in the quasi-static fading MAC[C]//2020 IEEE International Symposium on Information Theory. Piscataway: IEEE Press, 2020: 3019-3024.

[106] MARSHAKOV E, BALITSKIY G, ANDREEV K, et al. A polar code based unsourced random access for the Gaussian MAC[C]//2019 IEEE 90th Vehicular Technology Conference. Piscataway: IEEE Press, 2019: 1-5.

[107] PRADHAN A K, AMALLADINNE V K, NARAYANAN K R, et al. Polar coding and random spreading for unsourced multiple access[C]//2020 IEEE International Conference on Communications. Piscataway: IEEE Press, 2020: 1-6.

[108] ZHENG M F, WU Y P, ZHANG W J. Polar coding and sparse spreading for massive unsourced random access[C]//2020 IEEE 92nd Vehicular Technology Conference. Piscataway: IEEE Press,, 2020: 1-5.

[109] 牛凯, 许文俊, 张平. 面向 6G 的极化编码非正交多址接入[J].西安电子科技大学学报, 2020, 47(6): 5-12, 29.
NIU K, XU W J, ZHANG P. Polar coded non-orthogonal multiple access to 6G wireless systems[J]. Journal of Xidian University, 2020, 47(6): 5-12, 29.

[110] DOMMEL J, UTKOVSKI Z, STAŃCZAK S, et al. Joint source-channel coding and Bayesian message passing detection for grant-free radio access in IoT[C]//ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2020: 8574-8578.

[111] VEM A, NARAYANAN K R, CHAMBERLAND J F, et al. A user-independent successive interference cancellation based coding scheme for the unsourced random access Gaussian channel[J]. IEEE Transactions on Communications, 2019, 67(12): 8258-8272.

[112] WU Q Q, ZHANG R. Towards smart and reconfigurable environment: intelligent reflecting surface aided wireless network[J]. IEEE Communications Magazine, 2020,58(1): 106-112.

[作者简介]

范平志（1955− ），男，四川广汉人，博士，西南交通大学教授，主要研究方向为大规模物联网、高移动无线通信、信息理论与编码技术。

李里（1983− ），男，四川成都人，博士，西南交通大学副教授、硕士生导师，主要研究方向为信道编译码技术、大规模随机接入技术等。

陈欢（1995− ），女，浙江湖州人，西南交通大学硕士生，主要研究方向为大规模随机接入技术、群组通信技术。

程高峰（1995− ），男，山西晋城人，西南交通大学硕士生，主要研究方向为大规模随机接入技术。

杨林杰（1991− ），男，四川遂宁人，西南交通大学博士生，主要研究方向为大规模随机接入技术。

汤小波（1992− ），男，四川达州人，西南交通大学博士生，主要研究方向为大规模随机接入技术、超奈奎斯特传输技术。