

ZTE中兴

基于网媒融合的沉浸式通信 技术蓝皮书

移动网络和移动多媒体技术国家重点实验室

2026年4月

版权声明

本蓝皮书版权属于中兴通讯股份有限公司，并受法律保护。转载、摘编或利用其它方式使用本蓝皮书文字或者观点的，应注明“来源：中兴通讯股份有限公司”。违反上述声明者，中兴通讯股份有限公司将追究其相关法律责任。

免责声明

本报告基于撰写人认为可靠的公开资料或实地调研信息编制，但撰写机构及作者不保证所载信息的准确性、完整性或时效性。任何人士或实体因信赖或使用本报告所载信息而导致的任何损失，撰写机构及作者不承担任何法律责任。

本报告旨在客观、中立地反映行业或法治发展状况，总结实践经验与教训。报告中的观点、分析及预测仅代表撰写组在特定日期的研究判断，不代表任何政府机构或组织的官方立场，后续有权随时进行调整、更新或废止。

贡献单位与贡献人

主编单位：中兴通讯股份有限公司

共同编写单位：（按单位笔画排序）

上海交通大学、电子科技大学、北京邮电大学、北京理工大学、东南大学、南京邮电大学、清华大学、深圳大学

编写人员：（按姓氏笔画排序）

王劲涛、王海明、冯大权、冯雨龙、冯钢、乔秀全、向际鹰、牟林、吕顺、孙耀、朱敏、朱进国、朱赞、许进、许威、刘彦宾、刘贻静、李瑞峰、陆平、陈刚、陈智勇、陈鸣镡、宋利、宋维涛、宋杰、张强、郁光辉、周世东、周亮、赵志勇、赵亚鹏、胡留军、段向阳、秦志金、夏树强、夏宏飞、翁冬冬、黄河、黄成、黄亚坤、程义超、谢大雄、谢峰、谢绍伟、鲍有能、潘啸、魏元庆

前言

全球数字经济已进入虚实融合、智能驱动的深度转型期，数字技术与实体经济的深度融合，已成为重塑全球产业格局、驱动经济高质量发展的核心动力。移动通信技术作为数字信息基础设施的核心底座，正迎来从 5G 向 6G 代际跃迁的关键窗口期，其技术演进与产业落地，将成为决定我国数字经济核心竞争力与新质生产力培育成效的关键因素。

5G 有效支撑了人与人、人与物的高速广域连接，其承载的主流媒介形式以平面音视频为主；沉浸式通信作为 ITU-R M. 2160（IMT-2030 框架）明确定义的 6G 六大重要场景之一，旨在实现更智能、自然、真实的多模态交互，推动信息交互范式从“在线”向“在场”的根本性跨越。沉浸式通信覆盖虚实融合生产生活、全息远程协作、工业数字孪生、智能体协同等典型场景，具备强交互、高沉浸、智能化、全域化的核心特征，对通信、感知、计算等能力提出了体系化的升级需求。为满足上述需求，沉浸式通信的发展亟需突破系列新型使能技术与理论体系：首先，内生智能网络将深度融入沉浸式媒体内容的生产、分发与交互全流程，需要 AI 内生网络、语义通信等网媒融合创新基础理论与算法核心支撑；其次，沉浸式内容创作已快速进入 AIGC 时代，端网边云协同大模型技术迅猛发展，需要构建适配场景需求的通感智算存融合、算网一体化等新型网媒融合架构；再次，为支撑沉浸式通信强交互、高沉浸、智能化的核心特征，需要突破网络内生智能与确定性保障、多维感官交互与内容实时生成等网媒融合关键技术；最后，国际标准体系的统筹推进与前瞻布局，将为技术迭代与产业生态协同发展提供基础。

需要特别强调的是，沉浸式通信绝非局限于人类感官体验的升级，更涵盖智能体与智能体、人与智能体、智能体与物理世界之间的沉浸式交互——智能体作为具备专属传感系统与执行机构的核心载体，其感知维度（涵盖环境地理、材料特性、物理化学参数、电磁场等）与执行维度远超人类五官四肢，智能体间的沉浸式通信的必要性，源于复杂任务协同、跨区域资源调度、精准环境交互的核心需求，是 6G “万物智联” 时代的核心支撑，也是本蓝皮书区别于相关报告的核心亮点之一。

基于以上研究背景，中兴通讯作为牵头单位，联合清华大学、深圳大学、东南大学、上海交通大学、北京理工大学、电子科技大学、北京邮电大学、南京邮电大学等国内高等院校，邀请领域内顶尖专家学者共同完成本蓝皮书的编制工作。本蓝皮书从研究背景、基础理论、系统架构、关键技术、应用场景、标准推进、产业发展七大维度出发，对 6G 沉浸式通信技

术开展系统性分析与研究，立足“人-智能体-物理世界”三元交互视角，构建更完整、更具前瞻性的沉浸式通信理论与技术体系，旨在为领域内科研工作者、工程技术人员、行业从业者与政策制定者提供理论参考、技术指引与实践依据。

全文共设有五个核心章节。第一章详细阐述沉浸式通信研究背景与战略意义，系统分析其核心内涵、技术特征与面临的核心技术挑战；第二章围绕沉浸式通信的核心挑战，创新性提出网媒融合概念，介绍网媒融合基础理论与创新范式；第三章主要阐述沉浸式通信体系架构、关键技术等研究进展；第四章详细剖析沉浸式通信的典型应用场景与产业赋能；第五章对未来技术演进与产业发展趋势进行展望，并提出相关工作倡议。

蓝皮书仍有诸多不足之处，恳请各界批评指正。

目录

第一章 沉浸式通信研究背景与技术瓶颈.....	1
1.1 沉浸式通信研究背景与战略意义.....	1
1.1.1 下一代无线通信对沉浸式通信定位.....	1
1.1.2 沉浸式通信国际研究进展.....	2
1.1.3 我国开展 6G 沉浸式通信研究意义与进展.....	4
1.2 沉浸式通信的核心内涵与技术特征.....	5
1.2.1 沉浸式通信内涵概述.....	5
1.2.2 沉浸式通信技术特征.....	6
1.2.3 沉浸式通信智能化升维.....	9
1.2.4 沉浸式通信关键性能指标.....	10
1.3 沉浸式通信技术研究瓶颈.....	13
1.3.1 基础理论与算法构建瓶颈.....	13
1.3.2 网络标准技术演进瓶颈.....	15
1.3.3 媒体资源适配协同瓶颈.....	18
1.3.4 其他研究瓶颈.....	19
第二章 面向沉浸式通信的网媒融合基础理论与范式.....	22
2.1 面向沉浸式通信的网媒融合概述.....	22
2.2 面向沉浸式通信的网媒融合基础理论.....	23
2.2.1 面向任务/语义的通信理论.....	23
2.2.2 基于网媒融合联合编码技术.....	26
2.3 面向沉浸式通信的网媒融合新范式.....	30
2.3.1 顶层设计-用户语义导向的网媒融合架构.....	30
2.3.2 中枢技术-网络与媒体双向感知.....	31

2.3.3 底层资源-通感智算存多维资源置换	33
第三章 面向沉浸式通信的网媒融合体系架构与关键技术	35
3.1 面向沉浸式通信的网媒融合总体架构	35
3.1.1 三空间四能力总体架构	35
3.1.2 端网边云分布式部署架构	37
3.2 面向沉浸式通信网侧协同关键技术	38
3.2.1 AI 内生网络	38
3.2.2 算网融合架构	41
3.2.3 确定性网络	44
3.3 面向沉浸式通信媒侧适配关键技术	48
3.3.1 语义编解码技术	48
3.3.2 沉浸式媒体编解码	50
3.3.3 通信感知一体化	54
3.3.4 全息呈现与全感官交互	58
第四章 基于网媒融合的沉浸式通信应用场景与产业赋能	65
4.1 典型沉浸式通信应用场景	65
4.1.1 文娱与消费：重塑感官体验边界	65
4.1.2 工作与协作：打破物理空间壁垒	66
4.1.3 教育与个人：普惠与高效并行	70
4.2 产业生态与赋能路径	71
4.2.1 产业链生态	71
4.2.2 XR 智能终端	74
4.2.3 产业协同与赋能路径	76
4.2.4 标准化与国际合作	77

第五章 未来展望与工作倡议	81
5.1 未来发展展望	81
5.2 未来工作倡议	83
参考文献	84
缩略语表	92

第一章 沉浸式通信研究背景与技术瓶颈

本章首先梳理了国内外在沉浸式通信领域的战略定位与标准化进程，明确了其作为下一代通信核心方向的战略价值；其次，通过剖析其技术内涵与特征，揭示了沉浸式通信系统从数据带宽驱动向体验与智能驱动演进的本质，突破传统仅聚焦人类体验的局限，延伸到智能体、机器的感知与交互维度；最后，聚焦于制约沉浸式通信发展的关键瓶颈，如网络架构极限、多模态数据处理及服务一致性保障等核心难题。本章的工作旨在提纲挈领地指出当前技术生态中的薄弱环节，为后续章节中针对具体应用场景的技术选型、架构优化及创新方案设计提供明确的研究方向与切入点。

1.1 沉浸式通信研究背景与战略意义

沉浸式通信作为下一代移动通信系统的核心应用场景与关键使能技术，正从愿景探索阶段迈向标准化与产业化的关键时期。在国际标准层面，以 3GPP、ITU 为代表的标准化组织已将沉浸式通信纳入 5G-Advanced 及 6G 的演进蓝图，通过定义全新服务化架构与能力集，为多模态交互与全息感知奠定基础。在学术研究层面，欧美日韩等发达国家依托其先发优势，在基础理论、核心算法与原型系统上持续深耕，但现有研究多仅聚焦于人类感官体验的升级，尚未形成“人-智能体-物理世界”三元交互的完整体系；而我国凭借庞大的市场需求与完整的产业链，在应用场景创新与关键技术攻关上同样取得了显著突破，正逐步从跟随者转变为引领者，尤其在智能体协同、通感智算存融合等领域的前瞻布局，有望构建更具完整性与先进性的沉浸式通信技术体系。

1.1.1 下一代无线通信对沉浸式通信定位

移动通信技术代际升级的本质，都是通过传输能力的颠覆性突破催生核心应用场景迭代，最终实现信息交互范式的变革。从 1G 到 5G，移动通信系统完成了从模拟语音到数字宽带、从人与人通信到万物互联的代际跨越，其承载的主流媒介始终以平面化的音视频信息为主。6G 技术将突破传统通信的能力边界与应用范畴，随着大模型、人工智能技术的发展，以 OpenClaw 为代表的智能体应用开启从“万物互联”到“万物智联、数字孪生”的全新发展

阶段^[1]，实现信息交互范式从“平面信息在线传递”到“全感官空间在场交互”的根本性跨越，实现人与人、人与智能体、智能体与智能体之间的通信。如图 1，沉浸式通信作为 ITU-R M. 2160（IMT-2030 框架）明确定义的 6G 六大核心应用场景之首^[2]，正是实现 6G 范式跃迁的核心载体之一。

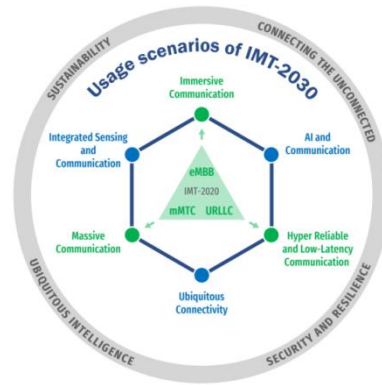


图 1 IMT-2030 定义的 6G 应用场景

更进一步地，沉浸式通信的内涵，远不止于人类感官维度的逼真复现，更本质地体现在智能体与机器之间“状态空间的无缝融合”——这正是 6G 通感智算存一体化的核心内涵。对人类而言，沉浸指向感官的真实；对智能体而言，沉浸指向内在世界模型的同步。每个智能体都内嵌关于物理规律、因果结构与意图偏好的隐变量，而真正的沉浸式通信，不再是传输“数据”，而是通过交换不同空间维度的词元，实时对齐并融合各智能体的意图隐变量、世界模型隐变量、认知负荷隐变量与因果结构隐变量，统一可执行的因果模型。由此，通信系统使智能体不再“接收外部信息”，而是直接嵌入一个共同的或者多个不同维度的、可预测的、可干预的隐状态空间，并使智能体在其中行动与演化。简言之：人类沉浸，是感官的延伸；智能体沉浸，是因果的共融。6G 的使命，正是让不同智能体的世界模型实现实时对齐与动态演化，从而开启真正的群体智能沉浸时代。

沉浸式通信的核心是通过多感知主体（人、智能体）的多模态信息全流程采集、传输、处理与交互，打破物理空间、感知边界与主体壁垒，提供人-人、人-智能体、智能体-智能体、智能体到物理世界之间的多维度信息融合、虚实空间无缝衔接、AI 智能服务、具身交互的真实交互体验，其中智能体凭借其更宽广、更精密的传感与执行能力，成为沉浸式通信向产业深度渗透的核心载体。

1.1.2 沉浸式通信国际研究进展

6G 移动通信技术作为未来数字信息基础设施的核心底座，是大国科技博弈的核心制高点。沉浸式通信作为 6G 研发的核心方向，全球主要经济体也均在抓紧布局，但现有国际研究多聚焦于人类感官体验（如 XR、全息通信），尚未形成涵盖智能体、机器感知的完整沉浸式通信体系，这也为我国实现技术引领提供了重要机遇。

2025 年 12 月，美国白宫发布《国家安全总统备忘录（NSPM-8）》，明确指出 6G 是未来十年尖端技术运行的基础^[3]。早在 2022-2024 年，美国电信行业解决方案联盟（ATIS）Next G Alliance 发布的《Roadmap to 6G》、《Multi-Sensory Extended Reality (XR) in 6G》、《North American 6G Roadmap Priorities》多份报告中^{[4][5][6]}，就将沉浸式体验列为 6G 六大核心能力之一，细化了面向 6G 时代的沉浸式技术，并指出其潜在的社会影响是深远的，但核心仍聚焦于人类多感官体验的升级，未涉及智能体间的沉浸式交互与机器感知维度。2025 年 3 月，美议员向国会提交《2025 年美国沉浸式技术领导力法案》，旨在确保美国在该领域领导地位。2025 年 6 月，美国国会研究服务局（CRS）发布的报告指出，可信人工智能、XR 等沉浸式技术的近期进展扩大了其在美国军方的应用范围，国防部已将相关技术列为国防部 14 个关键技术领域之一^[7]，其应用场景仍以人类视觉体验使用为主。

欧洲同样希望确保其在沉浸式通信领域的技术主权。欧洲委员会明确提出通过 6G 技术实现 XR、全息通信等更具挑战性的应用，并通过“地平线欧洲（Horizon Europe）”计划，持续投入数亿欧元研究智能交互与下一代互联网的结合^[8]。2023 年欧盟启动 Hexa-X-II（6G 旗舰项目第二阶段），开展 6G 关键使能技术研究。此外，欧盟还于 2021 年设立智能网络和服务联合伙伴项目（SNS JU 项目），支持全息通信、数字孪生在内的多个沉浸式通信技术项目研究。欧洲企业也在持续推进沉浸式通信技术研究，2024 年，欧洲 LSE 咨询报告整合爱立信、Meta、宝马等多家领军企业的观点，指出沉浸式技术对欧洲工业数字化转型至关重要，要确保欧洲在沉浸式技术领导地位^[9]。2025 年 11 月，诺基亚与爱立信联手开展 6G 视频技术研究；2026 年 3 月，诺基亚与德国弗劳恩霍夫集成电路研究所（Fraunhofer IIS）合作在世界移动通信大会（MWC）上展示了沉浸式语音/音频互通实时通话演示，同时配套展示 6G-XR 沉浸式空间音频体验，其核心仍围绕人类感官体验展开。

除美欧外，韩国、日本等主要经济体也将沉浸式通信作为 6G 研发的核心赛道，形成了全球竞速的发展格局。2023 年，韩国政府提出《K-Network 2030》计划，提出要确保掌握全球领先的 6G 技术。韩国政府还曾发布《沉浸式经济发展战略》，旨在开启延伸至虚拟领域的全新经济范式^[10]，重点聚焦消费端的人类沉浸体验。2024 年 8 月，日本政府发布《Beyond

5G 推进战略 2.0——支持 AI 社会的下一代信息和通信基础设施》，明确提出打造低延迟、高可靠性和低功耗的下一代信息和通信基础设施，实现通信、智能、算力及感知融合以及泛在连接，虽提及感知融合，但未明确延伸至智能体间的沉浸式交互。

整体来看，沉浸式通信是 6G 技术的标志性应用场景，是 6G 技术演进的核心牵引方向之一，全球围绕该赛道的技术攻关、专利布局、标准制定已进入白热化竞速阶段。但现有国际研究普遍多聚焦于人类感官体验升级，未涵盖智能体、机器的感知与交互需求，未形成“人-智能体-物理世界”三元协同的完整体系。这种研究缺口，恰恰为我国提供了技术引领的重要机遇——本蓝皮书立足更完整的沉浸式通信定义，将智能体、机器感知纳入核心范畴，构建更具前瞻性与完整性的理论、技术与应用体系，有望推动我国在 6G 沉浸式通信领域实现从跟跑到领跑的跨越，抢占全球技术制高点。

1.1.3 我国开展 6G 沉浸式通信研究意义与进展

早在 2023 年工信部等五部门联合印发的《元宇宙产业创新发展三年行动计划（2023—2025 年）》中明确指出：“在生活消费和公共服务领域形成一批沉浸交互的标杆应用，通过场景建设带动关键核心技术的突破”^[11]。2025 年 3 月，政府工作报告首次提出“建立未来产业投入增长机制，培育 6G 等未来产业”。2025 年 10 月，“十五五”规划建议文件明确提出“推动 6G 等成为新的经济增长点”。6G 已跃升为国家中长期发展战略的核心组成部分。2026 年 3 月，政府工作报告又提出“打造智能经济新形态，深化拓展‘人工智能+’，促进新一代智能终端和智能体加快推广”。沉浸式通信覆盖虚实融合生产生活、全息远程协作、工业数字孪生、智能体协同等典型场景，具备强交互、高沉浸、智能化的核心特征，开展 6G 沉浸式通信研究既是推动 6G 技术发展的必然要求，也是深化扩展“人工智能+”的具体行动，更是我国抢占 6G 技术制高点、培育新质生产力的关键举措。

中国在 2019 年成立 IMT-2030(6G) 推进组，启动了对下一代通信技术的系统性研究。2021 年，IMT-2030(6G) 推进组发布《6G 总体愿景与潜在关键技术》白皮书，创新提出 6G 潜在业务应用呈现出沉浸化、智能化和全域化特点。2022 年，推进组发布的《6G 总体愿景和关键能力》白皮书提出，沉浸式通信包含了沉浸式云 XR、全息通信、感官互联等多个应用场景，并通过基于典型场景和潜在业务应用，分析了对 6G 网络通信、计算、智能、感知、安全等关键能力的全新需求^[1]。与国际同类研究相比，我们认为应该逐步将智能体、机器感知纳入

沉浸式通信的研究范畴，为构建更完整的体系奠定了基础。

6G 沉浸式通信作为典型的新质生产力代表，体现了数字经济与实体经济的深度融合，不仅能推动消费端体验升级，更能通过智能体间的沉浸式协同，赋能工业制造、医疗健康等千行百业的数字化转型，对我国数字经济高质量发展具有核心战略价值。我国依托完整的产业链优势、庞大的应用场景需求以及在 AI、通感融合等领域的技术积累，有望在沉浸式通信领域构建独特竞争优势，实现技术引领与产业落地的双重突破。

1.2 沉浸式通信的核心内涵与技术特征

6G 沉浸式通信并非 5G 通信的简单速率升级，而是以虚实融合，全感交互，全域可达为核心目标，打破物理空间、时间与感知体验的边界，重构人与人、人与智能体、智能体与智能体、智能体与物理世界的交互模式，核心是基于“任务定义的世界认知”，围绕具体任务需求，实现不同感知主体（人、智能体）与物理世界之间的精准感知、高效传输与自然交互。它是 6G 通感智算存深度融合技术的核心应用场景之一，也是推动通信从单一传输向信息综合服务转型的关键载体。其核心内涵是通过具身智能、智能交互等技术创新，让人类从“观看”、“聆听”的被动接收者，转变为“在场”、“参与”的主动交互者，让智能体实现精准的环境感知与高效的协同交互，实现“天涯若比邻”的真实沉浸体验与“万物智联”的协同高效，同时赋能千行百业的数字化、智能化深度转型，构建跨领域融合的全新产业生态。

1.2.1 沉浸式通信内涵概述

6G 沉浸式通信的核心内涵可概括为“三维升维、双向协同、全域适配”，区别于 5G 时代的浅层沉浸式体验，也区别于现有国际报告仅聚焦人类体验的局限，实现了从功能满足到体验升维、从单一主体到多元协同的根本性变革。具体可分为三个层面展开：

其一，感知维度升维：突破 5G 时代视觉、听觉为主的单一感官体验，整合多维度感官数据，打破物理与虚拟世界的壁垒；也包括智能体与机器的感知升维，基于任务定义的世界认知，根据具体任务需求，实现对环境地理、材料特性、物理化学参数、电磁场等多维度信息的精准感知，其感知深度与广度远超人类感官。这也是全感官化、虚实一体化与多元协同特征的核心落地体现。例如，在元宇宙场景中^[12]，用户可真实感知虚拟环境的温度、海风的触感，甚至咖啡的香气，既能获得全感官的沉浸反馈，又能实现虚实场景的无缝切换；而场

景中的智能体可同步感知环境的湿度、气压、电磁场变化，为用户提供精准的环境预警与场景适配服务，实现人与智能体、智能体与物理世界的协同感知。再如，工业数字孪生场景中，智能体可感知设备的振动频率、温度变化、材料损耗等细微参数，实时传输至云端与数据分析智能体，实现设备故障的提前预判与精准维护。

其二，交互模式升维：摆脱传统通信“发送-接收”的单向逻辑，依托 6G 极致低时延技术，实现人类、智能体、机器之间的低时延、高可靠双向交互，确保动作、表情、触觉、词元（Token）、环境感知数据等反馈信息的同步传递，达成“无卡顿、无延迟”的实时交互体验，满足交互实时化的核心要求。每个智能体除了有智能以外，也有它专属的传感系统去感知和认知周边环境（物理世界），也有执行机构去作用于周边环境（物理世界），从而与物理世界有其视角下的交互，其感知维度和执行维度比人的五官和四肢要宽广和精密得多。为了更好的完成智能体在物理世界中的任务，及智能体之间的协作、对抗等任务，面向智能体、人-智能体，智能体间的沉浸式通信，将具有更宽广和深远的应用前景。例如，全息会议中^[13]，参会者的手势、表情可实时同步，握手时能感知对方的温度与力度，实现“面对面”的真实沟通体验；远程手术中，医生的操作指令可实时传递至机械臂（智能体），同时接收机械臂反馈的触觉信息，实现毫米级精准控制；智能体协同巡检场景中，多个巡检智能体可实时共享环境感知数据、任务进度，协同完成复杂区域的巡检任务，实现智能体间的沉浸式协同交互。

其三，服务范围升维：从“局部覆盖”到“全域无缝适配”，从“单一主体服务”到“多元主体适配”。依托 6G 空天地一体化网络^[14]，打破地面网络的覆盖局限，实现海洋、沙漠、高空等偏远地区的沉浸式通信全覆盖，让任何场景、任何终端、任何智能体都能接入沉浸式服务。同时，适配个人消费、工业制造、医疗健康、教育培训、智能体协同等多领域需求，实现“一人一策、一业一景、一体一案”的个性化、场景化沉浸服务，推动沉浸式体验从娱乐社交向产业应用、智能体协同深度延伸。

1.2.2 沉浸式通信技术特征

6G 沉浸式通信与 5G 浅层沉浸核心区别集中体现在全感官化、交互实时化、虚实一体化、智能内生四大特征^[15]，同时与国际现有研究的核心区别，在于增加多元协同，“人-智能体-物理世界”的三元交互体系。五大特征相互支撑、协同发力，共同构成 6G 沉浸式体验的

核心支撑，也是技术落地与业务开展的核心导向。

1. 全感官化：重构沉浸体验的感官边界，延伸多元感知维度

全感官化是 6G 沉浸式通信的基础特征，核心是突破 5G 时代“视觉+听觉”的二元感官局限^[16]，同时打破“仅聚焦人类感官”的认知局限，实现“人类全感官+智能体多维度感知”的双重覆盖与协同反馈——对人类而言，实现视觉、听觉、触觉、嗅觉、味觉五大感官的全面覆盖，让虚拟体验具备物理世界的真实感官质感；对智能体而言，实现环境地理、材料特性、物理化学参数、电磁场等多维度的精准感知，满足不同任务场景的感知需求。其核心要求是多感知主体、多维度数据的精准采集、高效传输与高保真还原，既要通过专用传感器捕捉人类触觉压力、嗅觉分子、味觉信号等数据，也要通过智能体专属传感系统捕捉环境多维度参数，依托 6G 高速率、高保真传输能力，将这些数据实时传递至终端设备与智能体，最终实现人类全感官体验与智能体精准感知的协同落地。例如，在远程文旅场景中，用户不仅能看到景区的高清画面、听到现场的声音，还能通过触觉设备感知微风的触感，通过嗅觉设备闻到花香、草木香；而场景中的智能体可感知景区的温度、湿度、人流密度，为用户推荐最优游览路线，实现人与智能体的协同沉浸体验。在工业检修场景中，检修人员可通过触觉反馈感知设备的振动、温度，精准判断设备运行状态；智能体可同步感知设备内部的压力、电流变化，辅助检修人员定位故障点，提升检修效率与准确性。全感官化的落地，离不开感官数据保真度、峰值传输速率等关键性能指标的支撑，是 6G 沉浸式体验区别于 5G、区别于国际现有研究的核心标志之一。

2. 交互实时化：保障沉浸体验的自然性，支撑多元主体协同

交互实时化是实现沉浸式通信体验的核心保障，关键在于达成“无延迟、无卡顿”的多主体双向实时交互^[16]，摆脱传统通信“单向传输”的局限，让人类的动作、指令，智能体的感知数据、执行反馈，能够同步传递，确保沉浸式体验的连贯性与自然性。其核心要求是极致的端到端时延与高可靠性，需将端到端时延降至毫秒级（ $\leq 1\text{ ms}$ ），同时保障通信可靠性达到 99.9999%，避免因时延过高或信号中断导致沉浸式体验中断、智能体协同失效。例如，全息视频沟通交流中，交流双方的表情、手势、语气可实时同步，不出现“音画不同步”的问题，握手时能实时感知对方的温度与力度；虚拟协同办公场景中，多个用户与智能体可实时操作同一个虚拟模型，动作与反馈同步，仿佛身处同一办公空间；智能体协同作业场景中，巡检智能体与调度智能体可实时共享感知数据与任务状态，调度指令与执行反馈无延迟传递，

确保协同任务高效完成。交互实时化的实现主要依赖 6G 超高速率、低时延技术、高可靠性传输及多模态数据同步和分布式算力能力，是交互模式从“被动接收”向“主动参与”、从“单一主体”向“多元协同”转型的关键。

3. 虚实一体化：打破物理与虚拟的边界，实现三元世界联动

虚实一体化是沉浸式通信的核心目标，关键在于实现物理世界与虚拟世界的深度融合、无缝切换，让虚拟场景具备物理世界的真实规律，让物理场景可通过虚拟技术实现延伸与拓展，构建“虚实共生”的全新交互空间^[17]。同时让智能体成为连接虚实世界的核心载体，其核心要求是实现物理数据、虚拟数据、智能体感知数据的实时映射、协同联动，通过数字孪生、通感一体等技术，将物理世界的人、场、物，智能体的感知与执行状态，实时映射到虚拟空间，同时让虚拟空间的操作可反向作用于物理世界与智能体，实现“虚实互控、体物联动”。例如，在数字孪生工厂中，通过 6G 网络将工厂的生产设备、生产流程，以及巡检智能体的感知数据、执行状态，实时映射到虚拟空间，工作人员可在虚拟场景中模拟设备操作、排查故障，操作指令实时同步至物理设备与智能体，既降低了操作风险，又提升了生产效率；在智慧交通场景中，虚拟交通模型与物理交通场景、交通管控智能体实时联动，智能体可感知实时交通流数据，虚拟模型可提前预判交通拥堵、事故风险，通过虚拟指令引导物理交通流与智能体调度，实现交通的智能管控。虚实一体化的落地，需要峰值传输速率、连接密度等网络关键性能指标的支撑，同时依赖数字孪生、通感智算存融合等技术的协同发力。

4. 智能内生：赋能沉浸体验的智能化升级，支撑任务精准完成

智能内生是 6G 沉浸式通信的核心技术支撑，关键在于让智能不再是外部附加功能，而是贯穿网络全流程、深度融入沉浸式体验、支撑多元主体协同的内生能力，实现“场景自适应、服务个性化、交互智能化”。其核心要求是将 AI、大数据、云计算等数智技术与通信技术深度融合，贯穿数据采集、传输、处理、反馈全流程，实现网络资源的动态优化、多模态数据的智能解析、个性化服务的精准推送^[18]。例如，在多用户+智能体全息会议场景中，AI 可智能识别发言者，优先分配带宽资源，保障发言者的动作、语音同步，同时智能过滤背景噪音、优化画面质量，同步调度会议辅助智能体，实时整理会议纪要、推送相关资料；在个性化沉浸服务场景中，AI 可根据用户的使用习惯、偏好，智能推送适配的虚拟场景、感官反馈模式，同时调度环境感知智能体，实时适配场景参数，实现“一人一策”的个性化体验；在网络运行过程中，AI 可实时监测网络状态，智能调整信号路由、压缩数据体积，

动态适配人类体验与智能体感知的不同需求，确保用户体验稳定性与智能体任务可靠性。智能内生化的实现，依赖于通感智算存深度融合技术，是 6G 沉浸式通信从“技术可行”向“体验最优、任务高效”转型的关键。

5. 多元协同：构建“人-智能体-物理世界”三元交互体系

多元协同是本蓝皮书将现有国际同类研究“仅聚焦人类体验”的拓展，实现人类、智能体、物理世界三者之间的协同感知、协同交互、协同完成任务。智能体作为核心协同载体，凭借其专属传感系统与执行机构，成为连接人类与物理世界的桥梁，既能够精准感知物理世界的多维度参数，为人类提供更全面的环境信息与任务辅助，也能够接收人类指令，作用于物理世界，同时实现智能体之间的高效协同，完成人类难以独立完成的复杂任务。例如，远程医疗场景中，医生（人类）通过沉浸式终端接收患者的生理数据，向手术智能体发送操作指令，手术智能体执行手术操作，并实时反馈手术部位的触觉、压力等数据，同时环境监测智能体实时感知手术室的温度、湿度、无菌环境参数，确保手术安全，实现人-智能体-物理世界的三元协同；工业巡检场景中，多个巡检智能体协同作业，分别负责设备参数感知、环境安全监测、故障定位，将感知数据实时传输至云端与人类操作员，人类操作员通过虚拟终端实时查看巡检状态，向智能体发送调度指令，实现多元协同巡检。多元协同的落地，需要突破智能体间通信协议、多主体数据同步、资源协同调度等关键技术，是 6G 沉浸式通信向产业深度赋能的核心支撑。

1.2.3 沉浸式通信智能化升维

沉浸式通信的内涵，远不止于人类感官维度的逼真复现，更本质地体现在智能体与机器之间状态空间的无缝融合，这也是 6G 通感智算存一体化的核心内涵。

对人类而言，沉浸指向感官的真实。无论是高保真视频、空间音频还是触觉反馈，其目标都是让人的视觉、听觉、触觉等外感知通道获得“临场感”。然而，对智能体，比如，自动驾驶车、工业机器人、AI 数字员工等，没有皮肤、没有视网膜，其“沉浸”则建立在另一套底层逻辑之上：内在世界模型的同步。

每个智能体都内嵌一个关于物理规律、因果结构与意图偏好的隐变量空间。这些隐变量无法被直接观测或打包成传统数据包，却决定了智能体如何感知环境、预测未来、做出决策。例如，一辆自动驾驶车不仅需要知道“前方有障碍物”，还需要传递其对障碍物运动轨迹的

不确定度、对路面摩擦系数的信念、以及紧急制动时对后方车辆反应的因果预期。这些隐变量的交换和对齐，才是智能体之间真正的“通信”。

传统的通信系统传输的是显式的符号与感知数据：图像、点云、指令等。而在 6G 通感智算存一体化的架构下，沉浸式通信的本质发生跃迁，通过交换不同空间维度的词元(Token)，实时对齐并融合各智能体的意图隐变量、世界模型隐变量、认知负荷隐变量与因果结构隐变量，统一为可执行的因果模型。这里的词元并非自然语言的单词，而是高维隐空间中的基本单元。词元可以是一个因果关系的压缩表示、一个意图的潜在编码、或一个世界状态的抽象快照。通过通感算存融合，网络不再是被动的管道，而是主动的隐空间对齐引擎，它能够识别、映射、同步不同智能体内部状态的不确定度、注意力焦点、资源余量与因果图，实现智能体之间的“沉浸式通信”。

由此，通信系统使智能体不再只是单纯接收外部信息，而是使其嵌入一个共同的、或者多个不同维度的、可预测且可干预的隐状态空间，并使智能体在其中行动与演化。这意味着：当一个智能体执行动作时，它同时感知到其他智能体对这一动作的因果反事实推演；当一个智能体更新信念时，其他相关智能体的世界模型能够近乎实时地收敛到一致，实现无缝融合，即没有协议握手、没有显式询问，智能体仿佛生长在一个共享的因果场中。

人类沉浸，是感官的延伸；智能体沉浸，是因果的共融。6G 的使命，正是让不同智能体的世界模型实现实时对齐与动态演化，从而开启真正的群体智能沉浸时代。在沉浸式通信时代，通信不再是信息的搬运，而是智能的耦合；网络也不再是连接终端的介质，而是分布式智能体的共同意识场。

1.2.4 沉浸式通信关键性能指标

6G 沉浸式通信四大特征的落地、业务的规模化开展，需依托明确的关键性能指标(KPI)作为支撑，这些指标直接决定沉浸式体验的优劣与服务的可用性，围绕“沉浸体验、服务质量、场景适配”三大核心需求设定 KPI 指标，是 6G 沉浸式通信区别于 5G 沉浸式业务的核心标志^[19]。同时本蓝皮书从人类沉浸体验的场景进一步延伸到智能体沉浸交互。

1.2.4.1 沉浸式通信业务指标

以沉浸式通信典型应用场景为例，我们挑选更高要求的性能指标进行计算。各场景的性

能指标如表 1 所示：

表 1 沉浸式通信业务指标

典型场景	用户体验速率	端到端时延	可靠性	感官数据保真度
沉浸式多人协作	下行：81 Mbps (4K/ 60 fps/ 300 压缩比/12 比特每色) 上行：10 Mbps	45 ms, 其中业务时延 40 ms	99.99%	多模态数据的保真度 $\geq 99\%$
远程全息会议	下行：972 Mbps (16K/ 120 fps/ 300 压缩比/12 比特每色) 上行：300 Mbps (体积视频)	35 ms, 其中业务时延 30 ms	99.99%	多模态数据的保真度 $\geq 99\%$
沉浸式学习	下行：972 Mbps (16K/ 120 fps/ 300 压缩比/12 比特每色) 上行：100 Mbps (多模态数据)	35 ms, 其中业务时延 30 ms	99.99%	多模态数据的保真度 $\geq 99\%$
虚拟现实电影	下行：972 Mbps (16K/ 120 fps/ 300 压缩比/12 比特每色) 上行：100 Mbps (多模态数据)	35 ms, 其中业务时延 30 ms	99.99%	多模态数据的保真度 $\geq 99\%$ 、触觉反馈时延 ≤ 1 ms、压力分辨率 ≤ 1 kPa
全息演唱会	下行：972 Mbps (16K/ 120 fps/ 300 压缩比/12 比特每色) 上行：300 Mbps (体积视频)	35 ms, 其中业务时延 30 ms	99.99%	多模态数据的保真度 $\geq 99\%$
远程专家指导	下行：81 Mbps (4K/ 60 fps/ 300 压缩比/12 比特每色) 上行：81 Mbps (4K/ 60 fps/ 300 压缩比/12 比特每色)	45 ms, 其中业务时延 40 ms	99.9999%	多模态数据的保真度 $\geq 99\%$
基于 AI 助手的机房智能巡检	下行：81 Mbps (4K/ 60 fps/ 300 压缩比/12 比特每色) 上行：81 Mbps (4K/ 60 fps/ 300 压缩比/12 比特每色)	45 ms, 其中业务时延 40 ms	99.9999%	多模态数据的保真度 $\geq 99\%$
沉浸式远程诊疗	下行：972 Mbps (16K/ 120 fps/ 300 压缩比	35 ms, 其中业务时延 30 ms	99.9999%	多模态数据的保真度 $\geq 99\%$

典型场景	用户体验速率	端到端时延	可靠性	感官数据保真度
	/12 比特每色) 上行: 100 Mbps (多模态数据)			
智能体间协同巡检 (预留)	待细化 (根据巡检任务需求设定)	待细化 (建议 ≤ 5 ms)	待细化 (建议 $\geq 99.9999\%$)	待细化 (环境参数感知保真度 $\geq 99.5\%$)
工业数字孪生协同 (预留)	待细化 (根据生产场景需求设定)	待细化 (建议 ≤ 1 ms)	待细化 (建议 $\geq 99.9999\%$)	待细化 (设备感知数据保真度 $\geq 99.8\%$)

1.2.4.2 沉浸式通信网络指标

基于沉浸式通信业务指标，沉浸式通信网络需要满足如下指标要求：

1. 用户体验速率：作为全感官数据、全息数据传输的基础，6G 沉浸式通信要求用户体验速率达到 1-10 Gbps，可满足 8K/16K 全息视频、多通道触觉反馈、嗅觉数据等海量多模态数据的实时传输需求，有效避免因速率不足造成内容加载卡顿、感官反馈滞后等问题，是全感官化、虚实一体化特征落地的核心保障。

2. 端到端时延：为实现交互实时化，6G 沉浸式通信要求端到端时延降至毫秒级 (≤ 1 ms)，可确保手势、表情、操作指令与感官反馈的同步传递，彻底解决 5G 时代沉浸式交互中“声画不同步”、“触感滞后”等问题，保障交互的自然性与真实性。

3. 可靠性：针对多用户、多终端协同沉浸场景（如多人全息会议、工业元宇宙协同操作），通信可靠性需达到 99.9999%（六个 9），避免因信号中断、数据丢失造成沉浸体验中断，满足高密度场景下的协同沉浸需求，契合全域服务的核心要求。

4. 感官/感知数据保真度：作为全感官化与多元协同的核心指标，一方面要求人类视觉、听觉、触觉、嗅觉等多模态数据的保真度 $\geq 99\%$ ，其中视觉数据分辨率需达到 16K、帧率 ≥ 120 fps，确保全息影像的清晰度与流畅度；触觉反馈时延 ≤ 1 ms、压力分辨率 ≤ 1 kPa，实现真实的触感还原；嗅觉、味觉数据需精准还原气味、味道的核心特征。另一方面，要求智能体环境感知数据的保真度 $\geq 99.5\%$ ，确保智能体对环境地理、材料特性、物理化学参数等信息的精准感知，为任务落地提供可靠数据支撑，推动虚实一体化与多元协同的深度落地。

1.3 沉浸式通信技术研究瓶颈

从技术内涵与特征来看，沉浸式通信已不再局限于传统音视频传输的带宽与速率提升，而是深度融合了人工智能、感知交互、算力网络等跨学科技术，致力于构建虚实融合的无缝体验。然而，这一演进过程也面临着显著的研究瓶颈。下面本文将从基础理论与算法，网络标准技术演进，新兴媒体传输需求，以及能效、安全等其他方面进行全面阐述。

1.3.1 基础理论与算法构建瓶颈

为破解沉浸式通信对超高清视频、三维全息、空间音频等媒体数据的传输挑战，亟需在网络架构、传输机制与能力拓展等维度实现系统性创新，通过构建网络与媒体深度融合的功能体系，打通网络传输、媒体处理、终端呈现全链路协同，打造具备弹性容量供给、多维资源调度、智能适配优化的信息服务基础设施。这种网媒融合模式，能够针对性解决沉浸式通信中媒体数据体量巨大、实时交互要求严苛、终端异构性突出等问题，以技术融合突破体验瓶颈，以效率提升替代资源堆叠，不仅能够为通信行业注入新质生产力，更能全面支撑数字经济时代沉浸式办公、全息社交、虚拟临场等场景的高质量发展，推动通信服务从连接可达向体验卓越的跨越式升级。其核心挑战在于：如何在异质网络环境中实现媒体内容的高效表征、智能传输与价值适配？这需要一套涵盖联合编码、语义通信、差异化传输、双向感知与用户导向架构的基础理论与算法体系。

1. 从“比特传输”到“语义传递”的范式跨越：语义级评价测量体系的缺失

传统通信系统的核心目标是保证比特级的无误传输，其性能由香农定理所界定。然而，在6G时代的沉浸式媒体、工业互联网等场景中，传输的最终目的是传递信息的“意义”并完成特定任务^[20]。这要求通信范式从“比特保真”转向“语义有效”。核心挑战在于，目前缺乏一套公认的、量化的语义级评价测量体系。我们无法简单地用信噪比或误码率来衡量一个生成模型是否成功重建了发送端的意图，也无法精确评估在信道质量下降时，系统究竟丢失了多少“意义”。没有这样的测量体系，语义通信的研究就缺乏统一的目标和比较的基准，难以从概念走向工程实践^[21]。

2. “悬崖效应”的克服与“优雅降级”的实现：如何在极端的异质网络环境中保障传输质量？

传统通信系统遵循香农分离定理，将信源压缩与信道编码独立优化。这在媒体内容传输中导致了“悬崖效应”：当信道质量降至某个阈值以下时，接收端重建内容的质量会急剧恶化，甚至完全无法解码^[22]。在由不同网络、不同终端构成的复杂环境中，信道质量动态变化，这种效应尤为致命。核心挑战在于，如何设计一种新的编码传输机制，使其能够感知信道状态，并实现重建质量随信道质量下降而“优雅降级”，而非“悬崖式”崩溃，从而在资源受限时优先保障核心语义信息的送达。

3. 内容价值非均匀分布与网络资源非对称配置的错配：如何识别并优先保障“高价值语义单元”？

媒体内容的价值在时间和空间维度上呈现显著非均匀分布^[23]。一帧图像中的关键目标、一段视频中的关键事件、一篇报道中的核心观点，构成了信息的“高价值语义单元”。然而，传统通信网络对所有数据一视同仁，将大量资源用于保护对最终任务贡献甚微的比特，造成了资源的巨大浪费。核心挑战在于，如何让网络“理解”内容的语义价值，并据此实现通信资源的非对称配置？这首先要求我们能够对媒体内容的语义重要性进行实时、准确的量化评估，其次需要设计一套基于语义优先级的资源调度机制，在信道拥塞或资源受限时，能够主动丢弃或压缩低价值信息，确保高价值语义单元的可靠传输。

4. 网络与媒体的“双向感知”鸿沟：如何从静态配置走向动态协同？

要实现上述差异化传输，网络和媒体必须打破各自为政的壁垒。一方面，网络需要感知其所传输内容的语义特征与服务需求，以便做出更明智的资源调度决策。另一方面，媒体内容的编码和生成策略也需要感知网络的实时状态与资源约束，动态调整自身的压缩率和表征方式。核心挑战在于，如何建立网络与媒体之间的双向感知机制，形成一种动态协同的闭环。这需要建立媒体内容特征与网络资源需求的映射模型，同时实现编码策略对网络状态的自适应调整。

5. 以“智能用户”为中心的融合愿景与多维资源异构性的矛盾：如何构建一体化的资源底座？

网媒融合的最终指向是用户，而用户的范畴已从传统的人拓展至 AI 代理、物联网终端、工业机器人等广泛的智能主体。服务这些智能主体的语义需求，需要打通通信、感知、智能、计算、存储等多维资源，形成一个一体化的资源底座。核心挑战在于，这些资源在属性、粒度、时空分布上具有高度的异构性，如何将它们高效地协同起来，使语义信息能够在五大维

度之间无缝流动、协同处理与智能适配？这要求我们超越传统的网络架构，设计一种全新的、以语义为纽带的资源管理与编排体系。

网媒融合的理论演进与技术突破，正在重塑信息传播的基础逻辑。从联合编码的端到端优化，到语义通信的意义传递，再到差异化传输的资源适配、双向感知的控制闭环，最终指向用户语义导向的信息服务一体化架构。展望未来，6G 通信与生成式 AI 的深度融合将进一步推动这一进程。以用户语义为中心的语义通信网络，将能够理解内容的意义、感知网络的状态、洞察用户的需求。在这一图景中，网媒融合不再是技术与内容的简单叠加，而是信息传播生态的系统性重构——其最终目标，是让每一比特的传输都有其意义，让每一次连接都服务于人的需求。

1.3.2 网络标准技术演进瓶颈

针对 6G 沉浸式通信，RAN 面临如下五大核心挑战^{[24][25]}：

1. **极致吞吐量与小区容量的挑战。**随着沉浸式业务向高分辨率、高帧率演进，单用户速率需求呈爆炸式增长，而速率提升会显著降低单小区用户承载能力，甚至导致容量大幅下降。6G RAN 需在满足超高带宽需求的同时保障多用户并发接入，要求物理层技术与频谱利用效率实现质的突破。

2. **确定性极低时延与调度灵敏度挑战。**沉浸式游戏、动作姿态捕捉、分布式渲染等业务对端到端时延提出更为严苛的要求，而时延预算的压缩会成倍增加调度难度、降低系统容量，同时时延抖动也将严重影响用户沉浸体验。6G RAN 需构建高效的确定性调度机制，缓解高吞吐量场景下的排队积压与时延波动问题。

3. **高可靠性与移动性体验连续性挑战。**关键场景对连接可靠性要求极高，移动状态下的切换中断、速率波动会引发丢包及时延波动等问题，在高帧率场景下更易造成画面失真。6G RAN 须进一步优化移动过程中的速率中断、性能波动等问题。

4. **终端功耗与能效的平衡挑战。**沉浸式设备对续航高度敏感，6G 业务数据量激增会使终端射频链路长时间处于激活状态，对节能提出更高要求。6G RAN 需原生支持节能机制，在不牺牲用户体验的前提下降低终端无效能耗。

5. **沉浸式通信集成音视频、动作指令、控制信号及感知数据等多类数据流，各类数据**

流的同步传输是保障沉浸式体验的核心前提。在多用户沉浸式互动（如互动游戏）等场景中，还需实现不同用户间多流业务的高效协同。这就要求 6G RAN 进一步优化多流协同传输机制，兼顾单用户内部数据流同步与多用户间业务协同。

此外，沉浸式通信的业务形态正由单纯的音视频传输，演进为集成触觉反馈、动作位姿、环境感知及 AI 计算能力的多模态综合体验。系统通过跨域融合，将通信、感知、计算、存储与 AI 能力深度集成，提升用户体验。例如利用分离渲染技术，将沉浸式业务任务动态卸载至边缘或云端的服务托管环境（SHE），从而突破轻量化 XR 终端在算力、续航与散热方面的硬件瓶颈；在接入技术方面，需融合地面网络（TN）与非地面网络（NTN），实现泛在覆盖与持续稳定的体验质量保障。

沉浸式通信在 6G 体系下的跨域融合同样面临多维度严峻挑战。首先，通信、感知、计算、存储与 AI 能力的深度集成带来了极高的系统复杂性，现有架构中网络功能（NF）间的强耦合关系，限制了新特性的灵活扩展以及多厂商环境下互操作性的快速验证。其次，确保多模态数据（如音视频、触觉、空间位姿）在异构终端与网络域间的精准时间同步，并在极低时延约束下完成计算任务卸载与分离渲染，对系统确定性传输能力提出了严苛要求。同时，轻量化终端在续航、散热与高性能处理需求之间的矛盾，仍是制约业务泛在普及的核心硬件瓶颈。在安全与隐私层面，超高带宽场景下实时加解密算法的性能瓶颈，以及用户敏感生物特征信息和环境空间数据在不同管理域间的流动，引发了对数字身份认证与数据主权保护的深层担忧。此外，地面网络与卫星等 NTN 网络的深度融合，仍需在频谱协调、全球统一标准及跨行业协作方面打破既有的信息孤岛与管理壁垒。

未来通信将从人-人通信向人-智能体、智能体-智能体通信演进，从真身通信向化身、分身通信演进。要实现 6G 内生的实时沉浸式通信业务，离不开核心网的强力支撑，核心网在架构设计上面临如下挑战^[26]：

1. 网智融合挑战

当前网络采用外挂式智能架构，存在交互时延长、流程复杂等问题，难以实现网络的自动化与智能化运行。通过网智融合技术，可助力核心网架构简化。核心网不同平面之间采用意图通信替代复杂的过程控制，借助原生智能体实现设备智能自治；与泛在接入设备之间实现端网智算协同，达成终端轻量化目标。网元设备与实时 AI 推理设备一体化部署，便于实现多媒体多流低时延协同及媒体与智能的深度融合，有效减少东西向流量，降低组网复杂度。

基于网智融合模式，原本处于透明传输状态的媒体数据，在网络中不再透明。AI 媒体处理技术已从单一的音视频压缩算法、时延控制、QoS 保障，向全向智能加工方向演进，涵盖意图识别、语义提取与重建、内容生成、声音克隆、内容鉴伪、多语言翻译、智能降噪与丢包补偿、内容审核、空间智能计算、多模态媒体多流协同等多个维度。

2. 算网融合挑战

在 5G 网络架构下，算力资源利用效率偏低，网络缺乏对算力资源的有效管控，主要存在两方面问题：一是算力孤岛，现有网络中多级算力资源相互独立、形成孤岛，不仅造成资源浪费，还无法满足低时延业务需求；二是算网分离，算力与网络各自独立运行、分开管理，导致资源调度难以实现全局最优。

进入 6G 时代，AI、沉浸式 XR 等新场景催生了算力需求的爆发式增长，亟需实现算力资源利用效率最优化。泛在算力需具备动态资源分配与负载均衡能力，以降低运营成本；同时需突破单点算力性能极限，提升算力规模效能，当端侧算力资源不足时，需将部分算力需求卸载至网络侧，实现算力的灵活调度。

3. 泛在实时挑战

泛在实时通信以实现任何人、任何物、任何时间、任何地点的全域互联为目标，在终端泛在、接入泛在与模态泛在三大方向上，面临系统性的技术与架构难题，这些难题直接制约了通信的实时性、可靠性与规模化部署。

终端泛在带来了海量异构设备的协同困境。通信终端已从手机、电脑延伸至 AR 眼镜、XR 头显、穿戴设备、智能座舱等，各类设备的算力、功耗、硬件能力差异悬殊，统一适配难度极大；同时，终端碎片化导致操作系统、接口协议、媒体能力互不兼容，大幅增加了跨终端互联互通与状态同步的成本。此外，超大规模并发接入、设备频繁上下线与移动切换，进一步对设备管理、寻址、鉴权及资源调度提出了严苛要求。

接入泛在则带来异构网络的不确定性服务难题。通信接入涵盖移动网络、Wi-Fi、卫星等多种制式，空口环境复杂、信号波动剧烈。跨网切换易引发时延抖动、链路中断与丢包，难以满足沉浸式实时业务的稳定性要求。

多模态融合进一步加剧实时同步与语义协同挑战。通信从音视频扩展到文本、点云、触觉、姿态、XR 等多元数据，不同模态在带宽需求、时延要求、采样率、语义抽象层级上差

异巨大。海量高码率多模态数据对传输带宽、网侧算力与编解码效率提出更高要求，传统通信架构与协议难以支撑轻量化、低时延、高可靠的多模态实时协同，成为泛在实时通信规模化落地的关键瓶颈。

4. 安全可信挑战

智能体、大模型、AIGC、个人智能助理等技术的应用，面临更为复杂的安全风险。除传统的传输风险、安全漏洞、外部攻击、虚拟化风险外，还新增了内容伪造、动作劫持、输入输出偏差、决策不公正与行为不合规、隐私泄露等新型风险，例如声音克隆、大模型幻觉、模型越狱、训练数据被污染、智能体动作被劫持及提示词诱导等问题。核心网安全隔离、数字空间全域防护与端到端安全传输机制，可为泛在智媒服务提供端到端的安全可信防护。

1.3.3 媒体资源适配协同瓶颈

沉浸式通信中，边缘实时渲染和 AI 推理与终端低功耗的核心矛盾，本质是高算力、高存储、高带宽需求与边缘算力节点和终端有限的算力资源、能源、散热之间的结构性冲突；叠加沉浸式通信的多模态交互、3D 超写实数字人、体积视频等强实时渲染和 AI 推理场景，进一步对网络边缘算力和存储提出了严峻挑战^[16]。例如，3GPP TR 22.870 中定义的沉浸式通信业务场景，要求 XR 头显支持单目 8K 120 fps 的实时云渲染及推流，其下行带宽需达到 640 Mbps，端到端时延需控制在 50 ms 以内；针对体积视频点云传输，上行带宽要求甚至高达 1 Gbps。同时，6G 时代超逼真 3D 内容极大丰富，沉浸式通信业务爆发带来的高并发需求，要求边缘节点达到 Tbps 级带宽及小于 50 ms 时延，这与边缘节点及终端有限的算力、功耗与散热能力形成突出矛盾^[27]。

该挑战主要体现在硬件能效、终端与边缘约束两大层面。一是硬件能效存在上限（业内常称其为“能效墙”问题），传统冯·诺依曼架构下，数据搬运功耗占比超过 60%（业内常称其为“内存墙”问题），即便专用 GPU，其能效比（TOPS/W）也难以覆盖高负载需求，且动态功耗随算力呈线性上升；二是终端与边缘硬件约束，手机、XR 头显、边缘盒子等设备电池/供电有限、散热空间极小，无法持续满负载运行，易触发热节流（Thermal Throttling），导致渲染帧率骤降与 AI 推理时延变长。此外，异构算力协同还存在功耗叠加问题，边缘渲染、AI 推理依赖 CPU、GPU 与通信基带的协同工作，多模块并发高负载会形成功耗叠加效应（通信+计算+存储三重能耗），远超各模块独立功耗的简单叠加，显著超出系统总功耗预算；

而 6G 太赫兹/毫米波通信需持续波束成形、高功率射频，进一步挤占功耗预算，加剧了渲染算力与通信功耗的资源争夺。

存储层面同样面临高带宽、大容量需求与低功耗存储的双重瓶颈。一是内存带宽瓶颈（业内常称其为“存储墙”问题），实时渲染需高频读写纹理、几何、渲染目标、体积数据，带宽需求达数百 Gbps，但低功耗内存（LPDDR5X、SRAM）的带宽与能效存在权衡取舍，带宽不足会导致计算单元空转、造成功耗浪费；体积视频数据维度高、冗余大，片外存储访问（External Memory Access, EMA）功耗占比高，其不规则访问模式进一步恶化能效。二是存储容量与功耗、成本的矛盾，高分辨率 3D 模型、体积视频、渲染缓存需 GB 级至 TB 级存储，但大容量存储（如 eMMC、SSD）静态功耗高、读写功耗大，与低功耗目标相冲突；同时，边缘节点需实现本地缓存与实时解码渲染的协同，存储与计算的频繁交互，进一步加剧了存算分离架构带来的额外功耗开销。

为缓解上述算力、存储与功耗的矛盾，未来 6G 沉浸式通信需通过智能资源调度优化计算效率^[28]，并结合终端-边缘-云协同架构分担负载，实现能耗与性能的平衡。例如，终端、边缘节点可负责处理 3D 内容局部变化较大的渲染任务及快速响应的 AI 推理，云端承担背景等 3D 内容变动不大的部分及时延不敏感的 AI 推理，通过终端-边缘-云协同的分布式云渲染和分布式推理技术，显著降低边缘侧实时渲染和 AI 推理的算力、存储与功耗压力。

1.3.4 其他研究瓶颈

沉浸式通信在迈向商业化部署的过程中，系统还面临着来自底层架构、硬件极限以及法律法规等多维度的严峻挑战，具体体现在以下四个方面。

1. **能效瓶颈：**终端续航与计算负载的矛盾。沉浸式通信对高吞吐量和实时性的要求与终端设备的功耗限制之间存在显著冲突，主要表现为三点：

- **硬件物理限制：**轻量化 AR 眼镜等设备受体积和重量约束，电池容量极为有限，使其对功耗极度敏感，难以支撑高负载沉浸式业务的长时间运行。
- **计算与通信的能耗平衡：**运行 XR 渲染、AI 推理等任务会快速消耗电池电量，并引发严重的发热（Thermal）问题，直接影响用户体验；尽管通过计算卸载（Split Rendering）技术可减轻终端计算负担，但频繁的无线数据传输同样会产生大量通信功耗，难以实现能耗平衡。

- **传统节能机制的局限：**现有 5G 节能机制（如 C-DRX）在处理具有突发性、非固定周期特征的沉浸式业务流量时效率不足，需通过更精细化的流量特征感知（如下次突发时间 TTNB）开展协同优化，提升节能效果。

2. **安全挑战：**超高性能需求与新型威胁的博弈。沉浸式通信的实时交互特性，对安全处理机制提出了近乎苛刻的要求，主要面临四大安全难题：

- **加密算法的性能瓶颈：**现有加密算法在移动环境下通常仅能支持约 38 Gbps 的速率，而 6G 全息媒体流可能需要 100 Gbps 甚至更高的带宽，需开发更高效的加解密技术，满足高带宽场景下的安全传输需求。
- **时延与安全的冲突：**为保证运动到光子（MTP）的极致时延、避免用户产生眩晕感，深度包检测等安全防护措施必须在极短时间内完成，否则会破坏用户沉浸体验，形成时延与安全的核心博弈。
- **量子威胁与算法迁移：**随着量子计算机（CRQC）的发展，现有加密体系面临量子攻击威胁，需提升密码敏捷性，以支持向量子抗性算法的平滑迁移，保障安全体系的兼容性与稳定性。
- **身份与意图安全：**沉浸式应用中大量应用 AI 智能体，如何确保智能体身份的真实性，防止因意图理解错误或恶意伪造身份导致非法指令执行，是当前安全架构的突出短板。

3. **隐私保护困境：**深度感知与用户权益的平衡。沉浸式体验依赖对用户及周边环境的深度数据采集，这一特性带来了前所未有的隐私风险，主要体现在三个维度：

- **敏感生物识别数据风险：**为驱动高保真化身或实现注视点渲染，系统需采集用户眼动追踪、面部表情、心率等高度私密的生物特征数据，此类数据一旦泄露，将严重侵犯用户隐私。
- **环境隐私泄露风险：**空间计算和空间锚点技术要求设备对周围物理环境进行高精度 3D 建模，这一过程可能未经授权采集并暴露用户家中隐私信息、旁观者身份等敏感内容，引发隐私安全隐患。
- **服务感知与隐私限制的矛盾：**网络为优化调度效率，需感知应用层特征（如切片选择），但这一行为可能与用户隐藏应用 ID、保护流量特征的意愿相冲突，甚至可

能违反部分地区的网络中立性原则，形成服务优化与隐私保护的两难。

4. **标准化挑战：**复杂性冗余与跨域互操作性障碍。全球范围内统一标准的缺失、现有协议的过度复杂化，是沉浸式通信技术商业化落地的重要瓶颈，具体表现为四点：

- **协议栈过度臃肿 (bloated)：**5G 网络的标准化过程引入了极高的逻辑复杂性，导致终端实现成本高昂且互操作性测试 (IOT) 困难，限制了其在大规模商业中的普及。
- **媒体格式的碎片化：**沉浸式通信涉及多种 3D 视频格式、点云数据及触觉协议，MPEG、3GPP 和 IETF 等组织之间的协调尚不足以支撑一个完全开放且互通的全球生态体系。
- **厂商锁定与互通成本偏高：**许多 XR 新特性的实现依赖于特定厂商的非标逻辑，这给运营商跨厂商部署核心网功能（如 NF 发现与选择）带来了极大挑战。
- **监管与法律规则的差异：**沉浸式业务跨越国界开展时，面临不同国家在频谱管理、数据主权及数字身份认证等方面的规则不兼容问题。

第二章 面向沉浸式通信的网媒融合基础理论与范式

本章首先剖析了沉浸式通信场景下传统通信技术所面临的理论局限与工程瓶颈，由此引出开展网媒融合研究的必要性与紧迫性。在此基础上，系统阐述了面向沉浸式通信的网媒融合基础理论，重点整合了语义通信理论与联合编码理论的前沿成果。最后，基于上述理论框架，构建了一种全新的网媒融合范式：该范式以用户语义为顶层设计逻辑，以网络与媒体的双向感知为核心技术驱动，以通信、感知、计算、存储多维资源的协同调度为底层支撑。本章的研究工作为下一章探讨具体的关键技术提供了理论依据与架构指引。

2.1 面向沉浸式通信的网媒融合概述

受限于香农边界，传统频谱效率提升技术已逼近物理极限，系统复杂度与基站能耗持续攀升。若继续依赖资源堆叠的发展模式，不仅将带来巨大能源消耗与成本压力，更与国家可持续发展战略方针相悖。而沉浸式通信作为数字经济时代的核心应用场景，其极致用户体验高度依赖全感官沉浸式交互、低时延高保真传输、多维资源动态适配等核心诉求，存在 1.3 节中所描述的基础理论与算法、网络标准演进、媒体资源适配协同、能效以及安全等技术瓶颈。

这些问题并非孤立存在，而是相互耦合、彼此制约。例如，极致的交互感要求端到端时延低至毫秒级，但全感官数据的采集与渲染又将带来数据量的指数级增长，从而提高了对网络带宽与算力调度的要求；高保真的传输需要信道容量逼近理论极限，而多维资源的动态适配又要求在复杂多变的网络环境中实现资源与业务的精准匹配，这在传统“尽力而为”的网络与媒体服务分离架构下几乎是无解的矛盾。面对如此严苛且多维的性能约束，传统以“连接”为中心的架构逻辑显露出根本性缺陷：通信网络仅负责数据管道的搭建，媒体服务则独立部署于云端或边缘，二者在资源感知、状态协同与能力编排上长期割裂，无法形成对全感官数据流的高效承载与智能调度。正是这种网与媒分离的设计范式，在面对全沉浸、全维度的未来通信需求时，系统性暴露出技术能力的上限与架构的刚性瓶颈。

本章节提出以语义为核心，将网络传输与媒体内容深度融合的发展思路，突破通信即管道的固化思维，构建集通信、感知、智能、算力、存储多维能力于一体的统一调度与高效协

同架构，如图 2 所示，主要实现思路如下：

首先，通过深入挖掘网络传输与媒体内容产生、呈现之间的语义特征内在关联，推动系统从“单纯传输”向“传产一体”演进，基于媒体自适应编码实现内容边传边产，基于网络差异化传输实现管道边用边优，达成从网络数据搬运到与媒体内容共生的根本性转变；

其次，为有效支持媒体内容语义在终端、边缘与云端的协同产生与高效渲染，需要以通信能力为枢纽，深度融合底层感知、计算、存储与智能资源，实现具备全局感知与跨域优化能力的多维资源协同置换模型，并池化为整体，实现按需服务。

最后，为支持基于语义的网媒融合关键技术落地，通过提取网络传输算子与媒体处理算子间的共有语义特征，构建跨域统一表征与度量体系，实现端到端信息产生、处理、传输、服务模型。并在此模型基础上，实现异构资源动态感知与调度，最终达成网络传输、内容生成、呈现控制与用户意图深度协同的新型组网架构。

该创新思路为打破网络“管道化”提供了全新的理论框架与技术路径，有力推动了通信系统从“以连接为中心”的传统范式，向“以用户为中心”的新范式跃升，标志着信息网络从连接基础设施迈向提供语义服务和多维能力的关键转型。

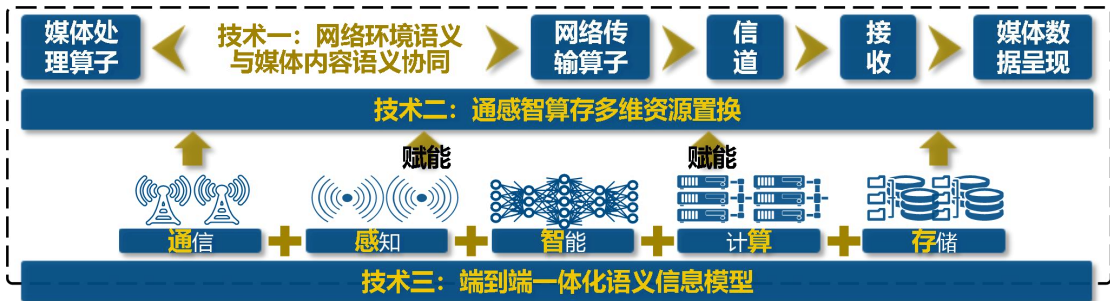


图 2 面向沉浸式通信的网媒融合关键技术研究思路

2.2 面向沉浸式通信的网媒融合基础理论

2.2.1 面向任务/语义的通信理论

Shannon 在经典信息论 (Classical information Theory, CIT) 中使用了玻尔兹曼熵函数来量化信息^[29]，将信息量定义为消除不确定性的程度。这一开创性工作奠定了经典通信的基础，但是在当时也引发了关于信息本质的学术争议：信息是否仅需关注数据的统计特征，

还是应当同时考虑其蕴含的语义？即 Shannon 所提的工程化假设是否能覆盖所有的通信情况。然而，历经七十余年的发展，在语义信息领域，至今仍未形成具有广泛共识的理论框架与量化方法，这成为制约智能通信与认知科学发展的关键理论障碍。事实上，在 Shannon 的奠基性论文发表不久后，通信理论便迎来了进一步的发展。Weaver 在其关于“通信的意义”一文中^[30]，将通信的概念划分为三个层级：

- LEVEL A: 通信符号传输的准确性有多高？（语法问题）
- LEVEL B: 所传输的符号在多大程度上传达了预期的意义？（语义问题）
- LEVEL C: 所接收的意义在多大程度上能产生期望的行为效果？（语用问题）

Weaver 认为 CIT 已较好地解决了语法问题，但由于其高度数学抽象性，忽视了另外两个关键层级。此外，要解决语用问题，必须先处理上一级的语义问题，因此语义问题的突破被认为是打破 CIT 局限的关键。其中，最核心的一环，是建立一种不受 CIT 工程假设约束的语义度量方法。

值得一提的是，语义概念虽然是由通信领域率先提出，但是却先被用于机器学习中，解释各种人工神经网络（Artificial Neural Network, ANN）输出数据的意义。因此，随着人工智能技术，尤其是深度学习（Deep Learning, DL）的不断发展^[31]，人们发现，当初有限的统计工具被极大地扩展，借助深度模型，现在可以描述一些非线性过程，使得通信领域也看到了突破工程化边界条件的希望，语义通信技术开始蓬勃发展起来。

语义通信的研究背景可以分为非 AI 的语义信息理论研究，以及基于 AI 的语义通信系统研究^[32]。但是，由于当前语义通信系统一般都基于深度学习模型搭建，而深度学习模型一般采用端到端的训练方式，防止训练得到的隐变量被映射到不同的隐空间中。因此，语义通信系统一般也采用发送端到接收端，且信源信道联合编码的方式实现。这样的实现方式与基于 CIT 的经典通信系统在范式上完全不同。如果我们认为信源消息是媒体内容，认为信道发送是网络传输，那么经典通信系统是基于 Shannon 的信源信道分离编码定理设计的，即采用了网媒分离范式设计。而对于语义通信，由于当前语义信息理论发展还不够完善，不存在分离定理，再加上其应用形式为深度学习模型，因此语义通信系统是基于联合信源信道编码方式实现的，语义通信与网络和媒体的融合密不可分。接下来我们首先阐述非 AI 的语义信息理论研究背景，之后描述基于 AI 的语义通信系统实现，最后论证语义通信与网媒融合架构之间密不可分的关系。

关于语义信息理论发展情况，最早由 Bar-Hillel 与 Carnap 于 1953 年奠基性地构建了首个语义信息模型^[33]，即经典语义信息理论（Classical Semantic Information Theory, CSIT）。该理论引入逻辑概率替代传统统计概率，通过熵函数量化命题的语义信息量。然而其核心缺陷在于 Bar-Hillel-Carnap 悖论（Bar Hillel-Carnap Paradox, BCP）：当处理矛盾命题时，语义信息量将发散至无穷大，这显然违背认知直觉。对此，CSIT 尝试通过“矛盾蕴含无限潜在信息”的哲学解释进行辩护，但并没有得到广泛认可，并且后续此类逻辑概率框架下的研究始终也未能突破 BCP 困境^[34]。进一步的理论突破出现在 2001 年，Floridi 提出的强语义信息理论（Theory of Strong Semantic Information, TSSI）揭示了 BCP 源于语义真性缺失^[35]。TSSI 创新性地构建了语义空间中的真性距离度量体系，通过衡量不同命题与理想真命题的真值距离规避 BCP。但该理论因公理化程度不足^[36]，在信息量计算公式的数学严谨性方面存在显著缺陷。随后，由 Simon D' Alfonso 在 2011 年完成逻辑概率与真性距离的理论融合^[37]，通过建立联合度量空间，既继承了 TSSI 的语义真性解释优势，又具备了 CSIT 的严格数学框架，最终导出兼具理论深度又可以避免 BCH 的语义信息公式。与此同时，Jie Bao 团队另辟蹊径^[38]，在保留 CSIT 逻辑概率框架基础上引入知识库分层结构，提出具有本地与公共知识库的语义通信模型。该模型虽通过条件概率扩展实现了语义背景的显式表达，但其本质仍属于 CIT 中条件熵理论的语义化延伸，没有触及语义真性度量的核心矛盾。

当前语义信息理论研究已突破传统数学框架，形成多学科交叉的创新格局。在系统论视角下，Kolchinsky 团队开创性地将语义信息定义为“物理系统维持自身存在的因果必需信息，这种信息对于系统随时间维持自身存在是因果必需的”^[39]。Kountouris 等则基于 AI 融合路径^[40]，提出语义信息的四维特征模型，整合过程动力学、稀疏表征、关联网络与生成式 AI 技术，构建了从信息产生、重构到传输的闭环理论框架。产业界代表华为则创新性地将人工神经网络的可存储状态数作为语义信息度量基准^{[41][42]}，通过 Körner 熵建立信息存储和处理模型^[43]。其强调语义与语用不可分割的理论立场虽偏离了 Weaver 原始框架，却与智能通信系统的工程实践深度契合，并提出了 Token 通信的概念。认知神经科学方向，石光明团队提出“物景-像景”信号二分理论^[44]，将语义定义为群体共识下的跨模态置信标注。通信理论改良派方向，张平、牛凯等通过引入同义映射算子^[45]，重构 CIT 框架下的语义信源编码、语义信道容量等核心概念，形成严格数学化的语义通信体系。

综上所述，从基础理论上，语义通信是允许数据失真，但可以通过建立先验信息（知识

库)使得语义保真的一种通信方式,其与数据保真的经典通信截然不同,不仅需要调和信源分布与信道分布,还需要考虑数据失真情况,发送端和接收端建立的先验信息,以及以人为本的数据服务质量等,因此需要联合考虑网络传输和媒体内容产生。从实现方式上,语义通信系统目前依靠需要端到端训练的深度模型,因此需要数据处理联合发送端与接收端,联合信源与信道实现,这是一种联合编码范式,与经典通信中信源信道分离编码截然不同,需要联合考虑网络传输过程中以及媒体内容准备传输过程中的数据处理方式,实现端到端信息服务模型。因此,语义通信的进一步发展需要网媒分离范式转向网媒融合范式,进一步地,在技术章节我们可以看到,网媒融合范式也需要语义通信的各项关键技术作为支撑,两者密不可分。

2.2.2 基于网媒融合的联合编码技术

Shannon 提出^[29],当不考虑复杂度时,信源信道分离编码(Separate Source-Channel Coding, SSCC)可以达到与联合信源信道编码(Joint Source-Channel Coding, JSCC)一样的效果。但是,实际通信系统并不严格满足上述两者等价的前提条件。实际通信场景中,信源往往采用失真编码方式,且信道不是离散无记忆信道。此外,根据实际通信系统使用情况,还需要考虑有限码长、多用户、时延要求等条件。这就使得基于 Shannon 理论搭建的通信系统在实际使用场景中无法达到最优的情况。

2.2.2.1 无失真情况下的分离编码最优性



图3 基于 Shannon 无差错传输理论的通信系统

其中 M 表示信源发出消息, \hat{M} 表示信宿接收消息, X 表示信道输入, Y 表示信道输出, $P(Y|X)$ 表示信道转移概率分布,或称其为信道条件分布。

Shannon 的理论基于其提出的三大定理,以及如图3所示的通信模型。这里采用 M 表示信源, $P(M)$ 表示信源分布, $H(M)$ 表示信源的熵, R_s 表示信源编码速率。当不考虑信

源符号所代表的具体含义，并且认为其属于典型集时，得到无损信源编码定理：

定理 1：对于给定的信源 M 和编码速率 R_s ，若 $R_s > H(M)$ ，则 R_s 是可达的；若 $R_s < H(M)$ ，则 R_s 是不可达的。

该定理表明，对于给定的离散无记忆信源，如果信源编码速率 R_s 超过信源熵，则存在无损编码方法。

对于信道，假设给定容量为 C 的离散无记忆信道 $\{X, P(Y|X), Y\}$ ，其中 X 表示信道输入， Y 表示信道输出， X 与 Y 构成联合典型集， $P(Y|X)$ 为信道条件分布，此时存在信道编码定理：

定理 2：若信道编码速率 $R_C < C$ ，则速率 R_C 是可达的。即对于任意信道编码速率 $R_C < C$ ，存在一个 $(2^{nR_C}, n)$ 的码字序列，当 $n \rightarrow \infty$ 时，其误码率 $P_e^{(n)} \rightarrow 0$ ；反之，若 $(2^{nR_C}, n)$ 码字序列的误码率 $P_e^{(n)} \rightarrow 0$ ，必有 $R_C < C$ 。

该定理表明，当所采用的信道编码，其编码速率小于信道容量时，可以借助编码的方法使得信道译码错误率趋近于 0，而且该错误率会随着码长的增加按照指数规律下降。

需要注意的是，上述无损信源编码定理不依赖于信道，信道编码定理也不依赖于信源分布，因此根据 Shannon 所述信源信道编码定理，即可得到分离编码最优的结果。

定理 3：如果 m_1, m_2, \dots, m_n 为有限字母表上满足渐进均分性和 $H(M) < C$ 的随机过程，则存在一个信源信道编码使得误差概率 $P_e^{(n)} \rightarrow 0$ 。反之，对于任意平稳随机过程，如果 $H(M) > C$ ，那么误差概率远离 0，从而不可能以任意低的误差概率通过信道发送这个信源。

定理 3 结合无损信源编码定理与信道编码定理，最终得到，如果对原始信源 M 进行传输，JSCC 与 SSCC 是等价的结论，即分离定理。但是，上述 Shannon 通信模型是很理想的抽象，虽高度概括了各类通信系统的本质，但远远不能刻画人类面临的实际通信问题。一方面，如果将人类语言作为信源，由于其具有非平稳性，不是各态历经的，进一步的，语言还具有模糊性，并非一定是确定的，概率论中没有有效的工具能够处理如此复杂的过程。另一方面，如果将人作为信宿，其发出消息和接收消息的空间是不同的，不同人还具有不同的消息判决

策略，这些空间和策略是动态变化的。事实上，Shannon 也曾试图采用概率方法对英语做近似表述^[46]，计算相应的语言熵，但其效能有限，未能获得进一步发展。综上所述，经典的信息理论远远无法刻画所有的通信过程^[47]。

2.2.2.2 数据失真情况下的联合编码最优性

语义通信就是为了解决上述困境而诞生的。本文认为，语义通信是允许数据失真，但是可以通过建立先验信息，即知识库，使得语义能够保真传输的通信方式^[45]。这是因为，语义通信更关注人，或者具有“智能”的机器对信息的理解与感受，而并不聚焦在数据是否无失真传输。当信源编码采用失真编码时，上一节中基于无损信源编码定理与信道编码定理无依赖条件，所得到的分离定理并不成立，此时描述信源编码过程的是率失真函数。本文用 $d(m, \hat{m})$ 来表示失真函数，其期望表示平均失真 $D = E[d(m, \hat{m})]$ ，如果用 P_D 表示满足平均失真 D 的任意概率分布， $I(M; \hat{M})$ 表示发出消息与接收消息的互信息，那么率失真理论的主要定理为：

定理 4： 对一个分布为 $P(M)$ 的独立同分布信源 M ，假设其信源译码后为 \hat{M} ，如果失真函数为 $d(m, \hat{m})$ ，则率失真函数 $R(D)$ 等于信息率-失真函数 $R^{(I)}(D)$ ，即：

$$R(D) = R^{(I)}(D) = \min_{P(\hat{m}|m) \in P_D} I(M; \hat{M})$$

是在平均失真 D 下可达的最小信息速率。反之，对任意满足平均失真 D 的率失真码，必然有 $R \geq R^{(I)}(D)$ 。

当采用率失真函数描述信源编码过程，采用信道编码定理描述信道编码过程时，两者是否具有依赖性，就成为 JSCC 是否比 SSCC 更有优势的关键。实际上，Shannon 在提出率失真理论的同时，也对两者进行了对比^[48]。发现此时的 JSCC 依旧可以达到最优，但是他并没有对 JSCC 达到最优的条件，以及 SSCC 是否依然与其等价，进行进一步分析。后续人们进一步研究，对第一个问题做出了部分回答^[49]：

定理 5： 对于给定的信源分布 $P(M)$ ，信道条件分布 $P(Y|X)$ ，以及相应的单字母编码方式 (f, g) ，如果 $I(M; \hat{M}) > 0$ ，当且仅当失真函数满足：

$$d(m; \hat{m}) = -c \log_2 p(m | \hat{m}) + d_0(m),$$

其中 $c > 0$ ，并且 d_0 为 m 的函数时，通信系统能够达到最优传输效果。

上述定理说明，采用定理 3 与定理 4 描述的通信系统能够以 JSCC 方式达到最优通信效果，但是需要失真函数 d 满足固定形式。该形式中的第一项实际包含了信道传输造成的影响，说明即便是 JSCC 方式，也需要考虑信源失真与信道噪声之间的协调。但是，此时采用 SSCC 是否依然能与 JSCC 达到相同的效果，目前依旧没有定论。这是由于，一旦固定信道编码速率，去计算信源失真率，或者固定信源失真率，去计算信道编码速率，往往得不到解析解。没有解析解就没有确定的码率，即无法说明是否存在编码方式可以调和信源分布与信道条件分布，使得理论分析难以继续。

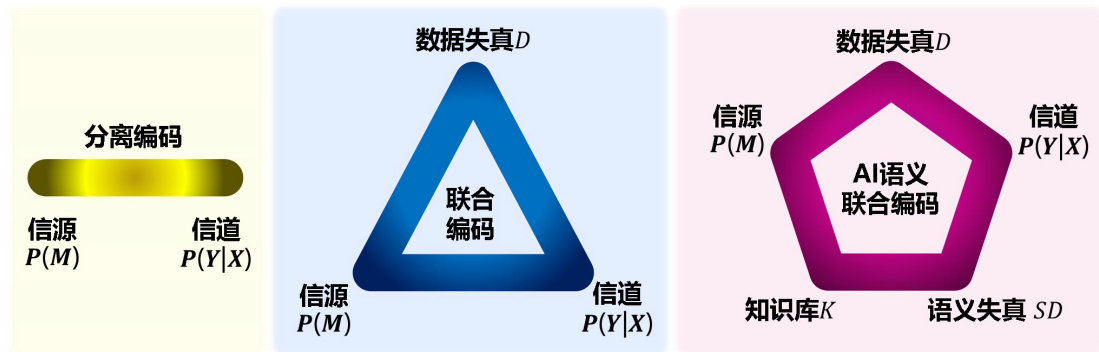


图 4 Shannon 三定理描述的分离编码形式，保真度准则下的通信方式，以及语义通信需要调和的不同参数

如图 4 所示，对于无损信源编码，定理 1、定理 2 与定理 3 共同指出，可以通过编码的方式对信源分布与信道条件分布进行调和，并且可以在信源端与信宿端分离编码，效果与联合编码等价，不会影响其最优性；对于有损信源编码，即保真度准则下的通信，定理 3、定理 4 与定理 5 共同指出，存在一定数据失真 D 的情况下，JSCC 方式依然可以对信源分布与信道条件分布进行调和，并且达到最优性能，但是失真函数需要满足固定形式，该形式与信源分布与信道条件分布均有关。进一步地，在目前工程上可以实现的情况下，能够获得最优码率的 JSCC 方式是可实现的。但是，SSCC 方式由于无法建立码率的解析表达式，是次优的实现方案；对于语义通信，因为其需要建立先验知识库 K ，因此是在建立先验知识库 K 的条件下，设计最优编码的方案，此外或许还可以根据通信的目的，产生一定的语义失真 SD ，因而其信源压缩或许可以更进一步，调和信源分布与信道条件分布的编码或许也更加容易设计，从而产生性能增益。但是，由于目前没有统一的语义通信理论，上述关于语义通信的结

中兴通讯版权所有未经许可不得扩散

论并未获得严格证明，需要进行实验验证。

2.3 面向沉浸式通信的网媒融合新范式

2.3.1 顶层设计-用户语义导向的网媒融合架构

在动态传输环境中，传统分离通信范式面临性能瓶颈，为此构建如图 5 所示的端到端一体化语义通信模型，旨在实现从“比特传输”到“语义任务保障”的根本范式转变。该模型通过引入收发端共享的语义知识库，并结合传输环境反馈构建智能闭环，形成一个具备感知、决策与自适应能力的智能通信系统。在发送端，通过“语义编码—波形适配—失真补偿—资源优化”的协同处理机制，将多模态信源中的核心语义特征智能映射为信道波形；接收端则执行相应的失真补偿、波形检测与语义解码，精准重构语义信息以驱动下游任务执行。通过通信技术与人工智能技术的深度融合，该系统赋予通信过程以语义理解与智能决策能力，从而在动态复杂的信道环境与多样化多模态任务中，实现高效鲁棒的语义信息传输。

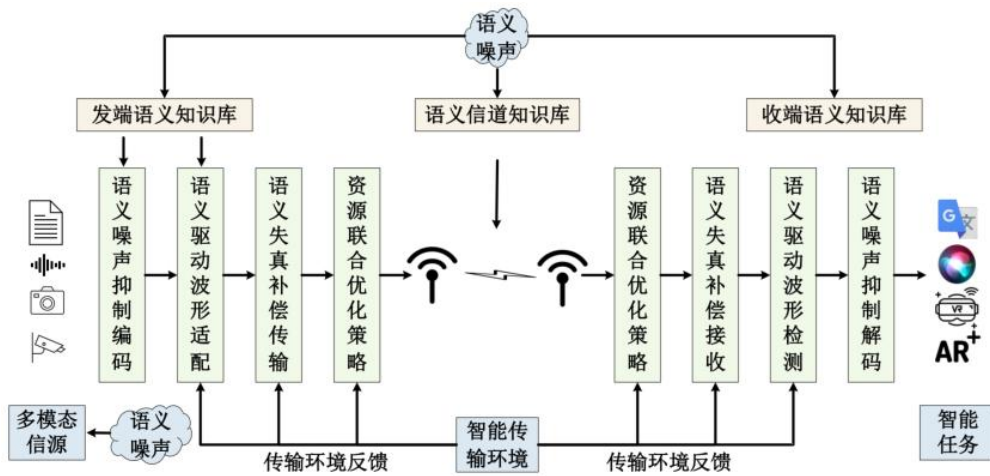


图 5 端到端一体化语义信息传输模型

通过构建用户语义导向的网络架构，实现了从传统技术指标驱动向用户体验价值驱动的根本性转变。该架构以用户为中心感知评价体系为核心基础，创新性地突破传统 QoS 单一维度局限，建立涵盖通信层、业务层和系统层的综合评估机制，通过跨层优化、资源协同矩阵和体验量化三大支柱，将系统优化目标从达成技术指标转变为创造可衡量的用户体验价值。

在组网架构层面，系统彻底颠覆以基站为中心的固定覆盖模式，提出一体化网媒融合网络架构。通过用户侧辅助的信道信息校准机制实现精准感知，采用上下行解耦与智能协作技

术构建动态资源池，并基于 SuperCell/VirtualCell 组网架构实现增强移动性管理。这套架构在分布式部署中融入集中式协同决策，既保持中心化调度优势，又具备动态资源适配能力，有效解决了多小区密集部署下的干扰管理和频繁切换等核心痛点。

整套技术体系通过感知评价与组网架构的深度耦合，形成了完整的“评估-决策-调度”闭环，使通信系统具备主动感知、智能响应和动态适应的服务能力，真正实现了从传输数据到传输服务能力的范式升级，为网媒融合提供了坚实的架构基础。

2.3.2 中枢技术-网络与媒体双向感知

为实现网媒融合这一关键目标，其技术核心在于构建网络与媒体之间语义特征的双向自适应智能机制。该系统通过实时感知网络状态（如带宽、时延、抖动）与媒体内容的语义特征（如场景重要性、视觉焦点），对传输策略与内容的生产、呈现进行一体化联合优化，从而彻底打破传统通信中网络与媒体相互割裂的局面，推动系统从被动的“管道化”传输向主动的“传产一体”模式演进。

在高清直播、XR 等高性能媒体服务场景中，传统编码方式因缺乏对传输环境与用户意图的动态感知能力，难以应对网络波动、终端异构与内容复杂度提升带来的多重挑战。为突破这一瓶颈，需构建以网络为中枢、以语义为核心的媒体自适应编码体系，使编码过程能够实时响应信道变化、理解内容价值、适配终端能力并响应用户期望，从而实现单向码流输出到语义驱动适配的根本性转变。

- **终端语义建模与能力适配：**系统主动识别并解析终端设备的硬件语义（如屏幕属性、解码器能力、算力水平），动态构建终端能力画像。基于此，编码器可实时生成与终端语义模型相匹配的码流，在保障优质体验的同时，避免因能力不匹配造成的解码延迟与资源浪费。
- **内容语义分析与智能码控：**通过深度视觉分析模型，提取视频内容中的运动强度、纹理细节、场景重要性等深层语义特征。基于特征语义理解，系统能够实施内容感知的码率分配与参数决策，对高语义价值区域（如快速运动主体、复杂纹理区域）进行精准码率倾斜，在带宽约束下最大化视觉保真度与语义完整性。
- **用户语义感知与个性化编码：**构建基于用户行为分析的语义偏好模型，精准识别用户对画质、流畅度、互动时延等体验维度的差异化需求。编码策略可依据用户语义

画像进行动态调优，实现从“通用流推送”到“个性化语义流定制”的升级，显著提升用户体验满意度。

- **网络状态语义化感知与预测：**系统通过轻量化探针与 AI 信道估计，实时提取并解析网络状态的关键语义（如可用带宽、抖动、丢包率及未来趋势），构建动态网络画像。编码器据此提前进行码率预降级或抢占式提升，变被动响应为主动预防、规避卡顿，实现网络语义与编码决策的闭环控制。

在网络应用高度复杂的今天，云游戏、远程医疗等场景对实时性与可靠性提出了极致要求。传统“尽力而为”的传输模式，由于无法识别业务内涵与用户意图，在面对突发流量时极易引发资源挤兑与关键业务失效，导致用户体验出现“雪崩效应”，形成服务降级与用户流失的恶性循环。因此，网络亟需从被动的“比特管道”向主动的“语义理解者”演进，即具备对业务内容与用户意图的深度感知能力，并以此为基础，结合实时信道状态与体验质量（QoE）需求，实现智能化的差异化传输。

- **语义级意图识别与价值评估：**引入语义通信的核心思想，超越传统的深度包检测，通过智能意图识别模型，对业务类型、用户行为及内容本身的语义价值进行精细建模与理解。例如，在远程手术中，系统能识别出“手术刀精准位移指令”比“背景生命体征监测数据”具有更高的语义紧迫性与价值密度，从而为差异化调度提供精准的决策依据。
- **网络资源动态重构与语义调度：**基于语义识别结果，引入动态网络切片技术与跨区域资源联合调度算法。网络不再是静态的，而是能根据语义业务的价值等级与体验需求进行实时重构与弹性扩展。对于高价值语义流（如医疗控制指令），系统可瞬间构建一个具备独占带宽、超低时延保障的虚拟网络切片，确保其资源不被抢占，显著提升网络对关键业务的响应敏捷性与确定性。
- **传输策略与语义意图的精准匹配：**面向远程医疗、交互式云游戏等高优先级业务，建立了一套语义触发的动态策略调度机制。当系统识别到关键语义（如“金融交易确认”、“VR 帧同步信号”）时，会立即触发策略引擎，为其开放专属传输通道与前向纠错等增强机制，确保核心语义数据优先、可靠、低时延地传输，从本质上避免关键业务的失效。

两大技术模块共同形成了“网络感知媒体、媒体适应网络”的双向优化闭环，构建了完

整的网媒融合技术体系。网络侧为不同价值的媒体内容提供差异化传输保障，媒体侧则根据网络状态智能调整呈现方式，二者协同实现了从简单连通到品质服务的本质提升，为沉浸式通信等创新应用奠定了坚实的技术基础。

2.3.3 底层资源-通感智算存多维资源置换

传统通信范式以单一频谱资源堆叠为核心的发展模式，在应对新兴媒体对海量数据、极致速率和超低时延的多样化需求时已力不从心。其根本瓶颈在于：网络、计算、存储等资源相互割裂，无法理解并动态适配业务的深层语义，导致资源利用效率低下，难以保障高质量的沉浸式体验。为此，需要建立多维资源置换模型，旨在将通信、感知、智能、计算、存储等异构资源进行有机融合与统一调度。该模型的核心是引入“语义映射”机制，将抽象的业务意图（如“确保虚拟现实无眩晕”）转化为对资源的具体、可量化的调度策略，从而实现对业务语义的智能承载。这推动了系统能力从单纯的数据搬运，到网媒智能融合的根本性转变，是构建具备弹性容量、支持智能调度的信息服务基础设施的关键。该模型的构建依赖以下 5 个关键技术的突破，如图 6 所示。

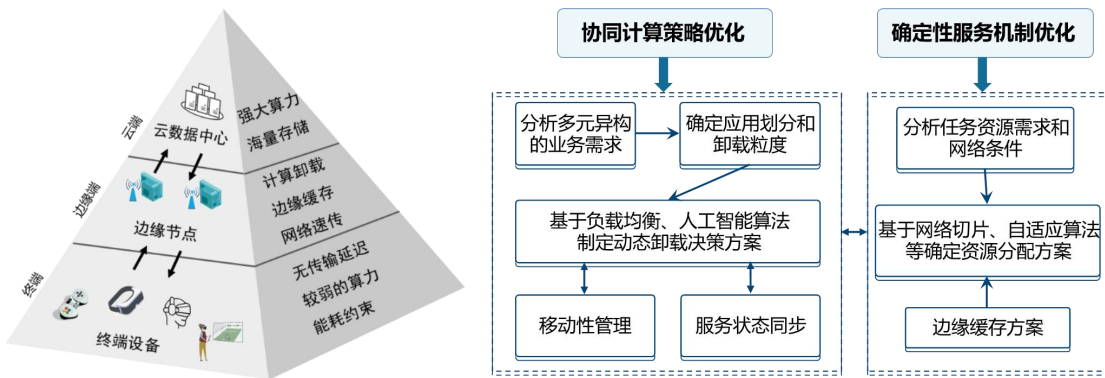


图 6 基于语义通信方式的异质异构资源感知与部署

- 异质异构资源统一表征技术：该技术旨在解决通信、计算、存储等物理资源在类型和度量上的异构性问题。其核心是通过数字孪生建立资源虚拟镜像，并引入“标准计算当量”作为统一度量衡，将不同维度的资源能力映射到同一量化体系中。最终形成一个统一的资源抽象层，将池化后的异构资源以标准服务接口形式提供给上层，实现从硬件异构性到服务统一性的转变，为跨域资源置换与协同奠定基础。
- 语义价值提取与优先级保障技术：该技术是实现从均等分配转向价值驱动分配的前

提。它利用深度学习或大模型，在数据传输前智能判别其承载的语义重要性（如文本的关键词、视频中的关键目标），并进行量化分级。基于此重要性权重，系统在物理层（功率/调制）、网络层（队列调度）实施差异化服务，甚至采用近似传输策略，确保在资源受限时高价值语义数据优先传输，最大化任务成功率。

- **自适应多维资源置换与协同调度技术：**该技术构建了一个实时感知网络与业务状态的智能决策系统。它采用多时间尺度框架：在长周期利用李雅普诺夫优化进行宏观资源预留与负载均衡；在短周期利用深度强化学习（如 PPO）进行实时精准决策。该技术核心是建立语义任务与多维资源（通信、计算、存储）的动态映射，联合优化语义压缩率、任务卸载比、发射功率等变量，实现资源在时空维度上的自适应协同与动态平衡。
- **端边云协同的智能任务决策框架：**该技术聚焦于协调云端大模型与终端小模型的高效协作。其核心是一个动态的协同推理框架，能根据网络条件、终端能力和任务复杂度，智能决策任务是在终端处理、卸载至云端，还是进行端边云分步协同处理。通过联合优化模型选择、部署位置与资源分配，并设计权衡时延、能耗与精度的综合效用函数，最终在满足多样化任务需求的同时，最大化端边云系统的整体效能与资源利用率。
- **知识迁移与增量更新技术：**该技术旨在应对动态环境与异构设备，实现模型的持续进化。它通过融合知识蒸馏与联邦学习，构建双向知识流：自上而下地将云端大模型知识蒸馏给终端轻量模型；自下而上地通过联邦学习聚合终端更新以优化全局模型。同时引入动态学习机制（如数据分布聚类）和应对灾难性遗忘的方法（如正则化），使系统能持续吸收新知识，并通过个性化联邦学习为每个终端适配兼具全局知识与本地特性的模型，实现整个系统认知能力的持续进化。

第三章 面向沉浸式通信的网媒融合体系架构与关键技术

本节在理论基础之上，面向沉浸式通信全感官化、交互实时化、虚实一体化、智能内生化的四化特征，结合物理空间、数字空间、意识空间的三元交互需求，提出三空间四能力网媒融合总体技术架构，并进一步给出端网边云分布式部署架构，形成从理论模型到工程实现的完整映射，为后续核心技术、应用场景与产业落地提供统一的体系化底座。

3.1 面向沉浸式通信的网媒融合总体架构

3.1.1 三空间四能力总体架构

面向沉浸式通信全域空间主体与逻辑功能的统一表达，本节提出三空间四能力架构。其中，三空间构成沉浸式通信的空间主体视图，对应物理空间、数字空间与意识空间的全域覆盖；四能力构成逻辑功能视图，实现从感知采集到智能处理的全流程能力支撑。三空间与四能力深度协同，共同构成 6G 沉浸式通信的总体架构体系。

三空间：空间主体视图

1. 物理空间：物理空间对应真实物理世界中客观存在的主体与环境，是沉浸式通信的现实来源与物理基底，并为沉浸式通信提供最原始的感知对象与交互锚点。主要包括：

- 物理主体：真实存在的人、生物、智能终端、交互设备、感知载体等。
- 物理环境：现实空间、地理场景、物理设施、环境状态等。

2. 数字空间：数字空间是数字世界中承载呈现、交互与计算的数字主体与数字环境，实现物理世界复刻、虚拟世界创造与虚实世界融合，支撑全息交互、具身交互等沉浸式体验，是沉浸式通信的核心载体，包含三类形态：

- 映射型数字孪生体：与物理实体一一对应、实时同步的数字镜像，包括数字人、数字生物、数字物体、数字场景等。
- 原生虚拟数字体：无直接物理对应、纯虚拟创作生成的数字人物、数字生物、数字

物体与数字环境等。

- 虚实混合数字体：映射型与原生虚拟型融合叠加形成的混合数字主体与混合空间。

3. 意识空间：主要由意识主体构成，意识主体是沉浸式通信的驱动源头与体验中心，包括人或智能体的感知、意图、决策与交互行为。意识主体通过多模态方式与物理空间、数字空间进行交互，驱动沉浸式业务的发起、执行与闭环，是本文定义的四化特征中交互实时化、智能内生化的核心体现。

三空间共同构成物理实体、数字孪生体、意识主体的闭环，是沉浸式通信各主体在架构层面的直接落地。

四能力：逻辑功能视图

四能力架构面向沉浸式通信全流程能力需求，实现从物理世界感知采集到数字世界呈现、再到意识主体交互的完整链路，同时与 2.3.3 节通感智算存融合理论形成映射。

1. 感知能力：感知能力是物理空间向数字空间转化的入口层，专注于对物理空间的全域、精准、同步采集，主要包括：

- 视觉感知：人脸、姿态、场景、空间结构采集，支撑数字人建模与全息重构。
- 听觉感知：语音、空间音频、声场信息采集，支撑沉浸式声景还原。
- 动作感知：肢体、手势、表情、微动采集，支撑实时驱动与具身交互。
- 环境感知：位置、距离、运动状态、通感一体化信息采集。
- 同步校准：多模态数据时间戳对齐、空间坐标统一、感官同步误差控制。

感知能力是通感智算存融合中“感知”能力的直接体现，为沉浸式体验提供高质量、高同步的原始数据。

2. 通信能力：通信能力是沉浸式数据传输的高速、低时延、确定性传输大动脉，面向 Tbps 级峰值速率、端到端时延 ≤ 1 ms 等极致 KPI 进行设计，实现多模态感知数据、点云、全息、动作指令等大带宽、高时延敏感业务的可靠传输，对应通感智算存融合中的通信能力。

3. 智算存能力：智算存能力是沉浸式通信的数字孪生生成与智能决策中枢，承担实时渲染、数字孪生建模、数字人驱动、语义理解、意图决策、模型推理等核心功能，面向语义熵、率失真优化等超越香农信息论新范式，实现从比特传输到语义传递的跃升，对应通感智

算存融合中的智能、计算与存储能力。

4. 安全管控能力：安全管控能力为沉浸式通信全域提供可信、安全、隐私保护的底座保障，实现身份可信、数据安全、隐私计算、权限管控与行为审计，覆盖物理空间数据采集、数字孪生运行、意识主体交互全流程，保障沉浸式业务安全合规。

三空间与四能力之间存在很强的协同关系，感知能力采集物理空间信息，通信能力保障实时高效传递，智算存能力生成并驱动数字孪生体，安全管控能力实现全域可信保障，最终由意识主体完成意图交互与体验闭环。三空间四能力架构将理论模型、功能区分、空间主体统一为有机整体，成为 6G 沉浸式通信所有技术与应用的核心架构。

3.1.2 端网边云分布式部署架构

在三空间四能力总体架构基础上，本节从工程部署与无线接入两个维度，给出可落地实现的分布式架构与空口方案，支撑沉浸式业务在真实网络环境下高效运行。

端网边云分布式架构是三空间四能力架构在泛在部署层面的具体实现，将感知、通信、智算存、安全能力分布式部署在终端、网络、边缘节点和云端中，实现资源最优协同与体验极致保障。

- 以终端为载体，承担本地感知采集、轻量智能预处理、数字孪生呈现、低功耗实时交互功能，是感知能力的主要入口与沉浸式体验的直接出口。
- 以网络为神经，承担各功能模块、各节点间数据高效实时传输、同步及保障的功能，是通信能力的主要实现载体和交互实时化的重要支撑。
- 以边缘节点为锚点，承担低时延孪生闭环、实时渲染、AI 推理、本地语义处理、业务分流与本地决策功能，解决边缘实时渲染、AI 推理与低功耗之间的矛盾，保障交互实时化需求。
- 以云端为中枢，承担全局数字孪生建模、大模型训练与更新、全域意图编排、跨域协同调度功能，实现智能内生与全域一体化管控。

端网边云协同实现通感智算存能力泛在部署，为沉浸式业务提供连续、高效、可扩展的部署底座。

3.2 面向沉浸式通信网侧协同关键技术

本节面向 6G 沉浸式通信对网络灵活适配、实时交互、全域协同的核心需求，从网络架构与关键技术层面构建一体化支撑体系。本节提出的 AI 内生网络、算网融合架构和确定性网络，是将内生智能^[50]和交互实时化深度融入网络架构设计、协议实现与运行控制全过程的原生网络形态，既是总体体系架构在网络域的工程化落地，也是基础理论从理论走向系统实现的关键载体，为沉浸式通信提供智能内生、算网融合、极致性能、语义保真的网络支撑。

3.2.1 AI 内生网络

AI 内生网络是以 AI 为架构基础、以意图为驱动、以语义与体验为优化目标的新一代网络范式，区别于 5G 及以前网络“AI 外挂”的模式，实现架构原生智能、控制全域智能、业务适配智能的统一。

- 在网络架构层面，采用分布式智能架构，与 2.2.2 节端网边云分布式部署架构同构设计，将智能能力下沉至空口、接入网、核心网、算力节点，形成端到端一体化智能架构。
- 在网络控制层面，以内生 AI 实现全网状态感知、资源调度、策略生成、故障处置的自动化与智能化，支撑网络自组织、自优化、自愈。
- 在空口和业务适配层面，遵循超越香农信息论新范式，以语义熵、语义保真、率失真优化为核心目标，而非仅以比特传输为评价准则。

AI 内生网络向上承接物理空间、数字空间、意识空间的三元闭环需求，向下支撑全感官化、交互实时化、虚实一体化、智能内生化的业务需求，是 6G 沉浸式通信网络架构的核心基石。

3.2.1.1 AI 内生空口设计

AI 内生空口是沉浸式通信无线接入层的原生设计范式，是超越香农信息论新范式、语义熵与率失真理论在空口层面的直接落地，将 AI 从外挂优化变为空口原生基因。

- 面向语义的空口优化目标：以语义保真、语义熵最优、率失真性能最优为核心目标，而非仅以传统比特误码率为优化准则，支撑多模态数据的保真度 $\geq 99\%$ 的沉浸式交

互需求。

- **空口原生智能能力：**基于 AI 实现信道压缩和估计、波束管理、资源调度、链路控制、干扰协调的自学习、自适应、自优化；原生支持通感智算存一体化，实现感知与通信联合优化；动态适配多模态感知数据的差异化传输需求，保障视觉、听觉、动作等数据的同步与实时性。

AI 内生空口与端网边云分布式智能深度联动，形成空口、网络、算力一体化协同，为三空间四能力总体架构提供高性能、高智能、高可靠的无线接入支撑，是 6G 沉浸式通信区别于 5G 的标志性能力之一。

3.2.1.2 AI 内生网络的架构设计

AI 内生网络围绕沉浸式通信全域升维协同需求，构建分布式、跨域融合、语义原生的智能架构，与三空间四能力总体架构和端网边云分布式部署架构深度对齐，具体表现为：

1. 分布式智能架构。采用端网边云协同的分布式智能架构：

- **端侧智能：**与感知能力协同，实现本地感知数据预处理、轻量推理、实时交互控制。
- **网络智能：**在接入网与核心网部署网络内生智能单元，实现空口、连接、传输的统一控制。
- **边侧智能：**承担低时延孪生渲染、本地语义处理、本地策略决策、智能接入控制与分流，支撑交互实时化。
- **云端智能：**负责全局模型训练、全域意图理解、数字孪生全局编排与跨域协同。

分布式智能架构从网络架构层面解决沉浸式业务大带宽、极低时延、高可靠、强同步与资源受限、无线信道不确定性的矛盾，缓解边缘实时渲染与功耗约束之间的冲突。

2. **跨域融合架构。**打破传统 RAN 与核心网、传输网与算力网、通信与感知的域间边界，形成通感智算存一体化网络架构：

- **网络-空口协同：**AI 内生网络与 AI 内生空口无缝协同，实现端到端质量保障。
- **网络-网络协同：**支持 RAN、核心网、传输网跨域协同，实现灵活拓扑、动态连接、确定性转发。

- **网络-算力协同：**将算力存储视为网络可调度资源，实现算网原生融合。
- **通信-感知协同：**支持通信和感知联合优化。

跨域融合架构直接解决第一章提出的跨域融合不足，RAN 与核心网适配困难等技术瓶颈。

3. **语义原生架构。**将语义作为网络架构的核心设计目标，在连接建立、路由选择、资源调度、质量保障中引入语义信息与意图表达，使网络从尽力传输比特升级为精准传递语义与意图，是超越香农信息论新范式在网络架构层面的直接体现。

3.2.1.3 AI 内生驱动的核心控制能力

AI 内生网络以内生 AI 为引擎，形成自组织、自优化、自愈三大核心控制能力，实现网络架构与功能的全域自治。

1. 自组织能力：网络动态生成与自动部署

- 自组织是网络在架构层面的自适应能力，实现网络即插即用、动态拓扑、按需组网。
- 支持网络节点自动发现、自动接入、自动拓扑生成，适应沉浸式场景多终端、高移动性、高保障的接入需求。
- 基于 AI 自动完成空口资源、传输链路、算力资源的初始分配与协同规划。
- 支持面向数字孪生场景的虚拟网络动态创建，实现物理网络与孪生网络的协同组织。自组织能力从架构与控制协同层面，解决沉浸式通信网络部署复杂、适配性差的问题。

2. 自优化能力：是网络运行态的核心智能，以沉浸式体验与语义保真为优化目标。

- **语义与率失真优化：**基于语义熵与率失真理论，动态调整编码、调制、调度与路由策略，在有限资源下最大化语义准确度与沉浸质量。
- **空口与网络联合优化：**与 AI 内生空口协同优化波束、链路、资源，保障端到端速率、时延等指标。
- **算网协同优化：**AI 驱动算力与网络资源联合调度，缓解边缘渲染压力，提升能效。
- **多模态业务差异化优化：**对视觉、听觉、动作、环境等感知数据提供差异化质量保障，降低多感官同步误差。

自优化能力直接支撑第一章提出的峰值速率、时延等关键 KPI。

3. **自愈能力**：故障自治与体验连续性保障。自愈是网络可靠性层面的内生能力，保障沉浸式业务不中断、不卡顿。

- 基于 AI 实现全网状态感知、异常检测、故障定位与预测功能。故障发生时，自动完成链路切换、路径重路由、节点替换、资源重调度，实现秒级甚至亚秒级恢复。
- 面向数字孪生与全息交互场景，保障三元闭环不中断、虚实映射不失步。自愈能力从网络架构与功能层面为沉浸式体验提供高可靠、高可用保障。

3.2.1.4 AI 内生网络与总体架构协同

AI 内生网络是 2.3 节网媒融合新范式在网络域的核心使能技术：

- 为物理空间的全域感知提供泛在、高效的接入与传输控制。
- 为数字空间的实时生成、渲染、同步提供架构、模型、数据、算力与存储支撑。
- 为意识主体的意图交互提供语义理解、意图驱动与体验保障。
- AI 内生网络是智算存能力与感知能力、通信能力、安全管控能力的深度耦合，实现四能力在网络域的自治落地。

同时，AI 内生网络与 AI 内生空口、端网边云分布式架构共同构成沉浸式通信全域智能底座，为后续确定性网络、算网融合架构、典型应用场景与产业赋能提供统一、开放、可演进的网络根基。

3.2.2 算网融合架构

算网融合是 6G 网络架构的核心驱动性技术之一，其核心理念是将云边端多样的算力通过网络化的方式连接与协同，实现计算与网络的深度融合及协同感知，达到算力服务的按需调度和高效共享。

算网融合通过统一管控计算、存储及网络资源，实现通信与计算资源的统一感知、动态编排与协同调度，不仅提高了资源利用效率和网络性能，还能够根据应用需求灵活调整资源分配，实现资源的最优配置和动态管理，为多样化的业务需求和复杂的业务场景提供灵活、

高效的服务支持。

算网融合核心突破在于将算力从资源转变为网络原生服务,通过实时感知网络状态和算力资源,动态调整资源分配,突破单一资源优化带来的性能瓶颈,为沉浸式通信的高带宽、低时延、高算力需求提供确定性保障,确保用户体验的流畅性。

3.2.2.1 算网融合核心特征

算网融合架构的首要特征是实现通信、计算、存储能力的统一度量与调度^[51]。具体包括以下四个方面:

- 统一度量与感知。建立基于计算、存储和网络等多维度的统一度量体系,为算力的感知、控制、管理和计费提供统一规则。实时感知网络外部或内部算力节点的服务状态和负载情况,快速选择计算节点以匹配需求,实现算力资源的精准发现与动态适配。
- 智能控制与管理。依据计算需求智能决策是否进行计算任务分解,并为分解后的任务分配和协调算力资源,实现计算任务的灵活编排。算网融合通过同时感知网络和算力,实现实时准确的算力发现,灵活动态的计算和连接服务的调度,实现全局动态最优,提供无处不在的计算和服务。
- 计算承载的控制与管理。提供计算数据交互的连接通道,同时满足通信 QoS 和计算 QoS 的双重约束,支持计算承载的创建、更新、删除等,实现计算承载的全生命周期管理,确保数据传输与计算执行的协同优化。6G 算网融合通过统筹计算时延和通信时延来更好地支持计算数据传输,实现二者的实时动态适配。
- 算力原生&服务化。算网融合将“算力”从资源属性转变为“网络原生服务”,通过统一服务接口与标准化能力开放,实现算力资源的按需调度、弹性伸缩与按量计费。用户无需感知底层资源分布与异构差异,即可获得“一点接入、即取即用”的社会级算力服务,支撑沉浸式通信等 6G 新业务的海量并发与极致体验需求。

3.2.2.2 算网融合技术架构

1. 总体架构组成

- **统一的算网融合控制:**负责通信和计算的统一调度控制、通信和计算的会话管理等,包括计算节点的管理、控制、发现、选择,计算需求的识别,计算任务的分解和卸载^[52]。
- **分布式计算节点:**分布部署于端、网、边、云侧;具备计算资源管理和计算执行能力,在算网融合控制调度下实现计算任务的执行^[53]。

2. 计算节点分布特征

6G 网络功能节点具有更强的分布式特征,利用广泛分布的通信和计算融合节点实现计算服务,可缩短计算数据传输时延并降低骨干网络的传输负载。具体包括:

- **终端侧:**轻量计算能力,支持本地推理和初步渲染,降低上云数据量。
- **无线接入网侧:**基站内置算力,实现实时数据处理和边缘推理,缩短服务时延。
- **边缘侧:**区域级算力中心,支持重载训练和复杂渲染,兼顾性能与成本。
- **核心云侧:**大规模算力集群,支持模型训练和全局调度,提供最强算力支撑。

3.2.2.3 算网融合关键技术

1. 算力度量与感知

算力的度量是算力网络的基础问题。基于计算、存储和网络等多个维度的统一度量可以为算力的感知、控制、管理和计费提供统一规则。算网融合架构需要实现对算力资源的实时感知和动态调度。通过算力感知技术,网络能够实时了解各节点的算力状态,包括 CPU/GPU 空闲率、内存余量、功耗等。基于这些信息,网络控制面可以进行智能的算力调度,将计算任务分配到最适合的节点,提供 QoE (体验质量) 保障,从而优化整体性能。这种算力感知与调度能力是实现沉浸式通信的关键,能够确保在任何时间、任何地点都能提供足够的算力支持^[54]。

2. 算网联合编排调度

将业务需求拆解为对计算任务和网络能力的要求,并实现联合编排调度,是实现算网融合的关键技术之一,其核心在于将该需求智能转化为双重约束,并通过统一编排引擎实现二者的协同优化。

6G 的核心目标是实现通信、感知、计算、AI 的内生融合，通过网络内生设计形成全域感知、高效计算、智能决策的一体化系统。演进方向上，构建 AI 原生的网络架构成为趋势。通过 AI 与网络的双向赋能，AI 优化资源调度策略、实现故障预判与自愈，网络则提供毫秒级算力互联能力，形成智能闭环。最终，算网联合编排调度将从人工配置迈向智能自治，为沉浸式通信等 6G 新业务提供确定性、弹性化的资源保障^[55]。

3. 端到端算网 QoS 控制

将现有 QoS 机制进一步扩展为具备算力能力的算网 QoS 机制。例如，扩展 QoS 参数，加入计算资源的类型、计算资源的大小及计算时延等级等。网络侧根据终端业务需求，进行算网 QoS 的会话建立、算力任务分配和资源选择，满足沉浸式通信对确定性体验的要求。

4. 计算的动态卸载和调度

在算网融合中，终端的位置、网络的时延和带宽，以及边侧和云侧的负载都呈现动态特性。因此，需要动态地执行计算任务的卸载和调度，具备对计算任务进行动态建模的能力，以便在移动性场景下有效执行计算任务，保障业务连续性。

5. 算网融合安全和隐私

保障计算任务在跨节点传输和执行过程中的数据安全与用户隐私，包括身份认证、数据加密、访问控制等机制，确保敏感数据在分布式环境下的可信处理。

3.2.3 确定性网络

3.2.3.1 现有确定性网络技术

时间敏感网络（TSN）由 IEEE 802.1 工作组标准化，旨在为局域网（LAN）内的时间敏感流量提供有界低时延和高可靠传输。TSN 的核心机制包括时间同步协议 802.1AS、流量调度整形协议 802.1Qbv、资源管理协议 802.1Qcc 和冗余可靠性传输协议 802.1CB 等。TSN 的设计前提是严格的时钟同步和拓扑可控，使其难以直接扩展至广域网环境。

确定性网络（DetNet）由 IETF 成立的工作组推动，旨在将 TSN 的确定性能力扩展到广域网场景。DetNet 基于显式路由和容错队列机制，仅需频率同步而非严格相位同步，降低了部署难度。其转发机制包括循环指定队列转发（CSQF）、确定性 IP（DIP）、标签化循环队列转发（TCQF）等演进方案。

5G 协议从 R16 开始引入 TSN，并在后续的 R17、R18 中持续增强。5G TSN 通过网络侧 TSN 转换器（network-side TSN translator, NW-TT）和终端侧 TSN 转换器（device-side TSN translator, DS-TT）完成 TSN 和 5G 之间的端到端同步，使 5G 基站能够利用时延敏感通信辅助信息（TSCAI），合理配置上行免授权、半持续调度、预调度等无线策略，使得空口资源调度能够更有效地匹配业务流特征，从而降低业务流的等待时延，提升无线调度效率。在 ToB 工业现场网等场景，5G 确定性网络已经具备较成熟的业务保障能力。

然而，当前的确定性网络技术主要面向垂直行业的远控/PLC 类较规律业务模型的时延敏感业务，部署环境也大多在局域可控场景。面向 ToC 沉浸式业务时，业务模型的快速变化、同一个流内不同报文的差异化保障要求、业务体验 QoE 的平稳度要求等都和传统垂直行业业务存在很大差异。同时，AR 等沉浸式业务需要考虑广域室外各种覆盖/干扰/负荷的复杂场景，这也对确定性网络技术提出了更高要求。

3.2.3.2 沉浸式业务特点与网络确定性需求

沉浸式业务呈现多元化形态，不同场景对网络 QoS 的需求各异，需要差异化的保障机制。

VR 大空间（如 VR 影院、虚拟实训基地）是典型的定点应用。用户在有限物理空间内自由移动，头显设备实时渲染高保真虚拟环境。这类业务的核心挑战在于：高用户密度下的容量保障与动作渲染同步的 QoS 保障。从容量维度看，单用户需求速率可达 100 Mbps 至 Gbps 量级，区域内并发用户数多。从 QoS 维度看，MTP 时延必须控制在 20 ms 以内，否则将引发严重眩晕感。这意味着网络需要在有限覆盖区域内同时满足高吞吐量和超低时延，对小区容量规划和资源调度提出极高要求。

数字人交互也需要远程沉浸式通信，通过三维重建、表情捕捉、语音同步、词元传输等技术，将真实人物的数字孪生以全息形态呈现。数字人传输需同时承载视频流（4K/8K）、深度图、音频流、控制信令、词元等多路数据，不同模态之间需保持亚毫秒级同步，否则将出现唇音不同步、肢体动作延迟、智能体响应缓慢等体验劣化问题。传统 QoS 机制孤立处理各流指标，难以满足跨模态协同要求。

AR 眼镜将智能体相关数字信息叠加至物理世界，应用场景涵盖导航、维修辅助、文旅导览等。与 VR 不同，AR 业务具有室外移动性、环境复杂性、终端轻量化、任务多样性四重特征。用户在室外移动时可能经历频繁小区切换，信号覆盖空洞和邻区干扰将直接导致虚拟

物体定位漂移或渲染卡顿。与此同时，AR 眼镜受限于功耗和散热，计算和通信能力远不及 VR 头显，这意味着部分处理需卸载至云端，导致智能体响应缓慢，影响沉浸式体验，进一步加剧了对无线链路低时延高可靠的依赖。

3.2.3.3 面向沉浸式业务传输的关键技术

3.2.3.3.1 网媒双向感知与 QoS/QoE 双层闭环机制

传统网络中，业务层与网络层相互隔离：业务层不了解网络资源状态，只能被动适应；网络层不感知业务精细化模型和体验需求，仅基于预设的 QoS 参数调度。这种架构难以支撑沉浸式业务对体验一致性的要求。

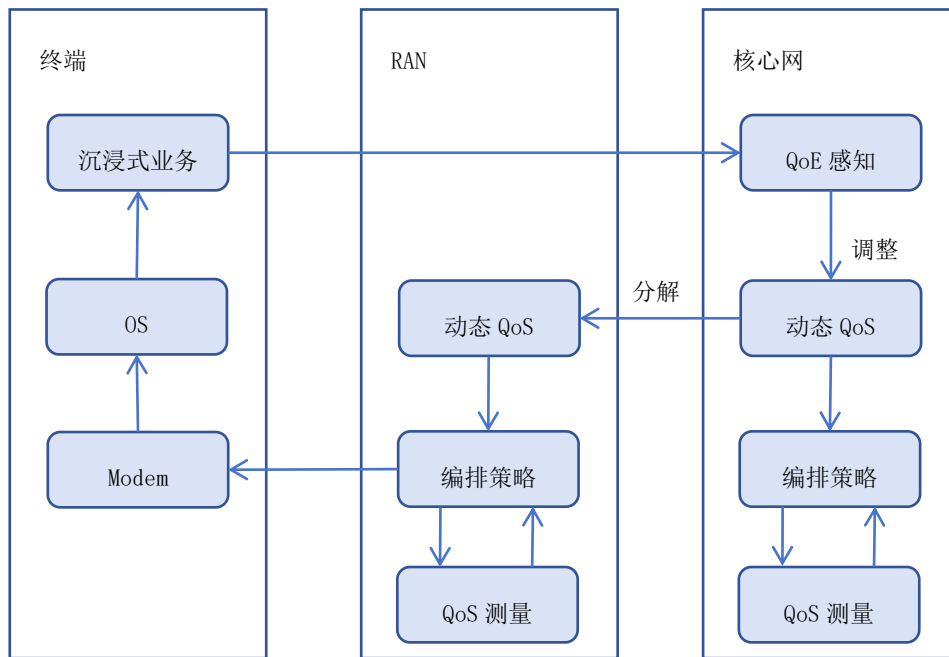


图 7 网媒双向感知和 QoS/QoE 双层闭环机制

如图 7 所示，网媒双向感知技术旨在打破这一壁垒。其核心思想是：在网络侧引入业务感知能力，通过媒体-网络协作交互，实时识别业务类型、精细化流特征及其保障要求，以快速调整 QoS 保障策略；在媒体侧引入网络感知能力，根据网络的拥塞指示和调度能力动态调整业务编码和算法。

在此基础上，通过 QoS（服务质量）/QoE（体验质量）双层闭环机制实现更深层次的网媒融合。内层闭环以网络 QoS 为对象，动态调整资源分配、调度策略；外层闭环以业务 QoE

为目标，监测用户主观体验，通过业务反馈和 AI 模型分析体验劣化的根因，两层闭环协同优化，最终实现网络 QoS 指标与业务 QoE 体验的一致性匹配，不再是孤立追求低时延或高带宽，而是确保网络策略的变化能够准确传递为用户可感知的体验提升。

3.2.3.3.2 广域泛在确定性保障技术

沉浸式业务从室内走向室外，从定点走向移动，要求确定性保障能力从局域网延伸至广域。然而，广域场景下的覆盖空洞、同频干扰、负荷波动等问题，将导致确定性流量的时延和抖动急剧恶化。对此，提出广域泛在确定性保障技术，如图 8 所示，包含以下核心要素：

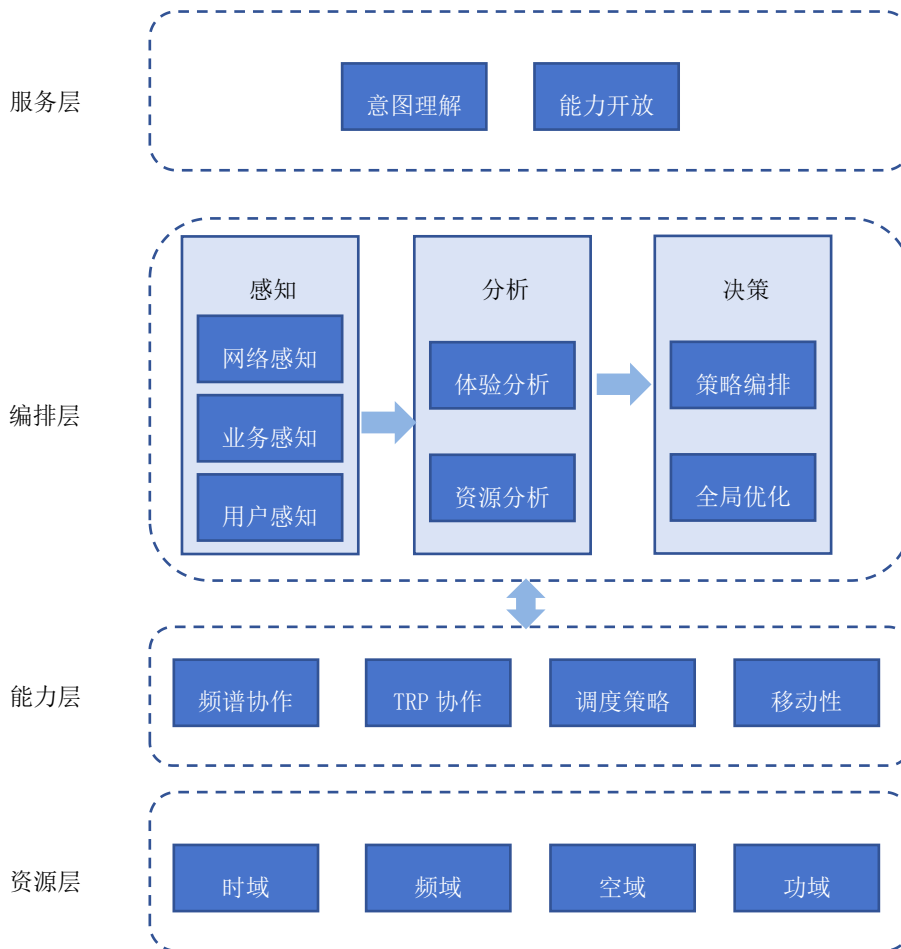


图 8 广域泛在确定性保障技术框架

用户为中心的多频多站协同技术：通过多站协同波束成形、小区间干扰协调、多载波协作灵活调配无线资源，确保在用户移动过程中无论在小区中心还是小区边缘都能够保持一致

性体验和平滑切换，满足沉浸式业务的确定性保障需求。

具备感知-分析-决策-执行闭环能力的保障智能体：将 AI 原生融入无线资源管理。首先，通过网络感知层采集小区/栅格的信道质量、干扰水平、业务状态等多维数据；其次，由智能体分析当前业务确定性需求与网络能力的匹配度，预测潜在的性能劣化点；再次，基于深度学习算法生成决策，如调整调制编码方案、触发多小区协作、预留资源等；最后，在毫秒级时间尺度内完成策略执行。通过此技术，无论用户在何地、以何种方式移动，网络都能通过主动式预测和保障，将覆盖/干扰/负荷的负面影响控制在业务容忍阈值以内，实现“泛在确定”的极致体验。

3.3 面向沉浸式通信媒侧适配关键技术

3.3.1 语义编解码技术

传统 LDPC、Polar 等编解码方式以比特级精确还原为目标，难以适配智能体之间以词元为核心的通信。这类方法过度追求无损传输，忽视语义内涵，导致效率低下且缺乏容错弹性。当前主流研究的语义编解码则力图实现从数据传输到知识传递的范式转换：通过提取高层语义特征，结合先验知识与任务导向优化，允许一定程度的比特误差，同时保障语义含义的准确还原；具备噪声鲁棒性、动态自适应能力及极低带宽占用等特性。由此，智能体可在符号空间中直接进行语义对齐，以意图触发代替逐字复述，从而从根本上契合多智能体协同与人机交互对高效、容错及认知级通信的需求，并符合未来沉浸式通信的研究方向。

3.3.1.1 基于神经网络的语义编解码技术

人工智能的快速发展不仅重塑了语义信息理论的研究范式，更推动了语义通信系统的工程化进程。最早的探索可追溯到 2003 年，Li 等人尝试使用全连接神经网络替代信源编码、信道编码与调制模块，构建端到端的联合信源信道编码系统^[56]，但由于当时的 AI 技术尚不成熟，实验效果并不理想。2019 年，Farsad 等人将信道融入基于 LSTM 的 Seq2Seq 模型中^[57]，用于文本传输，其性能超越了传统编码方案，并展示出基于人工神经网络编码的优越性，即语义相近的词句，其编码向量的距离也更近。2020 年，Deniz 等人提出基于卷积神经网络的 DeepJSCC 图像传输系统^[58]，实质上是一个采用自编码器损失函数训练的 DCGAN 模型，其传输性能与传统通信系统相当，但避免了传统数字通信在信道质量下降时出现的“悬崖效应”，

实现了重建质量的“优雅降级”。随后，DeepJSCC 系列不断扩展，研究了与经典通信方法衔接适配的各种算法，如引入反馈机制的 DeepJSCC-f^[59]、支持量化传输的 DeepJSCC-q^[60]、提升泛化能力的 DeepJSCC-1++^[61]，以及应对实际通信中的反馈与调制问题的基于 Transformer 架构的 JSCCformer-f^[62]。此阶段，人们主要探索基于人工神经网络的联合编码方式是否优于传统的分离编码方式，而语义的概念还没有得到广泛的普及。

2021 年，Xie 等人首次从语义角度解释 JSCC 性能增益^[63]，基于 Transformer 构建了 DeepSC 模型用于文本传输，将“语义”这一概念引入通信系统的增益解释，并采用 BERT 余弦相似度来衡量语义恢复程度，标志着语义通信系统的正式兴起。随后，基于原始的 DeepSC 模型，进一步研究了针对语音传输的语义通信模型，并将其称为 DeepSC-S 模型^[64]。更进一步地，他们还提出了解决图像与文本联合传输的 Mu-DeepSC 模型^{[65][66]}，将多种信息形式联合传输的 U-DeepSC 模型^[67]，试图用率失真函数解释语义通信模型框架^[68]等研究。此外，其他针对不同数据类型的语义通信模型也不断涌现。在图像传输领域，NTSCC（Nonlinear Transform Source Channel Coding）模型^[69]通过非线性变换与熵模型实现更高效的语义压缩。在视频通信方面，面向视频会议的 SVC（Semantic Video Conferencing）模型^[70]利用时序语义冗余实现极低码率下的高质量重建，并在此基础上，设计了一种适用于时变信道、融合新型语义差错检测器的增量冗余混合自动重传请求框架（SVC-HARQ）。该框架在比特资源消耗上具备灵活性，且能实现优异的传输性能，并设计了语义视频会议信道状态信息（SVC-CSI）反馈机制，通过该机制对关键点的传输资源进行合理分配，使系统性能得到显著提升。

随着大模型技术的爆发，语义通信也进入了以大模型驱动为特征的新阶段。2024 年，Jiang 等人提出基于 AI 大模型的多模态语义通信（LAM-MS）框架^[71]，利用多模态语言模型实现文本、图像、音频、视频四模态的统一语义表示与传输，通过可组合扩散模型（CoDi）构建共享语义空间，在保持语义一致性的同时显著降低传输开销。张文军院士团队则进一步提出基于生成扩散模型的生成式语义通信，通过传输部分参考信息，在高维流形上做低维限定，实现语义级信息恢复^[72]。

3.3.1.2 深度学习驱动的语义编解码技术

随着大模型与生成式 AI 技术向语义通信领域的深度渗透，深度学习驱动的语义通信体系正朝着多模态融合、端边云协同、高鲁棒性、低开销的方向演进，但其在实际落地过程中，

仍面临着语义表征泛化性、信道自适应能力、资源调度效率、端侧部署适配性等多方面的技术挑战。未来的研究将围绕大模型与语义通信的深度融合展开，聚焦核心技术瓶颈攻克，同时探索更贴合 6G、空天地海一体化通信、人机物互联等未来通信场景的技术范式，主要包括以下核心研究方向：

- **通用统一的语义表征体系构建：**当前多模态语义通信的语义空间构建仍存在模态间表征异构、场景适配性弱的问题，现有模型多针对特定模态组合或应用场景设计，难以实现跨领域、跨场景的通用语义表示。未来需攻克大模型驱动的通用语义 Token 技术，依托千亿级、万亿级大模型的涌现能力与跨模态对齐能力，打破文本、图像、音频、视频、传感信号等不同模态的数据壁垒，构建兼具语义紧凑性、表征唯一性、跨域泛化性的统一语义表征空间。
- **大模型与信道特性的深度协同适配：**大模型驱动的语义通信目前仍存在语义处理与信道传输脱节的问题，现有方案多将大模型的语义提取与传统的信道编码、调制解调分离开来，未充分利用大模型的认知能力挖掘信道时空特性，也难以在信道衰落、干扰、时延等动态变化场景下实现语义传输的自适应优化。未来需构建大模型驱动的端到端语义信道联合优化框架，让大模型深度参与信道感知、预测、编码与差错校正的全流程。
- **生成式 AI 与语义通信的深度融合创新：**以生成扩散模型为代表的生成式 AI 技术，为语义通信从语义还原向语义生成式恢复提供了核心支撑，目前生成式语义通信仍存在生成语义与原始语义的一致性把控难、高维流形低维限定的精度不足、生成效率低等问题。未来需深化生成式 AI 与语义通信的融合研究，推动语义通信从被动还原向主动生成的范式升级。
- **面向复杂场景的语义通信智能调控与安全防护：**未来语义通信的应用场景将从单一的端到端有线或无线通信，拓展至空天地海一体化、人机物互联、工业互联网、元宇宙等复杂异构场景，这些场景具有信道环境复杂多变、传输节点海量异构、语义信息类型多样、安全需求等级高等特点，对语义通信的智能调控与安全防护能力提出了更高要求。同时，大模型本身存在的对抗样本、隐私泄露、模型投毒等安全风险，也会向语义通信系统传导，带来新的安全挑战。

3.3.2 沉浸式媒体编解码

沉浸式媒体编解码是一种面向三自由度（Three Degrees of Freedom, 3DoF）、六自由度（6DoF）视角体验的技术体系，主要用于全景视频、多视角视频、三维几何数据等多模态沉浸式内容的高效压缩与还原。该技术针对传统二维编码在空间冗余复杂、数据量爆炸、交互性要求高三大痛点上的不足，实现了对沉浸式内容的有效处理，是支撑沉浸式多人协作、远程全息会议等沉浸式通信应用的关键底层技术。沉浸式媒体并非单一技术，而是一个包含多种媒体形态的生态系统，每种媒体形态都对编解码提出了独特的要求。

3.3.2.1 全景视频编码

全景视频是一种能够捕捉和呈现水平 360 度、垂直 180 度全方位视角的视频格式。观看者不再是被动的旁观者，而是可以自由动态地控制视角，在虚拟球体内部环顾四周，获得身临其境的沉浸感。业界主流的全景视频形式是 3DoF 的，支持用户交互性地切换观看视角，终端能够根据用户的观看视角动态渲染图像、视频及其相关联的音频。

全景视频编码与传输的整体技术架构主要由视频拼接与映射、视频编解码、存储与传输等技术构成。在现实场景中采集到的视频等媒体数据经过源端的全景视频拼接服务器进行拼接、旋转、投影等处理，成为完整视频画面，继而经过编码器编码得到全景视频内容源，利用内容分发网络完成系统层封装和媒体内容的分发；用户通过终端设备请求全景视频服务，其中传输协议访问模块完成全景视频索引文件的解析，终端根据传感器获取到的视窗元数据（如用户头动信息、观看方向等）请求相应的全景视频内容分片，并完成解封装、解码，最终根据视窗元数据完成视频画面的渲染和音频等其它媒体资源的播放。

3.3.2.2 多视角视频编码

多视角视频业务是指通过多个摄像机在不同角度同时拍摄，可以给用户提供不同视角的一种视频业务。具体来说，多视角视频编码的输入是一组由同步、预标定的摄像机阵列捕获的多视角视频流。每个视频流都包含纹理（Texture，即颜色信息）和对应的深度图（Depth Map，即每个像素到摄像头的距离）。系统并不直接传输用户可能看到的所有视角，而是只传输一个有限的、精心选择的子集（称为“基础视图”）。终端设备根据用户的 6DoF 位姿（位置与朝向）并通过深度图像的视点渲染技术，可以利用接收到的基础视图及其深度图，实时合成并输出用户当前请求的任意虚拟视角的画面。

为了用最小冗余保留重建任意视角所需的全部信息，在压缩效率与渲染自由度之间取得平衡，多视角视频编码的核心环节包括视角选择、冗余修剪、补丁拼接、视频编解码、元数据封装等。首先，以多视角纹理视频及对应深度图为输入，先选定一组基本视角作为锚点，其余为附加视角；再以基本视角为基础，按投影覆盖度将附加视角进行排序，形成层级依赖关系。其次，利用跨视角冗余修剪技术去除各视角间重复覆盖的像素区域，保留并将各视角独有的、不可替代的纹理与深度区域裁剪为补丁（Patch）。所有补丁按紧凑排列、避免重叠等布局规则拼接成 2D 图集帧，同时生成占用图用以标记图集内有效补丁区域与空白区域。随后，利用 HEVC、VVC 等 2D 视频编码器对已生成的图集视频序列（纹理与深度）进行高效压缩，并将元数据封装输出。最终，解码端将解析图集与元数据，反向映射重建多视角数据，并通过深度图像渲染合成用户指定的任意视角画面。

3.3.2.3 三维容积媒体编码

随着激光雷达、深度相机、摄影测量等技术的普及，高精度三维数据的获取成本大幅降低，三维容积媒体（点云、网格）已成为继文本、图像、视频之后的新一代核心媒体形态。关于数据表达方式，点云是由海量三维坐标点及对应属性（颜色、反射率、法线等）构成，可以分为密集点云与稀疏点云；网格是由顶点、边、面（通常为三角面）构成拓扑结构，适合表示连续曲面与虚拟人模型。关于技术路线，三维容积媒体编码可以分为基于视频的压缩与基于几何的压缩。

1. 基于视频的压缩

基于视频的压缩采用“降维”策略，将复杂的 3D 数据投影/映射为 2D 图像，复用现有成熟的 HEVC、VVC、AV1 等 2D 视频编解码技术。基于视频的点云编码是将点云媒体分割为多个局部近似平面的补丁，再将每个补丁正交投影到其局部平面上，生成占用图、深度图和属性图等。随后，利用 2D 视频编码器对降维打包后的平面图集进行压缩，且所有重建必需的几何与打包信息通过元数据流同步传输。基于视频的动态网格编码专门为动态、连通的 3D 网格设计，借鉴了基于视频的点云编码的成功经验，将动态网格的顶点位置和属性（例如几何偏移、颜色、法线）映射到 2D 参数域（称作 UV 展开图），生成 2D 图集视频序列，再使用 2D 视频编码器对其进行压缩。为了指导动态网格的解码和重建，相关元数据需要同步传输一个静态的“基础网格”（Base Mesh）及其 UV 展开信息（UV Atlas）。基础网格定义了拓扑结构（顶点连接关系），而逐帧的几何变化和属性变化则通过 2D 图集视频序列中的位

移图 (Displacement Map) 和属性图进行增量编码, 从而在保持拓扑一致性同时高效表示动态变化。

2. 基于几何的压缩

基于几何的压缩坚持“原生 3D”处理范式, 直接在三维空间对几何结构与属性进行压缩编码, 强调几何精度与无损/近无损能力。基于几何的点云编码, 一方面是对无序点云采用八叉树 (Octree) 递归细分空间, 仅编码存在点的体素; 另一方面是对有序点云利用空间相关性进行位置预测及残差编码。针对属性编码, 在已编码的几何结构基础上, 对每个点的颜色、法线、反射率等属性进行编码, 其常用方法包括预测 (利用邻近点属性) 和变换 (如 RAHT、Lifting)。

3. 基于 AI 的沉浸式媒体编解码

随着深度学习技术的快速发展, 基于 AI 的端到端沉浸式媒体编解码开始成为重要的研究方向。其核心思想是利用神经网络替代或增强传统编解码流水线中的关键模块 (如变换、量化、熵编码、环路滤波等), 实现更高效的率失真优化。在点云编码方面, 基于学习的点云几何压缩方法利用 3D 卷积神经网络或点云 Transformer 直接学习三维空间中的几何冗余, 已在部分场景下超越了 MPEG G-PCC 的压缩性能。在视图合成方面, 结合 NeRF/3DGS 的神经表示, 可以实现编码少量视角、解码端神经渲染任意视角的全新编解码范式, 从根本上改变了传统多视角视频技术编码所有视角的思路, 具有颠覆性的压缩效率潜力。

在 6G 算网融合与原生 AI 的赋能下, 沉浸式媒体编解码将向着三维化、智能化与协同化加速演进, 为构建“虚实融合”的新纪元奠定坚实的数据基座。

3.3.2.4 面向神经辐射场的体积视频编码

神经辐射场 (NeRF) 通过神经网络隐式地表达三维空间中的体密度与辐射颜色, 在逼真度与高自由度视角合成上表现卓越, 已成为体积视频的重要表达形态。然而, 当处理包含时间维度的动态视频序列时, 连续帧的独立表示会导致数据量呈指数级增长。面向神经辐射场的体积视频编码技术旨在解决动态长序列的存储与传输瓶颈, 其核心在于消除高维隐式特征空间中的空间冗余与时间冗余, 以满足流媒体传输与移动端渲染的标准要求。NeRF 体积视频压缩框架主要包含特征降维映射、显式残差建模与端到端联合优化等策略。第一种策略通过将三维或四维辐射场数据提取并规则化排列, 映射为 2D 特征图像视频流再调用 2D 视频编

解码器进行硬件加速解码^[73]。第二种策略则在原生特征域进行建模，利用运动网格记录三维空间位移，结合残差网格或紧凑的三平面（Tri-plane）结构对相邻时间戳之间的特征差异进行稀疏补偿，大幅降低帧间冗余^[74]。训练过程中将表示学习与特征压缩统一在率失真优化框架下，内置特征压缩子网络以进一步缩减体积。此外，为适应异构网络带宽与终端算力差异，建模过程还需引入分层渐进式编码与可变码率机制，使得单一神经网络模型能够输出多个不同精度级别的细节层次，终端可根据实际网络状况自适应请求并解码不同码率的体积视频分片^{[75][76][77]}。

3.3.2.5 面向 3D 高斯泼溅的体积视频编码

3D 高斯泼溅（3DGS）作为一种显式三维几何原语表达方式，通过优化数以百万计的带有特定参数（如位置、协方差、球谐函数等）的高斯椭球，结合可微光栅化技术实现了超实时的照片级渲染。但在动态体积视频（4DGS）场景下，由于每一帧的高斯点云属性随时间不断漂移变化且缺乏显式的帧间对应关系，使得动态高斯序列的存储体积极为庞大。面向 3D 高斯泼溅的体积视频编码技术，其核心挑战在于如何在保持极高渲染速度的前提下，通过冗余剔除与运动解耦机制实现高压缩比的流式传输。

为了适配现有移动端硬件的视频解码生态，部分技术方案将动态高斯的复杂多维属性降维映射并封装为标准的 2D 网格化视频流，通过强制的空间与时间一致性约束，使其能够被传统 2D 视频编码器高效压缩并在移动端渲染^[78]。在原生三维压缩路线上，则通过引入具备运动感知的动态高斯表示法，将物理场解耦为宏观尺度的刚性变换与微观尺度的局部残差形变，利用多分辨率运动网格和稀疏补偿高斯来精准捕捉帧间动态，并结合端到端的微型熵模型进行极限特征压缩。针对大规模流媒体分发场景，现代 4DGS 编码进一步融入分层渐进式架构，将高斯参数拆分为基础宏观运动层与增量式增强细节层，使得单个紧凑的码流文件即可支持多重码率输出，客户端能够根据实时网络状况动态截取并解码对应的画质层级，完成平滑的沉浸式交互体验^{[79][80]}。

3.3.3 通信感知一体化

沉浸式通信标志着移动通信从“尽力而为”的数据传输向“全真全感”互联的深刻转型，从而实现从信息传递到多感官体验的范式跃迁。智能体、扩展现实（XR）与全息通信共同构

成这一愿景的技术载体^[81]。根据 ITU-R 发布的 IMT-2030 框架,6G 将支持峰值速率 1-10 Tbps、用户体验速率 1 Gbps,较 5G 提升 10-100 倍,为全息通信的实时传输奠定带宽基础。同时,端到端时延需降至 0.1-1 ms,以支持触觉互联网和实时闭环控制,满足人类视觉 (> 20 ms 敏感)、听觉 (> 10 ms 敏感)与触觉 (< 1 ms 要求)系统的严苛生理特性。

在此背景下,通信感知一体化 (ISAC) 被定位为 6G 的关键使能技术,其战略价值体现在以下三大维度^{[82][83]}:

- 能力使能:通过感知获取的用户和环境信息支撑个性化、自适应的沉浸式服务;
- 效率使能:通过资源共享和协同优化提升频谱与能量效率;
- 体验使能:通过极低时延、高可靠通信与高精度感知保障流畅自然的沉浸体验。

3.3.3.1 通感一体化频谱资源

通信感知一体化通过同一信号与硬件平台融合高速传输与环境感知功能,为沉浸式体验提供高吞吐、低时延、高精度空间交互的能力。其中,毫米波、太赫兹与可见光技术特征对比如表 2 所示,可以发现它们在物理特性上天然互补,共同构成了支撑 6G 沉浸式通信的“黄金三角”频谱资源。在沉浸式通信的各类应用场景中,三者的技术优势通过精妙的协同设计得以充分发挥,实现“1+1+1>3”的系统效能。

表 2 面向沉浸式通信的多频段 ISAC 技术特征对比

技术维度	可见光 (Li-Fi)	太赫兹 (0.1-10 THz)	毫米波 (24-100 GHz)
核心优势	无频谱授权、高安全、照明通信一体化、亚毫米级定位	超宽带 (Tbps 级)、极窄波束、毫米级成像/手势识别	大规模 MIMO 成熟、兼顾覆盖与带宽、厘米级感知
典型速率	单链路 1 到 14 Gbps,多色 LD 可达 100+ Gbps	单链路 100 Gbps 到 1 Tbps,适合短距超高速	单用户 10 Gbps+,多波束复用百 Gbps 级
感知能力	3D 定位、姿态追踪、手势识别 (基于光场/TOF)	亚毫米级手势、隐蔽物体成像、生命体征监测	角度/距离/速度测量、人体轮廓、场景重构

技术维度	可见光 (Li-Fi)	太赫兹 (0.1-10 THz)	毫米波 (24-100 GHz)
覆盖范围	室内短距 (1 到 10 m), 依赖视距 (LoS)	近场 (1 到 10 m), 准光学传播, 大气窗口敏感	室内/场馆 (10 到 100 m), 可通过 RIS 补盲
沉浸式定位	室内 XR 终端、头戴设备 (HMD) 高精度锚点	近场全息交互、数字孪生实时映射、精密操控	场馆级全息通信、多人 XR 协同、低空终端

3.3.3.2 多频协同通感应用范式

1. 超高分辨率全息远程呈现

全息通信要求传输 8K/16K 3D 视频流 (需 > 100 Gbps), 同时实时捕捉用户微表情和手势 (需亚毫米级感知)。多频协同方案中, 下行链路利用太赫兹频段的超大带宽传输海量全息数据流, 确保画面无压缩、零延迟^{[84][85]}; 上行感知利用可见光进行高精度面部和手部动作捕捉, 可见光波长的散斑效应和多径特征能解析微米级的皮肤纹理变化^[86]; 辅助链路以毫米波作为控制信令通道和备用链路, 当太赫兹被遮挡或可见光阻断时迅速接管连接, 保证体验不中断^[87]。

2. 工业元宇宙与数字孪生实时映射

在大型工厂环境中实时构建厘米级精度 3D 动态模型, 需要多频协同: 毫米波基站覆盖整个车间, 提供基础的连接和大范围人员/车辆定位 (精度约 10 cm)^[87]; 太赫兹雷达通信节点在关键作业区利用其极高的距离分辨率 (毫米级) 对高速运动的机械臂进行实时建模, 检测微小裂纹或形变^[88]; 智能照明系统 (VLC) 作为绝对位置参考系, 为 AGV 小车和 AR 眼镜提供 < 1 cm 的绝对定位, 校正射频信号的累积误差^[86]。

3. 隐私保护的居家健康监测

在不使用摄像头的约束下实现老年人跌倒检测、睡眠呼吸监测, 需要多频协同: 毫米波穿透衣物和非承重墙, 监测大范围活动轨迹和跌倒事件; 太赫兹波对水分子极其敏感, 可非接触地透过衣物精确测量皮肤下血液流动引起的微小位移, 实现高精度心率和呼吸率监测; 可见光在特定区域 (如浴室门口) 形成光幕, 一旦光线被遮挡即触发警报, 作为最后一道防线。

3.3.3.3 关键技术突破

实现上节中多频段 ISAC 的高效运行，需突破以下核心技术体系：

1. 异构频谱融合的一体化波形设计

ISAC 的核心是解决通信高吞吐与感知高分辨的目标冲突，三大频段需针对各自物理特性设计融合波形：

- 可见光：采用时空联合调制（强度调制+主动空间调制），结合压缩感知的角投影成像，在 3 Gbps 通信下实现 39 Hz 帧率的 3D 定位；采用 DMT/OFDM 调制平衡照明与通信性能。
- 毫米波：多波束协同波形，通过多 RF 链生成独立通信/感知波束，在相干时间内分离时不变通信波束与时变感知波束；基于 OFDM 的雷达通信融合波形（如 OFDM-MIMO 雷达）。
- 太赫兹：超宽带脉冲/连续波融合波形，兼顾高速数据传输与高距离分辨率；考虑近场波束分裂效应的波形优化，适配短距精密感知。

2. 多模态感知融合与 AI 原生智能

- 跨频段数据对齐：解决不同频段采集数据在时空域的对齐问题，利用卡尔曼滤波或粒子滤波将毫米波的粗略轨迹、太赫兹的精细轮廓和可见光的绝对坐标融合^[89]。
- AI 驱动的语义感知：利用深度学习直接从原始 IQ 数据（射频）和光强序列（光学）中提取高层语义（如“有人在跳舞”而非单纯的“有物体移动”），降低回传数据量^[84]。
- 原生 AI 架构：将 AI 深度嵌入通信协议栈，支持实时信道估计、智能波束赋形与追踪、动态资源分配，实现从“反应式”到“感知驱动式”的范式转变^[89]。研究表明，AI 驱动的太赫兹（THz）系统可显著提升波束管理效率与链路可靠性^[84]。

3. 智能超表面与动态波束管理

- 宽频带智能超表面（RIS）：研发能同时响应毫米波、太赫兹甚至可见光的可重构智能表面，不仅能反射通信信号，还能将感知波束导向盲区，消除阴影效应^[89]。
- 极速波束追踪：利用低频段感知结果预测用户轨迹，提前控制高频段波束进行预对

准，实现无缝切换^[84]。

- 联合波束赋形：设计同时优化通信信噪比（SNR）和感知克拉美-罗下界（CRLB）的预编码矩阵，实现“一束多用”^[89]。

4. 近场通信感知与硬件器件创新

- 焦束成形（Beam Focusing）：不仅控制波束方向，还控制波束的聚焦深度，将能量精准汇聚在用户设备上，同时提高该点的感知分辨率。
- 光子集成电路（PIC）：利用硅光技术，在同一芯片上集成激光器、调制器、探测器，实现太赫兹信号的光学生成与探测，大幅降低功耗和体积。
- GaN 基同质集成光发射/接收芯片：单芯片实现通信与感知功能，降低系统体积与功耗。

3.3.4 全息呈现与全感官交互

3.3.4.1 全息与光场传输技术

在面向 6G 等下一代沉浸式通信的探讨中，产业界常提及的“全息通信”与物理学概念上的“光学全息”存在本质区别。

光学全息是利用光的干涉和衍射原理，直接记录并再现物体完整波前信息的技术。由于它在物理层面上还原了场景的真实光波场，能够提供包含所有深度线索的完美三维视觉体验，尽管光学全息被视为三维显示的终极形态，但在现阶段及可预见的未来内，极难实现大规模商业化。其核心阻碍在于严苛的硬件物理极限，物理再现高质量的全息图要求显示面板的像素间距达到亚波长级别；同时，实时生成计算全息图（CGH）所需的数据传输率和计算算力极其庞大，远超当前集成电路、空间光调制器（SLM）制造工艺以及现有通信网络的承载极限。

鉴于光学全息极高的物理与工程壁垒，当前 3GPP、ITU 等通信标准组织在讨论“全息通信”时，实际上是指通过高维度的三维数据采集、传输与计算机图形学渲染，在视觉体验上提供沉浸感的媒体技术。其实质是以三维点云和密集光场阵列为代表的媒体形态。通信领域的全息与光场传输，利用多角度摄像头阵列同步捕捉真实场景的几何结构（Depth/Geometry）与纹理外观（Texture）。它将真实世界转化为数字化的三维模型或密集视点集合，支持六

自由度（6DoF）。这意味着用户可以在虚拟的三维空间内自由改变位置与视线，“绕到物体背后”或贴近观察细节，获得极高的沉浸感与交互性。

这种技术路径依靠几何与纹理构建真实感也会产生大量数据，原始数据每分钟也可达TB级，因此，面向6G的通信全息传输，其核心挑战在于：如何在有限的网络带宽与端侧算力条件下，通过高效的三维数据压缩编码、智能视点自适应传输以及端云协同渲染，实现高保真、低延迟的沉浸式三维内容分发。

3.3.4.2 全感官交互技术

全感官交互技术是沉浸式通信的重要支撑技术，通过多模态感知与反馈硬件、实时信号处理算法以及跨模态融合机制的协同，构建覆盖触觉/力觉、嗅觉/味觉等多维感官的交互闭环，推动沉浸式通信由传统视听交互向多感官协同交互演进，从而提升人与数字空间交互的真实感、自然性与沉浸体验。

1. 触觉/力觉反馈技术

触觉/力觉反馈技术的核心是将数字物理属性精准转化为人体可感知的刺激，提升交互真实感。例如在沉浸式协同设计中，用户可通过触觉手套感知物体表面粗糙度、光滑度等材质特性。其技术架构主要由虚拟物理引擎、信号编码模块、反馈执行硬件三部分构成。虚拟物理引擎基于有限元分析（Finite Element Analysis, FEA）与离散元法（Discrete Element Method, DEM），实时计算虚拟物体的材质、运动学等参数，输出法向力、切向力、振动频率等力觉特征，力分辨率可达0.1 N，振动采样率100 Hz。信号编码模块采用脉冲宽度调制（Pulse Width Modulation, PWM）等技术，将力觉特征编码为硬件可识别的控制信号，并针对纹理、碰撞等场景优化信号形式，信号传输延迟低于20 ms，满足实时交互要求。反馈执行硬件以压电陶瓷驱动器、形状记忆合金（Shape Memory Alloy, SMA）执行器、电磁振动马达为核心，分别适配精细纹理、大力度碰撞、宽频振动等不同反馈场景。

在上述技术架构的基础上，当前重点聚焦于触觉信号的高效编码、跨模态融合处理与低延迟传输等关键问题。在编码与传输层面，围绕触觉语义压缩、可扩展编码与自适应实时流传输等方向持续探索，针对带宽受限与信道波动条件下触觉交互实时性与感知质量难以兼顾的核心痛点，形成了从语义编码到端到端传输优化的系统性研究方案^{[90][91][92]}；并从跨模态语义相关性与信息论视角，研究视觉先验辅助触觉编码的理论基础，为同步提升编码效率与

传输可靠性提供支撑^{[93][94]}。在跨模态融合与重建层面，聚焦于视觉与触觉特征的深度联合建模，探索视觉驱动的接触力推断、具身智能启发的触觉压缩，以及基于基础大模型的视触觉融合语义通信等前沿方向^{[95][96][97]}，并面向 XR 沉浸式场景研究生成式 AI 赋能的触觉传输新架构^[98]。研究覆盖从编码、传输到重建的完整技术链条^[99]。

2. 嗅觉/味觉模拟技术

嗅觉/味觉模拟通过分子精准释放与神经电刺激两种技术路径，复现虚拟环境中的化学感知，可应用于虚拟旅游、数字美食等场景，让用户感知环境气味与食物味觉。其核心技术体系包含气味合成、味觉刺激、环境联动控制三部分。气味合成系统采用微流控芯片与气相色谱-质谱联用（Gas Chromatography-Mass Spectrometry, GC-MS）技术，对微胶囊内气味前体物质精准配比释放。微流控芯片通道宽度 10 μm ，压电阀流量调节精度达 0.1 $\mu\text{L}/\text{min}$ ；GC-MS 实时监测气体成分与浓度，形成闭环反馈以调控配比，气味复现度超 90%。味觉刺激系统依据味觉受体电生理特性，借鉴经皮神经电刺激原理，采用舌面电刺激技术，通过舌面柔性电极阵列输出特定电信号，模拟酸、甜、苦、咸、鲜五种基本味觉，识别准确率达 85% 以上。环境联动控制集成计算机视觉、位置追踪及即时定位与地图构建（Simultaneous Localization and Mapping, SLAM）技术，实现场景化精准触发。以虚拟咖啡交互为例，系统识别目标并定位后，同步释放对应香气与味觉电信号，整体响应延迟控制在 50 ms 内，提升沉浸式体验。

在多模态感知融合层面，当前聚焦于跨模态显著性预测与多感官融合识别等方向。在显著性感知方面，探索融合视觉、触觉等多模态特征构建通用显著性预测模型，为嗅觉等多感官的联动触发提供精准的注意力定位依据^[100]；在数字孪生与多感官融合方面，将视觉、触觉、听觉与嗅觉等多模态统一纳入联合建模机制，为沉浸式环境中的场景理解与嗅觉精准触发提供技术支撑^[101]。

3. 多模态融合技术

多模态融合技术面向沉浸式通信中视觉、听觉、触觉等多维感知信息的协同处理需求，其核心在于打通不同模态之间的信息壁垒，实现统一语义理解、协同传输与联合呈现，从而支撑更加真实、自然和连续的沉浸式交互体验。与传统多媒体业务主要关注音视频同步不同，沉浸式通信中的多模态融合进一步强调不同感官信息之间的语义关联、优先级区分与协同恢复能力，可广泛应用于沉浸式社交、远程协作、数字孪生、XR 交互、人机协同等场景，为

用户提供由“看得见、听得到”向“可感知、可交互、可协同”演进的全感官体验。

在这一方向上，当前重点聚焦于跨模态语义建模、触视觉协同感知、语义级联合传输以及多业务流资源协同等关键问题。围绕视觉、触觉、语音等异构模态的语义关联挖掘，探索构建统一的跨模态语义表示框架，为沉浸式场景中的多模态信息融合、共享理解和一致呈现提供基础支撑^{[102][103]}。围绕触觉等高实时性模态在沉浸式业务中的传输需求，推进跨模态辅助恢复与联合重建能力，利用模态间冗余信息提升触觉等弱模态在受限网络条件下的恢复质量，从而增强交互连续性和感知一致性^{[104][105]}。面向沉浸式通信业务中“多模态、多任务、多链路”并存的复杂场景，进一步关注多模态语义信息的协同调度与资源分配问题，探索在带宽、时延、可靠性等多维约束下实现关键语义优先传输和多业务流联合保障，为高沉浸感业务提供更加稳定的传输支撑^[106]。同时，面向 6G 与元宇宙新型应用场景，研究生成式智能赋能的跨模态语义恢复机制，推动多模态融合由传统的数据拼接与同步处理，进一步向语义增强、智能恢复和自适应协同演进^[107]。在系统演进层面，探索联邦学习支持下的跨模态语义协同机制，以增强多终端、多节点环境中的分布式训练、自适应优化和隐私保护能力，为未来大规模沉浸式通信系统部署提供支撑^[108]。

总体上看，多模态融合技术将成为沉浸式通信由“多媒体传输”走向“语义交互服务”的关键支撑方向。围绕统一语义表示、跨模态协同恢复、智能资源调度和生成式增强等能力持续布局，有望进一步提升沉浸式业务在真实感、交互性、稳定性和可扩展性等方面的综合表现，并为全感官交互系统在教育培训、远程医疗、工业协作、数字文旅等领域的落地应用提供技术基础。

4. 技术挑战与发展方向

尽管全感官交互技术在触觉反馈、嗅觉/味觉模拟以及多模态融合等方面已取得重要进展，但面向沉浸式通信业务的大规模应用，仍面临多方面关键挑战。首先，在感知与反馈设备层面，触觉/力觉硬件仍存在体积较大、功耗较高、柔性不足，所导致的长时间佩戴舒适性有限等问题，制约了其在轻量化终端和日常沉浸式场景中的持续部署。同时，嗅觉与味觉模拟系统在气味成分库规模、释放精度、残留控制和个体感知差异适配等方面仍存在明显局限，仍难以实现复杂化学感知的稳定、细腻且可重复的实现。其次，在信息表示与协同处理层面，不同模态在采样频率、数据结构、传输时延和感知敏感度等方面差异显著，现有方法虽然能够实现数据级同步与业务级调度，但在统一语义表示、跨模态语义关联建模和跨模态

联合恢复方面仍不够成熟，尤其是在开放动态网络环境下，如何兼顾语义一致性、实时性与鲁棒性，仍然是多模态融合面临的核心难题。进一步地，在系统支撑与评价层面，当前全感官交互业务尚缺乏统一的多模态时间同步机制、跨模态语义对齐标准以及面向沉浸体验的质量评价体系，这使得不同设备、不同网络 and 不同业务场景之间的协同适配能力仍然有限。

面向未来，全感官交互技术的发展将主要沿着轻量化感知终端、智能化化学感知、多模态语义协同和体系化服务保障四个方向持续推进。在设备形态方面，柔性电子、微纳制造、低功耗驱动器件及新型功能材料的发展，将推动触觉与味觉反馈终端向轻量化、可穿戴、低功耗和高舒适度方向演进。在化学感知方面，合成生物学、微流控与智能控制技术的结合，有望提升气味与味觉生成的丰富度、精准度和动态调控能力，从而更好支撑文旅、教育、数字消费和医疗训练等沉浸式应用场景。在多模态协同方面，未来技术重点将由传统的数据同步进一步转向语义同步与任务驱动协同，即通过对视觉、听觉、触觉、嗅觉等模态的共享语义表示、重要性排序和跨模态生成式恢复，实现关键语义的高效传输与一致呈现。在网络与服务层面，随着边缘智能、联邦学习、生成式人工智能和确定性传输机制的持续发展，全感官交互系统将逐步形成“感知—编码—传输—恢复—评价”一体化闭环，为沉浸式通信业务提供更加自然、稳定和智能的全感官交互能力。

3.3.4.3 三维重建与实时渲染

实时神经渲染与动态场景重建技术依托 NeRF、3DGS 及 4DGS 等核心算法，深度结合 6G 通信的超低时延、超高带宽、通感算一体化特性，实现了超高精度、低延时的虚拟场景生成与动态更新，完美适配 6G 沉浸式通信的核心需求。相较于传统 3D 渲染技术，该技术体系无需依赖复杂的 3D 模型手工建模，可直接从图片、视频或稀疏传感器数据中重建 3D 模型和场景，解决了传统建模技术人工成本高、建模周期长无法满足 6G 沉浸式通信时代海量 3D 内容需求的痛点^[109]。

NeRF 是一种使用神经网络来隐式表达 3D 场景的技术。不同于传统的三维重建方法把场景表示为点云、网格、体素等显式的表达，NeRF 将场景建模成一个连续的 5D 辐射场隐式存储在神经网络中，只需输入稀疏的多角度带姿态的图像训练得到一个神经辐射场模型，根据这个模型可以渲染出任意视角下的清晰的图片。也就是构造一个隐式的渲染流程，其输入是某个视角下发射的光线的位置，方向以及对应的三维坐标，送入神经辐射场函数得到体密度和颜色，最后再通过体渲染得到最终的图像^[110]。

3DGS 是一种基于高斯分布的点云表示与渲染技术，核心思想是将三维场景建模为一系列可学习的高斯分布，每个高斯分布具有以下属性：位置（均值），高斯分布的中心坐标；协方差矩阵，控制高斯分布的形状（椭圆体的大小和方向）；颜色，即 RGB 值，表示该区域的颜色；不透明度，控制该高斯分布对最终图像的贡献权重。3DGS 以其高效的实时渲染能力著称，相比于 NeRF 隐式神经网络表示方法，3DGS 通过直接使用高斯分布来表征场景中的点云数据，避免了复杂的神经网络计算过程^[111]。在同等画质（PSNR/SSIM）下，3DGS 可实现 100FPS 以上的实时渲染速率，较 NeRF 提升 2-3 个数量级，训练时间约为 20-30 分钟（单 GPU，中等规模场景），显著优于 NeRF 的数小时训练周期。这种轻量级的设计使其能够在单 GPU 上快速运行。同时，3DGS 不依赖任何神经网络结构（如 MLP），因此其计算复杂度较低，对硬件资源的需求也较小。即使是在大规模场景中，也可以通过单张 GPU 实现高质量的重建和渲染^[112]。

4DGS 是在传统 3DGS 基础上引入时间维度，用于高效表示和渲染动态场景的前沿技术。其核心思想是将每个高斯椭球体在时空四维空间（ x, y, z, t ）中建模，即建模场景随时间变化的动态信息，实现对动态场景的连续建模与渲染。总之，4DGS 可以被理解为“随时间变化的 3DGS”，它不仅继承了 3DGS 的高效渲染优势，更实现了对复杂动态场景的高保真连续还原。Streaming GS 是针对持续输入的视频流实现在线流式重建的前沿技术。传统的 3D/4DGS 通常需要事先采集全量视角图像进行耗时的逐场景优化，难以应用于实时直播等对时效性要求极高的场景。相比之下，流式重建采用前馈推断与增量更新相结合全新范式，能够随视频流输入逐帧生成 GS 流，实现了从“离线全局优化”到“在线即时生成”的跨越。在此基础上，针对连续长视频流带来的数据堆积与显存溢出挑战，最新的方法引入了自适应密度控制与冗余压缩机制，有效遏制了计算和内存开销的膨胀^[113]。此外，部分前沿方法进一步打破了对预设相机参数的强依赖，能够实现在未知的相机轨迹下“边走边建”。这类具备自适应压缩能力且轻量的流式架构，不仅大幅提升了重建速度，更进一步契合了沉浸式通信对超低时延和即时反馈的要求。

在此基础上，进一步发展的“可驱动高斯”强调动态场景不仅能够被重建与回放，还能够被参数化控制与编辑。其中，一个重要方向是将网格（Mesh）与高斯表示相结合：以网格作为稳定的几何骨架、拓扑约束和驱动载体，通过骨骼动画、表情参数、物理仿真或控制点位移来驱动网格形变，再将这种形变同步传递至附着其上的高斯分布，从而实现高保真、可控的动态渲染。相比纯高斯时序建模，Mesh-Gaussian 混合表示兼具显式结构约束与高质量

视效表达的优势,一方面可借助网格保持人物、服饰或器物的整体几何一致性与时序稳定性,另一方面可利用高斯对细节纹理、半透明区域和复杂外观进行高效表达^[114]。因此,可驱动高斯有望成为数字人、虚拟角色、沉浸通信和交互式内容生产中的关键技术路径^[115]。

目前,在沉浸式通信场景中,4DGS 面临的核心挑战包括:

- 时空高斯参数量随时间线性增长导致的存储与传输压力;
- 动态拓扑变化(如人物出入场景)下高斯分布的自适应增删策略;
- 多用户协同场景下 4DGS 模型的一致性同步问题。

针对上述挑战,当前研究方向包括基于关键帧的增量式 4DGS 更新、时空高斯参数的压缩编码、以及与 6G 边缘计算结合的分布式 4DGS 渲染架构等。

以上技术结合 6G 边缘计算,在沉浸式通信中的应用包括:

- 3D 超写实数字人重建与驱动,支撑虚拟社交/大型活动/赛事的实时三维数字人直播;
- 体积视频实时重建与传输,实现远程全息会议中参会者的实时 3D 重建;
- 大规模真实场景的高保真重建,如数字孪生工厂的实时镜像、博物馆文物的超精细 3D 数字化、城市级场景的沉浸式漫游等。

第四章 基于网媒融合的沉浸式通信应用场 景与产业赋能

本章紧扣沉浸式通信全感官化、交互实时化、虚实一体化、智能内生化的四大核心特征，从前沿消费体验切入，聚焦民生与生产核心场景，最后剖析上下游产业生态。旨在通过具象化的应用场景，为技术落地、产业协同及价值释放提供体系化指引，助力沉浸式通信与各领域深度融合。

4.1 典型沉浸式通信应用场景

4.1.1 文娱与消费：重塑感官体验边界

1. VR 大空间



图9 VR 大空间沉浸式体验

当前，VR 大空间沉浸体验在全国多地加速落地，行业形成多元化发展格局。2025 年 4 月，中央电视台报道湖北省博物馆 VR 大空间项目，该项目由中兴通讯运用网媒融合关键技术，实现 100 名用户同时在场，与文物开展跨时空互动（如图 9 所示）。此外，上海推出“再见青花·海上丝绸之路沉浸式体验展”，依托网媒融合技术实现 4K@90Mbps 超高清串流，精准还原青花瓷釉色与纹理细节，观众可化身“青花精灵”，沉浸式体验郑和下西洋的历史历程。

政策层面，2025 年国家电影局明确将 VR 大空间作品纳入独立影片类别，推动其与电影产业深度融合，赋予“空间叙事”全新表现形态，支持观影者在虚拟场景中自由探索互动，重构电影传播模式。技术层面，8K 超高清、自由行走、多模态互动已成为行业核心技术标准。基于 AI 驱动的个性化内容生成技术，能够实现体验场景精准适配，打造身临其境的极致沉浸体验。上述技术的规模化落地，需要依托 6G 技术所具备的大带宽、低时延、多模态同步传输及网媒融合等核心能力支撑，助力我国虚拟现实电影产业高质量发展。

2. 全息演唱会

如图 10 所示，通过实时体积视频与云渲染技术，异地歌手的三维体积影像被实时投射于全息舞台。观众可通过 MR/VR 终端自由切换多视角观看，结合环绕声场，实现虚实交融的视听震撼。例如，异地数字化身与现场乐队共舞，其细腻的面部表情与动作细节通过体积视频精准还原，打造跨地域联动的娱乐新范式。



图 10 虚拟全息演唱会

4.1.2 工作与合作：打破物理空间壁垒

1. 沉浸式多人协作

身处异地的员工借助虚拟化身(Avatar)进入统一的 3D 虚拟办公间,依托 6G 网络(MTP < 10ms, 带宽 > 1Gbps)实现零延迟交互。通过空间音频、手势、触觉等多模态方式,在共享数字白板上研讨方案,并可对 3D 模型进行拆解演示。该模式在云端渲染,大幅降低终端算力负载,实现高品质沉浸式协同,彻底消除远程办公的孤立感,提升跨地域协作效率。

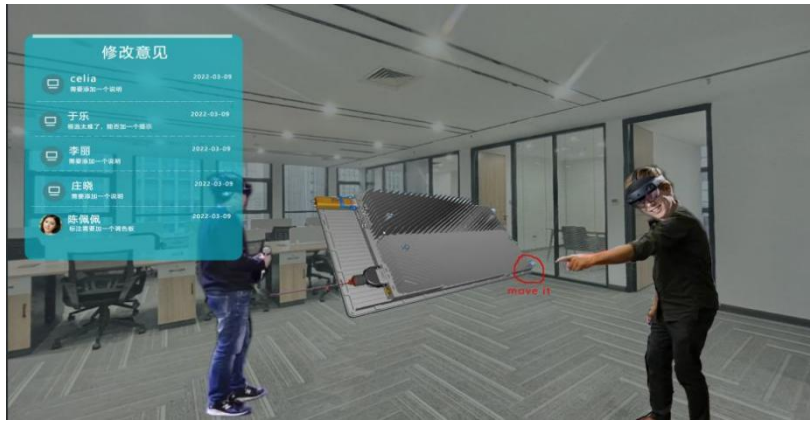


图 11 虚拟会议

如图 11 所示，在虚拟会议场景中，参会人员可通过自身虚拟化身，借助手势操作对服务器等 3D 模型进行拆解演示，在空间共享数字白板上研讨设计方案、实时批注修改，同时依托空间音频实现高清临场沟通，系统还可自动完成多语种实时翻译同步。相较于传统视频会议的桌面共享模式，沉浸式协同办公能够大幅缓解远程办公人员的孤立感，在缩减差旅成本、降低办公开支的同时，切实提升跨地域团队的沟通协作效率。

2. 远程全息会议

远程全息会议场景（如图 12 所示），通过融合 3D 超写实数字人与物/场三维重建、体积视频技术，构建出具有空间感知与物理交互能力的虚拟会议空间。该场景采用多视角摄像头阵列或深度传感器阵列，实时捕捉参会者的几何结构、纹理特征与动态动作数据。采用 3DGS 或 NeRF 等神经辐射场/实时三维重建算法，将多视角采集数据转化为可多角度交互的 3D 动态模型，支持 6DoF 视角切换，并可结合手势进行自然交互。参会者佩戴双目 16K/8K 分辨率的 VR 头盔与动作捕捉组件，驱动超写实 3D 数字人完成自然动作；基于沉浸式协作框架构建的虚拟会议室，支持多人实时协同操作。参会者可围绕 3D 模型进行产品演示、数据标注或实时修改。通过空间定位技术实现物理空间与虚拟场景的坐标系对齐，消除传统视频会议的二维平面化局限。



图 12 远程全息会议

3. 远程专家指导

在工厂现场业务支持场景中（如图 13 所示），现场工人佩戴 AR 眼镜实时采集并回传现场作业画面与空间信息，远程专家通过基于沉浸式协作框架构建虚实融合的沉浸协作空间。远程专家通过 XR 实时云渲染平台将设备的 3D 模型进行渲染、编码并推流到工人的双目 4K/2K 分辨率的 AR 眼镜上，AR 眼镜通过空间定位技术将 3D 模型与现场设备精准对齐，最终呈现虚实融合的叠加画面。远程专家与现场工人通过数字白板协同、桌面共享、文件共享、AR 标注（含动态点标注）、音视频通话等方式进行沉浸式交互，大幅降低操作失误率，确保现场执行精准高效。该应用能显著提高现场支持效率，缩短现场业务处理时间。



图 13 远程专家指导

4. 基于 AI 智能体的机房智能巡检

在机房 AR 巡检中融入 AI 自然语音交互理解、意图识别、任务拆解、工具调用等能力，构建基于 AI 智能体的智能巡检体系，从“被动响应”转化为“主动提醒”，从“被动答疑”

变“主动预警”，降低用户操作门槛、提升处置效率和作业规范，让远程专家聚焦复杂问题。



图 14 基于 AI 智能体的智能巡检与指导

巡检人员进入机房后（如图 14 所示），AR 眼镜以第一视角持续采集用户环境的视频流、音频流、IMU 惯性信号、眼动数据、对话文本及 6DoF 位置信息等多模态数据。通过边缘意图识别模型进行意图识别与理解，用户无需再手动点选任务工单和频繁确认，AI 可根据现场情境，自动识别匹配执行巡检任务并自动补充巡检结果，用户只需决策确认，巡检过程中，AI 实时监测操作步骤，提供规范的指引和误操作告警，同时针对不同用户的巡检习惯和现场问题，调用合适 Agent 工具提供差异化建议与服务，如发现故障时，AI 智能体依托大模型上下文关联，结合历史故障数据、维修案例，自动判别中小故障，推送对应维修手册；复杂故障再联动远程专家，减少专家占用。

5. 沉浸式远程医疗

沉浸式远程诊疗场景（如图 15）是沉浸式医疗典型的应用场景，该场景依托基于 MR/VR 的沉浸式协作框架，打破地域壁垒，实现异地专家医生与本地医生的沉浸式协作诊疗。本地医生与患者身处实体诊疗环境，通过头戴式双目 16K/8K 分辨率的 MR/VR 设备接入边缘计算平台；异地专家则以“数字化身”形式，在虚拟空间中与本地医生实现空间锚定与实时交互。

诊疗过程中，系统依托超大模型实时云渲染技术，将患者的 CT、MRI 等二维影像高精度重建为三维全息解剖模型，悬浮于诊疗空间内。异地专家与本地医生可同步从任意角度缩放、剖切模型，精准定位病灶位置与毗邻关系，并进行虚拟标注、测量与手术路径模拟。结合 3D 超写实数字人、多模态交互能力，异地专家能如同亲临现场，通过手势、语音与本地医生实时沟通诊疗思路，直观演示操作要点，共同制定个性化方案。



图 15 沉浸式远程诊疗

此模式不仅解决了传统远程会诊依赖二维影像、沟通抽象的痛点，更让优质医疗资源跨越空间限制，大幅提升复杂病例诊断与手术方案制定的精准度与效率，尤其适用于基层医院疑难重症会诊与复杂手术指导场景。

4.1.3 教育与个人：普惠与高效并行

1. 沉浸式学习

在沉浸式学习的应用场景中（如图 16 所示），基于 MR/VR 的沉浸式学习系统分为教师端与多个学生端。教师通过中控台 PC 终端，同步管控所有学生端 MR/VR 终端；同时启动自身高权限的 MR/VR 终端，对教学 3D 内容进行拆解、组装等操作，作为教学示范。教师端佩戴双目 16K/8K 分辨率的 MR/VR 终端具备更高管控权限（如场景编辑、全局同步、学生端锁定等），与学生端权限区分，可以对 3D 内容进行操作，如拆解与组装。教学场景 3D 内容由边缘服务器实时渲染、编码后，以流媒体形式推流至各学生端 MR/VR 终端。教师以此进行教学示范与讲解，引导学生在其 MR/VR 终端上完成学习、交互与协作任务。此外，在自主学习模式下，学生端可独立操作 3D 内容进行探究式学习，以强化动手实践与探究式学习，提升知识掌握效率。



图 16 沉浸式教学

2. 个人生活助理

基于用户数字分身的 AI 智能体，是个人生活的智能管家，以拟人化交互、精准执行能力，为用户高效处理各类日常繁琐的事务，解放时间精力。用户只需通过语音或文字指令，数字分身便会全程自主完成服务，无需手动操作。

日常中，用户数字分身的 AI 智能体可自动梳理收件箱，筛选重要邮件并智能回复常规咨询，标记紧急邮件提醒用户；节假日出行时，用户告知出行时间、目的地及偏好，数字分身会对比票务价格、余票情况，一键预订火车票，同步匹配周边高性价比酒店，结合用户喜好制定详细旅游计划，包含交通、景点、餐饮安排及应急方案。

此外，它还能联动个人日程，提前提醒出行时间、酒店入住信息，实时同步行程变动。全程无需用户分心，数字分身以高效、精准的服务，覆盖日常办公辅助、出行安排等核心场景，让用户摆脱繁琐事务，专注于自身核心需求，实现个人生活的便捷化、智能化自助管理。

4.2 产业生态与赋能路径

沉浸式通信产业是由芯片、算力、内容、终端、网络与行业应用共同构成的复杂生态系统，各环节高度协同、相互赋能。本章从核心产业链、内容生态、XR 终端产业、产业赋能路径四个维度，系统分析产业格局、发展现状与未来演进方向。

4.2.1 产业链生态

沉浸式通信产业链可划分为上游底层硬件（芯片+算力）、中游内容与平台、下游应用服务三大层级，其中芯片与算力决定体验上限，内容决定用户粘性，终端决定入口规模，共同构成产业规模化发展的基础。

4.2.1.1 芯片

沉浸式通信芯片主要面向 AI/AR 眼镜与 VR/MR 头显两类终端，两类场景技术要求差异显著，市场格局与国产化进度呈现明显分化。

1. AI/AR 眼镜芯片

AI/AR 眼镜强调轻量化、低功耗、基础空间感知与 AI 交互，芯片门槛相对可控，国产替代进展较快。

- 国际方面：高通骁龙 AR1 在中高端 AI/AR 眼镜中占据主流地位，生态成熟、兼容性强。
- 国内方面：以 XREAL X1 空间计算芯片、紫光展锐 W337、物奇微、恒玄等为代表的国产芯片已规模商用，广泛应用于 XREAL、界环等主流品牌，在功耗、体积与基础 AI 能力上逐步实现对海外方案的替代。

2. VR/MR 高端空间计算芯片

VR/MR 对算力、图像处理、多传感器融合、低延迟交互要求极高，长期由海外厂商主导。

- 高通：骁龙 XR2+ Gen2 采用 4 nm 工艺，支持单眼 4K、多摄并发、12 ms 光子到光子延迟，是 PICO、创维、Meta 等主流 VR/MR 设备的通用方案。
- 苹果：M2+R1 双芯片架构为 Vision Pro 自研方案，5 nm 工艺，M2 负责主运算，R1 专责定位与视觉实时处理，可处理 12 摄+ 5 传感+ 6 麦输入，实现 12 ms P2P 延迟，确立高端 MR 标杆。
- 国产突破：2025 年 11 月万有引力 X100 芯片发布，成为国内首款对标 M2+R1 的全功能空间计算芯片，支持 10 ms P2P 延迟、8K@120Hz 混合现实显示，标志国产高端 XR 芯片实现里程碑式突破。

3. 国产专用芯片矩阵

万有引力已形成覆盖多场景的芯片体系，进一步完善国产供应链：

- X100：高端 MR 空间计算主芯片，超低延迟、超高画质。
- EB100：反向透视专用处理器，实现类 Vision Pro 透视效果，并在机器人双目视觉、灵动脸等领域落地。
- VX100：AI/AR 眼镜专用，尺寸仅 4 mm×7 mm，小于高通 AR1，支持手势识别、端侧大模型交互、空间视频拍摄。

4.2.1.2 算力基础设施

沉浸式通信高度依赖实时渲染算力与 AI 推理算力，两类算力共同支撑高画质、低延迟、智能交互的业务体验。

1. 实时渲染算力

实时渲染负责 3D 场景光影、纹理、画面输出，是沉浸式视觉体验的基础。

- 主流方案：以英伟达 RTX 5080 为代表的消费级 GPU 为主，性价比显著优于专业计算卡，适合云渲染规模化部署。
- 国产替代：摩尔线程庐山系列支持硬件光追与主流渲染 API，可满足实时云渲染需求。
- 部署形态：采用 PC Farm 机架式集群，单机 2~4 路独立节点，算力物理隔离，避免多用户抢占导致卡顿；配备 CRPS 冗余电源与 IPMI 2.0 远程管理，支持机房大规模稳定运维。

2. AI 推理算力

AI 推理主要服务于 3D 数字人驱动、智能体交互、场景理解、多模态感知等核心能力。

- 轻量级任务：复用渲染节点 GPU，实现渲染 + 推理共址部署，降低成本。
- 高并发任务：采用英伟达 RTX A 系列、L4 低功耗计算卡，保证持续稳定推理负载。
- 国产方案：寒武纪 MLU370-X4 等智能加速卡已可适配沉浸式场景 AI 推理需求。
- 部署架构：以边缘算力为核心，就近提供服务，显著降低端到端时延，保障虚实融合实时性。

4.2.1.3 内容生态环节

XR 内容是沉浸式通信触达用户、形成商业闭环的核心抓手，当前生态仍处于“供给不足、结构单一、成本偏高”的发展阶段。

1. 内容应用分类

- ToB 行业内容：覆盖虚拟办公、教育培训、工业巡检、产品展示、文旅文博、数字孪生等，场景价值明确、商业化落地较快。
- ToC 消费内容：高度集中于游戏，健身、社交、短视频直播、云旅游、影视等品类供给明显不足，导致 XR 设备仍被贴上“游戏机”标签。

2. 内容产业痛点

- 爆款内容匮乏，用户粘性不足。
- 传统 3D 制作流程长、成本高、门槛高，制约生态扩张。

3. AIGC 带来结构性变革

- 2D 内容：字节跳动 Seedance 2.0 实现电影级视频生成。
- 3D 内容：3DGS 技术大幅提升建模与大场景重建效率。

未来多模态 AIGC 工具普及后，将实现“文本/图像一键生成 3D 内容”，显著降低创作门槛，推动内容品类全面丰富，为沉浸式通信规模化提供核心动力。

4.2.2 XR 智能终端

XR 终端（AR/AI 眼镜、VR/MR 头显）是沉浸式通信的核心入口与主要载体，正从“显示外设”向“空间计算平台”跃迁，成为继手机之后新一代智能终端。

4.2.2.1 终端发展主线：AI 原生+空间计算

当前 XR 终端的核心演进方向是 AI 深度内化，人工智能不再是附加功能，而是贯穿光学、芯片、系统、感知、交互的底层内核：

- 从“语音助手单点能力”升级为端侧大模型原生支持，接入通义千问、Deepseek等国产大模型。
- 具备环境理解、场景识别、多模态交互、自动化任务执行能力，智能化从“被动响应”走向“主动助理”。
- 产品在显示清晰度、重量、续航、佩戴舒适度持续逼近消费级可用标准，整体体验从“可用”进入“好用”阶段。

4.2.2.2 交互范式重构：从平面到空间

XR 终端彻底重构人机交互方式，形成区别于传统智能终端的“三新”特征，如表 3 所示：

表 3 新一代 XR 眼镜交互范式演进对比表

核心维度	传统智能终端	XR 眼镜智能终端	带来的直观感受
交互方式	单一模态平面交互（触控、语音）	空间认知协同（眼动 + 语音 + 手势 + 空间定位）	更自然：像真人一样面对面交流，无需记忆指令，在三维空间中实现无感交互
应用架构	流程驱动（人适应软件操作流程）	空间意图驱动（软件适配人的空间需求）	更省事：一句话搞定跨应用、跨空间的复杂任务，无需分步操作
设备关系	点对点简单连接（单设备数据传输）	全场景空间共生（跨设备服务无缝流转，虚实空间协同）	更无缝：服务跟着人走，不受设备、空间限制，实现全场景虚实融合体验

4.2.2.3 国内 XR 终端产业格局

中国已形成完整且具备全球竞争力的 XR 终端产业体系：

- 上游：核心元器件自主可控程度持续提升，国产替代加速。
- 中游：制造能力全球领先，供应链成熟稳定。
- 下游：消费级市场国产品牌占据主导；行业级市场与国际巨头同台竞争。

整体呈现“科技巨头全生态布局、专业厂商垂直深耕”的竞争格局，终端品牌快速崛起，中高端产品逐步实现对海外品牌的追赶。

4.2.2.4 未来 3-5 年终端发展趋势

形态持续轻量化、全天候佩戴化、消费级普及化。

场景从影音、游戏、办公，全面延伸至工业、医疗、教育、文旅、交通、社交、家居等全领域。

XR 眼镜将成为空间计算入口与个人随身智能中心，实现千行百业规模化落地，成为沉浸式通信最重要的硬件载体。

4.2.3 产业协同与赋能路径

为推动沉浸式通信从试点走向规模化商用，需要从技术、生态、场景、产业合作四个层面构建赋能体系。

4.2.3.1 技术赋能：补齐底层短板

加快高端 XR 芯片、国产 GPU、光学器件突破，实现核心硬件自主可控。

完善端—边—云算力架构，优化低时延、高带宽、高可靠的网络适配能力。

推动 AIGC 与 3D 内容生产工具标准化，降低内容创作门槛。

4.2.3.2 生态赋能：构建开放共赢体系

联合芯片、终端、内容、运营商、应用开发商共建产业联盟。

开放 SDK/API 接口，吸引开发者与创作者入驻。

建立内容扶持、流量分成、商用变现机制，丰富 ToC 与 ToB 内容供给。

4.2.3.3 场景赋能：以应用牵引产业成熟

ToC 优先突破健身、社交、直播、云旅游、影视等高频场景，打造爆款应用。

ToB 聚焦数字孪生、远程协作、工业巡检、实训教学、文旅文博等可复制场景。

以典型场景带动终端普及、内容繁荣与网络升级，形成正向循环。

4.2.3.4 产业赋能：协同推进规模化落地

加强产业链上下游协同，推动芯片—终端—内容—算力一体化方案成熟。

推动行业标准制定，提升设备互通、内容兼容、业务互操作能力。

依托运营商 5G/6G 网络部署，实现沉浸式通信业务全域覆盖与商业变现。

4.2.4 标准化与国际合作

4.2.4.1 沉浸式通信标准化进展

沉浸式通信的标准化工作在主要标准化组织 ITU、3GPP、MPEG 等的进展^[120]如表 4 所示：

表 4 沉浸式通信组织的工作内容与目标

组织	主要内容与目标
ITU	<p>ITU 负责顶层设计与全球规则制定：</p> <ol style="list-style-type: none"> 1. ITU-R M.2160 (2023 年 11 月)：正式确立 IMT-2030 框架，将“沉浸式通信”列为六大关键场景之一； 2. 2024-2027 年：定义技术性能要求 (TPR) 和通用评估准则； 3. 2027 年 2 月-2029 年 2 月：启动 IMT-2030 候选技术方案 (RIT) 征集； 4. 2030 年：发布 IMT-2030 最终技术规格。
3GPP	<ol style="list-style-type: none"> 1. R18 确立 XRM (XR 与媒体服务增强) 架构及 MeCAR (AR 媒体能力) 规范，定义分离渲染基本流程，已于 2023 年完成； 2. R19 重点研究基于 IMS 的分离渲染 (SR_IMS)、数字人化身通信 (AvCa11) 以及实时传输协议 (RTP) 的 5G 增强 (5G_RTP_Ph2)，已于 2025 年完成； 3. R20 启动 FS_6G_MED (6G 媒体研究) 和 FS_6G_ARC (6G 架构研究)，探索 6G 环境下的沉浸式媒体交付架构及 AI 媒体表征，计划完成时间为 2028 年。
MPEG	<ol style="list-style-type: none"> 1. MPEG (ISO/IEC JTC 1/SC 29/WG 03) 围绕沉浸式通信正推进以 MPEG I (ISO/IEC 23090 系列) 为核心的标准体系，目标支撑面向 6DoF 的体积视频、沉浸式音视频与真实感交互场景；

组织	主要内容与目标
	<p>2. 其中 MPEG-I 第 2 部分（ISO/IEC 23090-2, Omnidirectional Media Format, OMAF）已形成稳定版本并持续修订以支持沉浸式流媒体分发；第 5 部分（Visual Volumetric Video based Coding, V-PCC）已发布并迭代以提升点云压缩效率；第 12 部分（Immersive Video, MIV）已完成标准化并进入增强阶段（eMIV），支持多视点视频的高效合成与 6DoF 观看；第 29 部分已完成 Video based Dynamic Mesh Coding(V-DMC) 等新型体积内容编码，以增强对动态网格与真实人像通信的支持；</p> <p>3. 未来计划方面，MPEG 将在 2026 - 2028 年间持续完善 MIV 增强功能、体积视频与动态网格压缩工具链，并与 3GPP/ITU 的 IMT 2030（6G）需求对接，形成面向实时沉浸式通信（如远程协作、数字人、全息通信）的完整端到端媒体标准体系。</p>

沉浸式通信的未来应植根于全球统一标准体系与跨行业的深度协同^[121]。建议 3GPP 及其合作伙伴继续强化与 5G-MAG、DASH-IF 及 Metaverse Standards Forum 等行业协会的联系，确保通信协议与互联网媒体分发工作流的无缝衔接。国际合作应重点关注多厂商互操作性验证，通过开放参考实现和联合测试，降低沉浸式业务的全球部署门槛。同时，在 6G 愿景下，各国应在频谱管理、算力共享以及数字身份的可信认证等领域达成共识，共同构建一个绿色可持续、隐私受保护且无缝漫游的全球沉浸式数字生活空间。

4.2.4.2 沉浸式媒体技术标准化进展

为应对沉浸式媒体巨大的数据量、复杂的格式以及互操作性问题，国际标准化组织 MPEG 自 2015 年起启动了雄心勃勃的标准化工作，逐步构建了一套完整的沉浸式媒体技术栈（MPEG-I, ISO/IEC 23090）^[122]。

1. 全向媒体格式（Omnidirectional Media Format, OMAF）

MPEG 于 2016 年正式启动 OMAF 项目，并于 2018 年发布了第一版标准（ISO/IEC 23090-2），是 MPEG-I（沉浸式媒体）系列标准的第一个正式标准。

OMAF 定义了统一的投影格式（支持立方体投影 CMP、等距柱状投影 ERP），引入将全景视频划分为独立区域（Tiles）的概念，为视角自适应传输奠定了基础。同时，OMAF 也规范

了一个端到端的应用格式，定义了基于 DASH（HTTP 动态自适应流协议）或 MMT（MPEG 媒体传输协议）的完整传输框架。

2. MPEG 沉浸式视频 (MPEG Immersive Video, MIV)

在 OMAF 解决了 3DoF 问题后，产业界对支持用户自由移动的 3DoF+或 6DoF 体验提出了更高要求。为此，MPEG 于 2018 年启动 MIV 项目，并于 2021 年发布第一版标准（ISO/IEC 23090-12），其中利用多视角视频与深度图作为输入，通过视角选择、打包（Packing）和虚拟视点合成技术，支持在解码端实时渲染任意 3DoF+视角。由于底层框架和技术路线存在相似性，在 MIV 标准后续扩展中，强化了与 V-PCC、V-DMC 等标准的协同。MIV 标准本身不定义底层编解码器，而是明确依赖 VVC（H.266）等先进 2D 编解码器作为其视频压缩引擎。

3. 基于视频的点云压缩 (Video-based Point Cloud Compression, V-PCC)

点云作为表示 3D 空间最通用的数据格式，在自动驾驶、VR/AR、数字孪生等领域的应用需求激增。2017 年，MPEG 发起了点云压缩的通用测试模型征集活动，其中，“基于视频”（Video-based）和“基于几何”（Geometry-based）两大技术路线脱颖而出。V-PCC 的基本思想是“化 3D 为 2D”，巧妙地利用了数十年来发展成熟的、极其高效的 2D 视频编解码技术（如 HEVC、VVC）来解决 3D 点云压缩难题。V-PCC 的核心环节包括体素化与分块、几何投影、属性投影、视频编码、元数据封装等。V-PCC 是 MPEG-I Part 5 标准（ISO/IEC 23090-5）的技术核心，也是其中首个、最成熟的实例。MPEG-I Part 5 标准是一个分层的、模块化的标准框架，其全称是 Visual Volumetric Video-based Coding (V3C) and Video-based Point Cloud Compression (V-PCC)，其中 V3C 定义通用比特流、语法、解码流程；V-PCC 则是 V3C 在点云场景的具体实现，定义了点云专用的投影、映射、元数据格式。

4. 基于几何的点云压缩 (Geometry-based Point Cloud Compression, G-PCC)

与 V-PCC 同期启动，G-PCC 主张采用原生 3D 算法直接在三维空间中对点云的几何结构和属性进行建模与压缩。2023 年，G-PCC 标准第一版发布（ISO/IEC 23090-9），其核心是先几何、后属性的分层编码。几何编码是 G-PCC 的基础，负责压缩点的空间位置信息，其中八叉树（Octree）是 G-PCC 用于无序点云的主要几何编码工具，而预测+残差编码主要用于有序点云，例如按激光雷达扫描线顺序排列的点。在几何结构被压缩后，G-PCC 接着压缩附着在每个点上的属性，如颜色、法线、反射率等。由于是原生 3D 处理，G-PCC 能实现无损或近无损压缩，重建的 3D 模型几何结构极其精确，没有投影失真或空洞问题。

5. 基于视频的动态网格编码 (Video-based Dynamic Mesh Coding, V-DMC)

随着元宇宙和虚拟人经济的爆发，对可实时驱动、高保真的动态 3D 角色（即变形网格）的压缩需求变得极为迫切。2022 年，MPEG 启动了 V-DMC 项目，旨在为动态网格提供一个专用的、高效的压缩标准。V-DMC 的核心是 UV 参数化 (UV Parameterization)，它将动态网格序列的几何、拓扑、属性等信息映射到一个公共的 2D UV 展开图上，将顶点的动态属性（如位置偏移）转换为 2D 纹理视频，再用 VVC 等先进 2D 编码器进行压缩。当前，V-DMC 标准已进入最终国际标准草案 (Final Draft International Standard, FDIS) 阶段 (ISO/IEC FDIS 23090-29)，完成技术定型，即将正式发布。

6. V3C 数据的封装 (Carriage of Visual Volumetric Video-based Coding Data)

由于同样是基于视频的压缩技术，MIV、V-PCC、V-DMC 实际上可以共享通用码流结构，因此，MPEG 组织研究了此类 V3C 数据在主流媒体文件格式和流媒体协议中的封装规则，发布了 V3C 数据的封装标准 (ISO/IEC 23090-10)。该标准通过标准化的封装规则，使得复杂的 3D 体积视频内容能够像普通 2D 视频一样，被存储在 ISO BMFF/MP4 文件中，并通过 DASH、MMT 等流媒体协议进行自适应传输。未来如果出现更高效的基于视频的压缩技术，只需将其适配到 V3C 框架下即可，无需重新定义整个标准，极大地增强了标准的前瞻性和灵活性，同时也促进了 V3C 技术的产业落地。

7. G-PCC 数据的封装 (Carriage of Geometry-based Point Cloud Compression Data)

与 V3C 数据的封装标准相对应，MPEG 组织也研究了 G-PCC 数据基于 ISO BMFF/MP4 协议的存储与传输技术，针对 G-PCC 数据的特性进行了定制，发布了 G-PCC 数据的封装标准 (ISO/IEC 23090-18)。与 V3C 封装的多轨道、多提取器模型不同，G-PCC 的封装模型相对简单。因为 G-PCC 数据是一个自包含的、原生的 3D 压缩流，它不像 V3C 数据那样严重依赖外部的视频编码器。因此，通常只需要一个 G-PCC 数据轨道即可完整描述内容，无需复杂的提取器机制来关联多个视频轨道。G-PCC 数据的封装标准也兼容 DASH、MMT 等流媒体协议，支持网络自适应传输。

第五章 未来展望与工作倡议

5.1 未来发展展望

前文从基础理论、体系架构及关键技术等维度，系统阐述了基于网媒融合的沉浸式通信体系架构，以及网侧协同和媒侧适配的关键技术，试图厘清网络与媒体相互融合的初步态势。展望未来，沉浸式通信的演进将不再局限于单点技术的突破，而是呈现出以用户为中心的智能连接，多模态数据重构与传输，以及内容走向“开源”的复合趋势。本章从智能体终端、生成式大模型、生成式内容分发网络及环境感知与重构等研究热点的角度，勾勒出面向未来沉浸式通信的网媒融合技术发展展望：

1. 以用户沉浸式体验为中心的智能体终端

随着 Openclaw、Seedance 等 AI 智能化应用的不断普及，6G 时代的终端形态将发生根本性变革。全国人大代表、中兴通讯高级副总裁苗伟指出，相比 5G，6G 的连接对象将从“人+物”扩展到“人+物+智能体”，海量智能终端需要全域覆盖、永不失联的确定性网络连接。

在此背景下，智能体终端将成为网络能力向用户侧的延伸。未来的终端设备不再是单纯的通信管道端点，而是具备本地 AI 计算、环境感知与自主决策能力的“网络智能体”。基于此，6G 将构建人、物、智能体和谐共生的新型基础设施，网络也将从被动响应的管道升级为主动服务的智能体网络。

2. 基于生成式大模型的多模态数据重构与分发

传统语音、视频通话追求物理层像素级保真，然而沉浸式通信不仅要传输视听信息，还需同步触觉、力反馈、温度甚至嗅觉等多维感官数据。在面对对极致码率、多维感官同步的需求时，传统通信范式在能效与体验之间的深层矛盾日益凸显。生成式模型尤其是多模态生成式模型未来极有可能成为 6G 沉浸式通信的使能技术。通过融合视觉、听觉、触觉等感觉与上下文信息，可以构建统一的感官信息表征与重构机制，为全息远程呈现、沉浸式交互课堂等场景提供技术底座。

沉浸式业务对带宽与算力的极致需求，将推动传统内容分发网络进一步向生成式演进，催生新一代的生成式内容分发网络，其核心逻辑是基于用户指令的内容边缘生成，而非从服务器下载。传统 CDN 依赖内容的预先缓存与就近分发，而面向沉浸式通信的场景化内容（如

全息影像、三维场景)具有海量、动态、个性化的特征,预缓存模式难以奏效。未来趋势是将生成式 AI 能力下沉至网络边缘,根据用户请求与上下文实时生成个性化内容。

3. 语义驱动的沉浸式通信技术架构

语义通信与生成式模型的融合可带来几个数量级的传输效率提升,彻底突破传统数据传输管道的容量限制。其核心机制在于:发送端通过语义编解码器从原始数据中提取高价值语义信息,接收端利用生成式模型将语义信息映射为三维多视角视频表示。这种“语义提取+内容生成”的范式,将推动沉浸式通信从传像素向传语义、生内容的根本转变。

算网融合将为语义驱动的沉浸式通信关键技术提供基础设施支撑。6G 网络将实现通信与计算、存储的深度融合,通过动态资源分配支持边缘节点的 AI 推理任务。数字孪生体实时映射物理网络状态,大模型生成最优资源调度策略,推动通信大模型与数字孪生技术的协同演进,构建内容主动生成设计的分发网络。

4. 面向沉浸式通信的通感一体化与数字环境重构

未来 6G ISAC 将走向多频段深度融合架构,容量层(毫米波)支持热点区域高速率,极致层(太赫兹、可见光)在特定场景提供 Tbps 级速率和毫米级感知。人工智能驱动的自适应 ISAC 将实现从“人工优化”到“自主适应”的转变,神经网络辅助波形优化、强化学习驱动实时决策、边缘智能实现实时优化。器件与硬件革新方面,太赫兹硅基芯片量产、可见光 micro-LED 阵列普及、毫米波-RIS 一体化硬件将大幅降低系统成本与功耗。场景规模化落地方面,2030 年左右随着 6G 商用,毫米波 ISAC 将率先在场馆级全息通信中落地,可见光 ISAC 融入智能照明成为室内沉浸式体验标配,太赫兹 ISAC 在工业元宇宙、医疗仿真等精密场景实现突破。

综上所述,基于网媒融合的沉浸式通信技术将在多个维度持续演进:生成式大模型重构内容生成与传输的底层逻辑,智能体终端重塑人机物交互的边界,生成式内容分发网络实现边缘智能,通感一体化与数字环境重构实现虚拟场景搭建等。这些技术路径相互交织,共同推动沉浸式通信从视听传输走向多感官共生,从连接管道迈向智能协同的网媒融合,最终实现“天涯若比邻、虚实共空间”的未来愿景。

5.2 未来工作倡议

沉浸式通信正处于从技术突破向规模应用转型的攻坚阶段，在核心技术、标准体系、场景落地、生态协同等方面还面临诸多挑战。为推动沉浸式通信产业高质量发展，充分释放网媒融合关键技术价值，结合产业发展实际，提出以下未来工作倡议。

一是聚焦核心技术攻关，筑牢产业发展根基。建议产学研用各方加大沉浸式通信基础研究与应用研发投入，突破关键技术瓶颈并形成自主可控的技术体系；同时，加快原型验证，加速创新科技成果从实验室向产业端转化，为沉浸式通信产业创新发展提供坚实技术支撑。

二是加快标准体系建设，引领产业规范发展。依托 IMT-2030（6G）推进组、行业联盟等平台，推动建立覆盖技术、产品、应用、安全等全领域的沉浸式通信标准体系，产学研各方积极参与国际标准制定，推动中国方案、中国技术纳入国际标准，提升产业国际话语权。

三是推动场景融合应用，拓展产业发展空间。以场景创新牵引技术迭代与产业升级，聚焦全息通信、医疗健康、沉浸式教育、数字文创等重点领域，打造一批可复制、可推广的示范应用案例，推动沉浸式通信与千行百业深度融合。

四是构建协同生态体系，凝聚产业发展合力。建议产学研用各方深度合作，成立沉浸式通信行业生态联盟，统筹推进技术研发、标准制定、应用培育等工作；同时，加强产业链上下游相互协同，推动芯片、终端、网络、内容等环节深度融合，形成优势互补、协同发展的产业格局。

凝聚产业合力、共绘发展蓝图。最后，我们呼吁产学研用各方凝聚共识，共同推进技术突破、标准制定、应用创新，携手共建开放包容的沉浸式通信生态，实现产业链协同升级，让沉浸式通信的技术与场景向千行百业延伸，逐步打破数字鸿沟，最终助力我国数字经济向智能经济跃升。

参考文献

- [1] IMT-2030(6G)推进组. 6G 典型场景和关键能力[R]. 北京: IMT-2030(6G)推进组, 2023.
- [2] International Telecommunication Union. IMT for 2030 and beyond: Framework and overall objectives of the future development of IMT: ITU-R M.2160-0[S]. Geneva: ITU, 2023.
- [3] The White House. FACT SHEET: President Donald J. Trump Takes Action to Win the 6G Race[EB/OL]. 2025.
- [4] ATIS Next G Alliance. Roadmap to 6G[R]. Washington: Alliance for Telecommunications Industry Solutions, 2022.
- [5] ATIS Next G Alliance. Multi-Sensory Extended Reality (XR) in 6G[R]. Washington: Alliance for Telecommunications Industry Solutions, 2023.
- [6] ATIS Next G Alliance. North American 6G Roadmap Priorities[R]. Washington: Alliance for Telecommunications Industry Solutions, 2024.
- [7] SAYLER K M. Military Applications of Extended Reality[R/OL]. Washington: Congressional Research Service, 2025.
- [8] European Commission. Extended reality | Shaping Europe's digital future[EB/OL]. 2022.
- [9] London School of Economics and Political Science. Unleashing European Innovations for the Metaverse[R/OL]. London: LSE, 2024.
- [10] Ministry of Science and ICT, South Korea. Immersive Economy Development Strategy[EB/OL]. 2024.
- [11] 工业和信息化部办公厅, 教育部办公厅, 文化和旅游部办公厅, 等. 元宇宙产业创新发展三年行动计划(2023—2025年)[EB/OL]. 2023.
- [12] Mystakidis S. Metaverse[J]. Encyclopedia, 2022, 2(1): 486-497.
- [13] Langa S F, Montagud M, Cernigliaro G, et al. Multiparty holomeetings: Toward a new era of low-cost volumetric holographic meetings in virtual reality[J]. IEEE Access, 2022, 10: 81856-81876.
- [14] Shahid H, Amatetti C, Campana R, et al. Emerging advancements in 6G NTN radio access technologies: An overview[C]//2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit). IEEE, 2024: 593-598.
- [15] Shen X, Gao J, Li M, et al. Toward immersive communications in 6G[J]. Frontiers in Computer Science, 2023, 4: 1068478.
- [16] Shamsabadi A A, Yadav A, Gadallah Y, et al. Exploring the 6g potentials: Immersive, hyperreliable, and low-latency communication[J]. IEEE Vehicular Technology Magazine, 2025, 20(1): 74-82.
- [17] Zhou C, Hu S, Gao J, et al. User-centric immersive communications in 6G: A data-oriented framework via digital twin[J]. IEEE Wireless Communications, 2025, 32(3):

122-129.

- [18] Sehad N, Bariah L, Hamidouche W, et al. Generative AI for immersive communication: The next frontier in Internet-of-Senses through 6G[J]. IEEE Communications Magazine, 2024, 63(2): 31-43.
- [19] Shen X, Gao J, Li M, et al. Toward immersive communications in 6G[J]. Frontiers in Computer Science, 2023, 4: 1068478.
- [20] Sagduyu Y E, Erpek T, Yener A, et al. Will 6G be semantic communications? Opportunities and challenges from task oriented and secure communications to integrated sensing[J]. IEEE Network, 2024, 38(6): 72-80.
- [21] Wang Y, Han H, Feng Y, et al. Semantic communication empowered 6G networks: Techniques, applications, and challenges[J]. IEEE Access, 2025, 13: 28293-28314.
- [22] Strinati E C, Barbarossa S. 6G networks: Beyond Shannon towards semantic and goal-oriented communications[J]. Computer Networks, 2021, 190: 107930.
- [23] Strinati E C, Di Lorenzo P, Sciancalepore V, et al. Goal-oriented and semantic communication in 6G AI-native networks: The 6G-GOALS approach[C]//2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit). IEEE, 2024: 1-6.
- [24] 3GPP. Study on 6G Use Cases and Service Requirements; Stage 1: 3GPP TR 22.870[S]. 3GPP, 2026.
- [25] 3GPP. Study on XR (Extended Reality) Evaluations for NR: 3GPP TR 38.838[S]. 3GPP.
- [26] 3GPP. Study on 6G Use Cases and Service Requirements (V2.0.0): 3GPP TS 22.870[S]. 3GPP, 2026.
- [27] Wang H, Dong H, El Saddik A. Immersive multimedia communication: state-of-the-art on eXtended reality streaming[J]. ACM Transactions on Multimedia Computing, Communications and Applications, 2025, 21(7): 1-33.
- [28] Alhashimi H F, Hindia M N, Dimiyati K, et al. Survey on AI-enabled resource management for 6G heterogeneous networks: recent research, challenges, and future trends[J]. Computers, Materials, & Continua, 2025, 83(3): 3585.
- [29] Shannon C E. A mathematical theory of communication[J]. The Bell System Technical Journal, 1948, 27(3): 379-423.
- [30] Weaver W. Recent contributions to the mathematical theory of communication[J]. ETC: A Review of General Semantics, 1953, 10(1/2): 136 - 157.
- [31] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436 - 444.
- [32] Feng Y, Xu J, Liang C, et al. Decoupling source and semantic encoding: An implementation study[J]. Electronics, 2023, 12(13): 2755.
- [33] Bar-Hillel Y, Carnap R. Semantic information[J]. The British Journal for the Philosophy of Science, 1953, 4(14): 147 - 157.
- [34] Bacchus F I. Representing and reasoning with probabilistic

knowledge[C]//Proceedings of the 13th International Joint Conference on Artificial Intelligence. Chambéry, France, 1993: 227 - 234.

[35] Floridi L. Outline of a theory of strongly semantic information[J]. Minds and Machines, 2004, 14(2): 197 - 221.

[36] Floridi L. Is semantic information meaningful data?[J]. Philosophy and Phenomenological Research, 2005, 70(2): 351 - 370.

[37] D' Alfonso S. On quantifying semantic information[J]. Information, 2011, 2(1): 61 - 101.

[38] Bao J, Basu P, Dean M, et al. Towards a theory of semantic communication[C]//2011 IEEE Network Science Workshop. West Point, NY, USA, 2011: 110 - 117.

[39] Kolchinsky A, Wolpert D H. Semantic information, autonomous agency, and non-equilibrium statistical physics[J]. Interface Focus, 2018, 8(6): 20180041.

[40] Kountouris M, Pappas N. Semantics-empowered communication for networked intelligent systems[J]. IEEE Communications Magazine, 2021, 59(6): 96 - 102.

[41] Orlitsky A, Roche J R. Coding for computing[C]//Proceedings of the 36th Annual Symposium on Foundations of Computer Science. Milwaukee, WI, USA, 1995: 502 - 511.

[42] Yuan D, Guo T, Bai B, Han W. Lossy computing with side information via multi-hypergraphs[C]//2022 IEEE Information Theory Workshop (ITW). Mumbai, India, 2022: 344 - 349.

[43] Harangi V, Niu X, Bai B. Generalizing Körner's graph entropy to graphons[J]. European Journal of Combinatorics, 2023, 114: 103779.

[44] Shi G, Gao D. On semantic information[J]. Chinese Journal of Information Theory, 2025, 1: 1 - 10.

[45] Niu K, Zhang P. A mathematical theory of semantic communication[J/OL]. arXiv preprint, 2024.

[46] Shannon C E. The redundancy of English[C]//Cybernetics; Transactions of the 7th Conference. New York: Josiah Macy, Jr. Foundation, 1951: 248-272.

[47] Wang Y, Li H. Information Theory and Coding Theory[M]. Beijing: Higher Education Press, 2013: 338.

[48] Shannon C E. Coding Theorems for a Discrete Source With a Fidelity Criterion[C]//Claude E. Shannon: Collected Papers. IEEE, 1993: 325-350.

[49] Gastpar M, Rimoldi B, Vetterli M. To code, or not to code: lossy source-channel communication revisited[J]. IEEE Transactions on Information Theory, 2003, 49(5): 1147-1158.

[50] 李文璟, 喻鹏, 张平. 6G 智能内生网络架构及关键技术分析[J]. 中兴通讯技术, 2023, 29(5): 2-8.

[51] 3GPP. Study on 6G Use Cases and Service Requirements (V2.0.0): 3GPP TS 22.870[S]. 3GPP, 2026.

- [52] 王卫斌, 周建锋, 毛磊. 6G 通感算智融合架构和场景赋能探索[J]. 移动通信, 2024, 48(3): 21-25.
- [53] 陈新宇, 牛娇红, 陆光辉. 基于 6G AIaaS 的分布式网络框架及技术方案[J]. 移动通信, 2023, 47(6): 110-114.
- [54] 王辰, 白雪茜, 张强, 等. 面向实时通信的边缘智能关键技术研究[J]. 电信科学, 2025, 41(11).
- [55] 段向阳, 杨立, 夏树强, 等. 通感算智一体化技术发展模式[J]. 电信科学, 2022, 38(3): 37-48.
- [56] Rongwei R L, Lenan L W, Dongliang G D. Joint source/channel coding modulation based on BP neural networks[C]//Proceedings of the International Conference on Neural Networks and Signal Processing. Nanjing, China, 2003, 1: 156-159.
- [57] Farsad N, Rao M, Goldsmith A. Deep learning for joint source-channel coding of text[C]//2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Calgary, AB, Canada, 2018: 2326-2330.
- [58] Bourtsoulatze E, Kurka D B, Gündüz D. Deep joint source-channel coding for wireless image transmission[J]. IEEE Transactions on Cognitive Communications and Networking, 2019, 5(3): 567-579.
- [59] Kurka D B, Gündüz D. DeepJSCC-f: Deep joint source-channel coding of images with feedback[J]. IEEE Journal on Selected Areas in Information Theory, 2020, 1(1): 178-193.
- [60] Tung T Y, Kurka D B, Jankowski M, et al. DeepJSCC-Q: Constellation constrained deep joint source-channel coding[J]. IEEE Journal on Selected Areas in Information Theory, 2022, 3(4): 720-731.
- [61] Bian C, Shao Y, Gündüz D. DeepJSCC-1++: Robust and bandwidth-adaptive wireless image transmission[C]//2023 IEEE Global Communications Conference (GLOBECOM). Kuala Lumpur, Malaysia, 2023: 3148-3154.
- [62] Wu H, Shao Y, Ozfatura E, et al. Transformer-aided wireless image transmission with channel feedback[J]. IEEE Transactions on Wireless Communications, 2024, 23(9): 11904-11919.
- [63] Xie H, Qin Z, Li G Y, Juang B H. Deep learning enabled semantic communication systems[J]. IEEE Transactions on Signal Processing, 2021, 69: 2663-2675.
- [64] Weng Z, Qin Z. Semantic communication systems for speech transmission[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(8): 2434-2444.
- [65] Xie H, Qin Z, Li G Y. Task-oriented multi-user semantic communications for VQA[J]. IEEE Wireless Communications Letters, 2022, 11(3): 553-557.
- [66] Xie H, Qin Z, Tao X, et al. Task-oriented multi-user semantic communications[J]. IEEE Journal on Selected Areas in Communications, 2022, 40(9): 2584-2597.
- [67] Zhang G, Hu Q, Qin Z, et al. A unified multi-task semantic communication system with domain adaptation[J/OL]. arXiv preprint, 2022.

- [68] Qin Z, Tao X, Lu J, et al. Semantic communications: Principles and challenges[J/OL]. arXiv preprint, 2021.
- [69] Dai J, Wang S, Tan K, et al. Nonlinear transform source-channel coding for semantic communications[J]. IEEE Journal on Selected Areas in Communications, 2022, 40(8): 2300-2316.
- [70] Jiang P, Wen C K, Jin S, et al. Wireless semantic communications for video conferencing[J]. IEEE Journal on Selected Areas in Communications, 2023, 41(1): 230-244.
- [71] Jiang F, et al. Large AI model empowered multimodal semantic communications[J]. IEEE Communications Magazine, 2025, 63(1): 76-82.
- [72] Wu T, Chen Z, Lu G, et al. Joint source-channel-generation coding: From distortion-oriented reconstruction to semantic-consistent generation[J/OL]. arXiv preprint, 2026.
- [73] Wang L, Yao K, Guo C, et al. VideoRF: Rendering dynamic radiance fields as 2D feature video streams[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024: 470-481.
- [74] Wang L, Hu Q, He Q, et al. Neural residual radiance fields for streamably free-viewpoint videos[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023.
- [75] Zheng Z, Zhong H, Hu Q, et al. JointRF: End-to-end joint optimization for dynamic neural radiance field representation and compression[C]//2024 IEEE International Conference on Image Processing (ICIP). IEEE, 2024.
- [76] Hu Q, Zhong H, Zheng Z, et al. VRVVC: Variable-rate NeRF-based volumetric video compression[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(4): 3563-3571.
- [77] Zheng Z, Zhong H, Hu Q, et al. HPC: Hierarchical progressive coding framework for volumetric video[C]//Proceedings of the 32nd ACM International Conference on Multimedia. New York: ACM, 2024: 7937-7946.
- [78] Wang P, Zhang Z, Wang L, et al. V³: Viewing volumetric videos on mobiles via streamable 2D dynamic Gaussians[J]. ACM Transactions on Graphics (TOG), 2024, 43(6): 1-13.
- [79] Hu Q, Zheng Z, Zhong H, et al. 4DGC: Rate-aware 4D Gaussian compression for efficient streamable free-viewpoint video[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2025: 875-885.
- [80] Zheng Z, Wu Z, Zhong H, et al. 4DGCPro: Efficient hierarchical 4D Gaussian compression for progressive volumetric video streaming[C]//Advances in Neural Information Processing Systems (NeurIPS), 2025.
- [81] Xu H, et al. Bridging Light and Immersion: Visible Optical Interfaces for Extended Reality[R]. 2026.
- [82] RF 技术社区. 从 eMBB 到沉浸式体验, 6G 如何满足全真全感互联需求? [EB/OL]. 2025.

- [83] Huawei. 6G ISAC-OW Extends the Frontier of Spectrum for Wireless Communication Systems[R].
- [84] Shome S, et al. An extensive review of THz communication in 6G: Facilitating technologies with edge computing and native AI[J]. Franklin Open, 2025, 13: 100434.
- [85] Chaccour C, et al. Joint Sensing and Communication for Situational Awareness in Wireless THz Systems[C]//ICC 2022 - IEEE International Conference on Communications. 2022: 3772-3777.
- [86] Li Z, et al. Adaptive visible light integrated sensing and communication for cable-free virtual reality[J]. 光子学研究(英文), 2025, 9: 2453-2463.
- [87] 人民邮电报. 北京移动首发‘双频通感立体网络’[N]. 人民邮电报, 2025.
- [88] RF 技术社区. 超越太赫兹: 6G 通信中的全频谱融合技术[EB/OL]. 2025.
- [89] Chinese Science Bulletin. Multimodal sensing-based dual-band co-aperture reconfigurable Intelligent metasurface communication system[J]. Chinese Science Bulletin, 2026.
- [90] Wang R, Chen K, Li S, Chen M, Zhao T. Efficient Semantic Codec for Real-time Vibrotactile Transmission[C]//Proceedings of the 33rd ACM International Conference on Multimedia (MM '25). Dublin, Ireland, 2025: 12092-12101.
- [91] Song J, Gao Y, Wu D, Zhou L. Scalable Tactile Coding[J]. IEEE Transactions on Multimedia, 2026.
- [92] Song J, Gao Y, Wu D, Zhou L, Mao S. Adaptive Live Tactile Streaming With Scalable Coding for Immersive Communications[J]. IEEE Transactions on Mobile Computing, 2026, 25(4): 5133-5145.
- [93] Gao Y, Wu D, Zhou L, Sari H. Cross-Modal Tactile Coding[C]//Proceedings of ICC 2025 - IEEE International Conference on Communications. Montreal, Canada, 2025: 5091-5096.
- [94] Yuan Z, Wu D, Zhou L, Sari H. On the Benefits of Cross-Modal Communications: From Source and Channel Coding Perspectives[J]. IEEE Transactions on Communications, 2025, 73(11): 10156-10171.
- [95] Chen M, Xie Y. Cross-Modal Reconstruction for Tactile Signal in Human - Robot Interaction[J]. Sensors, 2022, 22(17): 6517.
- [96] Lu H, Tan X, Chen M, et al. Cross-Modal Haptic Compression Inspired by Embodied AI for Haptic Communications[J]. IEEE Transactions on Multimedia, 2025, 27: 4996-5008.
- [97] Wang Z, Chen M, He X, et al. Visual-Tactile Fusion for Multimodal Semantic Communication with Foundation Models[C]//IEEE VTC2025-Spring, 2025: 1-5.
- [98] Zhang C, Zhang Z, Chen M, et al. A GAI-Based Haptic Transmission Architecture for Extending Headset Lifespan in Haptic-Enhanced XR[C]//ICC 2025 - IEEE International Conference on Communications, 2025: 1268-1273.
- [99] Wang Z, Chen M, Liu Q. A Review on Multimodal Communications for Human-Robot Collaboration in 5G: From Visual to Tactile[J]. Intelligence & Robotics, 2025, 5(3):

579-606.

[100] Liu H, Gao Y, Zhou S, Zhou L. More is Not Always Better: Toward General Cross-Modal Saliency Prediction for Immersive Communications[J]. IEEE Transactions on Multimedia, 2026.

[101] Zhou T, Zhang X, Kang B, Chen M. Multimodal Fusion Recognition for Digital Twin[J]. Digital Communications and Networks, 2024, 10(2): 337-346.

[102] Li A, Wei X, Wu D, Zhou L. Cross-Modal Semantic Communications[J]. IEEE Wireless Communications, 2022, 29(6): 144-151.

[103] Wang W, Liu M, Chen M. Cross-Modal Alignment for Multi-Modal Semantic Communications[C]//2023 IEEE Global Communications Conference (GLOBECOM). 2023: 5871-5876.

[104] Wei X, Shi Y, Zhou L. Haptic Signal Reconstruction for Cross-Modal Communications[J]. IEEE Transactions on Multimedia, 2022, 24: 4514-4525.

[105] Gao Y, Ni S, Wu D, Zhou L. Edge-Based Cross-Modal Communications for Remote Healthcare[J]. IEEE Journal on Selected Areas in Communications, 2022, 40(11): 3139-3151.

[106] Wang L, Yin A, Jiang X, et al. Resource Allocation for Multi-Traffic in Cross-Modal Communications[J]. IEEE Transactions on Network and Service Management, 2023, 20(1): 60-72.

[107] Chen M, Liu M, Wang C, et al. Cross-Modal Graph Semantic Communication Assisted by Generative AI in the Metaverse for 6G[J]. Research, 2024, 7: 0342.

[108] Huang R, Qiu C, Chen M, et al. Federated-Learning-Enabled Cross-Modal Semantic Communication for 6G[J]. IEEE Internet of Things Journal, 2025, 12(20): 41608-41624.

[109] Zhu Y, Huang Y, Qiao X, et al. AIVRTeach: Empowering Virtual Reality Classrooms with AI-Generated Content[J]. International Journal of Human-Computer Interaction, 2026: 1-27.

[110] Zhang C, Huang Y, Zhu Y, et al. PortaCap: Portable Volumetric Video Capturing System for Metaverse Interaction[J]. IEEE Transactions on Mobile Computing, 2026.

[111] Huang Y, Zhao Y, Chen Z, et al. Intent-Driven Cognitive XR Networks: Multi-Agent Orchestration for Immersive Communication[J]. IEEE Journal on Selected Areas in Communications, 2025.

[112] Bai B, Qiao X, Zhao H. UncertainGS: Uncertainty-aware indoor reconstruction via Gaussian splatting[J]. Neurocomputing, 2025: 132136.

[113] Zhang C, Cuan L, Yang Y, et al. MonoFlex3D: Low-Latency Glasses-Free 3D Teleconference from a Single Camera[C]//Proceedings of the 3rd ACM Workshop on Mobile Immersive Computing, Networking, and Systems. 2025: 1-8.

[114] Zhu Y, Wang H, Huang Y, et al. HOLOCOM: Toward Immersive Communication and Interaction with Holographic Video[J]. IEEE Wireless Communications, 2025, 32(2):

120-129.

[115] Huang Y, Bai B, Zhu Y, et al. ISCom: Interest-aware semantic communication scheme for point cloud video streaming on metaverse XR devices[J]. IEEE Journal on Selected Areas in Communications, 2023, 42(4): 1003-1021.

[116] Guo H, Weng D, Su M, et al. TagSplat: topology-aware Gaussian splatting for dynamic mesh modeling and tracking[J/OL]. arXiv:2512.01329, 2025.

[117] Chen Y, Liang B, Guo H, et al. PS-GS: Gaussian splatting for multi-view photometric stereo[J]. Computers & Graphics, 2026, 136: 104561.

[118] Yang Q, Dongye X, Guo H, et al. Minigs: efficient 3D Gaussian splatting with full factors weighted pruning for scene representation[J]. Multimedia Systems, 2026, 32: 66.

[119] Dongye X, Guo H, Luo L, et al. LoDAvatar: hierarchical embedding and selective detail enhancement for adaptive levels of detail Gaussian avatars[J]. Virtual Reality, 2025, 29: 178.

[120] 王欣晖, 周星月, 朱进国. 6G 网络架构探讨[J]. 信息通信技术与政策, 2023, 49(9): 13.

[121] 未来移动通信论坛. 6G 沉浸式通信场景需求与关键技术研究白皮书[R]. 北京: 未来移动通信论坛, 2025.

[122] 3GPP. Study on 6G Use Cases and Service Requirements: 3GPP TR 22.870 v2.0.0[S]. 3GPP, 2026.

缩略语表

缩略语	英文全称	中文全称
1G	1st Generation Mobile Communication Technology	第一代移动通信技术
2D	Two Dimensional	二维
3D	Three Dimensional	三维
4D	Four-Dimensional	四维
5D	Five-Dimensional	五维
3DGS	3D Gaussian Splatting	三维高斯泼溅
4DGS	4D Gaussian Splatting	四维高斯泼射
3DoF	Three Degrees of Freedom	三自由度
6DoF	Six Degrees of Freedom	六自由度
3GPP	3rd generation partnership project	第三代合作伙伴计划
5G	5th Generation Mobile Communication Technology	第五代移动通信技术
5G-MAG	5G Media Action Group	5G 媒体行动小组
6G	6th Generation Mobile Communication Technology	第六代移动通信技术
AI	artificial intelligence	人工智能
AGV	Automated Guided Vehicle	自动导引车
AIGC	Artificial Intelligence Generated Content	人工智能生成内容
ANN	Artificial Neural Network	人工神经网络
AR	Augmented Reality	增强现实

缩略语	英文全称	中文全称
ATIS	The Alliance for Telecommunications Industry Solutions	美国电信行业解决方案联盟
AV1	AOMedia Video 1	开放媒体联盟一代视频编码
BCP	Bar Hillel-Carnap Paradox	Bar-Hillel-Carnap 悖论
CRLB	Cramer-Rao Lower Bound	克拉美-罗下界
CIT	Classical information Theory	经典信息论
C-DRX	Connected-Mode Discontinuous Reception	连接态非连续接收
CMP	Cube Map Projection	立方体贴图投影
CoDi	Composable Diffusion	可组合扩散模型
CPU	Central Processing Unit	中央处理器
CRPS	Cloud Redundant Power Supply	云冗余电源
CRQC	Cryptographically Relevant Quantum Computer	密码学相关量子计算机
CRS	Congressional Research Service	美国国会研究服务局
CSIT	Classical Semantic Information Theory	经典语义信息理论
DASH	Dynamic Adaptive Streaming over HTTP	基于 HTTP 的动态自适应流媒体
DASH-IF	DASH Industry Forum	DASH 行业论坛
DEM	Discrete Element Method	离散元法
DL	Deep Learning	深度学习
DMT	Discrete Multi-tone	离散多音频调制
GC-MS	Gas Chromatography-Mass Spectrometry	气相色谱-质谱联用
GPU	Graphics Processing Unit	图形处理器
EMA	External Memory Access	片外存储访问
eMMC	embedded Multi Media Card	嵌入式多媒体卡

缩略语	英文全称	中文全称
ERP	Equirectangular Projection	等距柱状投影
FDIS	Final Draft International Standard	最终国际标准草案
FEA	Finite Element Analysis	有限元分析
FVV	Free-Viewpoint Video	自由视点
GB	Gigabyte	吉字节
Gbps	Gigabits per second	吉比特每秒
G-PCC	Geometry-based Point Cloud Compression	基于几何的点云压缩
HMD	Head-Mounted Display	头戴式显示器
HEVC	High Efficiency Video Coding	高效率视频编码
HTTP	HyperText Transfer Protocol	超文本传输协议
IETF	Internet engineering task force	互联网工程任务组
IMT	International Mobile Telecommunications	国际移动通信
IOT	Interoperability Test	互操作性测试
IPMI	Intelligent Platform Management Interface	智能平台管理接口
ISAC	Integrated Sensing and Communication	通信感知一体化
ITU	international telecommunication union	国际电信联盟
ITU-R	international telecommunication union radio communication sector	国际电信联盟无线电通信部门
JSCC	Joint Source-Channel Coding	联合信源信道编码
kPa	kilopascal	千帕斯卡
KPI	Key Performance Indicators	关键性能指标
LAM-MSc	Large AI Model Empowered Multi-modal Semantic Communication	大模型赋能的多模态语义通信

缩略语	英文全称	中文全称
LAN	Local Area Network	局域网
LPDDR	Low Power Double Data Rate	低功耗双倍数据率
LSE	The London School of Economics and Political Science	伦敦政治经济学院
MIMO	Multiple Input Multiple Output	多输入多输出
MPEG	Moving Picture Experts Group	动态影像专家小组
MTP	Motion- to- Photon	运动到光子时延
MWC	Mobile World Congress	世界移动通信大会
NeRF	Neural Radiance Field	神经辐射场
NF	Network Function	网络功能
NSPM	National Security Presidential Memoranda	国家安全总统备忘录
NTN	Non Terrestrial Network	非地面网络
NTSCC	Nonlinear Transform Source Channel Coding	非线性变换信源信道编码
MIV	MPEG Immersive Video	MPEG 沉浸式视频
MLP	Multi-layer Perceptron	多层感知器
MMT	MPEG Media Transport	MPEG 媒体传输协议
OFDM	Orthogonal Frequency Division Multiplexing	正交频分复用
OMAF	Omnidirectional Media Format	全向媒体格式
P2P	Photon-to-Photon	光子到光子
PC	Personal Computer	个人计算机
PIC	Photonic Integrated Circuit	光子集成电路
PPO	Proximal Policy Optimization	近端策略优化

缩略语	英文全称	中文全称
PSNR	Peak Signal-to-Noise Ratio	峰值信噪比
PWM	Pulse Width Modulation	脉冲宽度调制
QoE	Quality of Experience	体验质量
QoS	Quality of Service	服务质量
RAHT	Region Adaptive Hierarchical Transform	区域自适应层次变换
RAN	Radio Access Network	无线接入网
RGB	Red Green Blue	红绿蓝（三基色）
RIS	Reconfigurable Intelligent Surface	可重构智能超表面
RTP	Real-time Transport Protocol	实时传输协议
SHE	Service Hosting Environment	服务托管环境
SLAM	Simultaneous Localization and Mapping	即时定位与地图构建
SMA	Shape Memory Alloy	形状记忆合金
SNS_JU	Smart Networks and Services Joint Undertaking	智能网络和服务联合伙伴企业
SNR	Signal-to-Noise Ratio	信噪比
SRAM	Static Random-Access Memory	静态随机存取存储器
SSCC	Separate Source-Channel Coding	信源信道分离编码
SSD	solid-state drive	固态硬盘
SSIM	Structural Similarity Index	结构相似性指数
SVC	Semantic Video Conferencing	语义视频会议
TB	Terabyte	太字节
TENS	Transcutaneous Electrical Nerve Stimulation	经皮神经电刺激
TN	Terrestrial Networks	融合地面网络

缩略语	英文全称	中文全称
TTNB	Time To Next Burst	下次突发时间
TSN	Time-Sensitive Networking	时间敏感网络
TSSI	Theory of Strong Semantic Theory	强语义信息理论
V-DMC	Video-based Dynamic Mesh Coding	基于视频的动态网格编码
VLC	Visible Light Communication	可见光通信
V-PCC	Video-based Point Cloud Compression	基于视频的点云压缩
V3C	Visual Volumetric Video-based Coding	三维体积视频编码
VR	Virtual Reality	虚拟现实
VV	Volumetric Video	体积视频
VVC	Versatile Video Coding	通用视频编码
XR	Extended reality	扩展现实
XRM	Extended Reality and Media service	扩展现实与媒体服务
Wi-Fi	Wireless Fidelity	无线保真