智简编码传输
赋能高时效稳健通信

戴金晟

泛网无线通信教育部重点实验室

国家人工智能产教融合创新平台

Beijing University of Posts and Telecommunications

被关注的问题

序号	问题类型	具体问题	当前结论
1	应用场景	高时效智简编码传输当前主要用来解决哪些场景的通信问题?	
2	技术领域	智简编码传输究竟是压缩问题，还是传输问题?	
3	技术领域	联合信源信道编码（JSCC）是否等同于语义通信?	
4	技术性能	基于 AI 的 JSCC 性能是否比工程上已应用的 SSCC 标准“好”？	
5	科学问题	智简传输系统的核心增益来源在哪？是否会超越香农限?	
6	技术问题	JSCC 落地应用会面临哪些阻力？如何克服?	
7	开放问题	智简传输与智简网络是怎样的关系?	
8	开放问题	大模型等 AI 新技术，和智简通信有什么关联?	

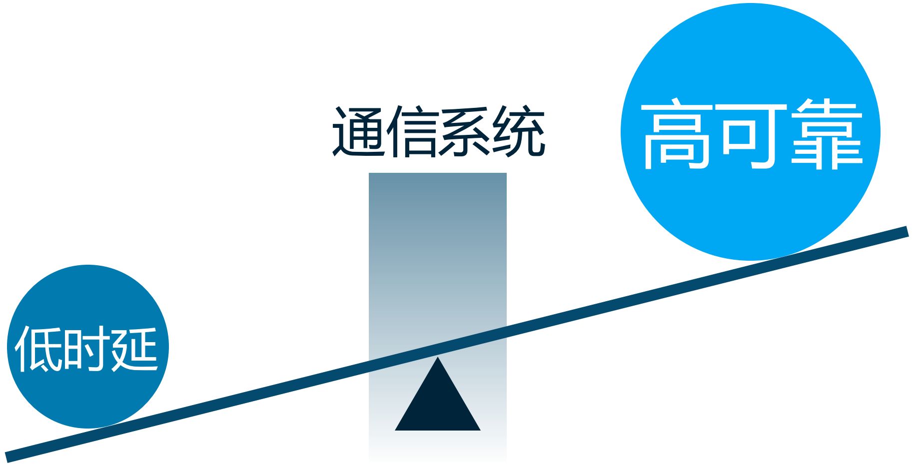
信源侧：高实时、大容量通信业务快速增长

❐ 新实时通信 RTC 业务带来新需求

- ➢ 下载与存储业务：“高效率压缩 + 高可靠传输”可以较好支持
- ➢ **交互会话业务：高实时、低开销**
- ➢ **直播推流业务：高实时、高并发**

现实情况：网络/信道状态高动态变化
大容量业务难以实现高实时稳健传输

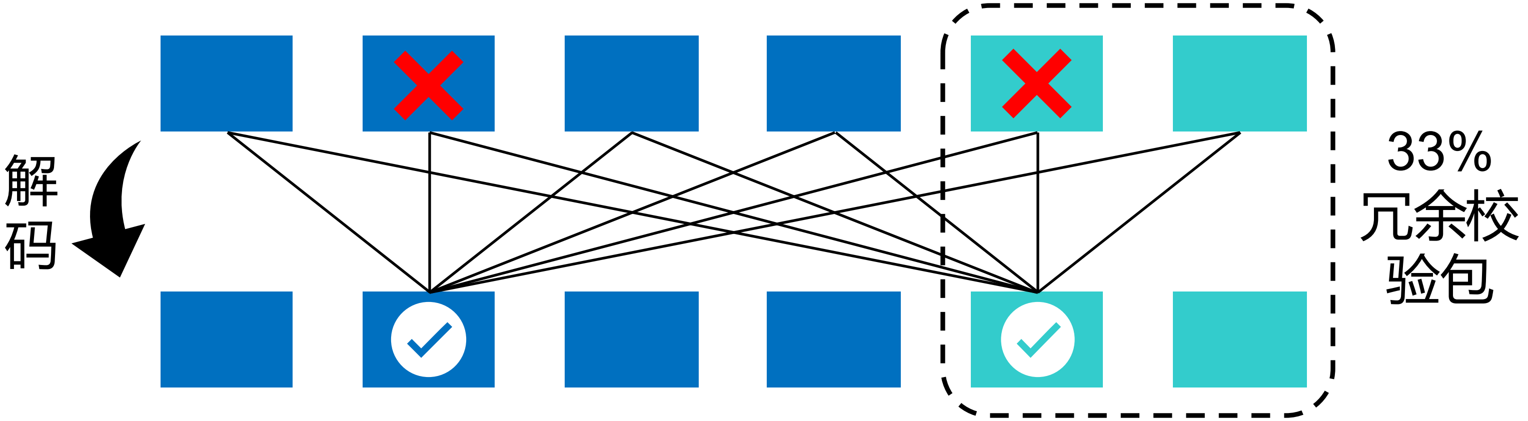
低时延与高可靠是固有矛盾



网络/信道侧：前向纠错 FEC 带宽开销高

❒ FEC 理论

➢ 若提前预知**网络传输丢包率**，可以通过**包级别 FEC 对抗网络丢包**



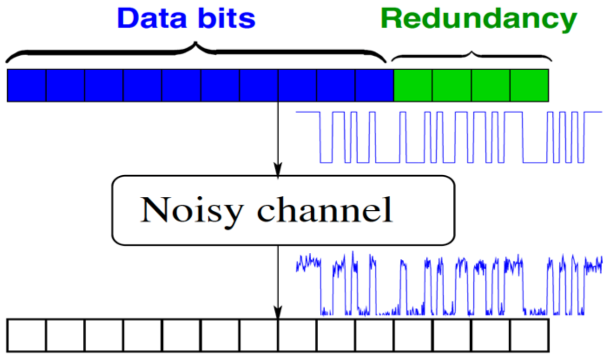
理论最优 FEC 定理

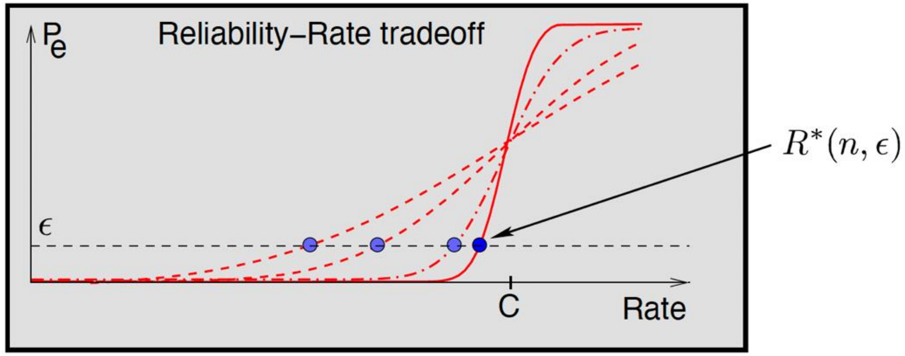
➢ **加 $r\%$ 冗余校验包，理论上最多可以保证对抗 $r\%$ 的网络传输随机丢包**

➢ 若提前预知**信道传输信噪比**，可以通过**比特级 FEC 对抗信道衰落**

$$R^*(n,\epsilon) = C - \sqrt{\frac{V}{n}}Q^{-1}(\epsilon) + \mathcal{O}\left(\frac{\log n}{n}\right)$$

$$R = \frac{k}{n} \leq R^*(n,\epsilon)$$





网络/信道侧：前向纠错 FEC 带宽开销高

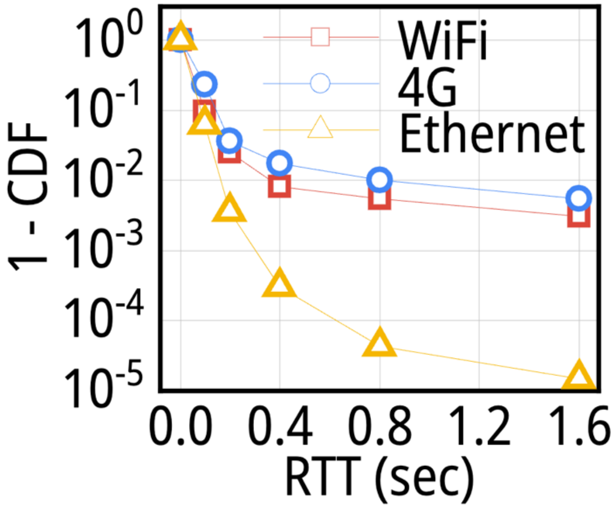
❐ 发送端前向纠错 FEC 增加带宽冗余：FEC 编解码效率低

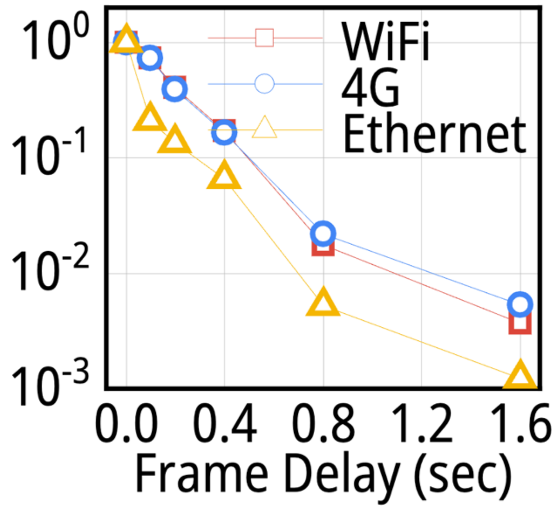
FEC 编码类型	典型应用场景	抗丢包恢复能力	解码复杂度	RTC 适用性
Reed-Solomon (RS)	深空通信	理论最优 (MDS)	$O(n^2)$ 或 $O(n^k)$ (高解码复杂度)	基本不适用 (受包数量和计算资源限制)
LDPC	WiFi，DVB	接近理论最优	$O(n)$ 或 $O(n\log n)$	特定设计，部分场景适用
XOR 奇偶校验	RTC (ULP-FEC, Flex-FEC 等)	无法接近理论最优	$O(k)$ (只与数据包数量 k 有关)	**目前 RTC 主要应用的 FEC 编码方式**

- RTC 常用的 **ULP-FEC** 和 **Flex-FEC** 编码方式采用**包级别 XOR 奇偶校验**，虽然性能不能接近理论最优，但**复杂度低，对 RTC 应用友好**
- **结论：实用化 FEC 编码效率不足，无法做到“加 $r\%$ 抗 $r\%$”，即便确知网络传输丢包率，实际加入冗余量也远大于丢包率**

网络/信道侧：后向纠错 ARQ 时延难保障

❐ 接收端后向纠错 ARQ 增加**时间冗余**：**尾部时延难降低**





Narayanan et al. (2020) [51]	Tail latency of the 5G hop does not improve against 4G, and could be around 200ms.
Daldoul et al. (2020) [22]	802.11ax (a.k.a. WiFi 6) has an average WiFi-hop latency of >30ms with 30 interferers.
Bhartia et al. (2017) [15]	More than one quarter of 802.11ac access points sufferfrom a latency of >100ms at the last hop.

无线网络传输时延：5G 相较于 4G 的平均时延显著降低，但尾部时延没有改善

- ➢ ARQ 的时延需求与业务类型高度关联，RTT 与应用场景和网络状态高度关联
- ➢ 一个正确到达却**“迟到”的数据包**，对 QoE 的损害与一个丢失的数据包无异
- ➢ 一次 ARQ 使数据包交付时间增加至少一个 RTT，**“三向延迟”合计易超过 150 ms**，部分时延要求苛刻的**交互式 RTC 业务“难应用 ARQ 技术”**

[R] Meng, Zili, et al. "Achieving consistent low latency for wireless real-time communications with the shortest control loop." *Proceedings of the ACM SIGCOMM 2022 Conference*. 2022.

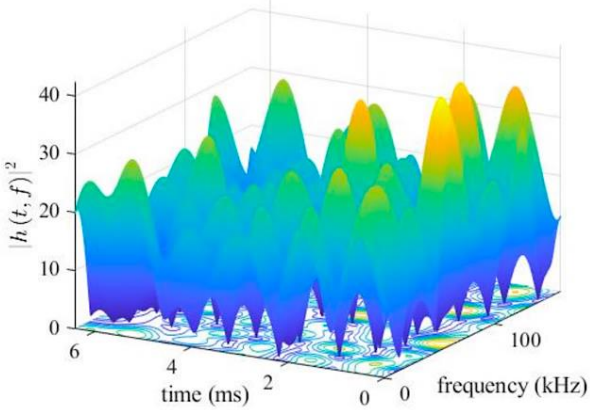
高时效稳健通信需求迫切

动态弱网下，通信系统参数调控滞后环境变化，业务难以实现高时效稳健传输

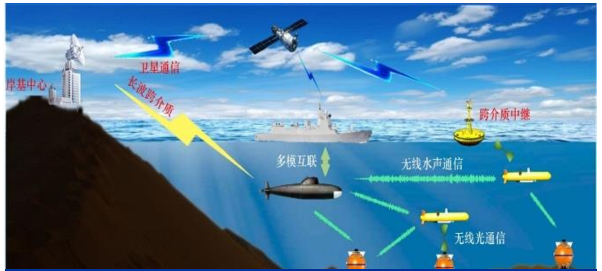
信源侧：高实时、大容量通信业务快速增长

业务传输
稳健性弱
时效性差

网络/信道侧：带宽不足、状态波动大，难以准确预测



工业无线网络：海量接入、碰撞多、干扰强



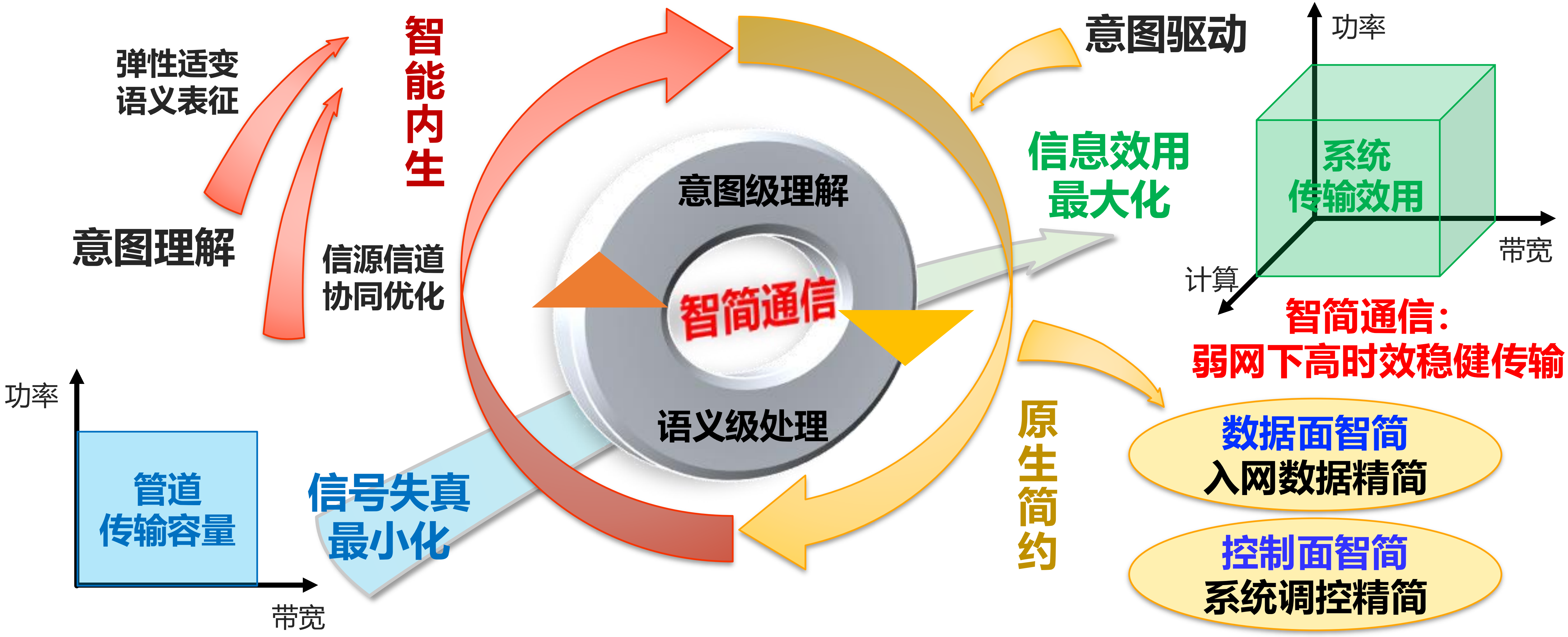
水下通信：弱信道、强干扰、高动态

偏远山区通信：弱信道、高动态

大流量密集通信：带宽低、网络拥塞

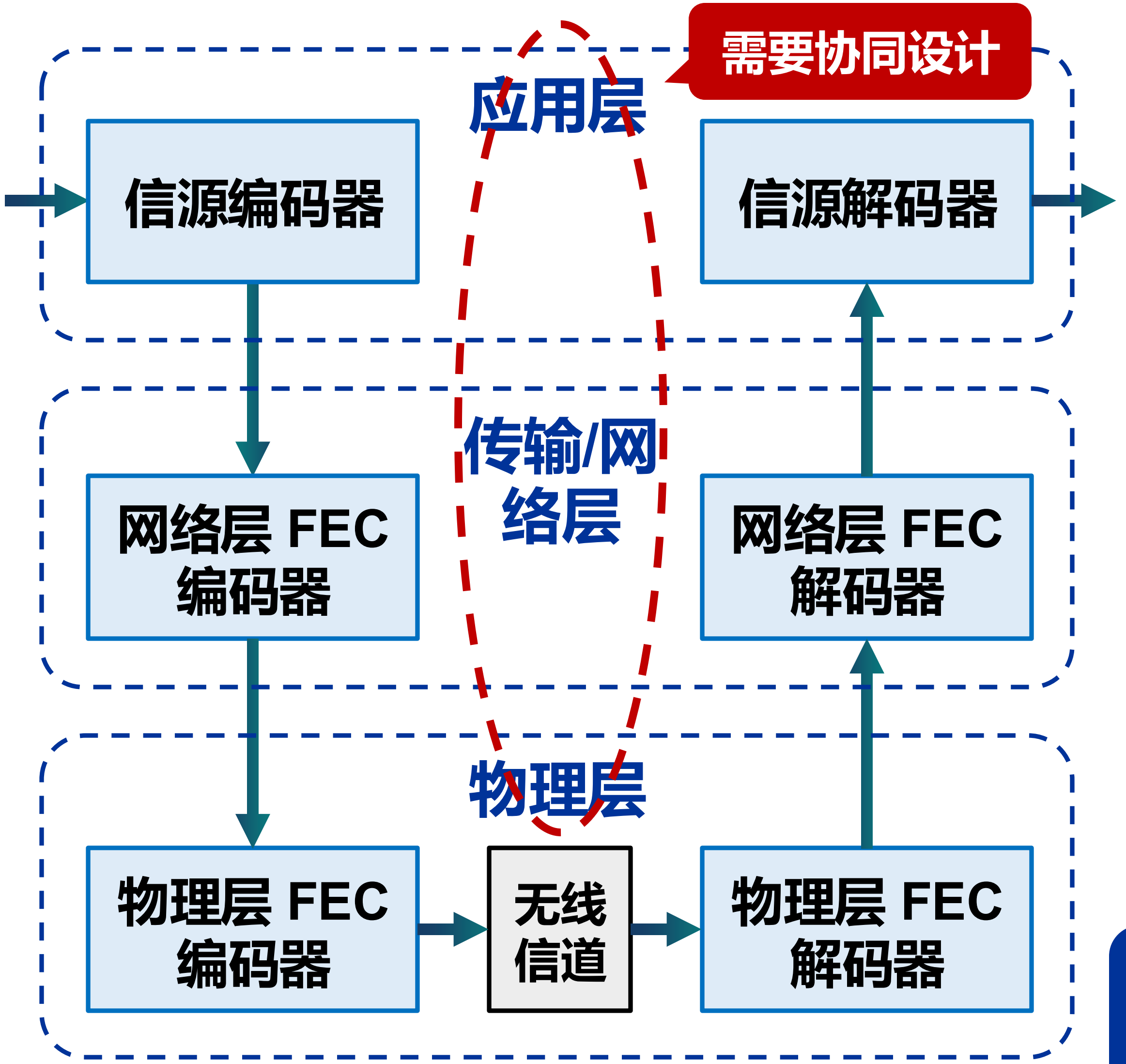
传统通信发展进入瓶颈，动态弱网下的高时效稳健传输需求引领通信理论与技术创新

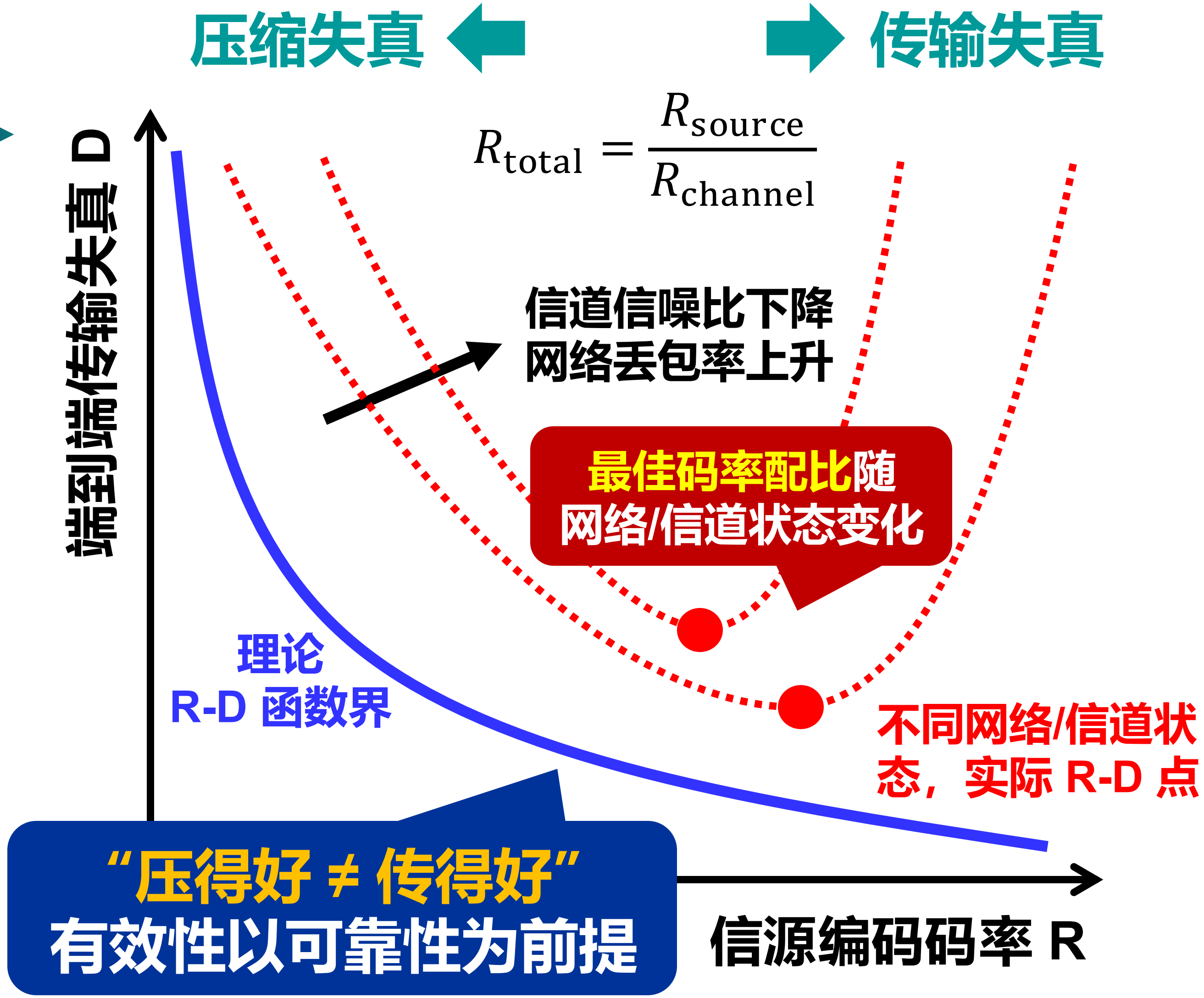
智简通信的设计目标



为建立“高时效、高可靠、高有效”的信息通信系统奠定理论与技术基础

端到端传输失真 = 压缩失真 + 传输失真

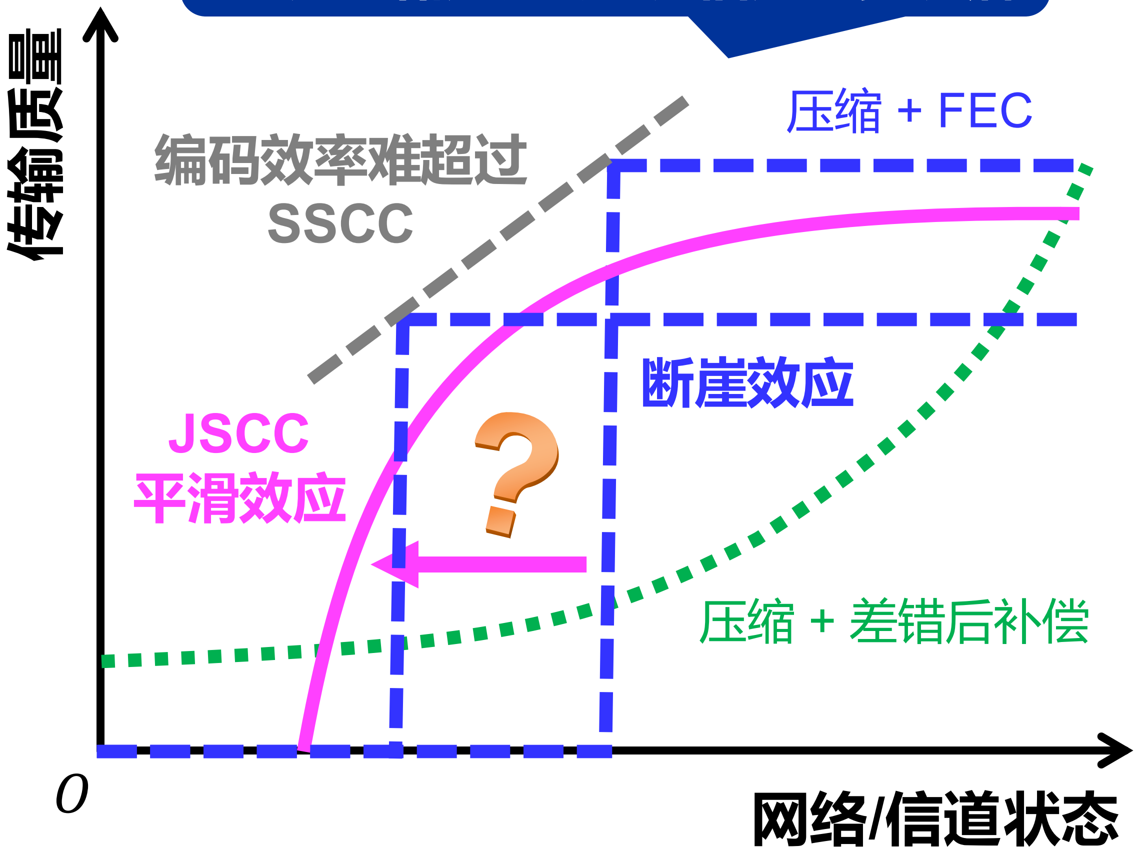




关键难点 — 高效压缩 + 强韧容错

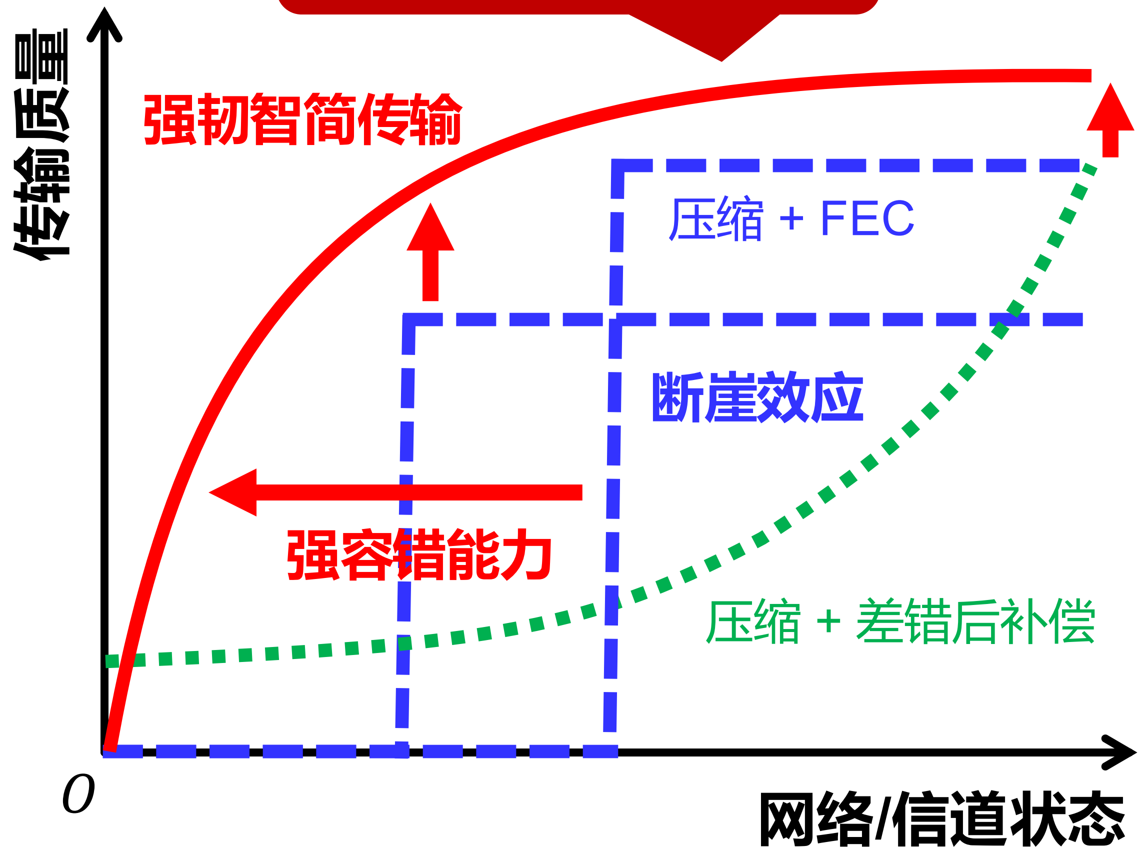
❒ 现实挑战：“压缩能力”与“容错能力”难兼得

目标：原生具备传输韧性
对抗网络随机丢包、信道时变衰落



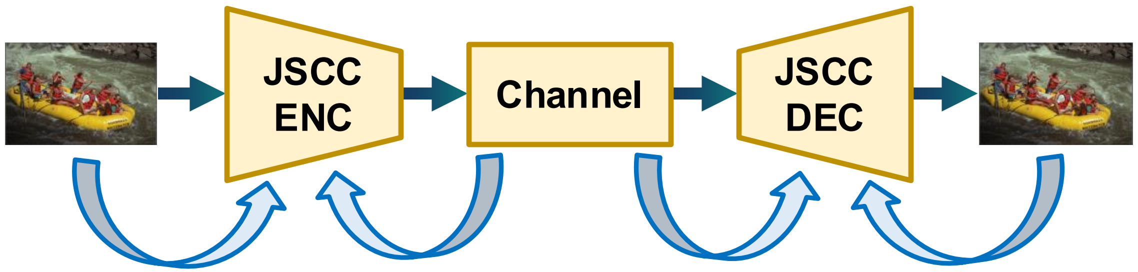
设计目标

挑战：既要高压缩效率，又要强容错能力



实验观察：JSCC 性能不容易超过 SSCC

带有噪声瓶颈训练的自编码器

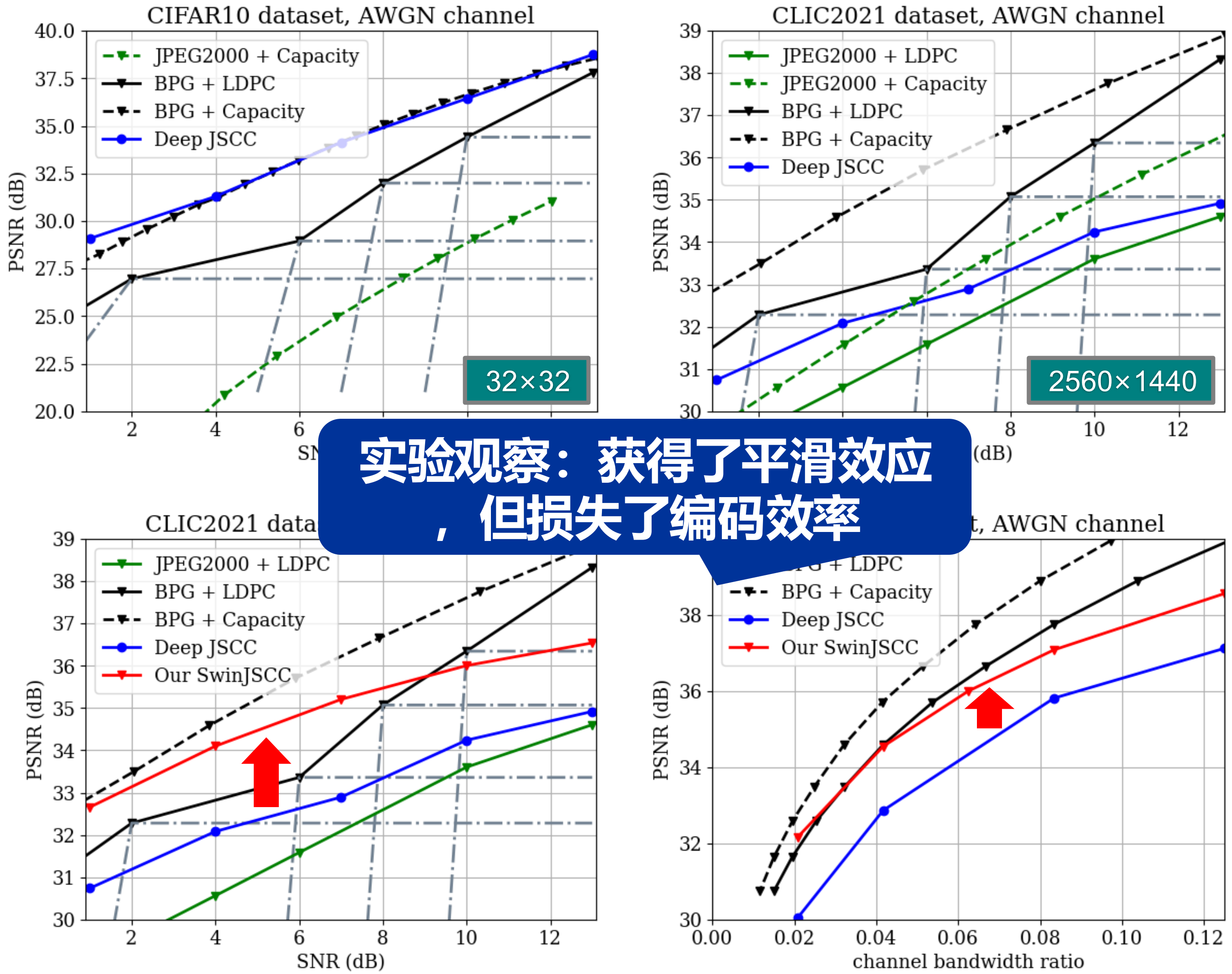


信道状态感知的端到端联合优化

端到端联合优化能否真正解决问题?

[1] E. Bourtsoulatze, D. Burth Kurka and D. Gündüz, "Deep Joint Source-Channel Coding for Wireless Image Transmission," in *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 567-579, Sept. 2019, doi: 10.1109/TCCN.2019.2919300.

[2] K. Yang, S. Wang, J. Dai, X. Qin, K. Niu and P. Zhang, "SwinJSCC: Taming Swin Transformer for Deep Joint Source-Channel Coding," in *IEEE Transactions on Cognitive Communications and Networking*, doi: 10.1109/TCCN.2024.3424842.

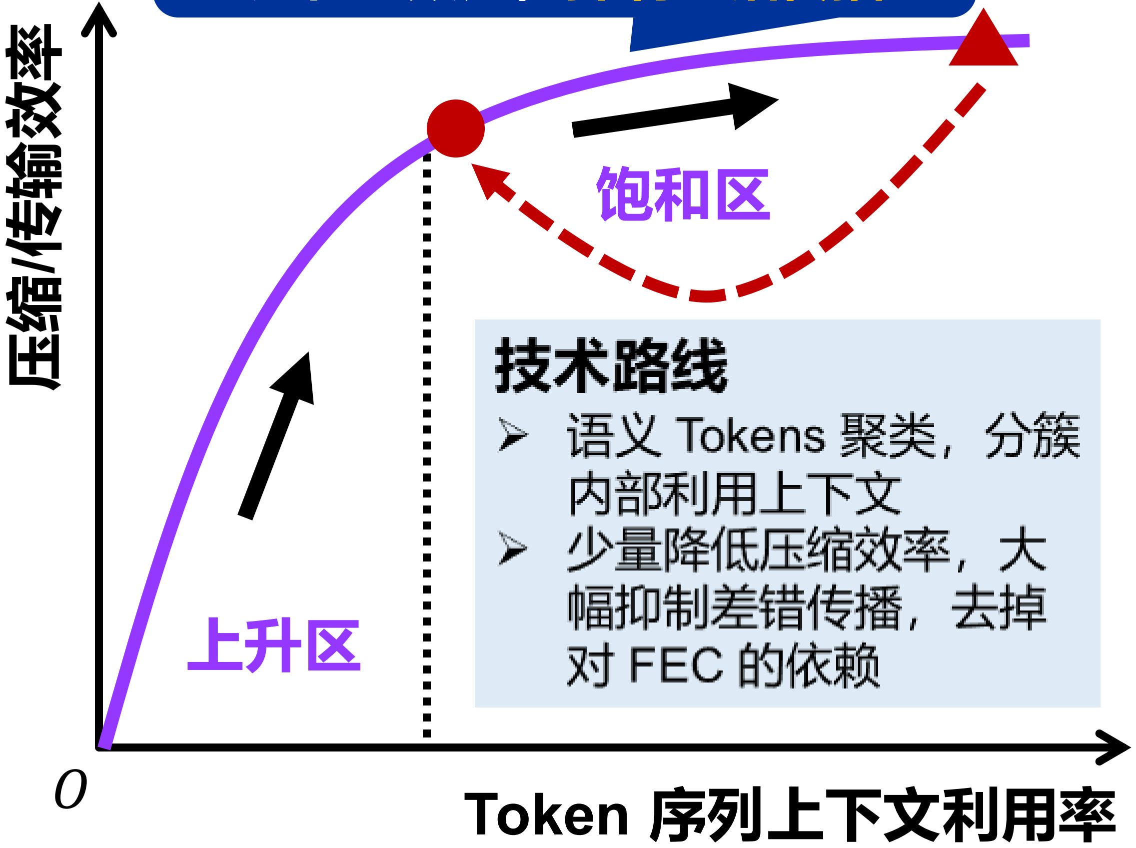


实验观察：获得了平滑效应，但损失了编码效率

可行性分析 — 能够利用数据冗余获得容错能力

“被动添加校验冗余→主动利用数据冗余”范式转变，实现强韧稳健传输

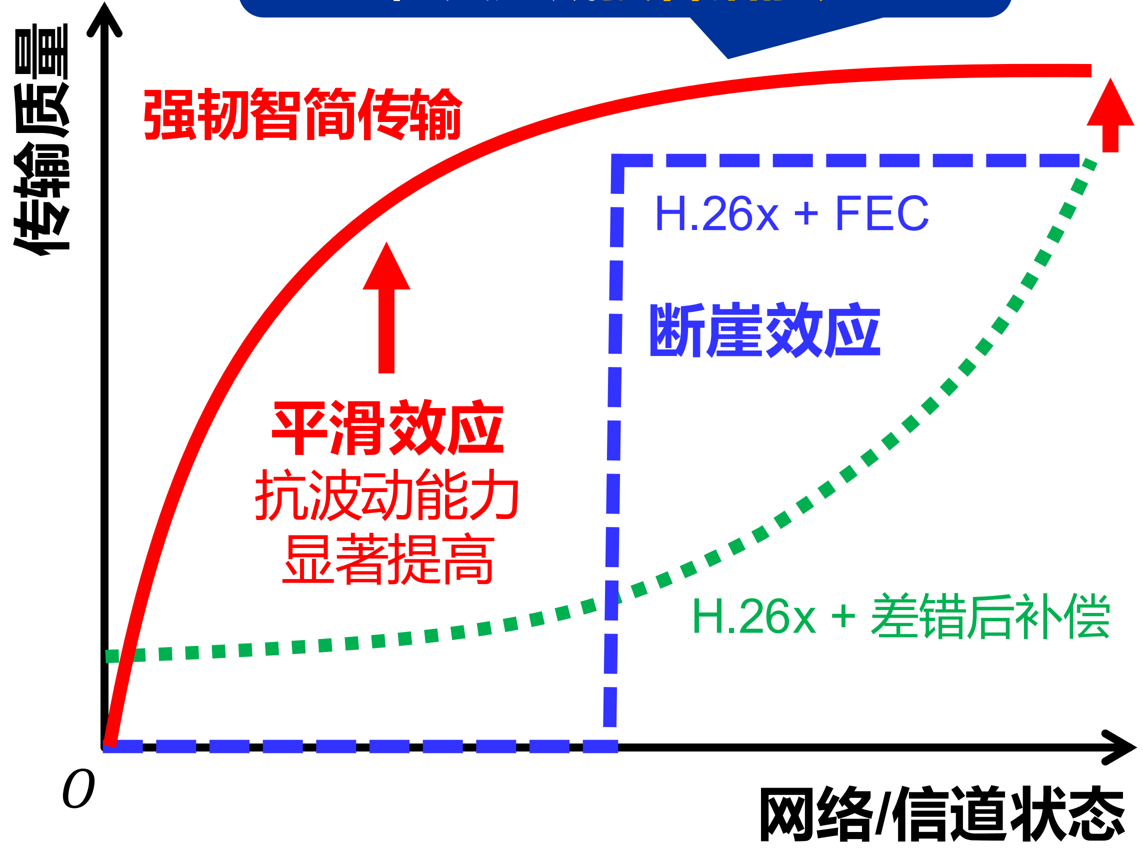
可行性分析：主动利用“上下文饱和”效应，抑制差错传播



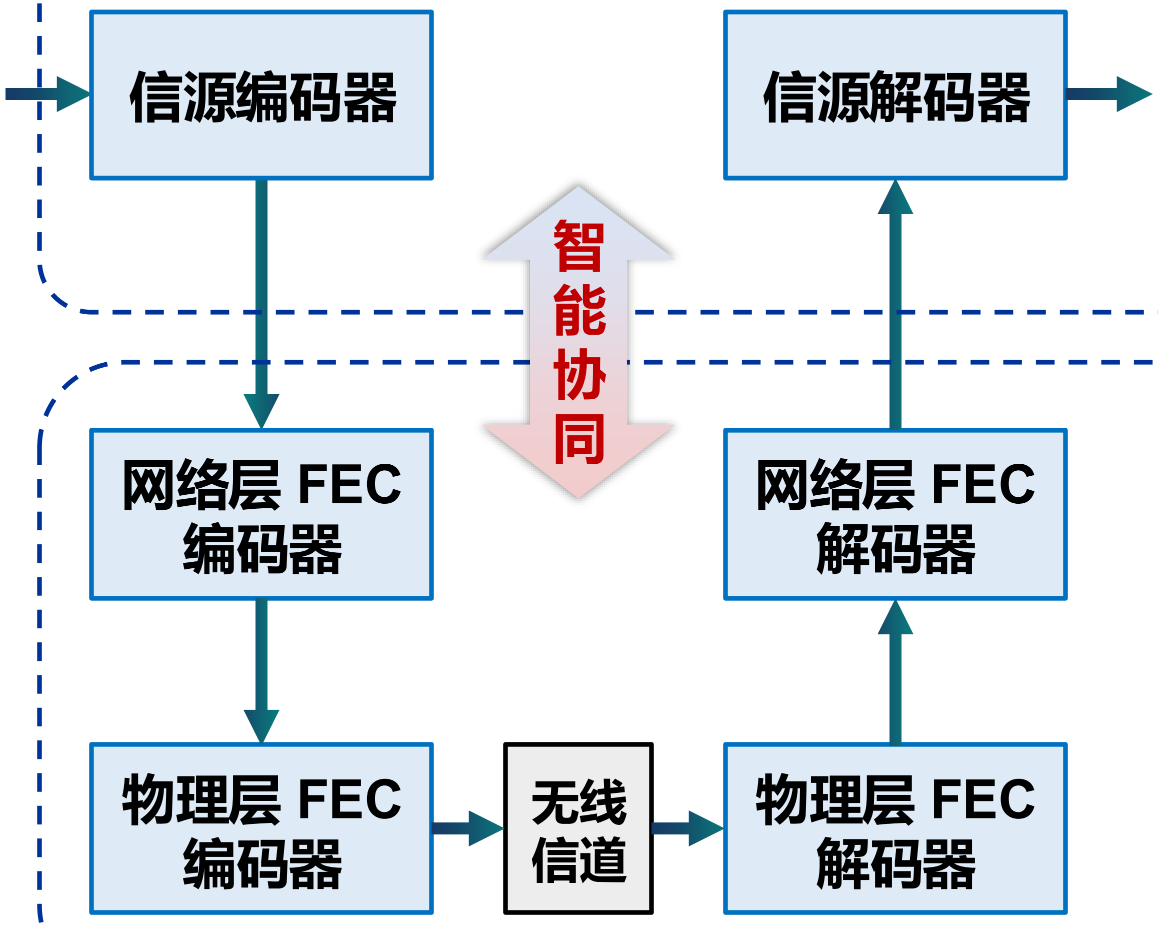
技术路线
- ➢ 语义 Tokens 聚类，分簇内部利用上下文
- ➢ 少量降低压缩效率，大幅抑制差错传播，去掉对 FEC 的依赖

达成目标

设计目标：既实现高压缩效率，又达成强容错能力



智简思路 — 主动利用先验，对抗传输损伤



传统：编解码器主要面向压缩效率联合优化

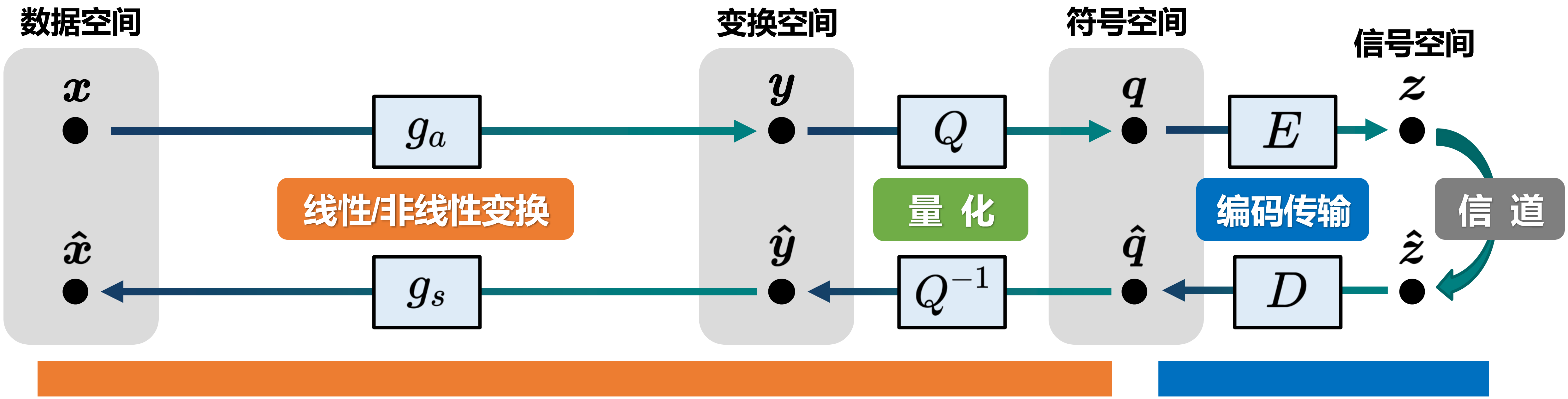
智简：同时向压缩效率和差错恢复联合优化

传统：与信源编解码器交互浅，以无失真传输应用层数据包为“硬目标”

智简：与信源编解码器交互深，柔性适配应用层业务需求与差错恢复能力，允许“有损传输”，有效缓解通信带宽与时延压力

传统编码传输 — 总体流程

❐ 传统编码传输系统



变换空间压缩信息量

$H(\boldsymbol{x}) \gg H(\boldsymbol{q})$

信息量有较大损失

【体现信源数据压缩能力】

传输失真

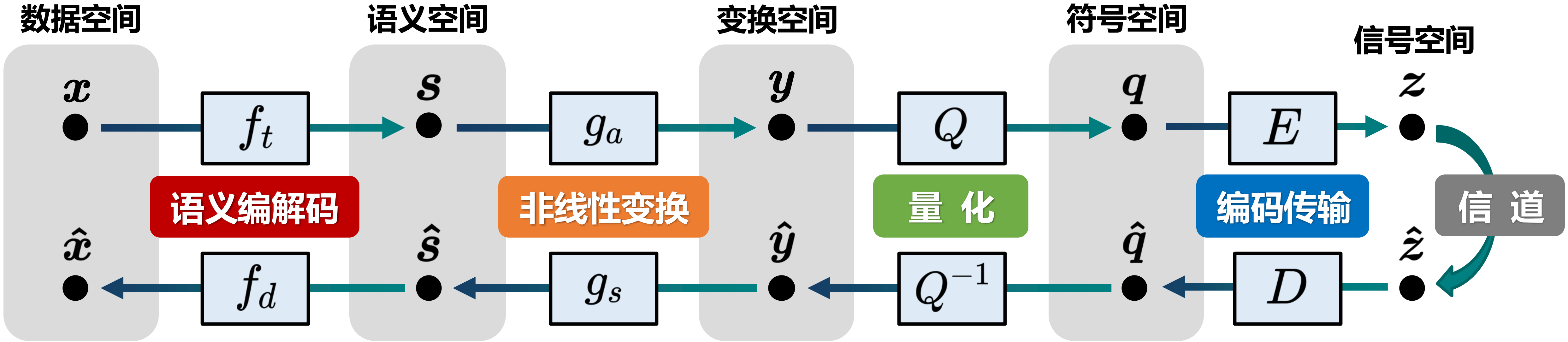
$H(\boldsymbol{q}|\hat{\boldsymbol{z}}) > 0$

信息传输有失真

【符号丢失或失真】

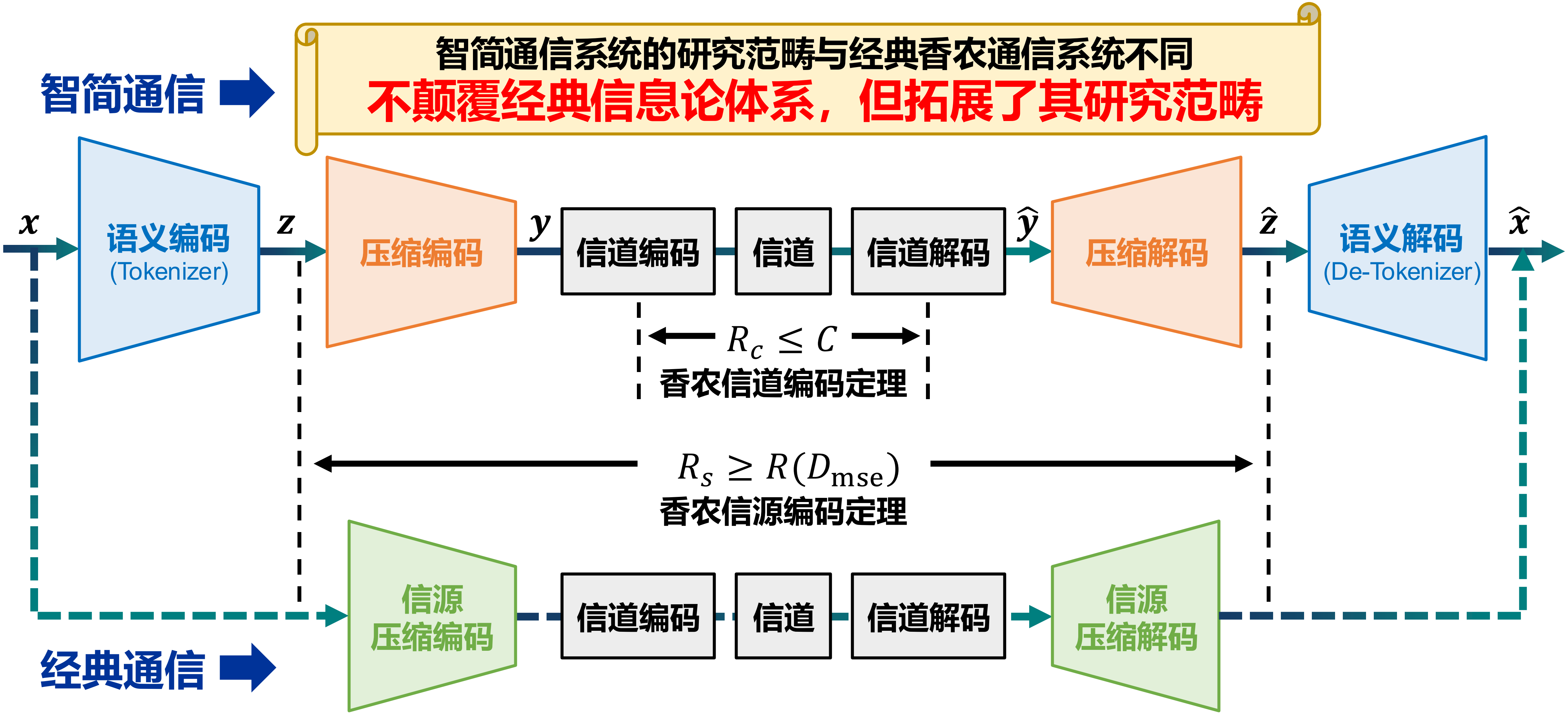
智简编码传输 — 总体流程

❒ 联合信源信道编码 JSCC ≠ 语义通信



语义空间压缩维度	变换空间压缩信息量	传输失真
$\dim(\boldsymbol{x}) > \dim(\boldsymbol{s})$	$H(\boldsymbol{s}) \gg H(\boldsymbol{q})$	$H(\boldsymbol{q}\|\hat{\boldsymbol{z}}) > 0$
信息量略有损失	**信息量有较大损失**	**信息传输有失真**
【$H(\boldsymbol{x}) > H(\boldsymbol{s})$】	【体现Token信息压缩能力】	【Token丢失或失真】

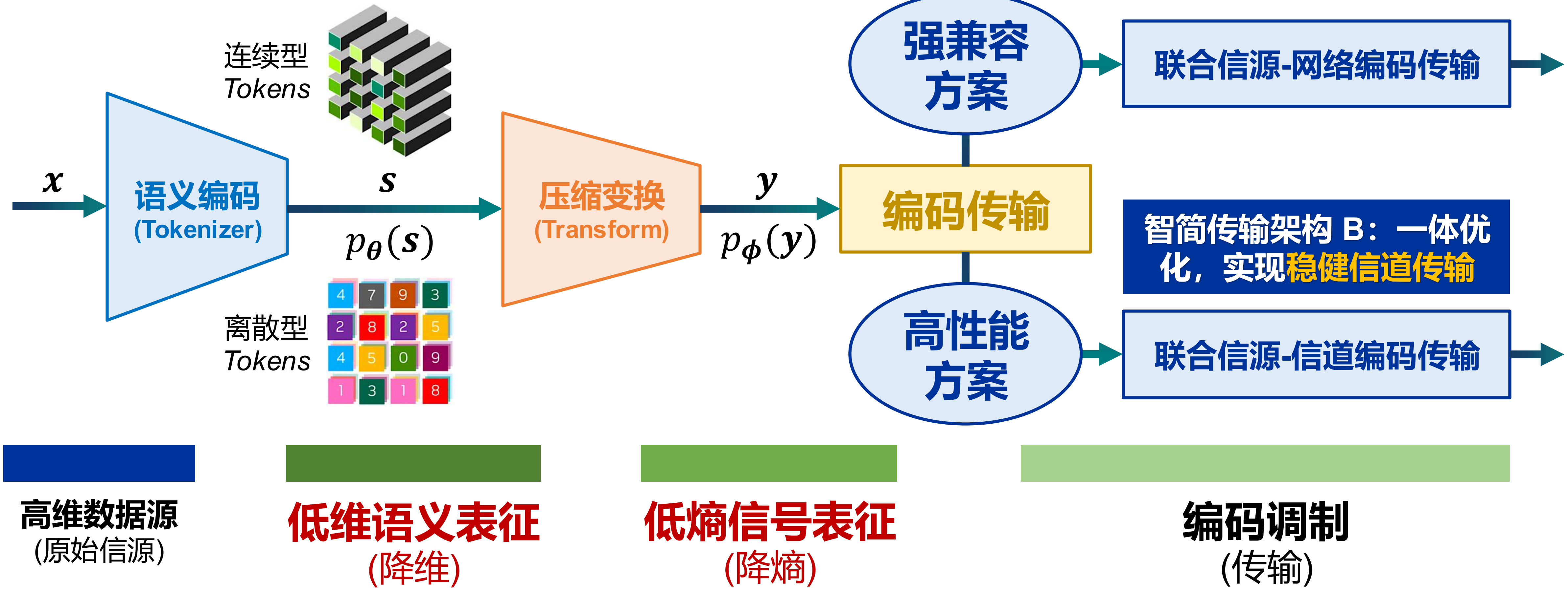
智简编码传输 vs. 经典编码传输



智简编码传输 — 具体实施方案

❐ 两类系统实施方案

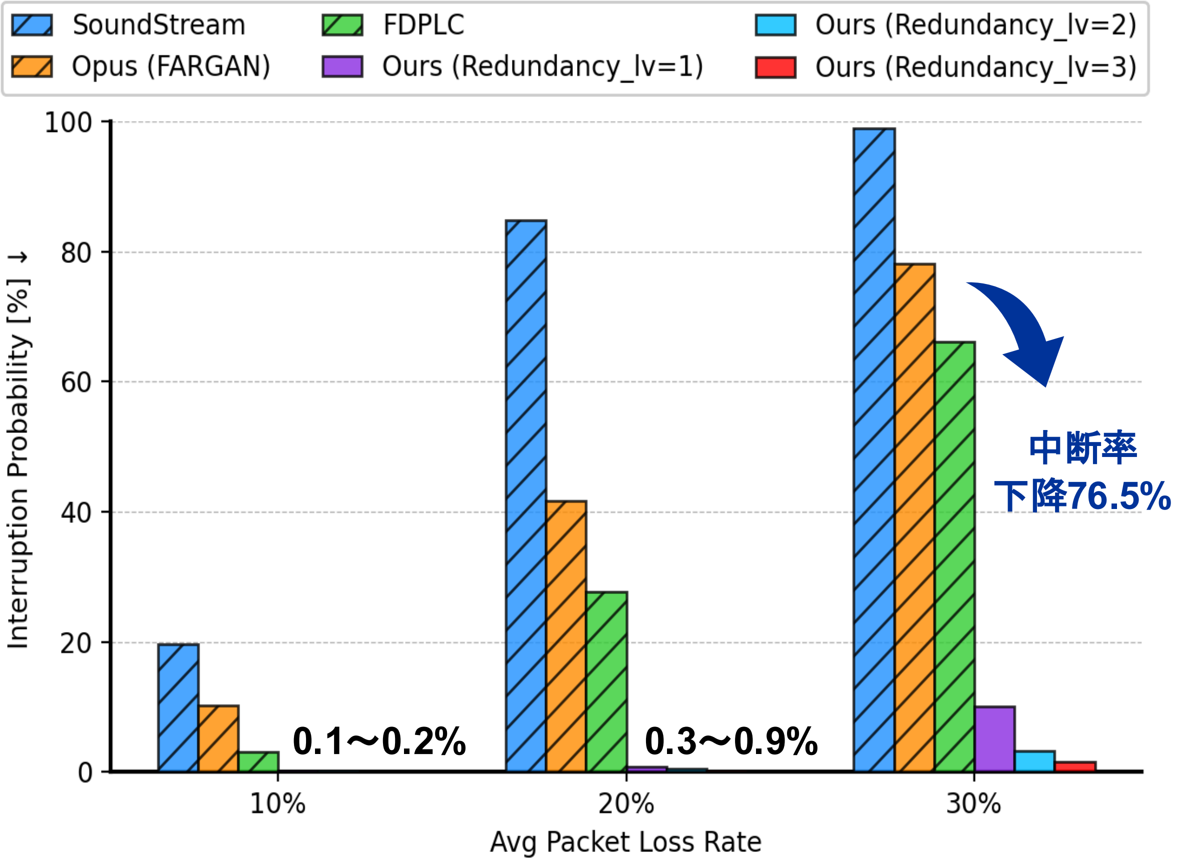
智简传输架构 A：功能协同，实现**强韧网络传输**

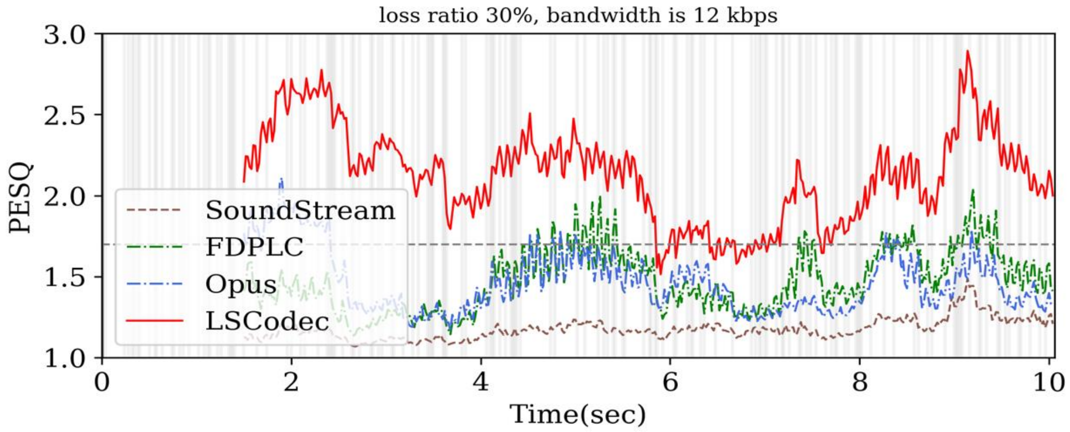


音频/语音的“强韧网络传输”

❐ 动态随机丢包网络环境下，实现高实时、强韧性音频/语音传输

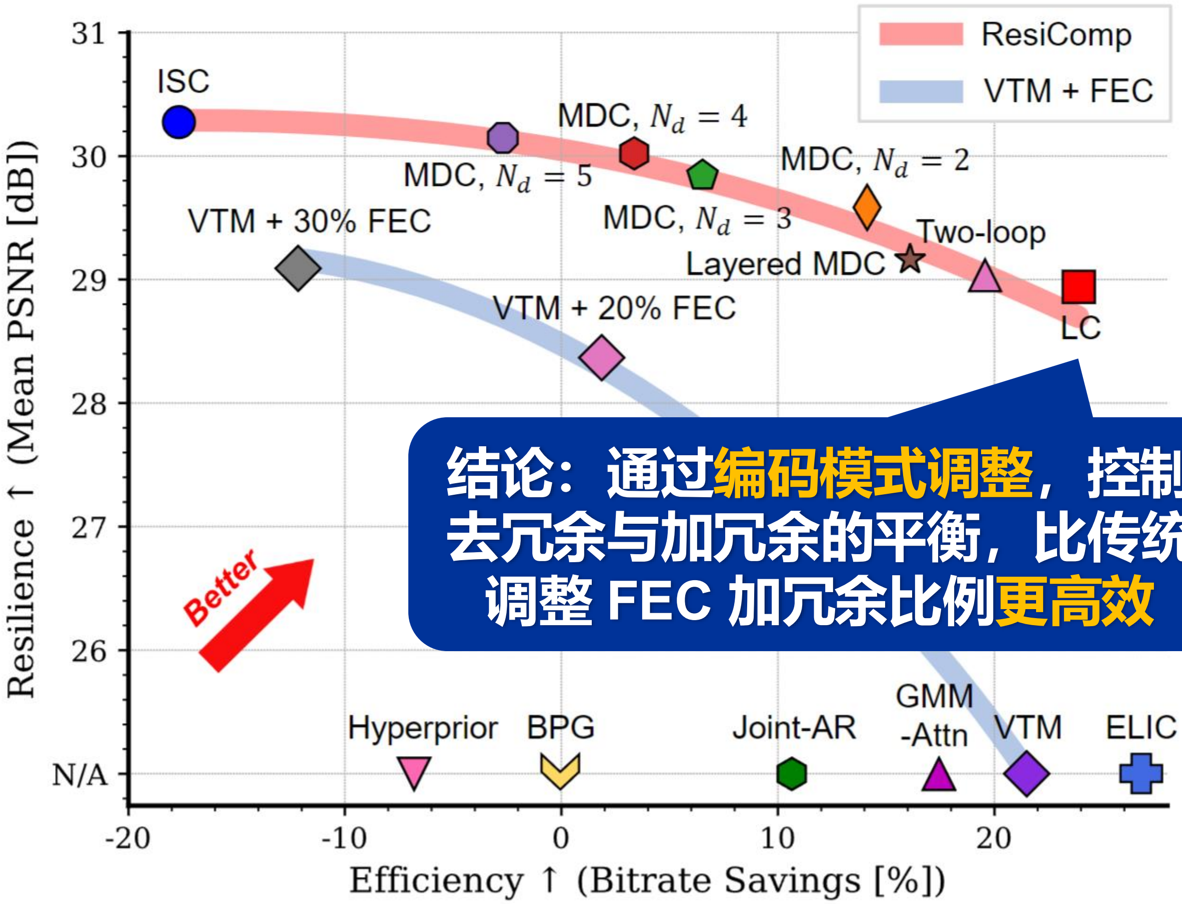
➢ 主要特点：低带宽开销、高保真恢复、强韧抗网络传输随机丢包、低重传开销



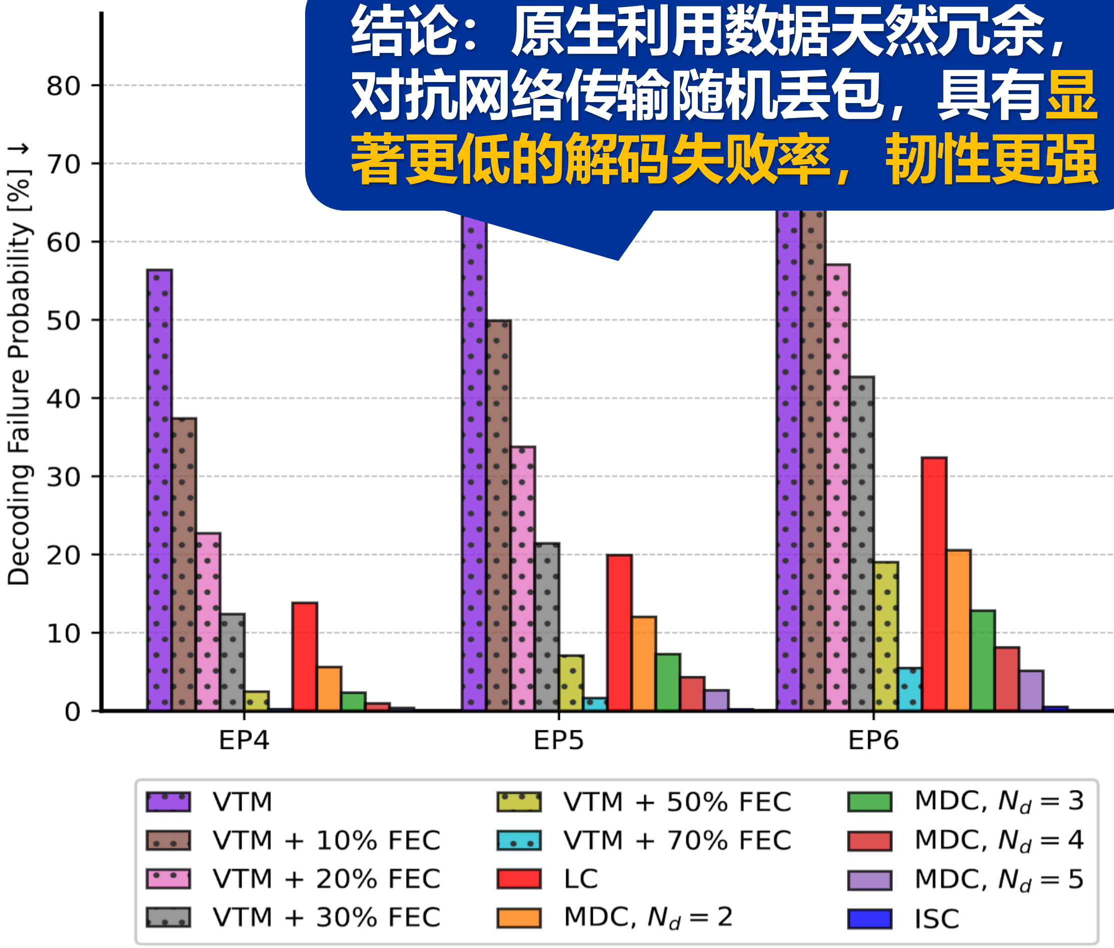
配置	网络传输平均丢包率：30%，总码率：12kbps				
方案	SoundStream	FD-PLC	Opus	智简传输 (本方案)	原音频
音频					
传输性能					

视频/图像的“强韧网络传输”

❐ 动态随机丢包网络环境下，实现高实时、强韧性视频/图像传输



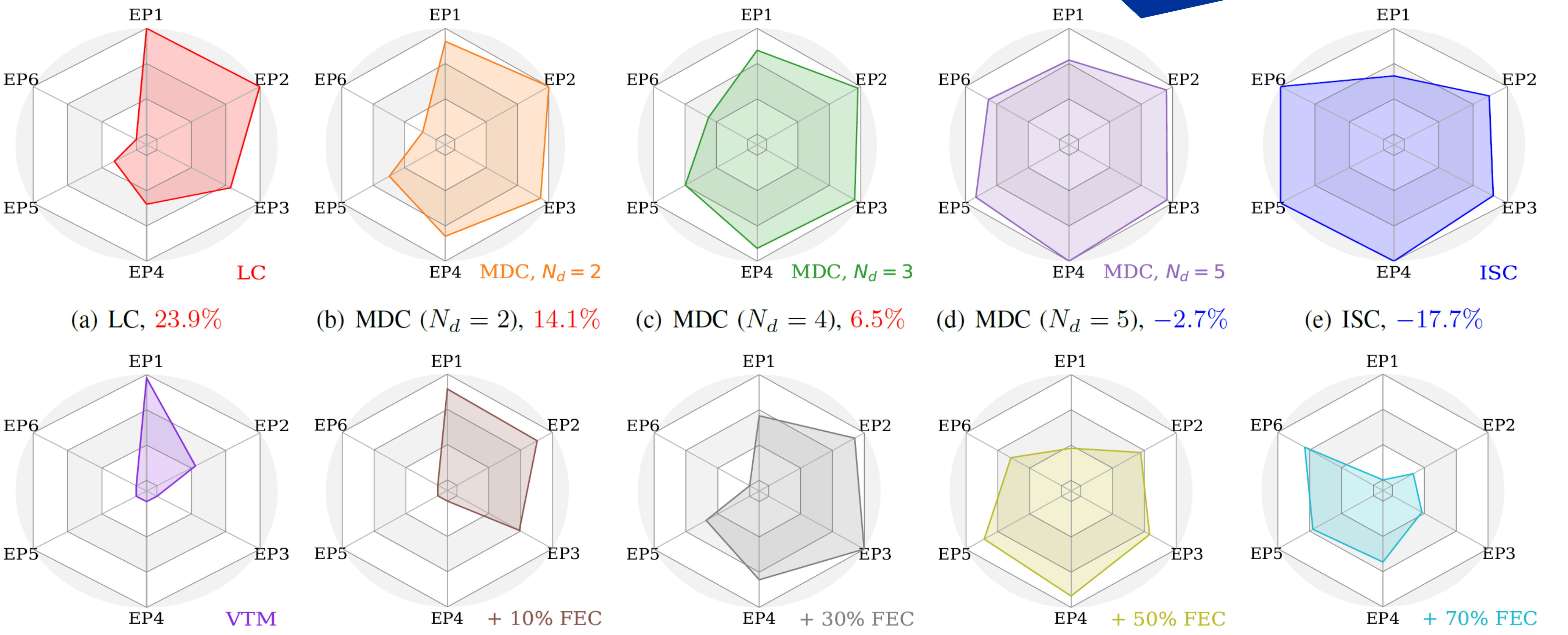
结论：通过编码模式调整，控制去冗余与加冗余的平衡，比传统调整 FEC 加冗余比例更高效



结论：原生利用数据天然冗余，对抗网络传输随机丢包，具有显著更低的解码失败率，韧性更强

视频/图像的“强韧网络传输”

结论：原生抗丢包的 ResiComp Codec 能够适配更广的丢包分布

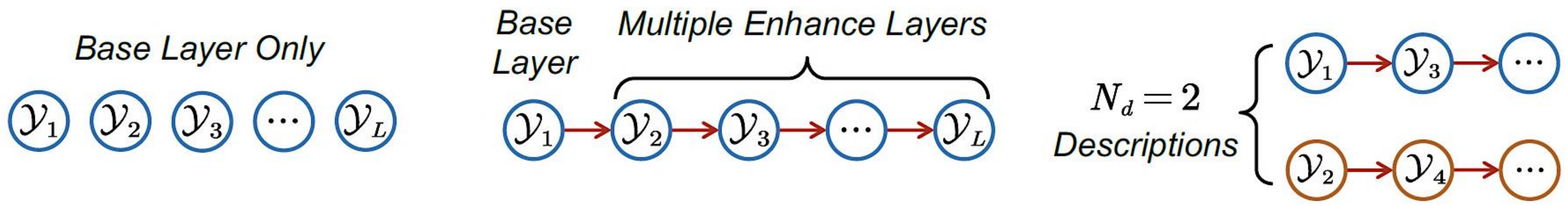


(a) LC, 23.9% (b) MDC ($N_d = 2$), 14.1% (c) MDC ($N_d = 4$), 6.5% (d) MDC ($N_d = 5$), −2.7% (e) ISC, −17.7%

(f) VTM, 21.5% (g) + 10% FEC, 12.8% (h) + 30% FEC, −12.2% (i) + 50% FEC, −57.0% (j) + 70% FEC, −161.7%

视频/图像的“强韧网络传输”

❒ 有效性与可靠性的权衡靠不同编码模式进行调节



(a) Intra-slice coding (ISC) (b) Layered coding (LC) (c) Multi-description coding (MDC)

结论：数据天然冗余 容错性价比高

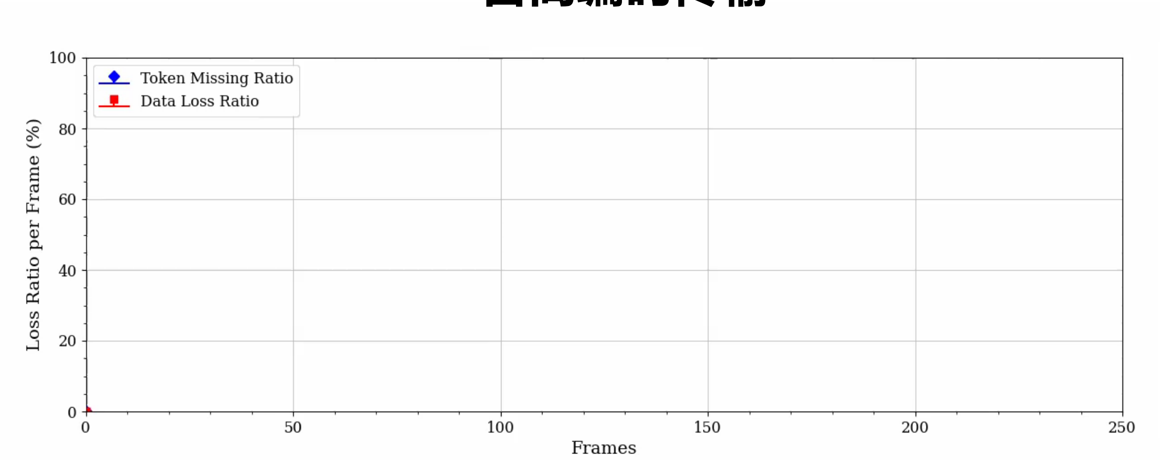
❒ 不同编码模式的综合性能

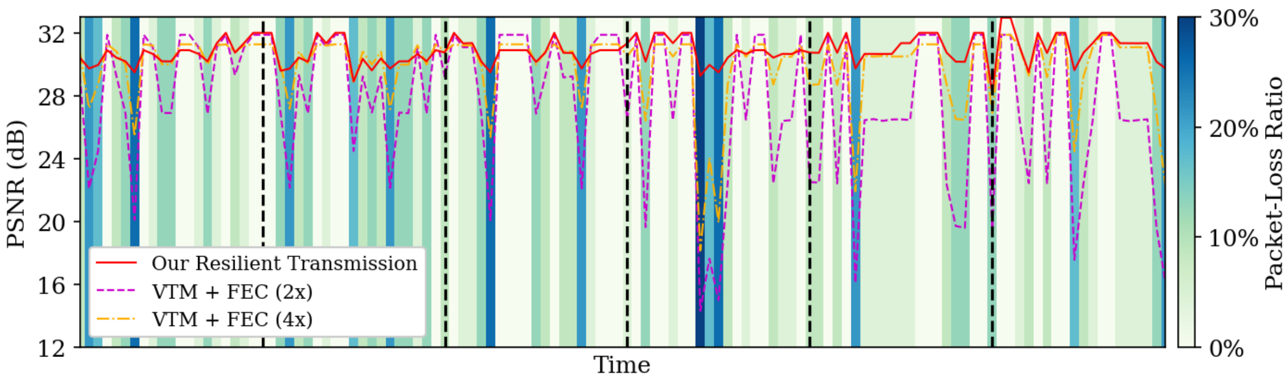
ResiComp 编码模式	抗随机丢包能力	编码效率 (BD-rate vs. H.265)
Slice 独立编码 (ISC)	**强**，无丢包差错传播	17.7%
多描述编码 (MDC)	**较强**，每个描述可独立恢复，部分描述丢失仍可解码	−14.1% ~ 2.7% (描述数量降低，效率降低)
分层渐进编码 (LC)	**弱**，前缀重要，初始包丢失影响较大	−23.9% (≈H.266)

视频/图像的“强韧网络传输”

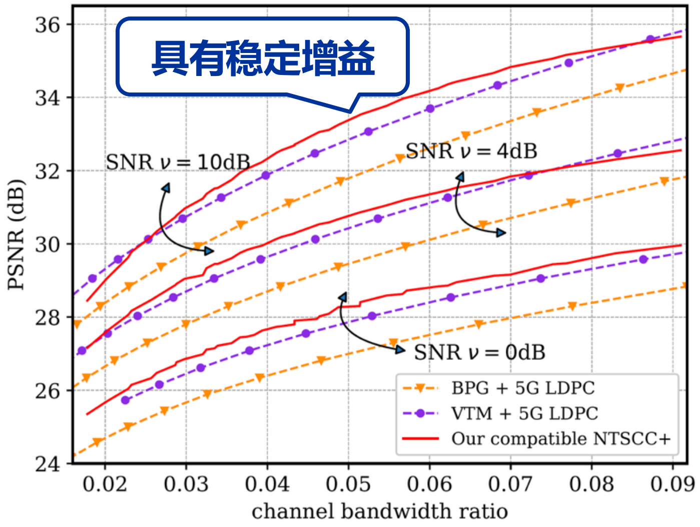
智简编码传输

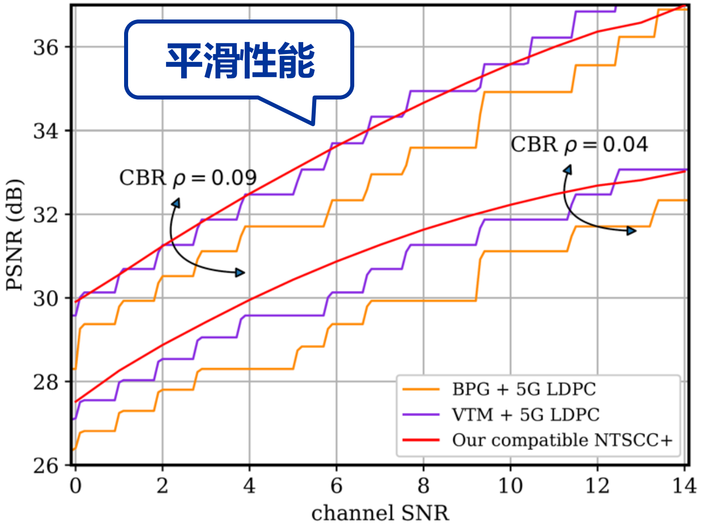
传统编码传输





视频/图像的“稳健信道传输”

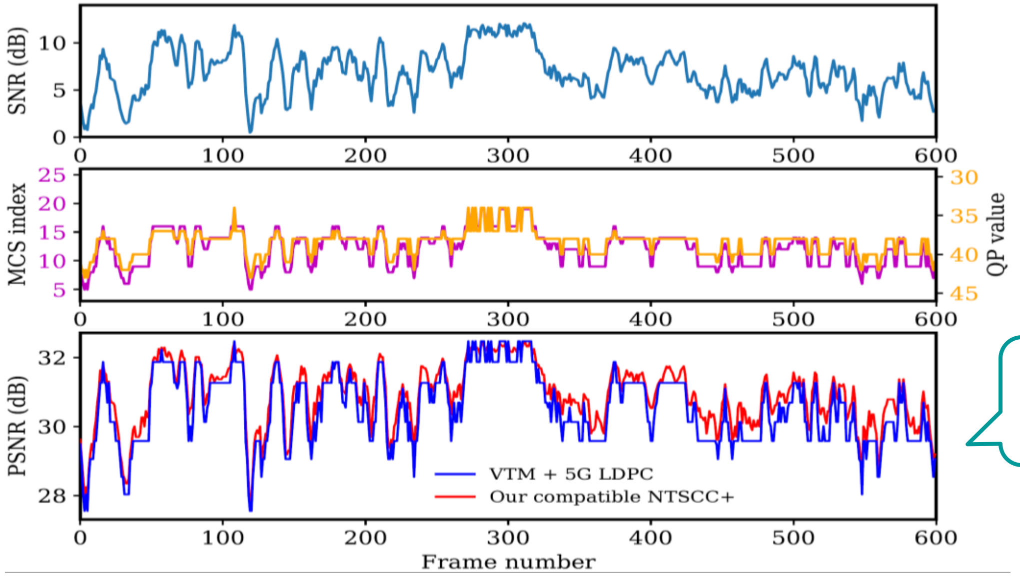




相较于 “VTM + 5G 传输” 的带宽开销节省

Coded transmission scheme	Kodak	CLIC21
JPEG + 5G LDPC	186.09%	586.31%
JPEG2000 + 5G LDPC	94.24%	323.11%
BPG + 5G LDPC	23.93%	55.43%
VTM + 5G LDPC	**0%**	**0%**
Ballé *et al.* + 5G LDPC	31.79%	47.71%
Minnen *et al.* + 5G LDPC	11.94%	11.65%
He *et al.* + 5G LDPC	6.86%	8.90%
Deep JSCC	74.87%	158.88%
NTSCC	1.33%	23.06%
Our NTSCC+	**–5.86%**	**2.94%**
Our compatible NTSCC+	–1.74%	20.94%
Our NTSCC++	**–14.25%**	**–4.70%**

编解码时延较为平衡



高动态衰落信道获得“稳健性能”

综合性能对比

Transmission scheme	BD-ρ (%, ↓)	Params (M)	End-to-end latency (s)	
			Encoding	Decoding
Minnen + 5G LDPC	11.94	14.13 †	0.046	1.7
He + 5G LDPC	6.86	26.60 †	0.087	0.15
VTM + 5G LDPC	0	—	65.8	0.289
NTSCC	1.33	32.94 ‡	0.036	0.015
Our NTSCC+	–5.86	57.36 ‡	0.058	0.023
Our compatible NTSCC+	–1.74	60.29	0.061	0.027
Our NTSCC++ (1 step)	–9.77	57.36 ‡	0.6	0.023

视频/图像的“高保真稳健信道传输”

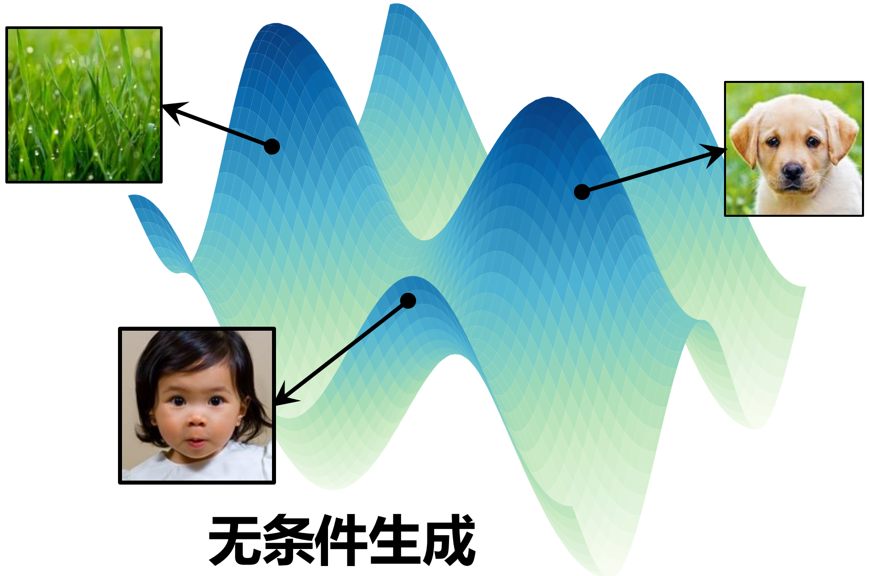
❐ 条件控制生成：解码新范式

信源编码 → 信道编码 → 无线信道 → 条件控制生成器
(接收信号视为控制条件) →

$$\boldsymbol{y} = \underbrace{\mathcal{H}(\mathcal{M}(\mathcal{E}_c(\mathcal{E}_s(\boldsymbol{x}))))}_{\mathcal{A}(\boldsymbol{x})} + \boldsymbol{n} = \boxed{\mathcal{A}(\boldsymbol{x})} + \boldsymbol{n}$$

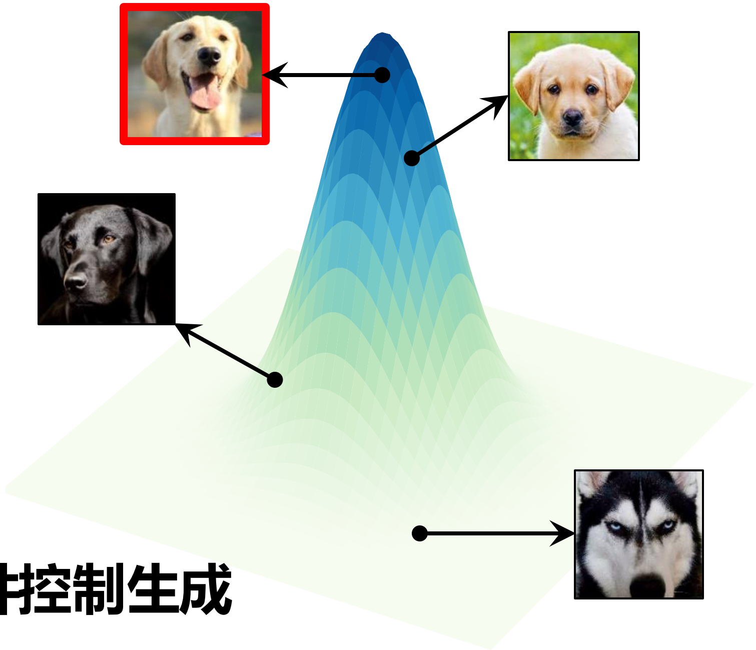
特点：非线性映射、时变、深衰落(强噪)

信源 $\boldsymbol{x}$



无条件生成

接收信号引导



条件控制生成

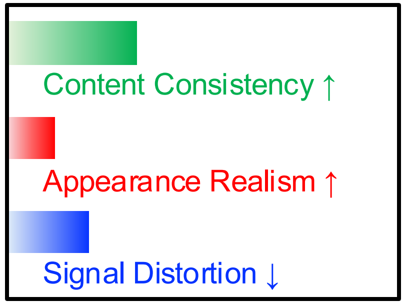
视频/图像的“高保真稳健信道传输”

信道带宽比 ≤ 0.02
(低带宽开销)

VTM + 5G LDPC
+ QAM + AMC
(当前应用方案)

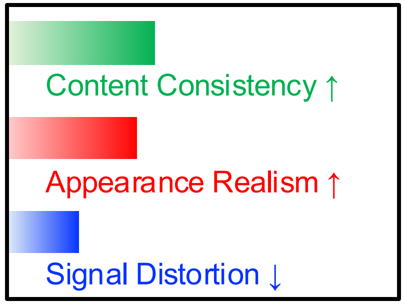
SNR_{AMC} = 2dB	SNR_{AMC} = 4dB	SNR_{AMC} = 2dB	SNR_{AMC} = 0dB	SNR_{AMC} = –2dB	SNR_{AMC} = –2dB
SNR = 6dB	SNR = 4dB	SNR = 2dB	SNR = 0dB	SNR = –2dB	SNR = –4dB

性能指标



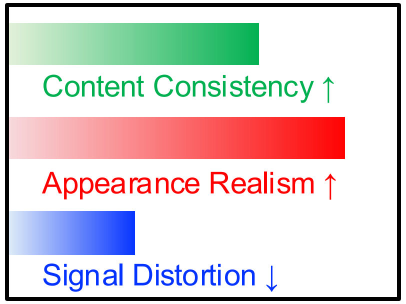
感知优化
JSCC 传输
(确定性解码)

SNR = 6dB	SNR = 4dB	SNR = 2dB	SNR = 0dB	SNR = –2dB	SNR = –4dB



生成式
JSCC 传输
(随机性解码)

SNR = 6dB	SNR = 4dB	SNR = 2dB	SNR = 0dB	SNR = –2dB	SNR = –4dB



被关注的问题

序号	问题类型	具体问题	当前结论
1	应用场景	高时效智简编码传输当前主要用来解决哪些场景的通信问题?	**解决“三高”难题**
2	技术领域	智简编码传输究竟是压缩问题，还是传输问题?	**主要是传输问题**
3	技术领域	联合信源信道编码（JSCC）是否等同于语义通信?	**不等同**
4	技术性能	基于 AI 的 JSCC 性能是否比工程上已应用的 SSCC 标准“好”？	**不一定，不同评价维度不一样**
5	科学问题	智简传输系统的核心增益来源在哪？是否会超越香农限?	**多级先验知识；不会超过**
6	技术问题	JSCC 落地应用会面临哪些阻力？如何克服?	**跨层协同难：浅协同容易，深度协同难**
7	开放问题	智简传输与智简网络是怎样的关系?	**降低网络负载，释放网络调控压力**
8	开放问题	大模型等 AI 新技术，和智简通信有什么关联?	**开放问题**

已发表相关学术论文

① J. Dai, S. Wang, K. Tan, Z. Si, X. Qin and P. Zhang, “Nonlinear Transform Source-Channel Coding for Semantic Communications,” in ***IEEE Journal on Selected Areas in Communications***, vol. 40, no. 8, pp. 2300-2316, Aug. 2022. **(ESI HC)**

② S. Wang, J. Dai, Z. Liang, K. Niu, Z. Si, C. Dong and P. Zhang, “Wireless Deep Video Semantic Transmission,” in ***IEEE Journal on Selected Areas in Communications***, vol. 41, no. 1, pp. 214-229, Jan. 2023. **(ESI HC)**

③ J. Dai, S. Wang, K. Yang, K. Tan, X. Qin and P. Zhang., “Toward Adaptive Semantic Communications: Efficient Data Transmission via Online Learned Nonlinear Transform Source-Channel Coding,” in ***IEEE Journal on Selected Areas in Communications***, vol. 41, no. 8, pp. 2609-2627, Aug. 2023.

④ S. Wang, J. Dai, X. Qin, Z. Si, K. Niu and P. Zhang, “Improved Nonlinear Transform Source-Channel Coding to Catalyze Semantic Communications,” in ***IEEE Journal of Selected Topics in Signal Processing***, vol. 17, no. 5, pp. 1022-1037, Sept. 2023.

⑤ J. Dai, X. Qin, S. Wang, L. Xu, K. Niu and P. Zhang, “Deep Generative Modeling Reshapes Compression and Transmission: From Efficiency to Resiliency,” in ***IEEE Wireless Communications***, vol. 31, no. 4, pp. 48-56, August 2024.

⑥ K. Yang, S. Wang, J. Dai, X. Qin, K. Niu and P. Zhang, “SwinJSCC: Taming Swin Transformer for Deep Joint Source-Channel Coding,” in ***IEEE Transactions on Cognitive Communications and Networking***, vol. 11, no. 1, pp. 90-104, Feb. 2025. **(ESI HC)**

⑦ S. Yao, J. Dai, X. Qin, S. Wang, S. Wang and K. Niu, “SoundSpring: Loss-Resilient Audio Transceiver With Dual-Functional Masked Language Modeling,” in ***IEEE Journal on Selected Areas in Communications***, vol. 43, no. 4, pp. 1308-1322, April 2025.

⑧ S. Wang, J. Dai, X. Qin, K. Yang, K. Niu and P. Zhang, "ResiComp: Loss-Resilient Image Compression via Dual-Functional Masked Visual Token Modeling," in ***IEEE Transactions on Circuits and Systems for Video Technology***, vol. 35, no. 7, pp. 7181-7195, July 2025.

⑨ S. Wang, J. Dai, K. Tan, X. Qin, K. Niu and P. Zhang, "DiffCom: Channel Received Signal Is a Natural Condition to Guide Diffusion Posterior Sampling," in ***IEEE Journal on Selected Areas in Communications***, vol. 43, no. 7, pp. 2651-2666, July 2025.

敬 请 批 评 指 正

DeepComm Research Group

"All credits go to my students"

daijincheng@bupt.edu.cn