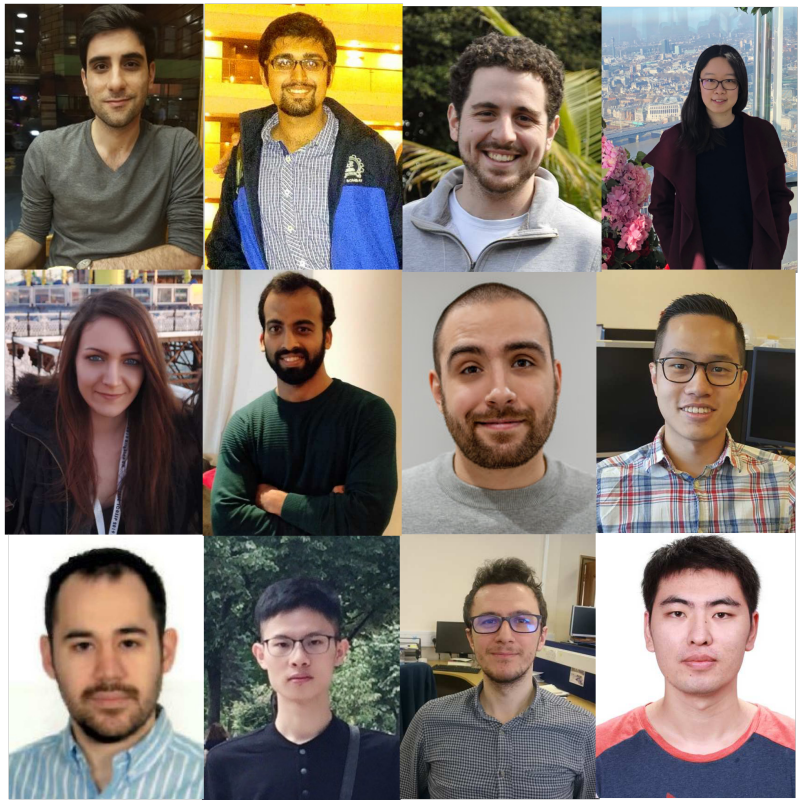
Beyond Transmitting Bits: Semantic and Goal-Oriented Communications

Deniz Gündüz

Imperial College London, UK
University of Modena and Reggio Emilia, Italy

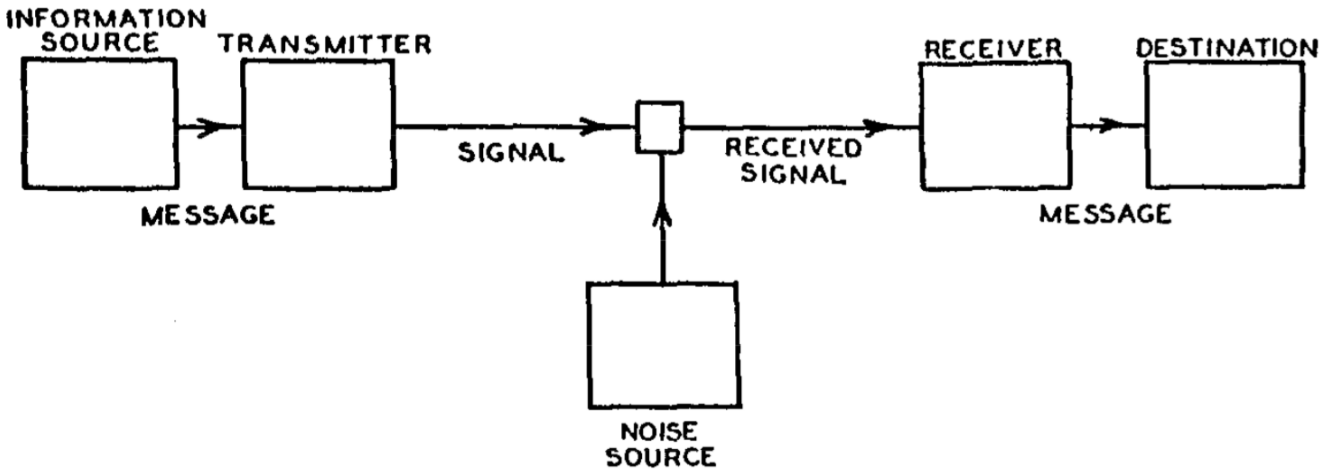
10 November 2021
One World Signal Processing Seminar Series

This work received support from European Research Council (ERC) through grant BEACON and from the UK EPSRC through project CONNECT.



WE ARE HIRING!

Multiple postdoctoral researcher and PhD student positions available!



The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning;* that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

Shannon, "The Mathematical Theory of Communication," in *The Bell System Technical Journal,* Jul- Oct., 1948.

Shannon and Semantics

> The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

- **Common belief:** Shannon intentionally excluded semantics from information theory.
- Is that really the case?

1.2. Three Levels of Communications Problems

Relative to the broad subject of communication, there seem to be problems at three levels. Thus it seems reasonable to ask, serially:

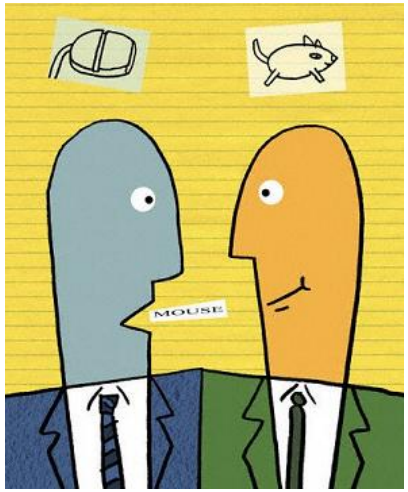
LEVEL A. How accurately can the symbols of communication be transmitted? (The technical problem.)

LEVEL B. How precisely do the transmitted symbols convey the desired meaning? (The semantic problem.)

LEVEL C. How effectively does the received meaning affect conduct in the desired way? (The effectiveness problem.)

W. Weaver, "Recent contributions to the mathematical theory of communication," in *The Mathematical Theory of Communication,* University of Illinois Press, Urbana, 1964.

What is Semantics?



- Semantics deals with 'meaning'
- Formal/ mathematical definition is elusive
- Many attempts: Carnap and Bar-Hillel,52, Bao et al.'11, Basu et al.'14, Guler et al,'18 (model-based approach)
- Basic idea: In communications, goal is to convey the meaning; exact reconstruction is not needed.

R. Carnap, Y. Bar-Hillel et al., "An outline of a theory of semantic information," RLE Technical Reports 247, Research Laboratory of Electronics, MIT, Oct. 1952.

J. Bao et al., "Towards a theory of semantic communication," IEEE NSW, 2011.

P. Basu, J. Bao, M. Dean, and J. Hendler, "Preserving quality of information by using semantic relationships," *Pervasive Mobile Computing*, 2014.

B. Guler, A. Yener, and A. Swami, "The semantic communication game," IEEE Trans. Cogn. Comm. Networking, 2018.

Rate-Distortion Theory

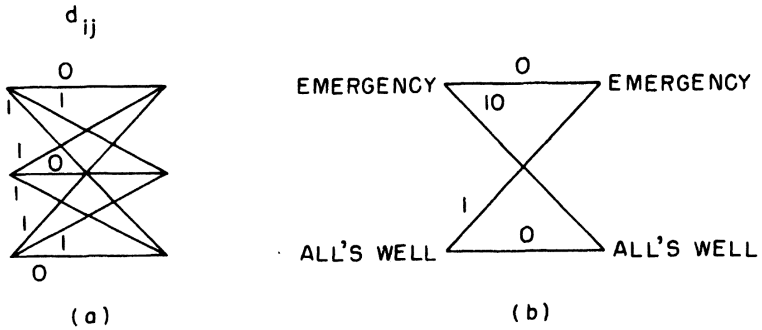
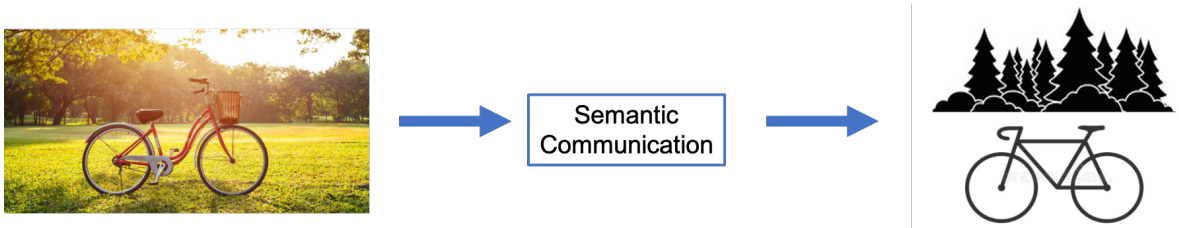


Fig. 1.

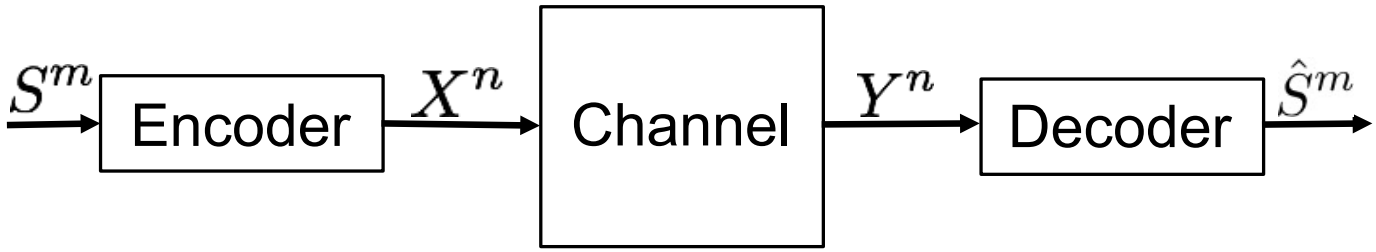
m = I HAVE HEARD THE MERMAIDS SINGING...

Z = I H?VT HEA?D TSE B?RMAIDZ ??NGING...

- Shannon theory does not require exact reconstruction of messages, channel coding theory does!
- Reconstruct messages within some fidelity measure.
- Rate-distortion theory studies the trade-off between communication rate and the end-to-end quality.



Shannon, Coding theorems for a discrete source with a fidelity criterion, *Institute of Radio Engineers, International Convention Record,* 1959.



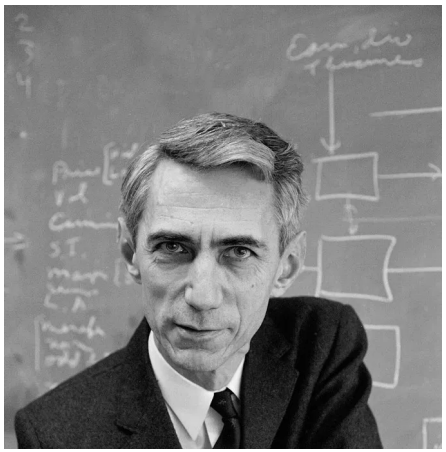
- Source: i.i.d. samples from p_S
- Channel: memoryless with $p_{Y|X}$
- Encoder: $f^{m,n} : S^m \to X^n$
- Decoder: $g^{m,n} : Y^n \to \hat{S}^m$
- Source-Channel Rate: $\frac{n}{m}$
- Additive distortion measure: $d(s, \hat{s})$:

$$d(S^m, \hat{S}^m) = \frac{1}{m} \sum_{i=1}^{m} d(S_i, \hat{S}_i)$$

- A rate- distortion pair (r, D) is achievable if there exists a sequence of encoders and decoders with $\frac{n}{m} \leq r$ and $\lim_{m,n\to\infty} E[d(S^m, \hat{S}^m)] \leq D$.

Shannon's Source-Channel Separation Theorem

- First compress the source
- Match quantized bits to the optimal channel code
- No loss of optimality

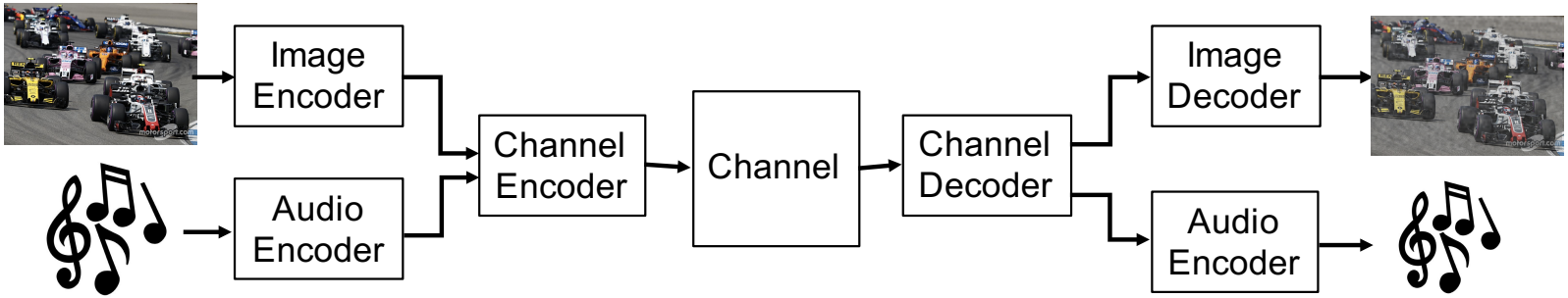


Separation Theorem

(Lossless) Rate r is achievable iff $H(S) \leq rC$
(Lossy) For given rate r and distortion measure $d(\cdot,\cdot)$, the minimum achievable distortion is given by $D(rC)$,

where $D(R)$ is the distortion-rate function of the source, and C is the capacity of the channel.

Separation has been instrumental thanks to modularity

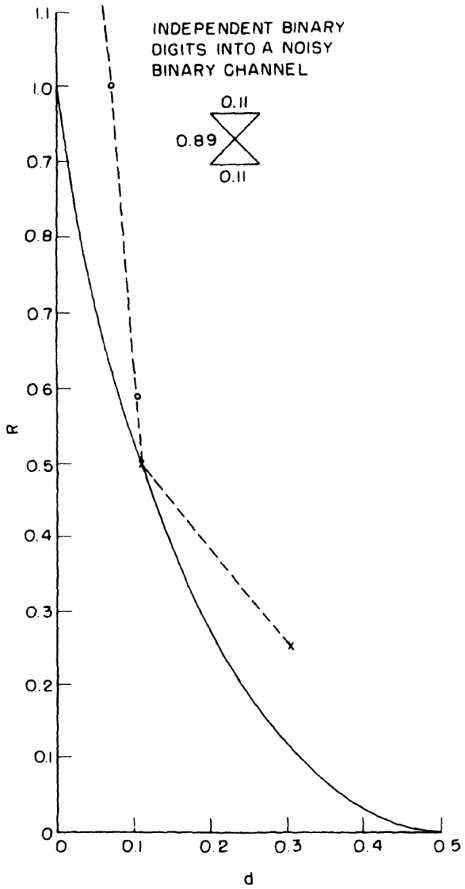


6G will rarely deliver sources for reconstruction!

Latency cost of modularity is too much for most ML applications!

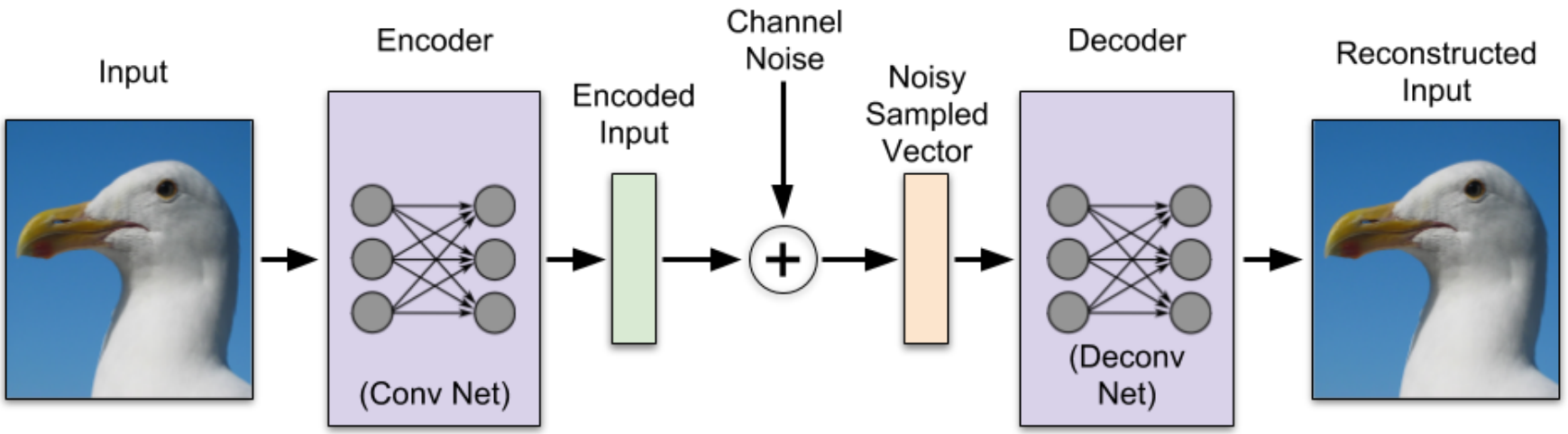
Can We Beat the Shannon Limit?

- Consider Bernoulli 1/2 source with error probability per letter as distortion measure
- Binary symmetric channel with capacity 1/2
- Uncoded tranmission is optimal
- "Any symmetric binary channel will have one point that can be attained exactly by means of straight transmission."

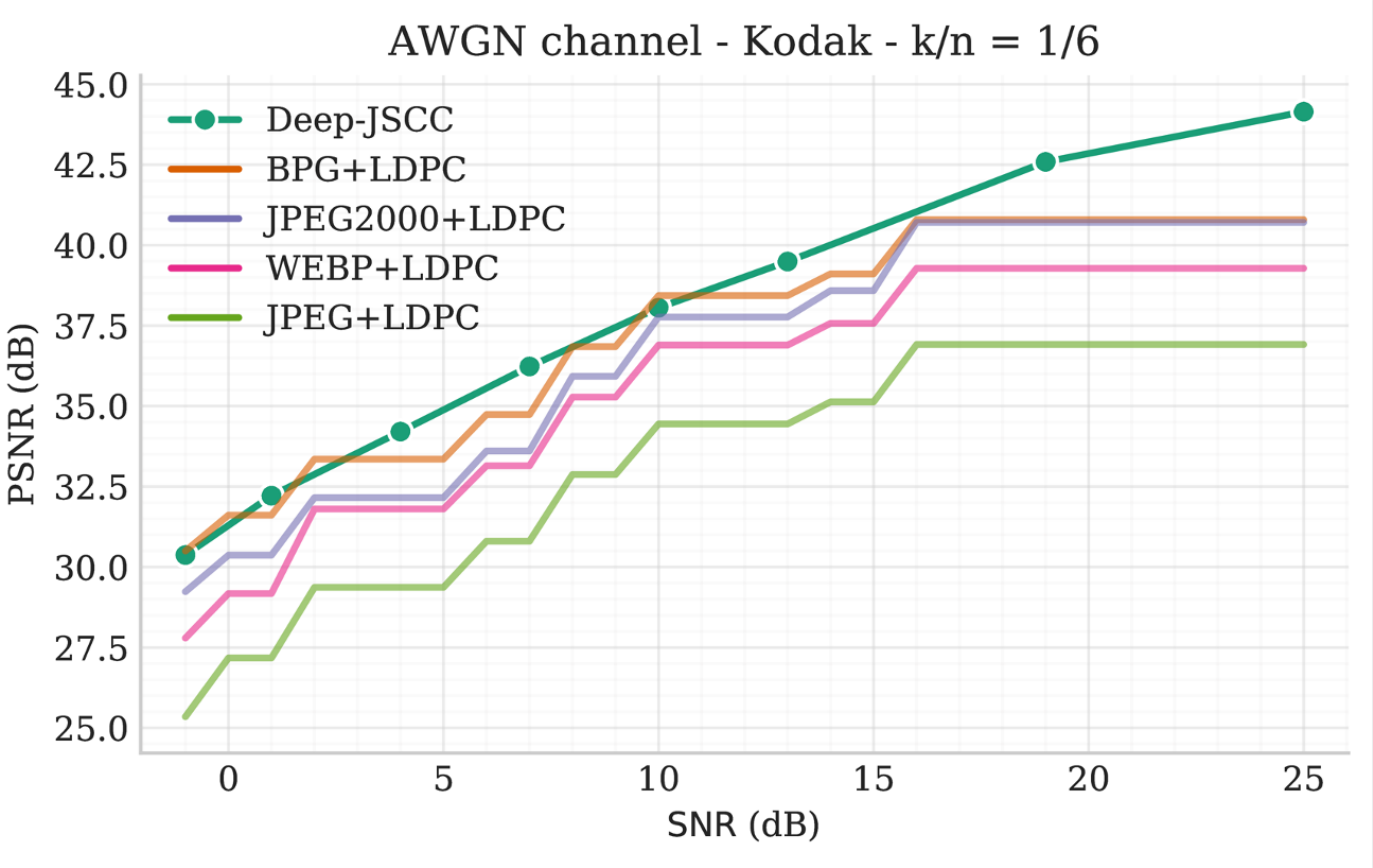


C. E. Shannon, "Coding Theorems for a Discrete Source With a Fidelity Criterion," *Institute of Radio Engineers, International Convention Record,* vol. 7, 1959.

- Deep neural networks can extract complex semantic relationships
- Forget about compression, channel coding, modulation, channel estimation, equalization, etc.
- Deep neural networks for code design



Deep Joint Source-Channel Coding (DeepJSCC)

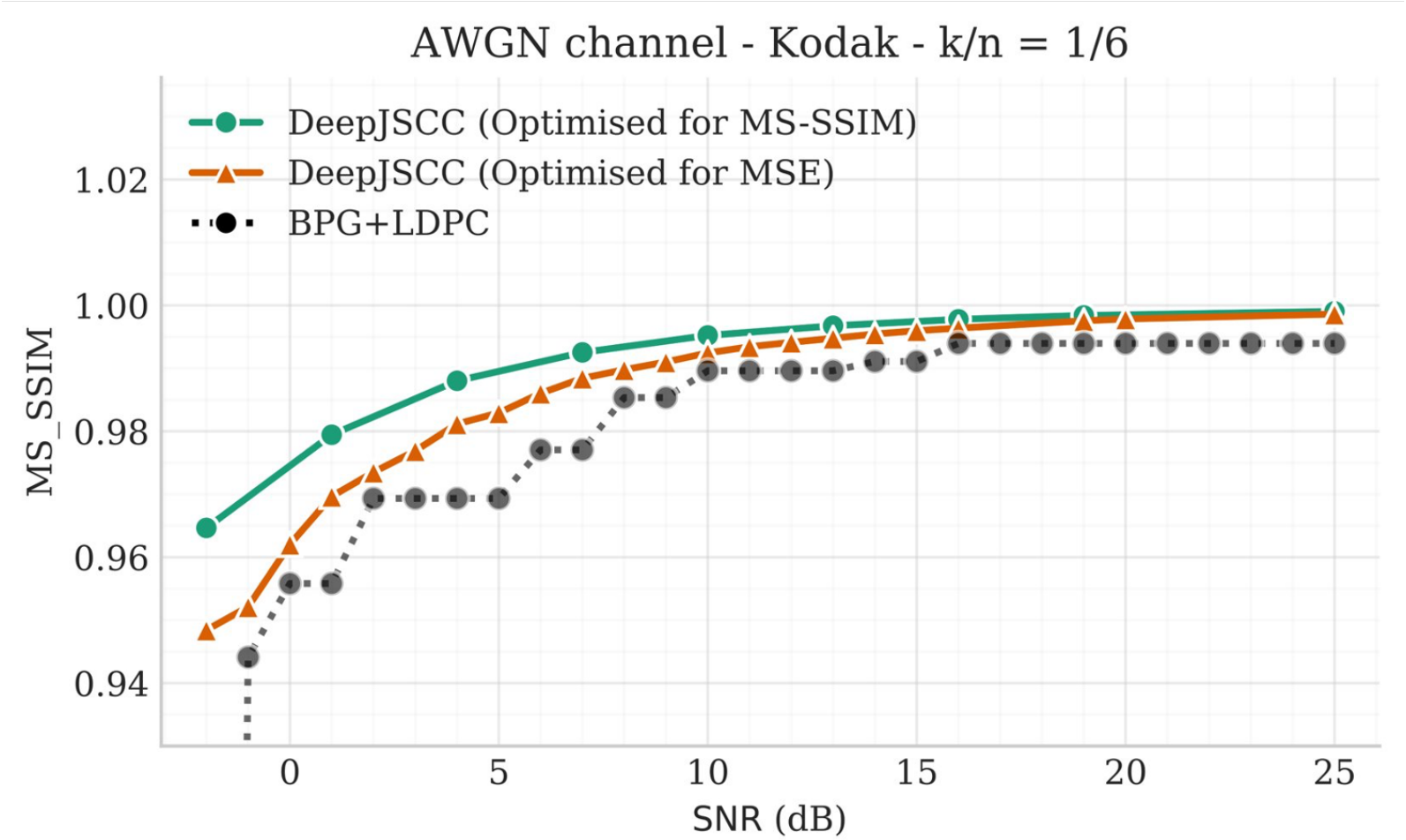


Superior performance compared to state-of-the-art digital schemes!

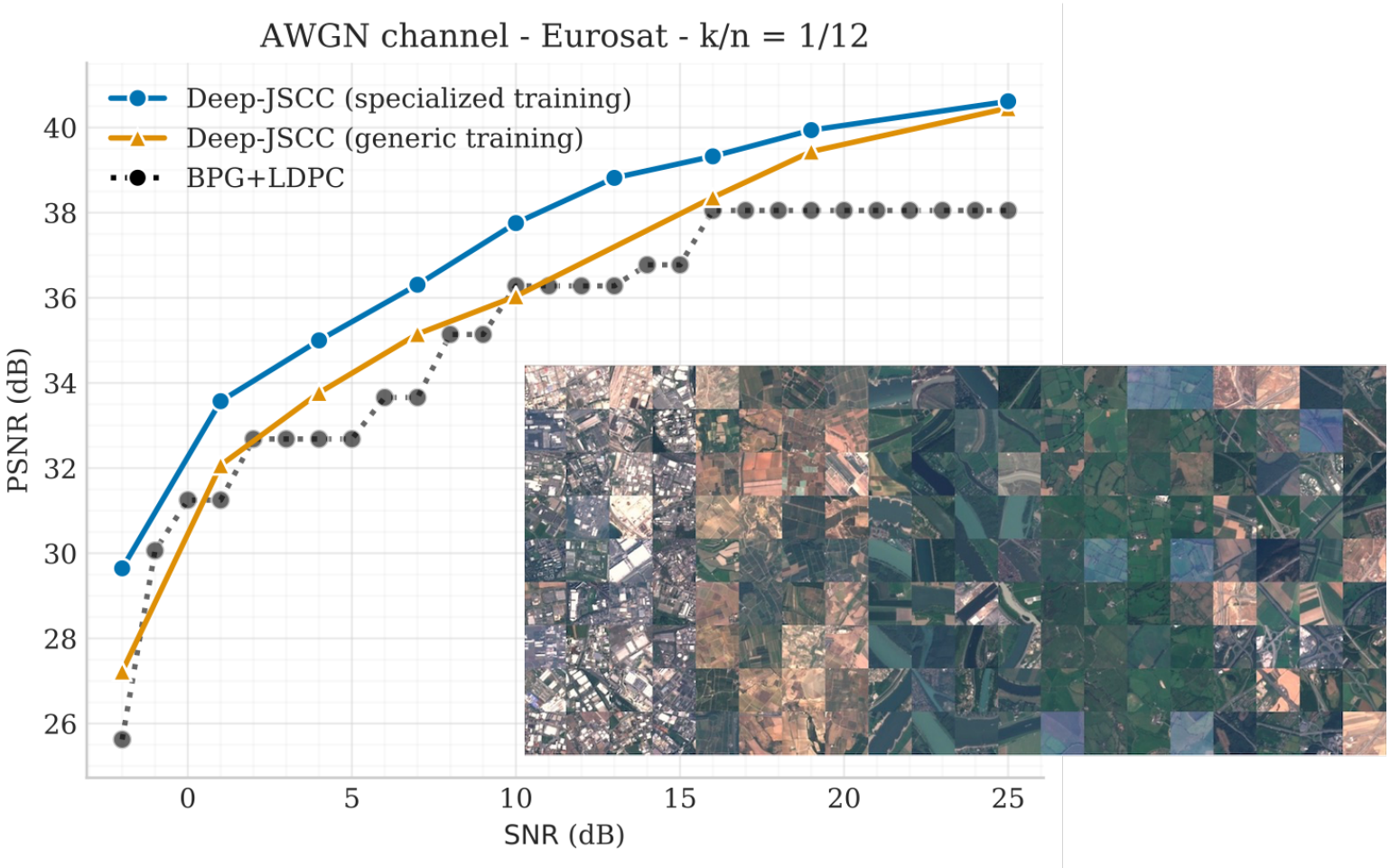
Image size (n): 768 x 512 x 3 $\sim$ 1,2M Blocklength (k): $\sim$200K

Bourtsoulatze, Burth Kurka and Gündüz, Deep joint source-channel coding for wireless image transmission, IEEE Trans. on Cognitive Comms. and Networking, 2019.

Burth Kurka and Gunduz, Joint source-channel coding of images with (not very) deep learning, Int'l Zurich Seminar (IZS), Feb. 2020.

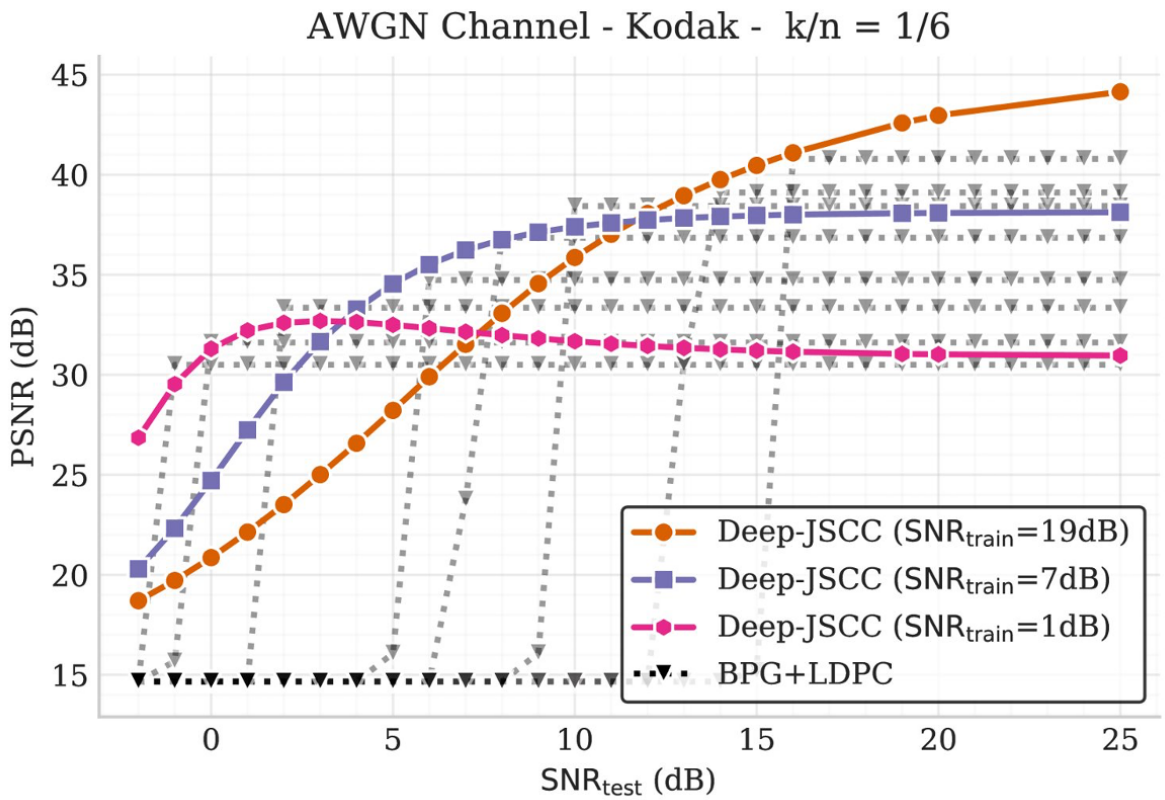


E. Bourtsoulatze, D. Burth Kurka and D. Gündüz, Deep joint source-channel coding for wireless image transmission, IEEE Transactions on Cognitive Communications and Networking, vol. 5, no. 3, pp. 567 - 579, Sep. 2019.



Burth Kurka and Gunduz, Joint source-channel coding of images with (not very) deep learning, Int'l Zurich Seminar (IZS), Feb. 2020.

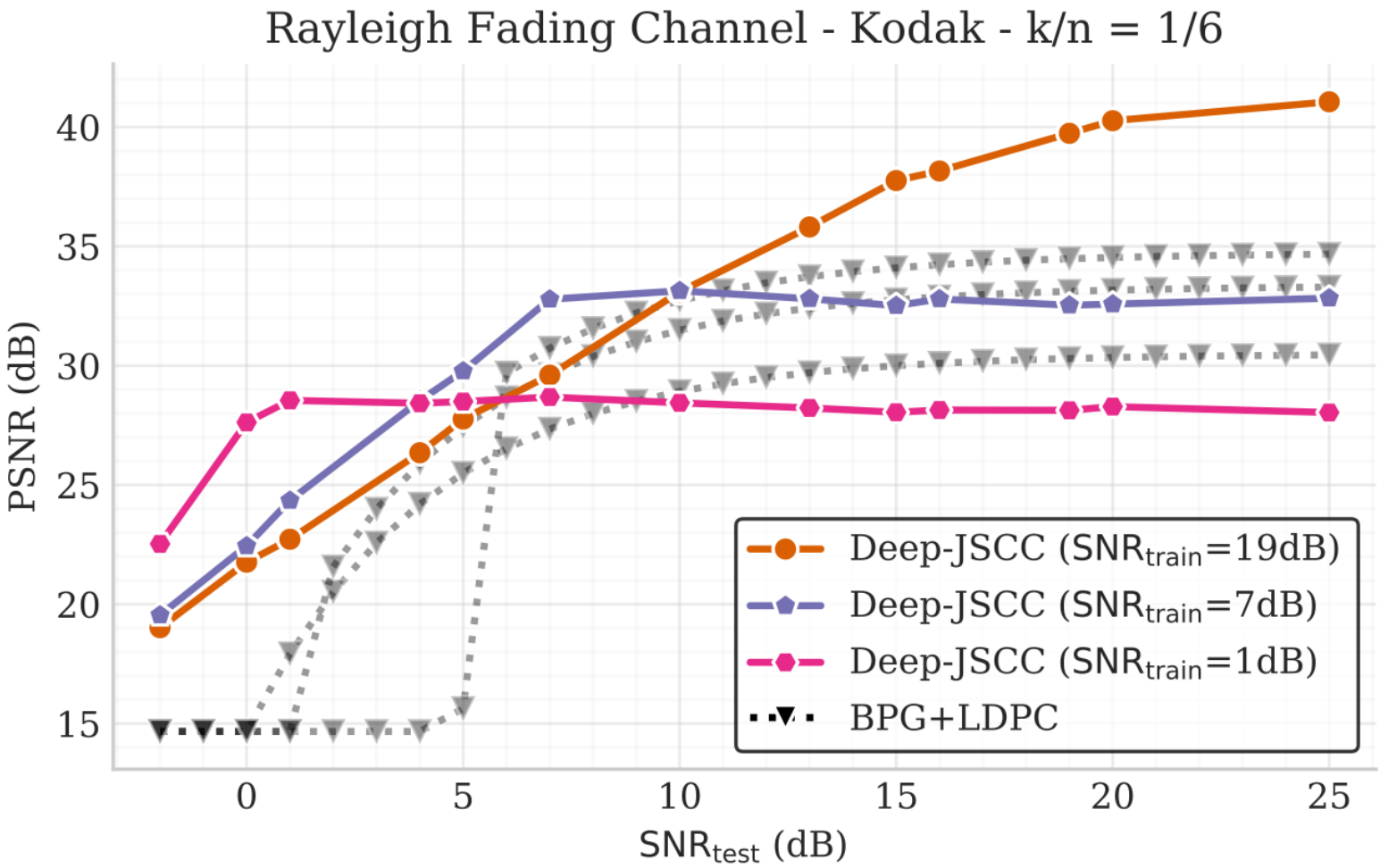
DeepJSCC - PSNR vs. Test SNR



- Provides graceful degradation with channel SNR!
- More like analog communications than digital.

Burth Kurka and Gunduz, Joint source-channel coding of images with (not very) deep learning, Int'l Zurich Seminar (IZS), Feb. 2020.

Deep JSCC over a Rayleigh Fading Channel

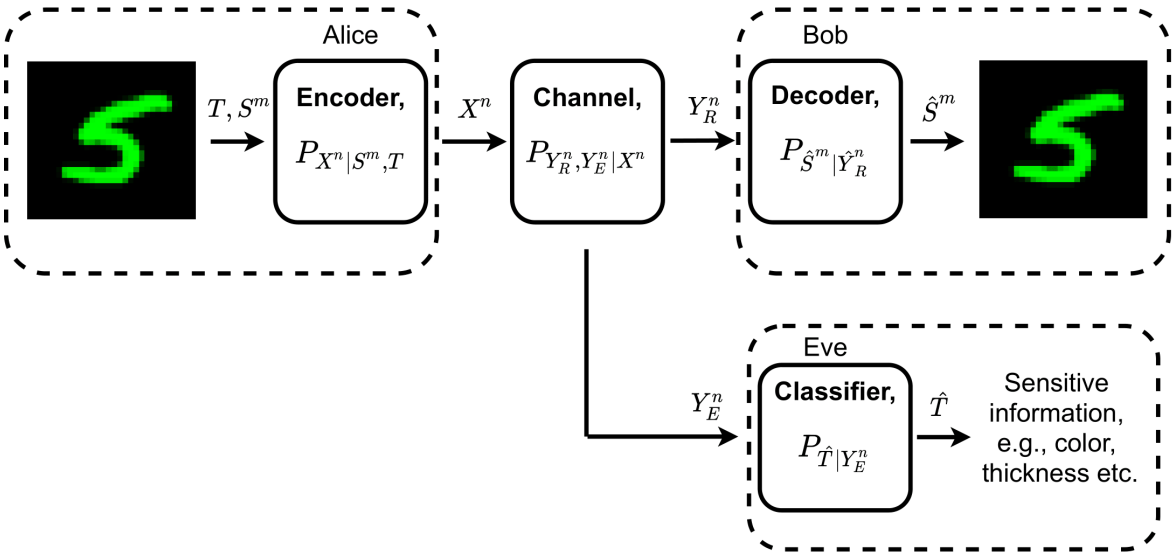


No pilot signal or explicit channel estimation is needed!

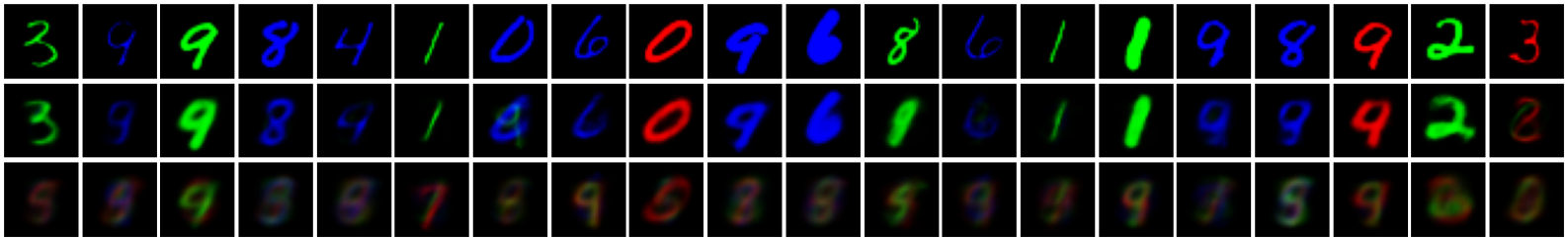
Burth Kurka and Gunduz, Joint source-channel coding of images with (not very) deep learning, Int'l Zurich Seminar (IZS), Feb. 2020.

Sample Images

Original	BPG + LDPC	DeepJSCC
		
	PSNR: 24.096 MS-SSIM: 0.794	PSNR: 25.932 MS-SSIM: 0.831
		
	PSNR: 24.449 MS-SSIM: 0.872	PSNR: 26.095 MS-SSIM: 0.924
		
k/n = 1/24, SNR = 1dB	PSNR: 22.279 MS-SSIM: 0.779	PSNR: 24.408 MS-SSIM: 0.907



- Generative network for stochastic encoding



Erdemir, Dragotti and Gunduz, Privacy-aware communication over the wiretep channel with generative networks, arXiv, Sep. 2021.

Semantic-Aware Text and Speech Communication

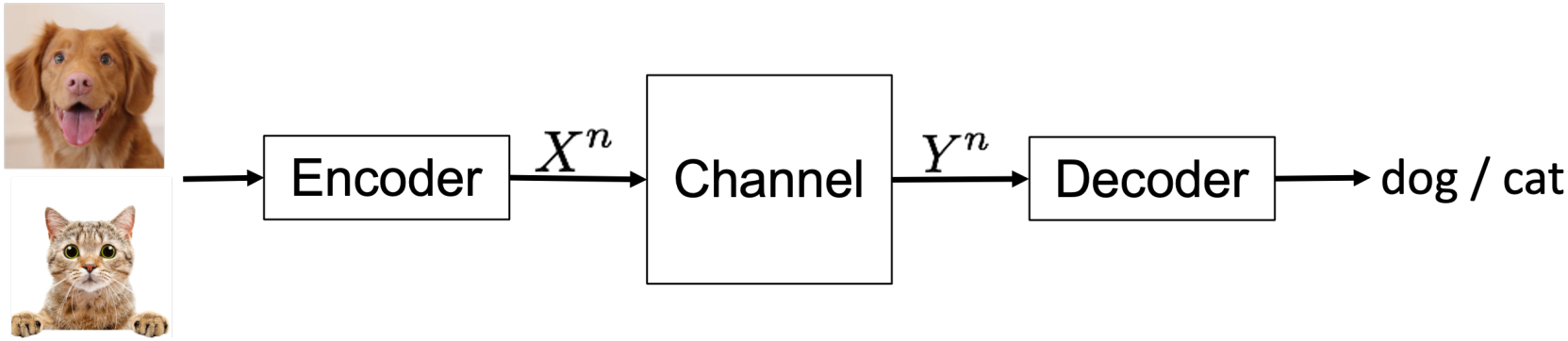
- Text and speech are other complex sources with complex semantic relations
- Originally: Knowledge-based methods
- Recently significant advances in 'statistical methods': deep learning, transformers
- JSCC for text and speech

Farsad, Rao and Goldsmith, Deep learning for joint source-channel coding of text, *IEEE ICASSP,* 2018.

Weng, Qin and Li, Semantic communications for speech signals, arXiv, 2020.

Xie, Qin, Li, Juang, Deep learning enabled semantic communication systems, arXiv, 2020.

Remote Inference over a Noisy Channel



- Machines often do not want to reconstruct the source signal. Instead they want to make inference based on data.
- Is separation theoretically optimal?
- What are the practical approaches?

Null hypothesis $H_0 : U^k \sim \prod_{i=1}^{k} P_U$, Alternate hypothesis $H_1 : U^k \sim \prod_{i=1}^{k} Q_U$.

Acceptance region for H_0: $\mathcal{A}^{(n)} \subseteq \mathcal{Y}^n$

Definition

Type-2 error exponent κ is (τ, ϵ) **achievable** if there exist k, n, such that $n \leq \tau \cdot k$, and

$$\liminf_{k,n\to\infty} -\frac{1}{k} \log\left(Q_{Y^n}(\mathcal{A}^{(n)})\right) \geq \kappa$$

$$\limsup_{k,n\to\infty} -\frac{1}{k} \log\left(1 - P_{Y^n}(\mathcal{A}^{(n)})\right) \leq \epsilon$$

$$\kappa(\tau, \epsilon) \triangleq \sup\{\kappa' : \kappa' \text{ is achievable}\}$$

Hypothesis Testing over a Noisy Channel

Null hypothesis $H_0 : U^k \sim \prod_{i=1}^{k} P_U$, Alternate hypothesis $H_1 : U^k \sim \prod_{i=1}^{k} Q_U$.

$$E_c \triangleq \max_{(x,x')\in\mathcal{X}\times\mathcal{X}} D(P_{Y|X=x}||P_{Y|X=x'})$$

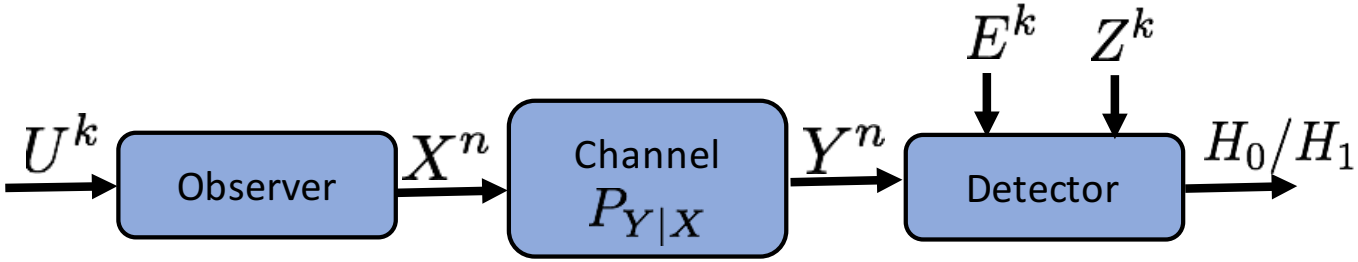
$$\kappa(\tau,\epsilon) = \min\left(D(P_U||Q_U), \tau E_c\right)$$

Making a decision locally at the observer, and communicating it to the detector is optimal.

More interesting when computational complexity constraints are imposed.

Sreekumar and Gunduz, Hypothesis testing over a noisy channel, IEEE ISIT, Jul. 2019.

Jankowski, Gunduz and Mikolajczyk, Joint device-edge inference over wireless links with pruning, IEEE SPAWC, Jul. 2020.



$$H_0 : (U^k, E^K, Z^K) \sim \prod_{i=1}^{k} P_{UEZ}, \qquad H_1 : (U^k, E^K, Z^K) \sim \prod_{i=1}^{k} Q_{UEZ}.$$

- Problem open for general Q
- Let $\kappa(\tau) = \lim_{\epsilon \to 0} \kappa(\tau, \epsilon)$

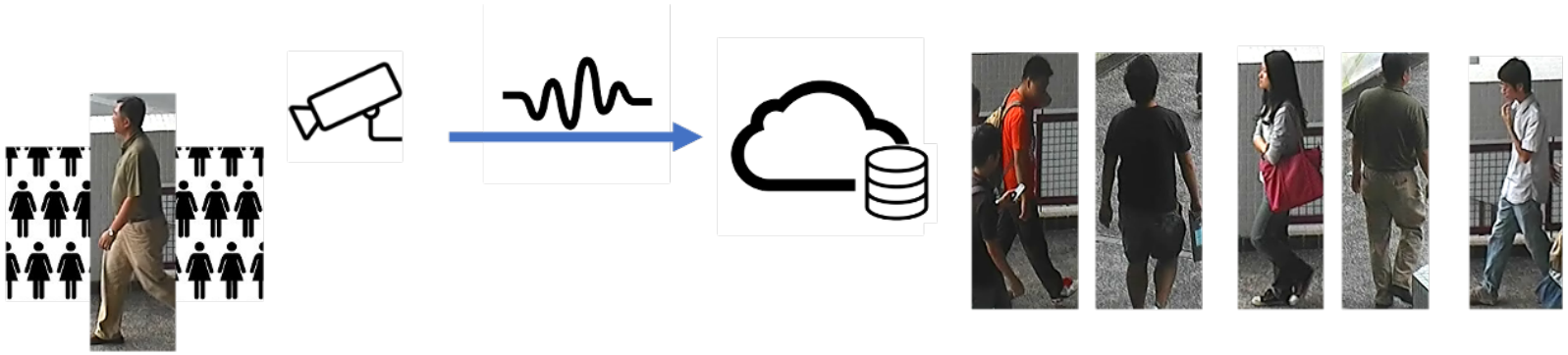
Testing Against Conditional Independence: $Q_{UEZ} = P_{UE} P_{E|Z}$

$$\kappa(\tau) = \sup \left\{ \begin{array}{c} I(E;W|Z) : \exists\ W \text{ s.t. } I(U;W|Z) \leq \tau C(P_{Y|X}), \\ (Z,E) - U - W,\ |\mathcal{W}| \leq |\mathcal{U}| + 1. \end{array} \right\},\ \tau \geq 0.$$

Optimal performance achieved by a separation-based scheme.

S. Sreekumar and D. Gündüz, Distributed hypothesis testing over discrete memoryless channels, IEEE Transactions on Information Theory, Apr. 2020.

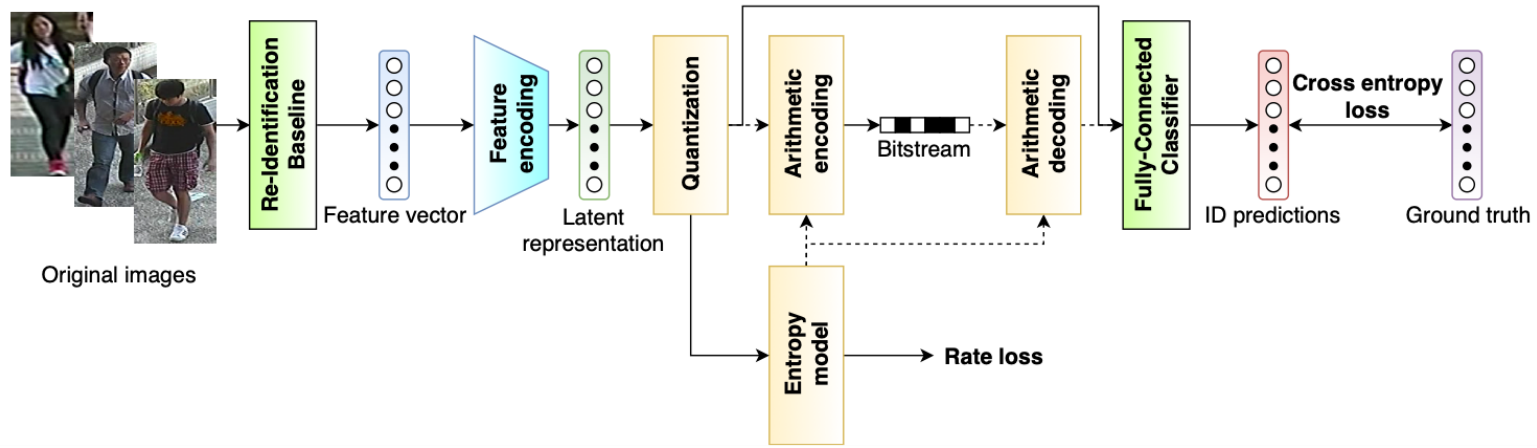
Image Retrieval at the Edge



Goal: match an image from a wireless camera with another image in a large database (person, car, animal, etc.)
Standard approach:

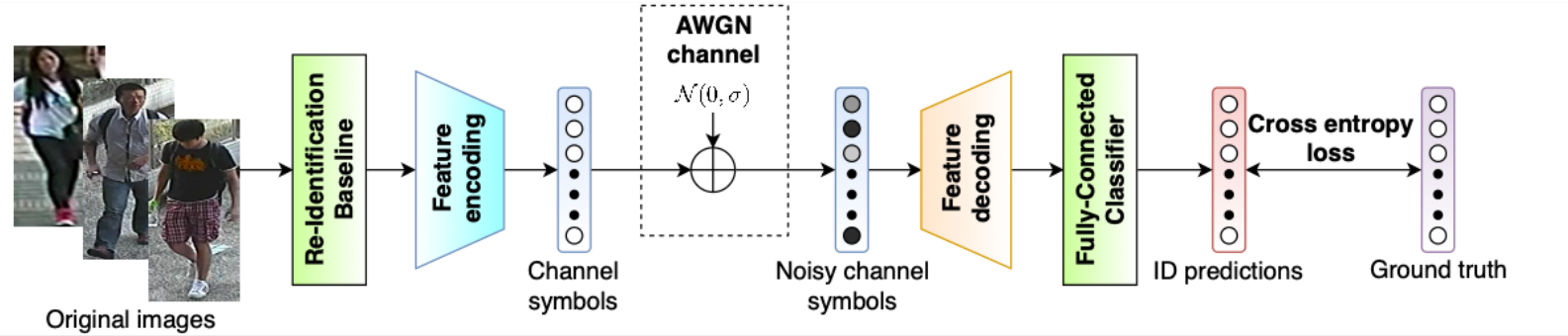
- Transmit images to the cloud
- Apply best retrieval algorithm in the cloud
- Determine features most relevant for re-identification over image database
- State-of-the-art: Deep neural nets (e.g., ResNet-50)

Retrieval-Oriented Compression



- Transmit only features relevant for retrieval
- Apply entropy coding to quantized features: **model-based machine learning**

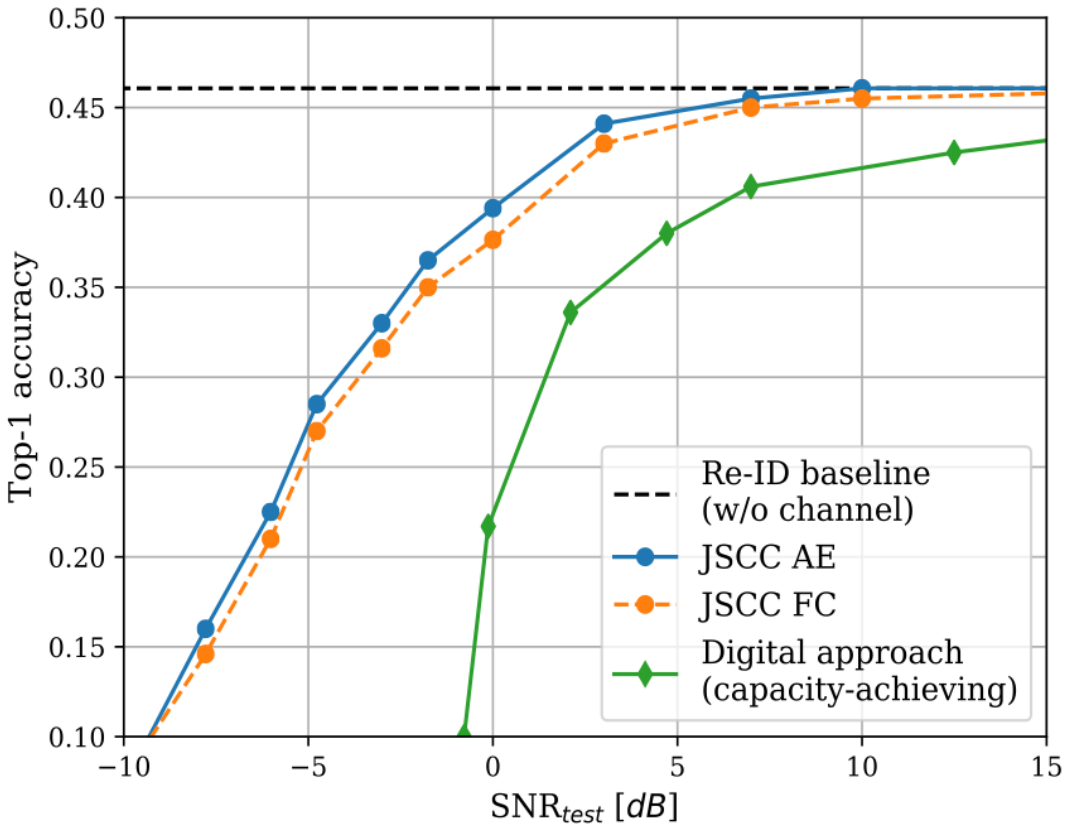
M. Jankowski, D. Gunduz and K. Mikolajczyk, Wireless image retrieval at the edge, IEEE Journal on Selected Areas in Communications (JSAC), Jan. 2021.



- Feature transmission is a joint source-channel coding problem
- Separation is suboptimal in general
- Joint approach provides better performance as well as 'graceful degradation'

M. Jankowski, D. Gunduz and K. Mikolajczyk, Wireless image retrieval at the edge, IEEE Journal on Selected Areas in Communications (JSAC), Jan. 2021.

Person Re-identification Over Noisy Channels



- CUHK03 dataset: 14096 images of 1467 identities taken from two camera views
- 256x128 coloured images

M. Jankowski, D. Gunduz and K. Mikolajczyk, Wireless image retrieval at the edge, IEEE Journal on Selected Areas in Communications (JSAC), Jan. 2021.

AirNet: Neural Networks in the Air

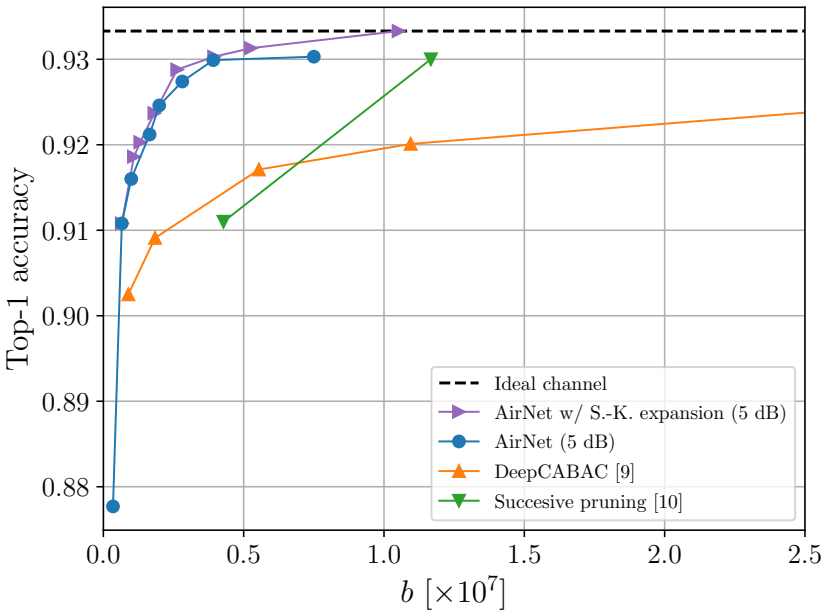
- Neural networks (NNs) may need to be transmitted, or stored over lossy 'analog' media (neuromorphic hardware)
- Classification will be done with the noisy NN
- **Conventional approach**: Compress NN weights, use channel coding against errors
- **Proposed approach**: Pruning (for bandwidth reduction) + noise injection during training + knowledge distillation

---

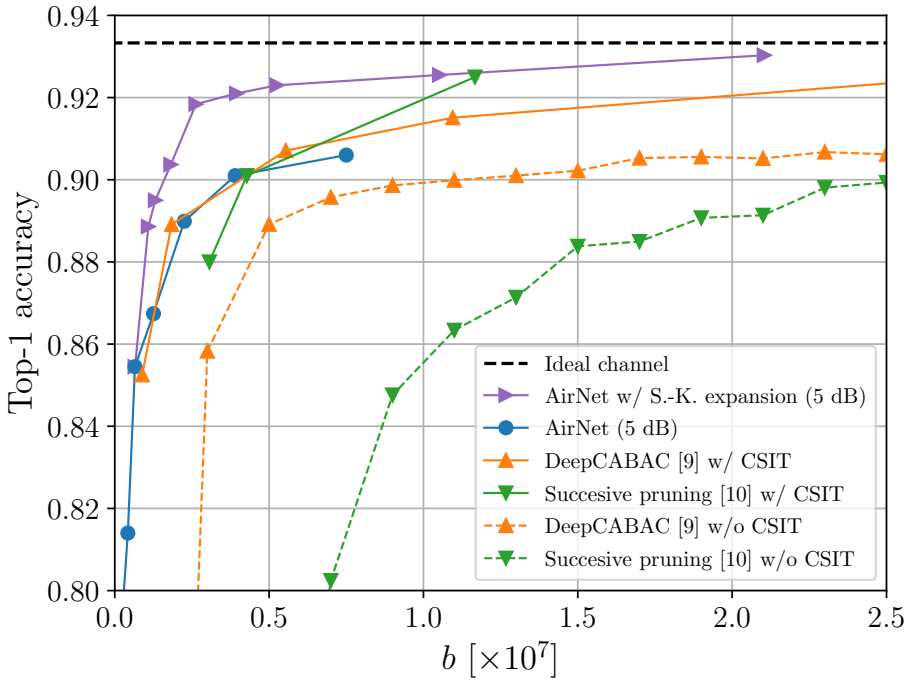
Jankowski, Gündüz, and Mikolajczyk, AirNet: Neural network transmission over the air, arXiv, 2021.

Shao, Liew and Gündüz, Denoising noisy neural networks: A Bayesian approach with compensation, arXiv, 2021.

AirNet over AWGN Channel



AWGN, SNR = 5 dB



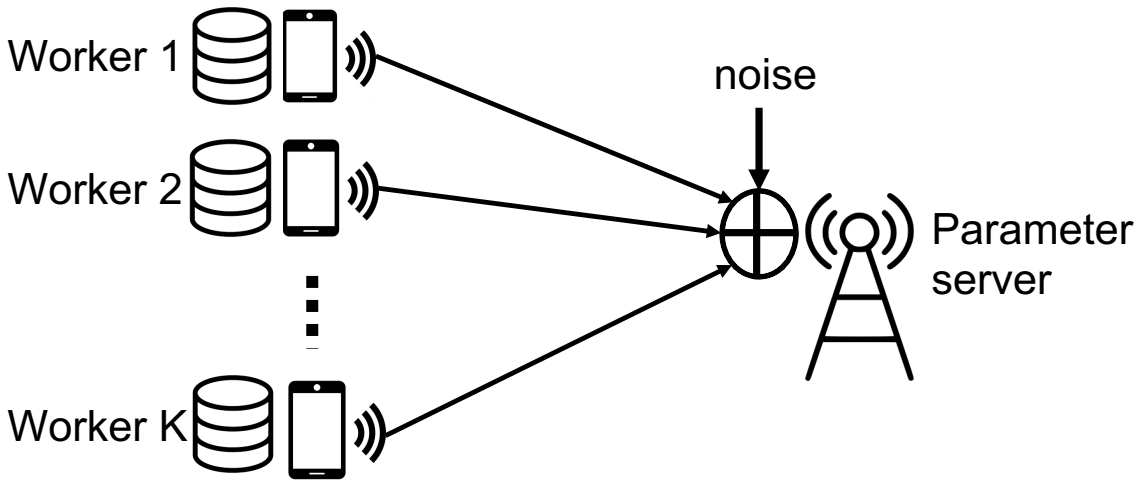
Fading channel, Average SNR = 5 dB

- Small-VGG16 for CIFAR-10 classification
- Observation: Better to prune more, and then introduce redundancy through SK mapping

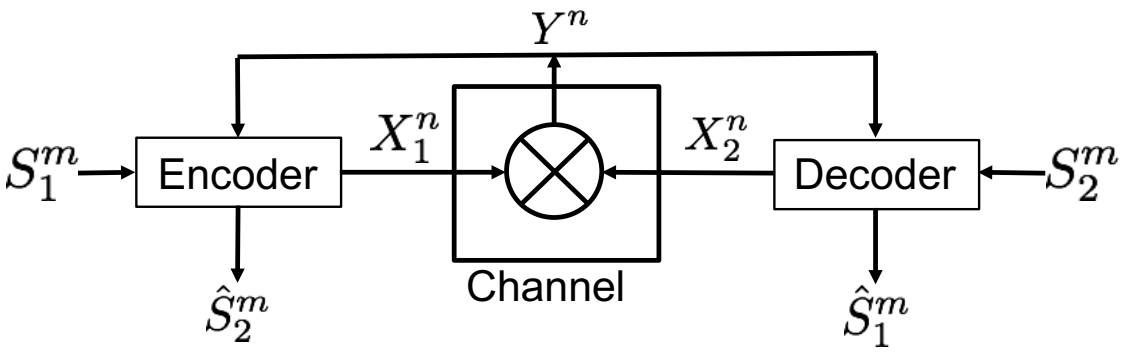
---

Jankowski, Gündüz, and Mikolajczyk, AirNet: Neural network transmission over the air, arXiv, 2021.

- Communication is bottleneck in distributed learning
- ML literature focuses on reducing the number and size of gradient informaton transmitted from each worker
- Underlying channel ignored
- In edge learning, wireless channel is limited in bandwidth and may suffer from interference



- Separation does not hold for multi-user channels



- Binary two-way multiplying channel: $X_i \in \{0,1\}$, $i = 1, 2$

$$Y = X_1 \cdot X_2$$

- Capacity still open: Shannon provided inner/ outer bounds
- Consider correlated signals S_1 and S_2:

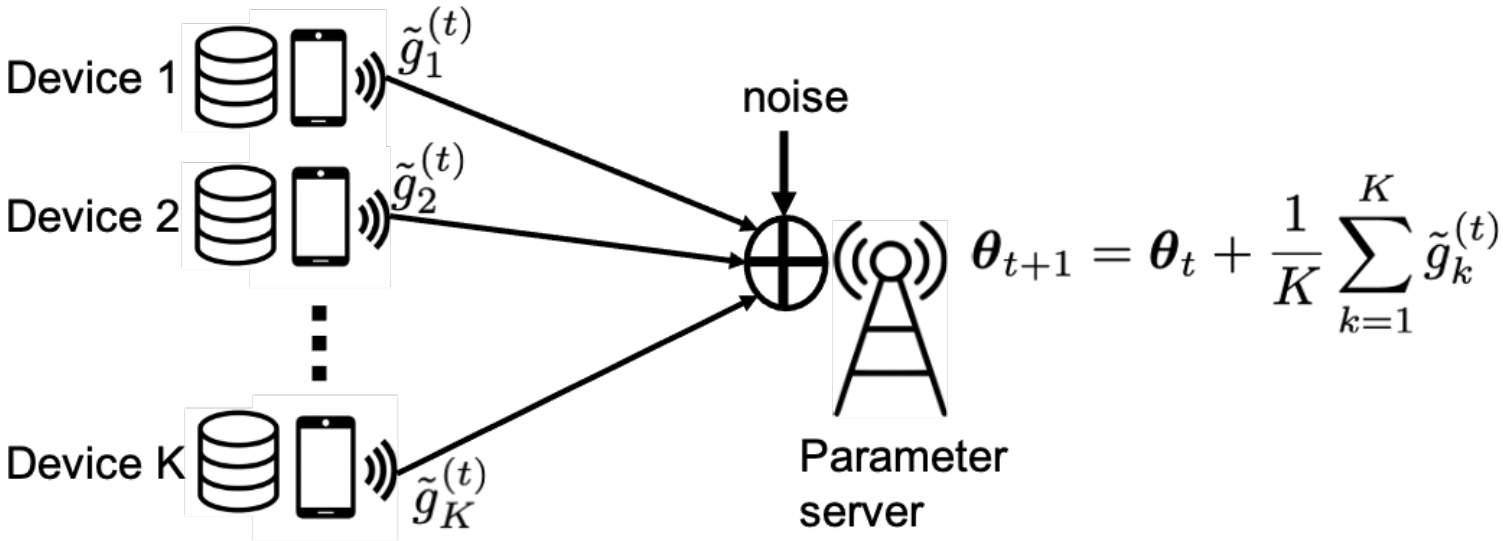
	0	**1**
0	0	0.275
1	0.275	0.45

- With separation, they need to exchange rates

$$H(S_1|S_2) = H(S_2|S_1) = 0.6942 \text{ bpss}$$

C. E. Shannon, **Two-way communication channels**, *in Proc. 4th Berkeley Symp. Math. Satist. Probability,* vol. 1, 1961, pp. 611-644.

FEEL as a Distributed Computation Problem

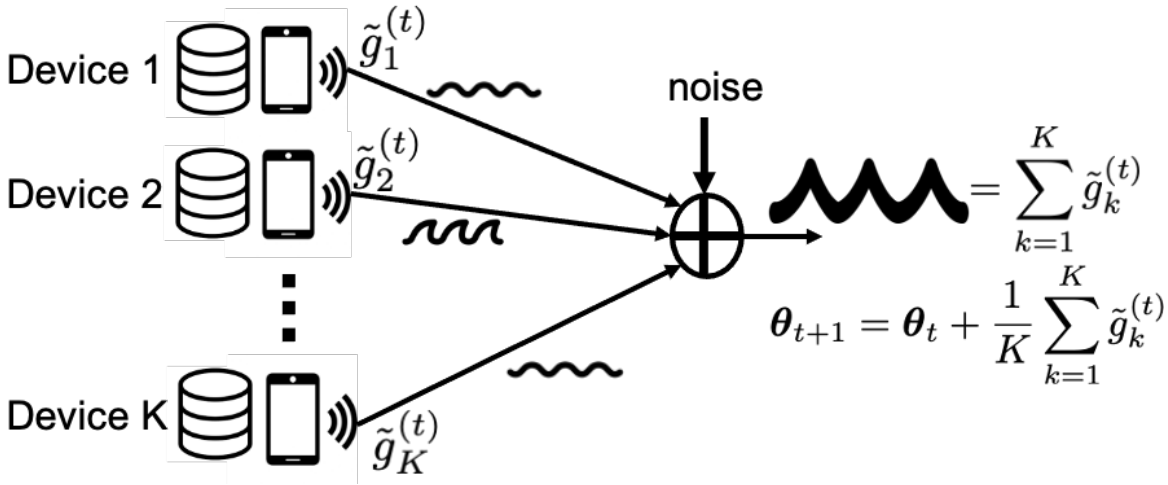


- A distributed joint source-channel coding problem
- Goal: Compute the average of the sources
- Proposed approach: Simultaneously transmit gradients in an uncoded fashion: over-the-air computation (OAC)

Goldenbaum, Boche and Stańczak, Nomographic functions: Efficient computation in clustered Gaussian sensor networks, IEEE Trans. on Wireless Comms., April 2015.

Amiri and Gunduz, Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air, *IEEE Trans. Signal Proc.*, 2020.

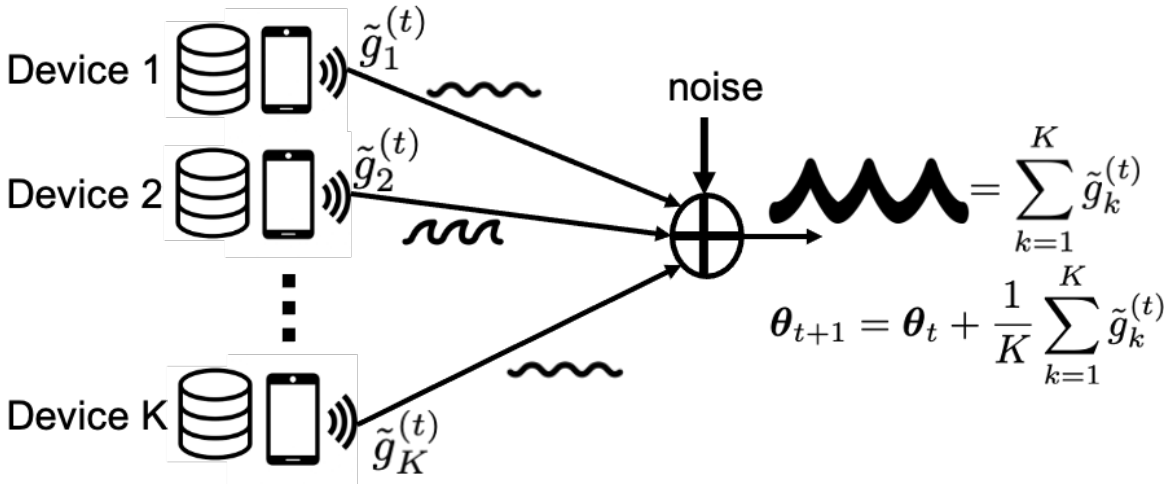
FEEL with Over-the-Air Computation (OAC)



- Exploit interference rather than avoiding it
- Benefit from the signal superposition property of wireless channels by over-the-air computation
- Turns air into a computation medium
- Challenges:
  - Channel variations can lead to bias
  - Gradient dimension can be very large: VGG Net ∼140 million, ResNet ∼26 million parameters
  - Sending each weight individually introduces significant delay

Amiri and Gunduz, Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air, *IEEE Trans. Signal Proc.*, 2020.

FEEL with Over-the-Air Computation (OAC)

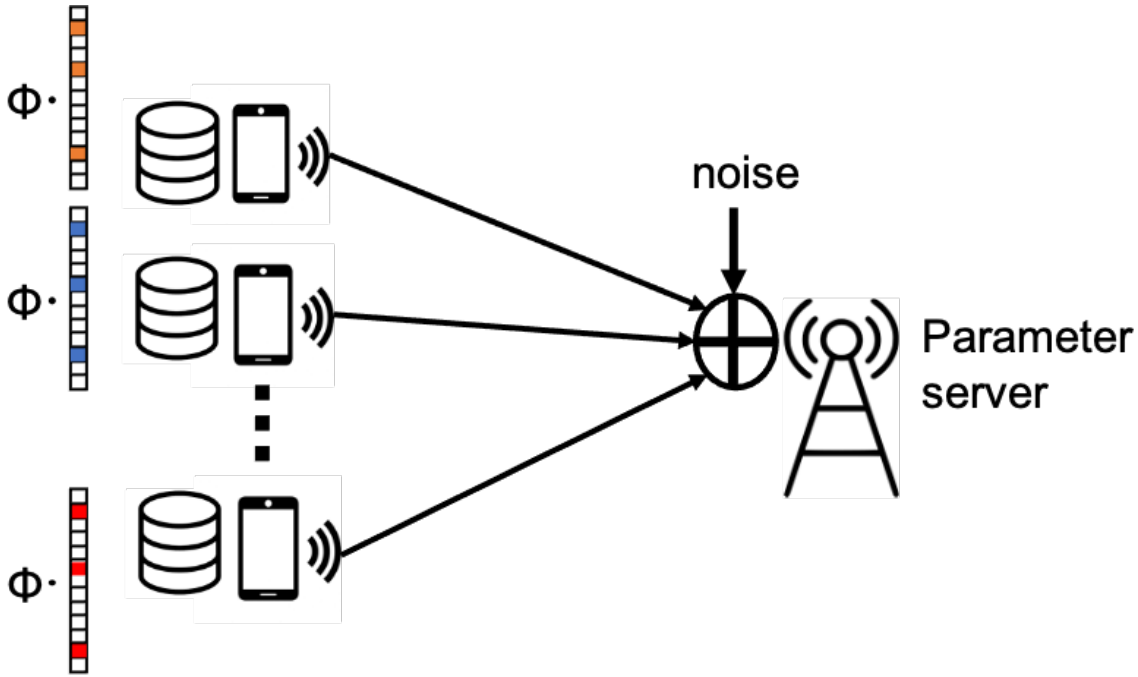


- Exploit interference rather than avoiding it
- Benefit from the signal superposition property of wireless channels by over-the-air computation
- Turns air into a computation medium
- Challenges:
  - Channel variations can lead to bias
  - Gradient dimension can be very large: VGG Net ∼140 million, ResNet ∼26 million parameters
  - Sending each weight individually introduces significant delay

---

Amiri and Gunduz, Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air, *IEEE Trans. Signal Proc.*, 2020.

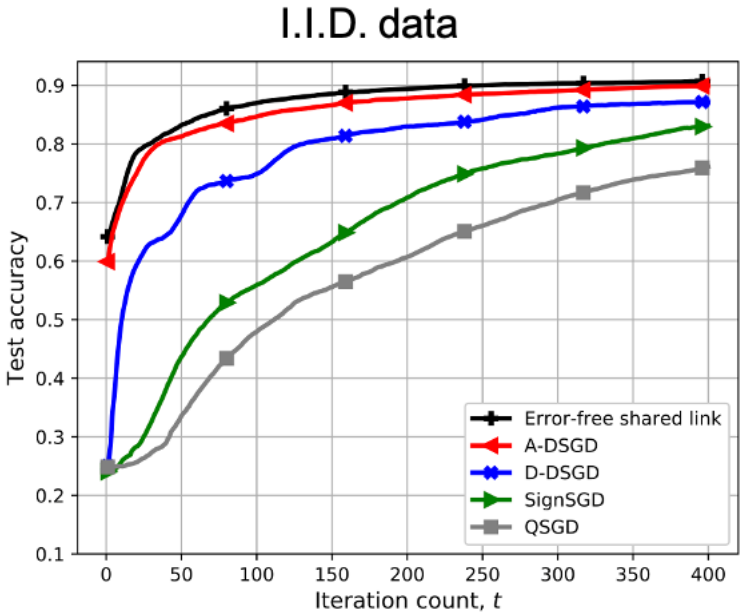
FEEL with Sparsified Over-the-Air Computation (OAC)

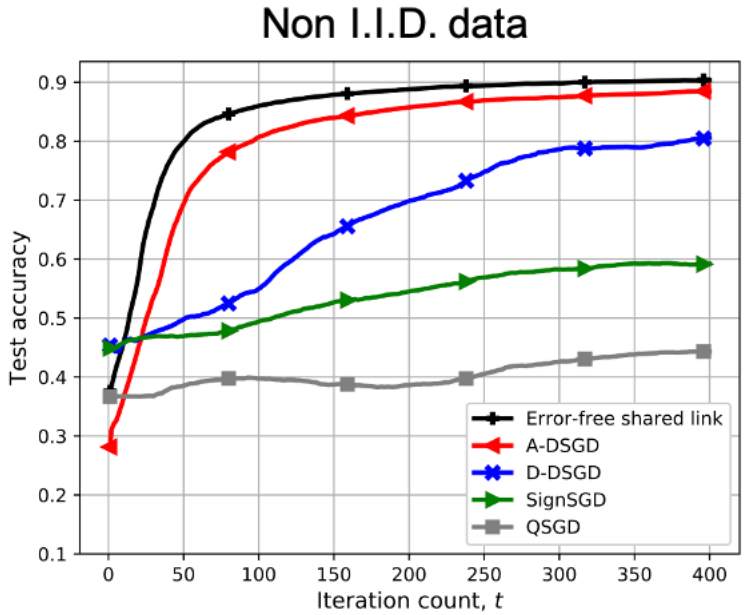


- Thresholding to sparsify gradient estimates
- Use pseudo-random linear projection to reduce bandwidth
- Approximate message passing (AMP) based decoding at parameter server
- No need to send indices of sparse elements

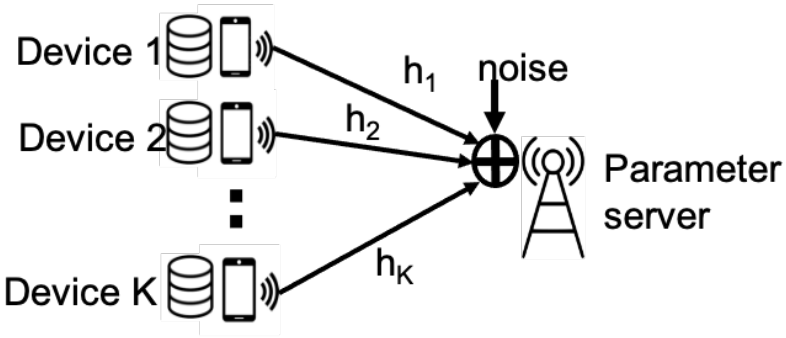
Amiri and Gunduz, Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air, *IEEE Trans. Signal Proc.*, 2020.

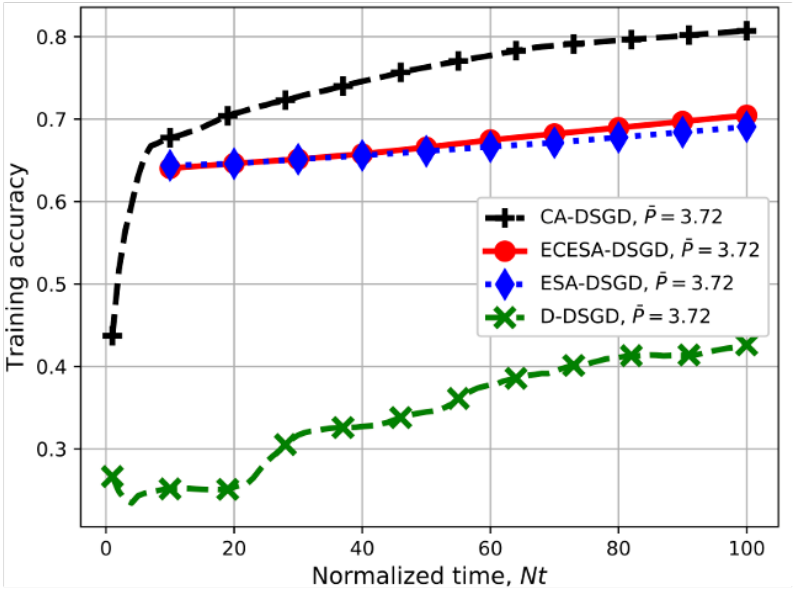
- Distributed MNIST classification (single layer with 10 neurons, ADAM optimizer)
- Parameter vector size $d = 28 \times 28 \times 10 + 10 = 7850$
- $P_1 = 127$, $P_2 = 422$
- Bandwidth: d/2 symbols,
- Sparsity level: d/2





Channel gains known: requires channel inversion



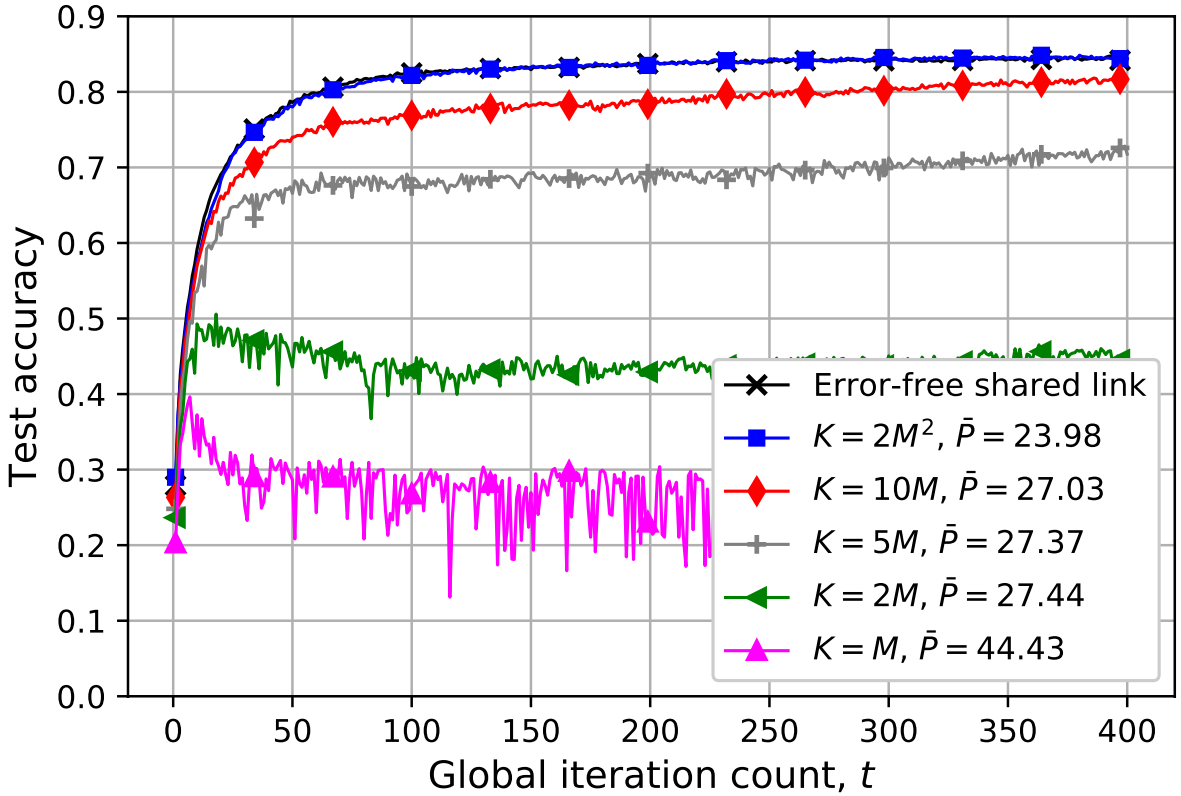


Amiri and Gunduz, Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air, *IEEE Trans. Signal Proc.*, 2020.

Zhu, Y. Wang, and K. Huang, Broadband analog aggregation for low-latency federated edge learning, IEEE Trans. Wireless Commun., 2020.

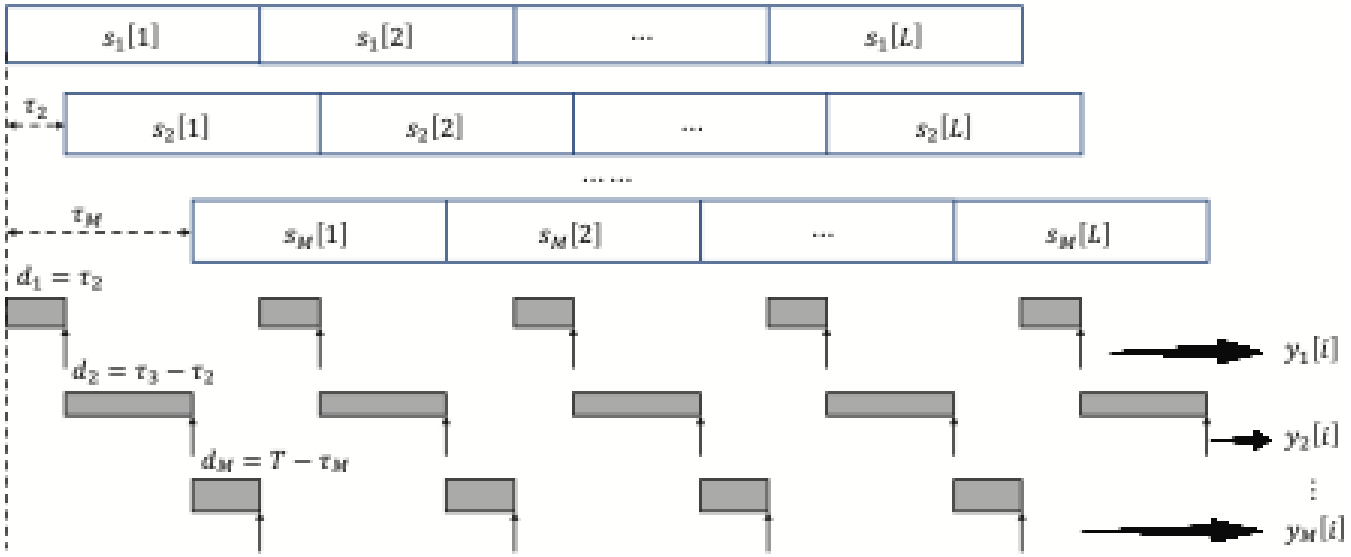
Yang *et al.*, Federated learning via over- the-air computation, IEEE Trans. Wireless Commun., 2020.

- Channel gains unknown at the transmitters
- Employ multiple antennas at the parameter server to resolve uncertainty
- $M = 20$ devices



Amiri, Duman, Gunduz, Kulkarni and Poor, Blind federated edge learning, *IEEE Transactions on Wireless Communications*, Aug. 2021.

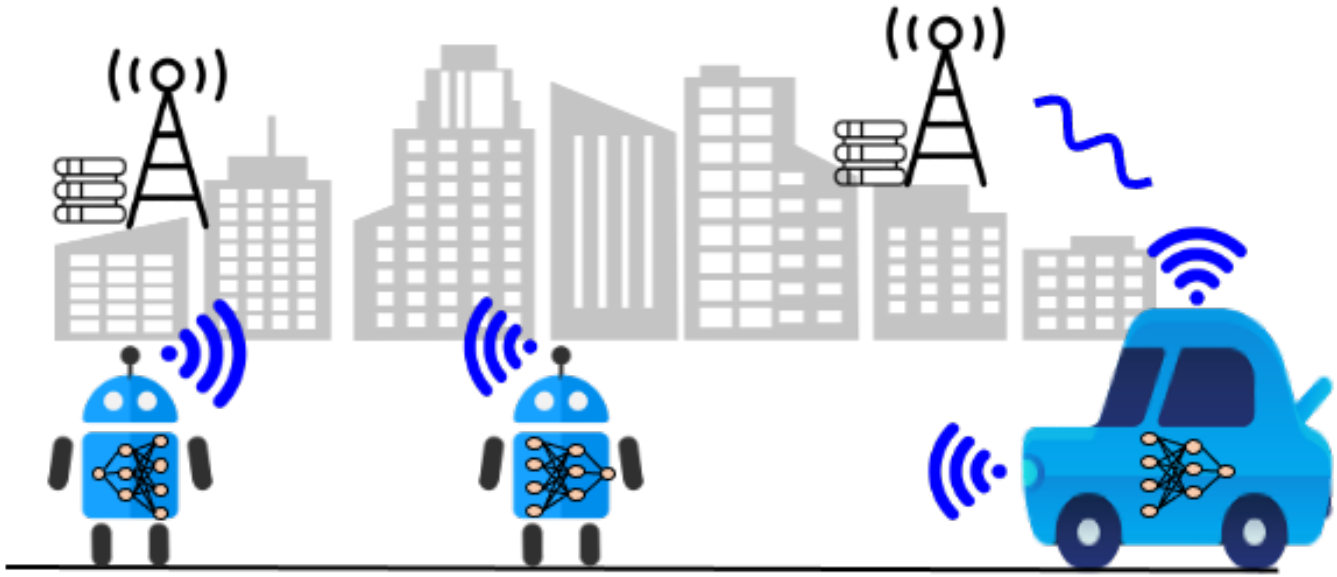
FEEL with Misaligned Over-the-Air Computation



- Match-filter received signal with M filters of delay
- Estimate sum through an analog message-passing procedure: SP-ML estimator
- Boils down to estimating individual samples. Suffers from error propagation.
- Alternative: aligned sample estimator (use only the M-th sample)

Shao, Gunduz, and Liew, **Federated edge learning with misaligned over-the-air computation**, *IEEE Trans. on Wireless Comms.*, to appear..

Effective Communications

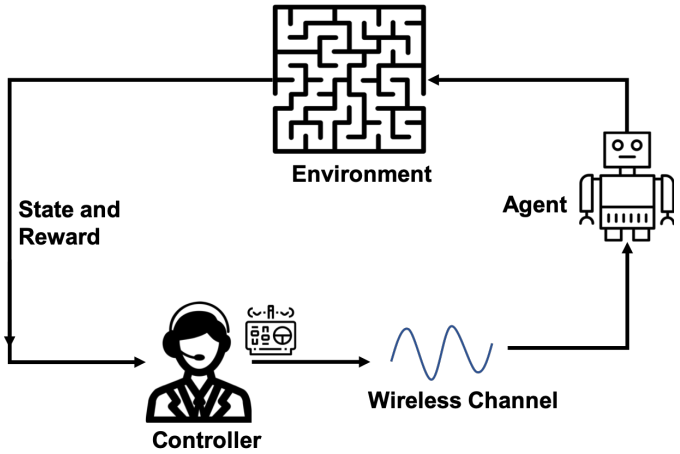


- Level C: Effectiveness Problem (Weaver and Shannon):

How effectively does the received meaning affect conduct in the desired way?

- Communication to enable achieving a goal

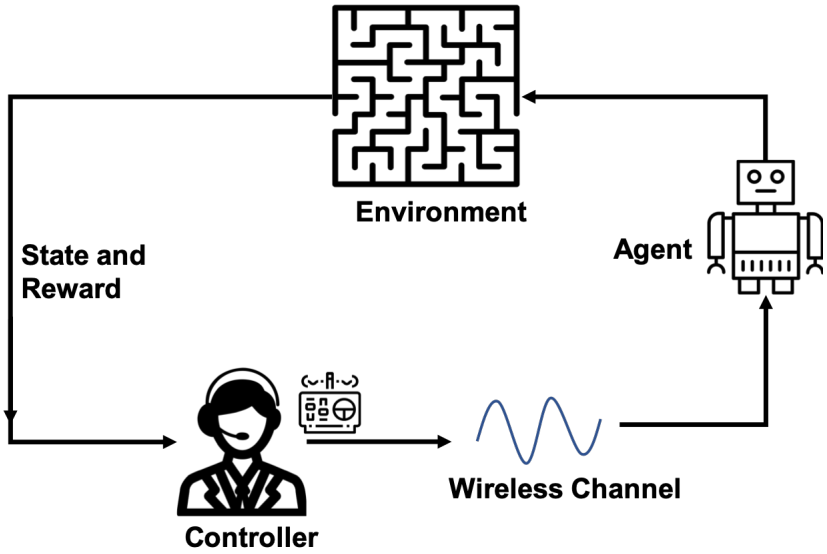
Remote-Controlled Markov Decision Process (MDP)

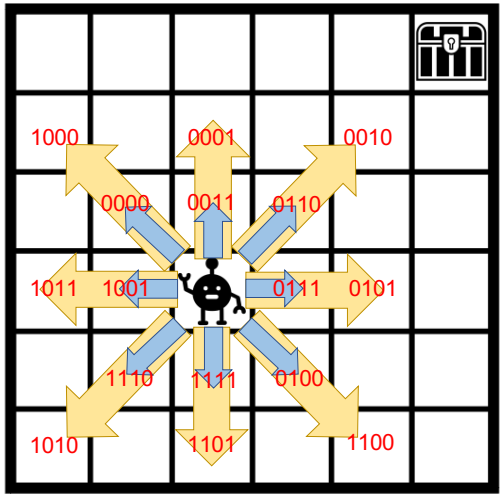


- Create a two-agent version of an MDP
- Controller observes the state and reward, but agent takes the action
- A noisy communication channel in between
- Agent can depend solely on the received signal, or may have some limited observation of the system state

Tung, Kobus, Roig Pujol, and Gunduz, "Effective Communications: A Joint Learning and Communication Framework for Multi-Agent Reinforcement Learning over Noisy Channels," IEEE JSAC, 2021.

Example: Grid World

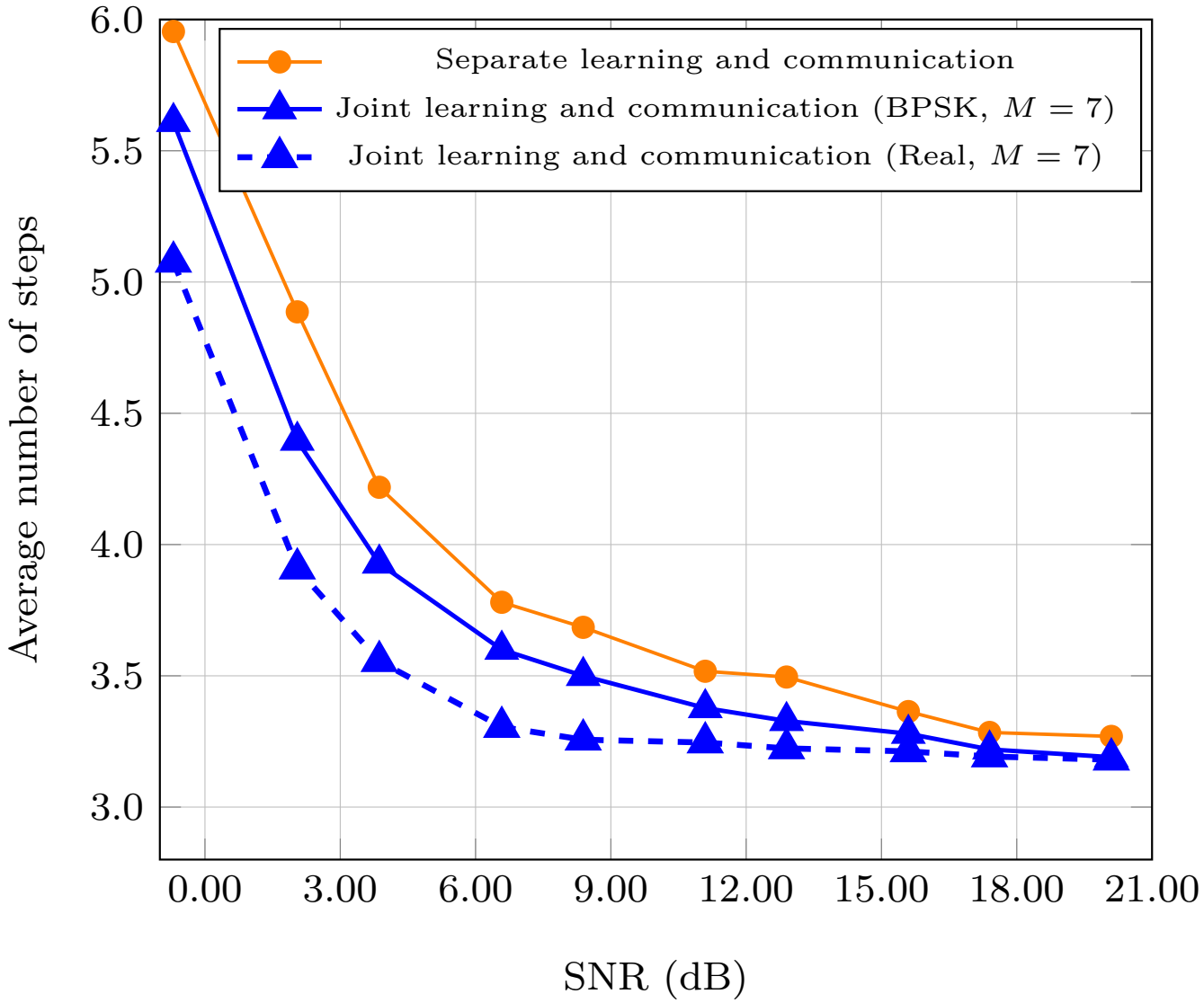




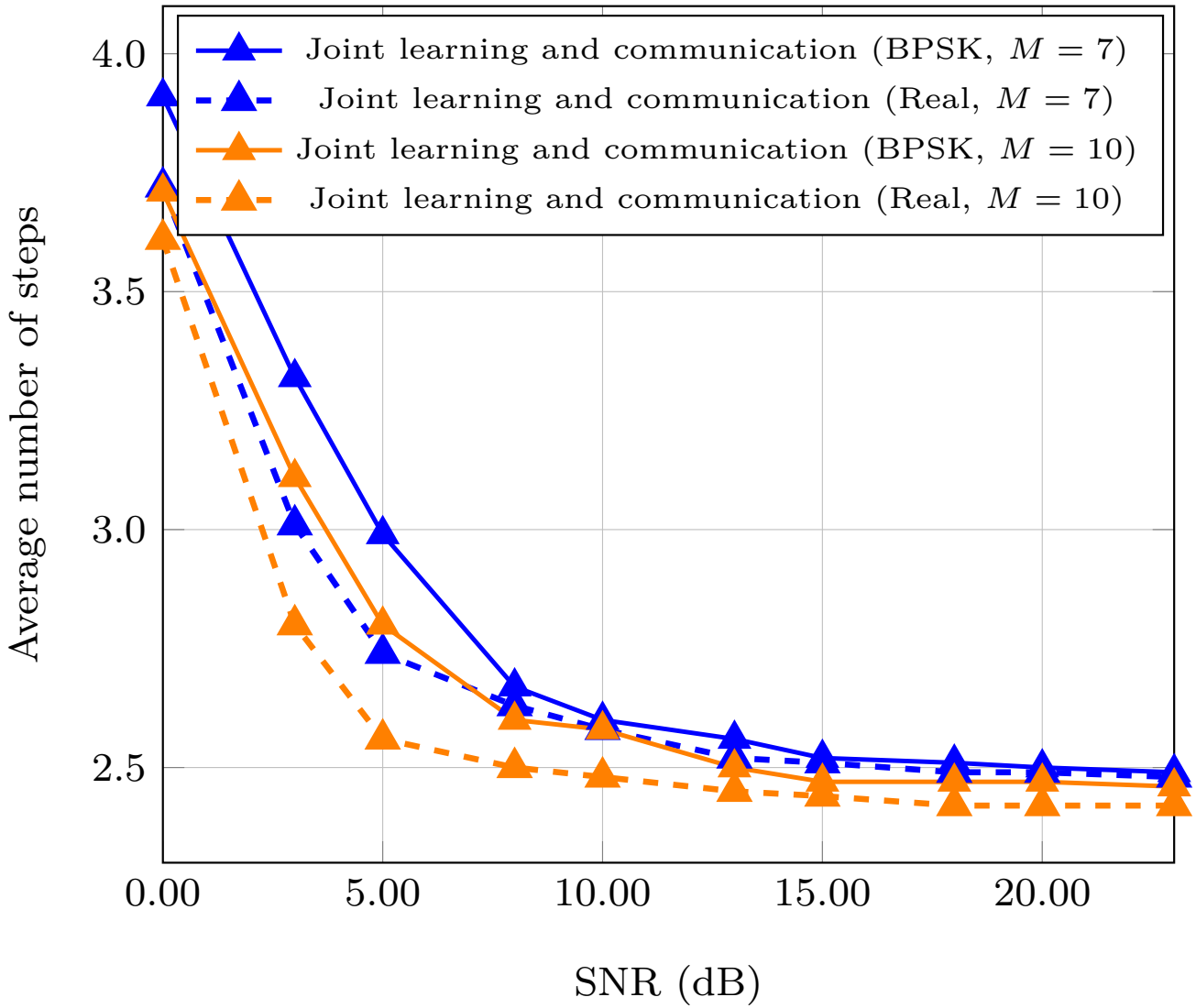
- $L \times L$ grid world
- Agent can take 16 actions (1 or 2 steps in every direction)
- Arrives at a random neighbouring cell w.p. δ
- Find treasure at a random location as fast as possible
- Channels:
  - Binary symmetric channel (BSC) with cross-over probability p_e
  - Binary input AWGN channel
  - AWGN channel with average power constraint

Tung, Kobus, Roig Pujol, and Gunduz, Effective Communications: A Joint Learning and Communication Framework for Multi-Agent Reinforcement Learning over Noisy Channels, IEEE JSAC, 2021.

AWGN Channel ($\delta = 0.05$)



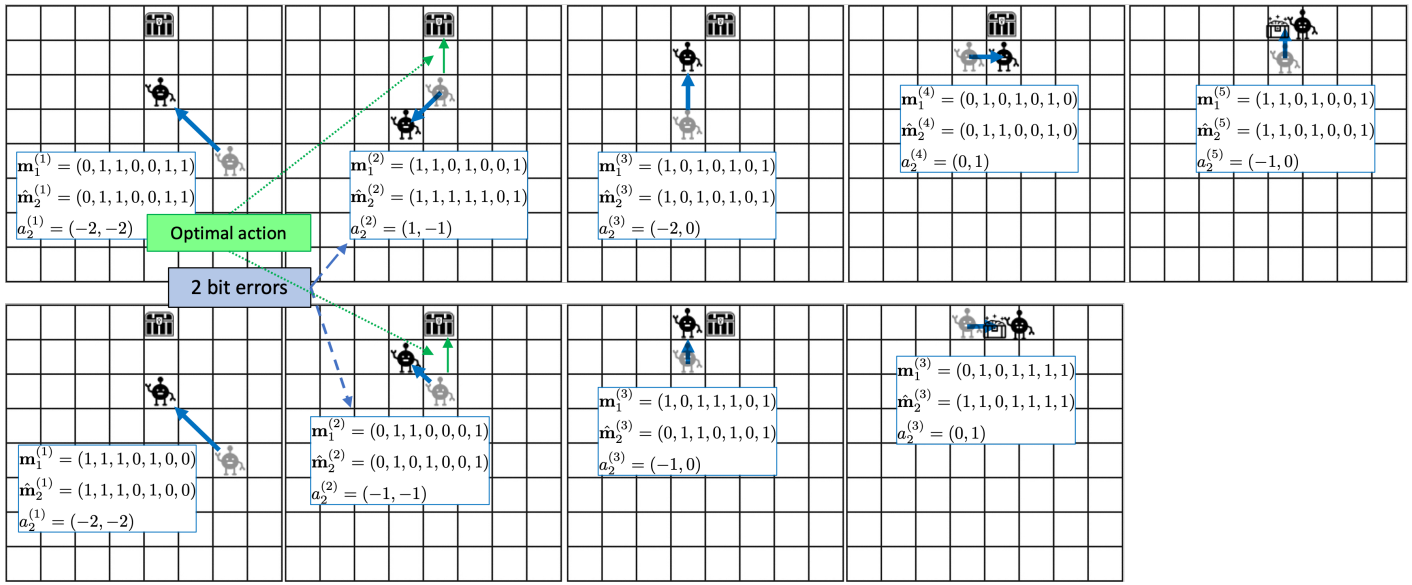
Effect of Bandwidth (AWGN channel , $\delta = 0$)



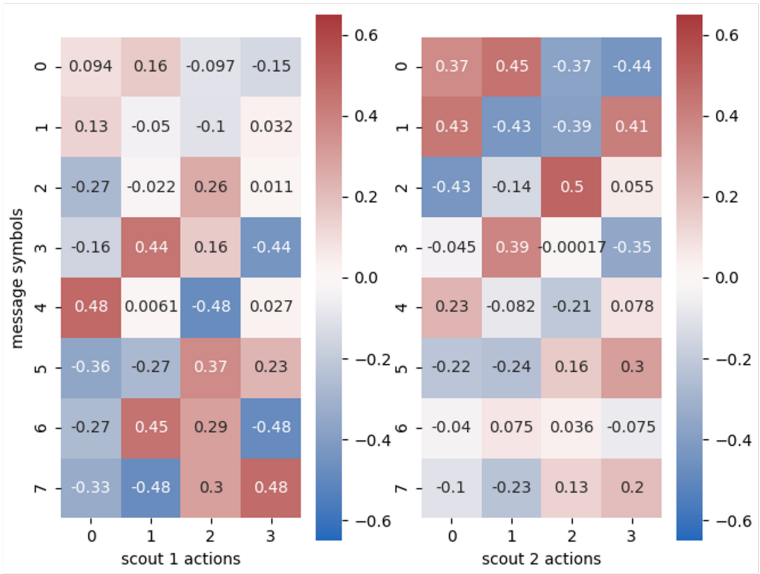
Example: Separate vs. Learned Communications

Separation

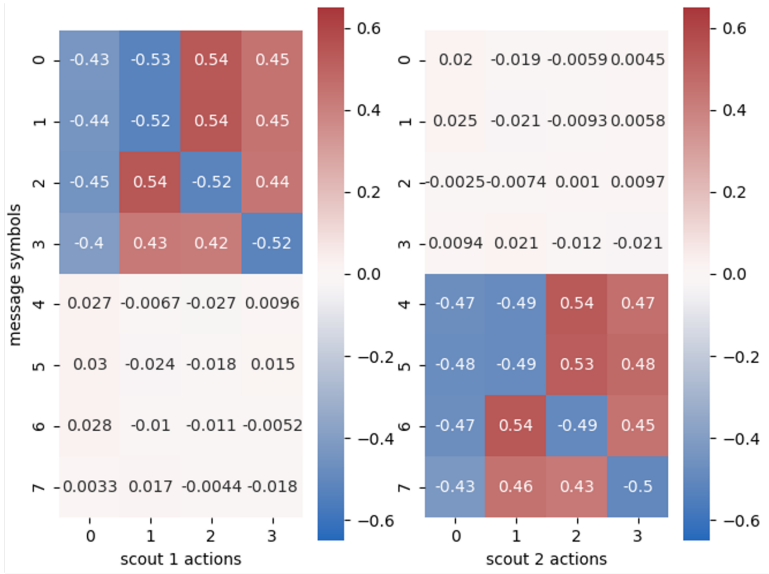
Learnt



Tung, Kobus, Roig Pujol, and Gunduz, "Effective Communications: A Joint Learning and Communication Framework for Multi-Agent Reinforcement Learning over Noisy Channels," IEEE JSAC, 2021.

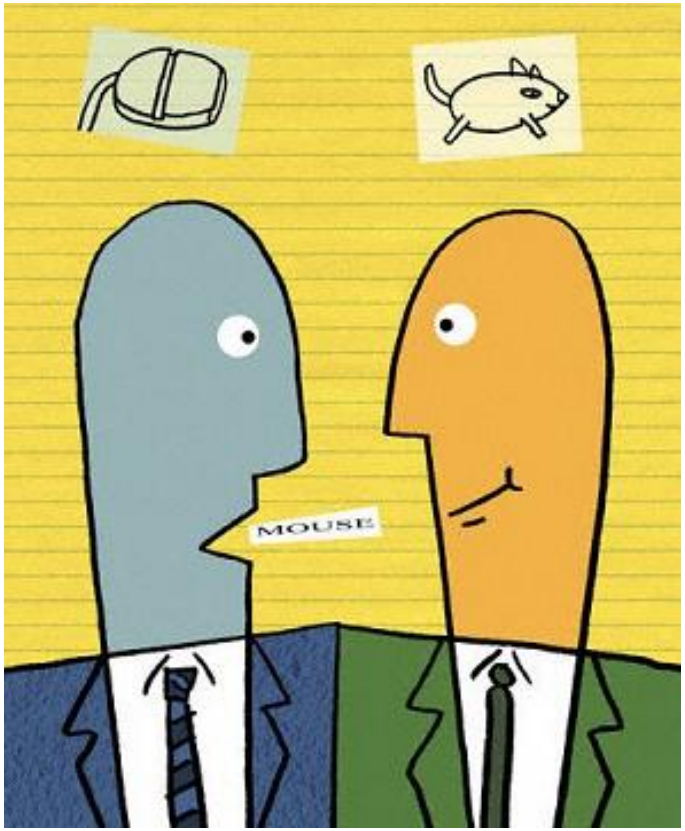


Learned Codewords



Expert Codewords

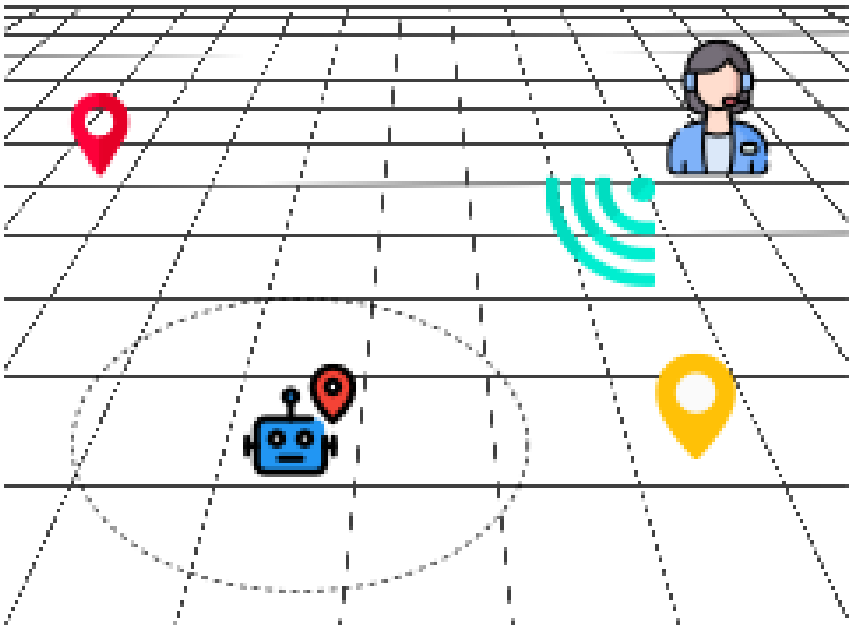
Context-Aware Communications



- Context is important
- Do transmitter and receiver share the same context?
- Various ways to model context: side information, mismatched distortion measures, etc.

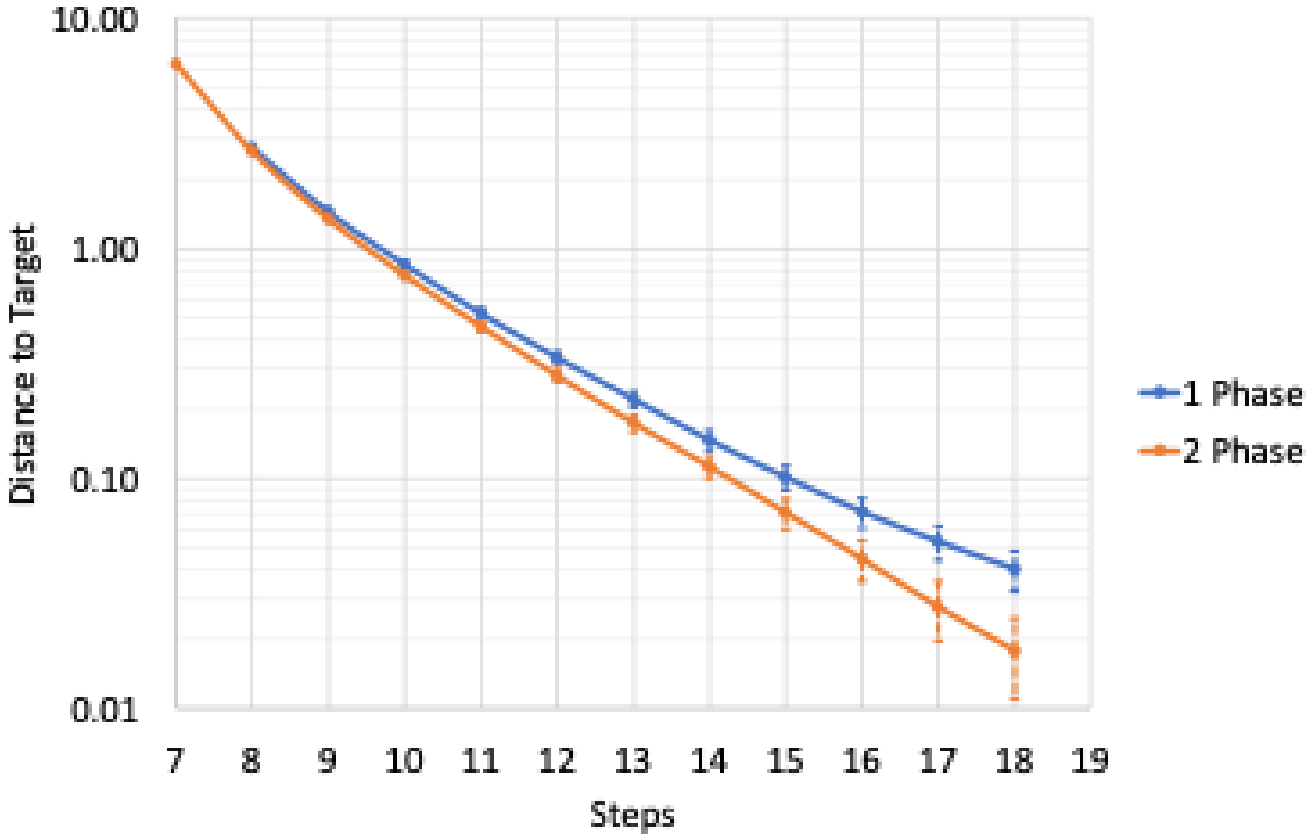
Juba and Sudan, "Universal Semantic Communication," STOC, 2008.

Context-Aware Effective Communications



- Consider the effectiveness problem, e.g., remote-controlled MDP
- Assume agent also knows its state, but not the target; receives a signal when it is within a certain distance from the target
- What should we transmit? Target location or the action?

Tung, Kobus and Gunduz, "Context-Aware Effective Communications," Asilomar, 2021.



- We need both: Context-dependent communications!
- Start by transmitting target location (Phase 1), and switch to transmitting action when near the target.

Tung, Kobus and Gunduz, "Context-Aware Effective Communications," Asilomar, 2021.

Conclusions

- Semantic/ goal-oriented communications is an emerging area of research for 6G and beyond communication networks
- Often used "Beyond Shannon" argument is questionable. On the contrary, Shannon theory, in particular JSCC, provides a framework to study these problems, but we need to go beyond memoryless source/channel and additive distortion measure assumptions
- Practical code design and implementation mostly open
- Machine learning can help us design practical JSCC schemes that can beat state-of-the-art
- JSCC is essential for modern communication systems with extremely low latency and low power requirements, particularly for distributed inference/ training applications
- Proposed a new approach to effectiveness problem, which combines reinforcement learning with communications.

Forthcoming:

IEEE Journal on Selected Areas in Communications (JSAC), Special Issue on Beyond Transmitting Bits: Context, Semantics and Task-Oriented Communications

Forthcoming:

Y. Eldar, A. Goldsmith, D. Gündüz and H. V. Poor, *Machine Learning and Wireless Communications*, Cambridge University Press.