Machine learning for and with Wireless Communications

David Burth Kurka

Information Processing and Communications Lab
Department of Electrical and Electronic Engineering
Imperial College London, London UK

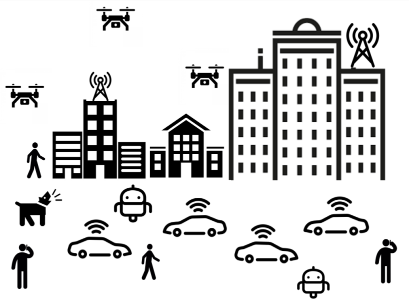
d.kurka@imperial.ac.uk

Imperial College London

Information Processing and Communications Laboratory

- Part of Intelligent Systems and Networks Group: dedicated to machine learning for robotics, networking, machine vision, multi-agents systems, neural processing, etc.

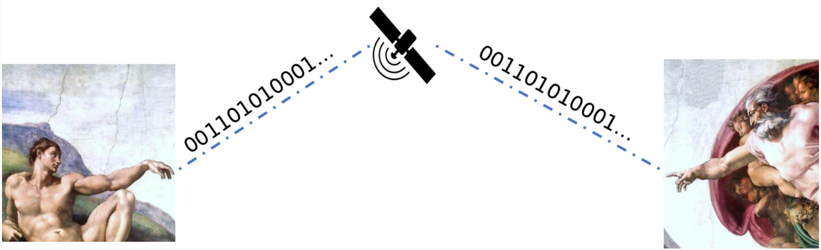
Future Autonomous Systems



- Intelligence is the key for future autonomous systems …
- …and so is communication
- Our goal is to push the boundaries of **communications** using **intelligence**, and to develop better intelligence with more efficient **communications**.

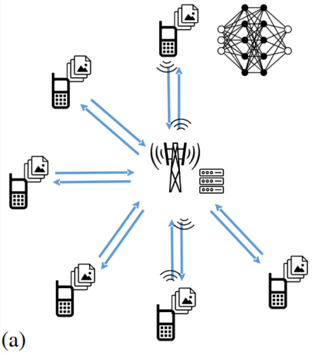
Tactile Internet

- *"Internet network that combines ultra low latency with extremely high availability, reliability and security"* (ITU)
- Next generation Internet of Things: human-machine and machine-machine interaction: haptic interaction with visual feedback
- Augmented reality (AR), virtual reality (VR), automation, robotics, remote education, telepresence, ...
- 1ms round trip delay?



IPC-Lab - Selected Projects

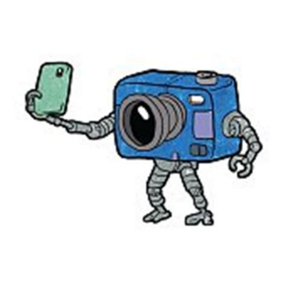
- **Deep Joint Source-Channel Coding**
- Distributed Computation
- Federated Learning in the Wireless Edge
- DL for massive MIMO training
- Adversarial ML for Secure Communications
- Hypothesis Testing Over Noisy Channels
- Privacy Aware Learning
- Deep RL for Wireless Resource Optimization
- Smart Meter Privacy
- Proactive Content Delivery



Deep Joint Source Channel Coding

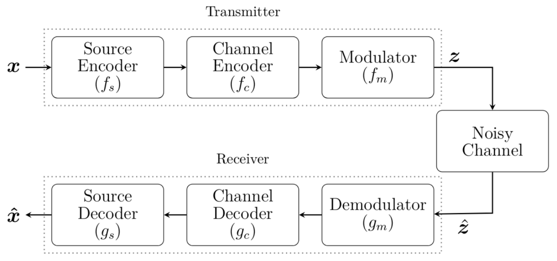
Motivation

- Increasing applications involving transmission and consumption of images in human-human, human-machine (e.g. AR, VR, telepresence) and machine-machine (e.g. pattern recognition, automation) interactions
- Particularly, how to deal with image/video transmission under extreme low latency, small bandwith and energy constraints?



Wireless Image Transmission

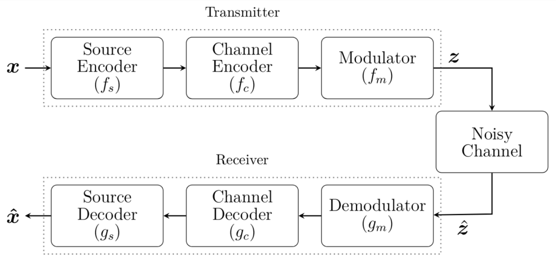
How information is transmitted over a noisy channel?



Shannon's Separation Theorem (1948):

- First compress underlying source into bits; Then, transmit bits over noisy channel reliably
- Highly efficient compression algorithms (e.g. JPEG, JPEG2000, BPG) and near-optimal channel codes (LDPC, Turbo, Polar)

Wireless Image Transmission

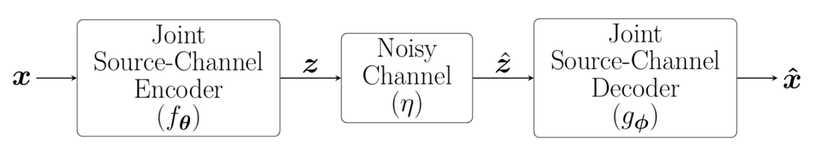


Challenges brought by modern requirements:

- Optimally holds only for infinite blocklength and complexity
- Design assumes a specific channel quality, being vulnerable to changes, variations or non-ergodic channels
- No separation theorem for multi-user networks: when broadcasting to many users, target the worst one

Proposed System - Joint Source Channel Coding

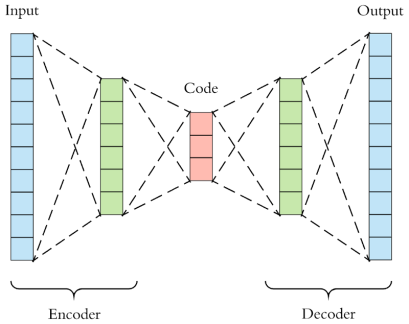
Can we learn to do better?



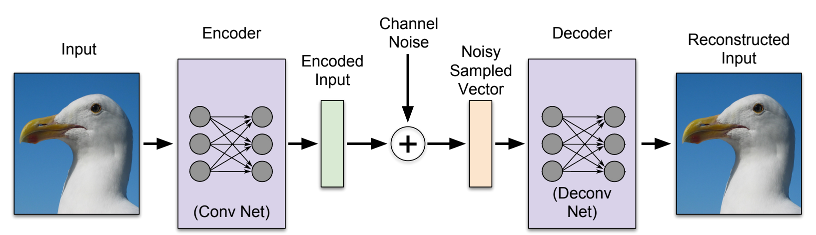
- Can we learn a direct mapping from pixel values to channel input symbols?
- Can this optimal mapping be learned directly from data, without the need of prior models?

Autoencoders

- Unsupervised learning: two neural networks trained together
- Similarities to digital communication systems
- Successfully applied in compression; design of channel code; blind channel equalization; etc.



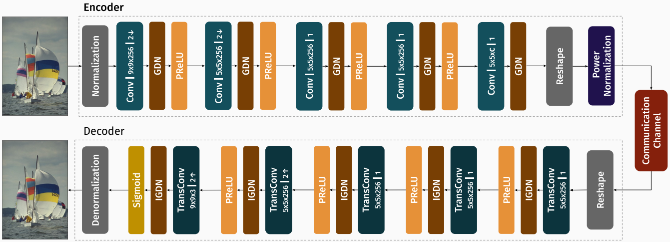
Deep Joint-Source Channel Coding



- Noisy channel added as non-trainable layer
- Directly mapping to channel: no bits conversion, speeding up the process and exploiting channel coding compression
- Coherent mapping: similar content stay close to each other

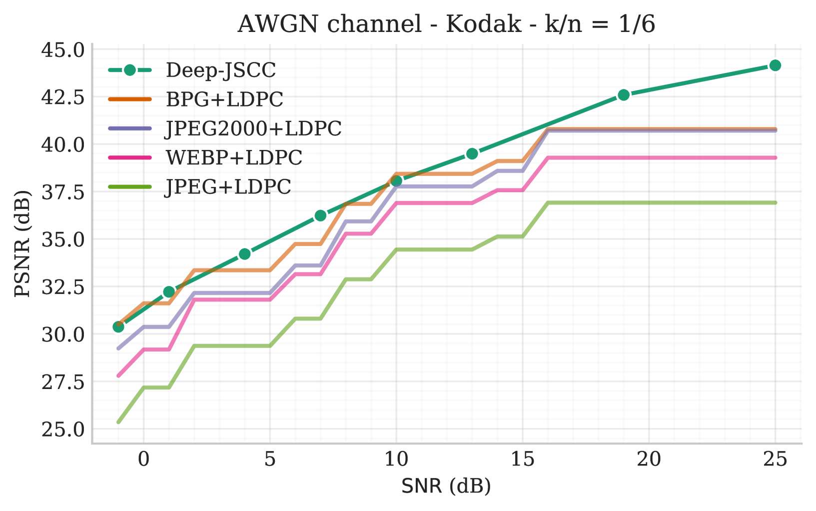
E. Bourtsoulatze, D. Burth Kurka and D. Gunduz, **Deep joint source-channel coding for wireless image transmission**, IEEE Trans. on Cognitive Comm. and Networking, Sep 2019.

Deep Joint-Source Channel Coding



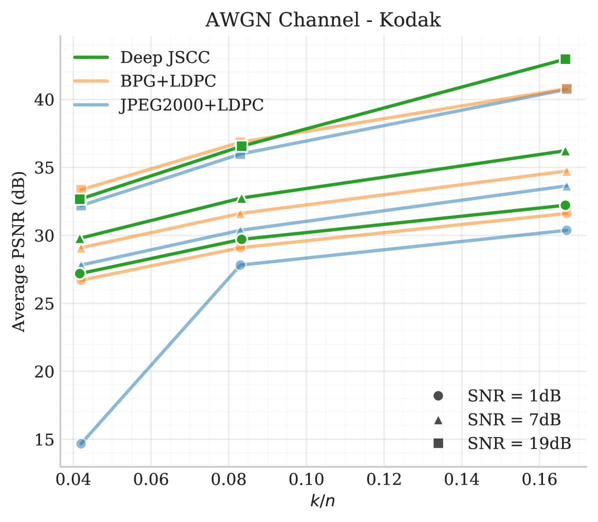
- ✓ Low-delay
- ✓ Low-energy
- ✓ Excelent performance
- ✓ Robust to channel changes
- ✓ Variable Bandwidth

DeepJSCC



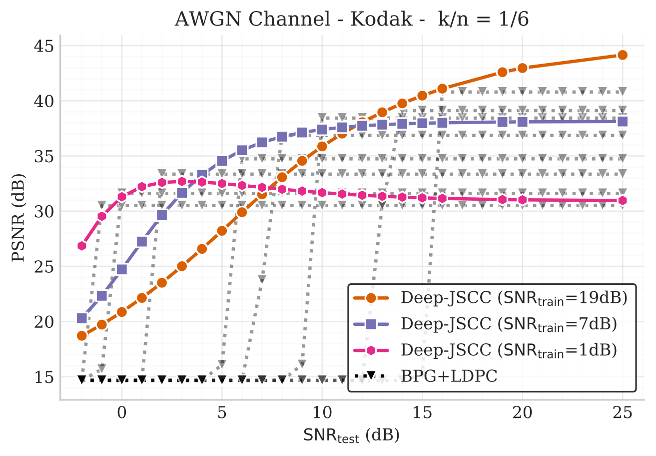
Superior performance compared to state-of-the-art digital schemes!

DeepJSCC - Performance by Bandwidth



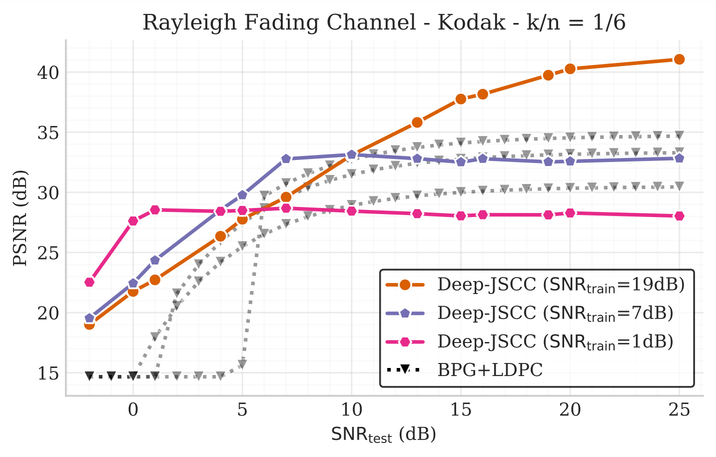
Consistently better in low-bandwidth, low-SNR regime (e.g. IoT)

DeepJSCC - Analog Behaviour



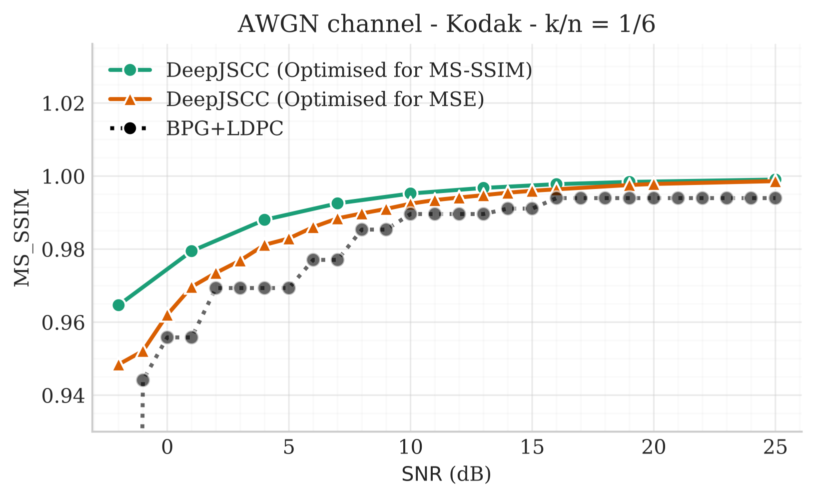
- Not affected by the "cliff effect": graceful degradation
- Resembles analog communications!

DeepJSCC - Rayleigh Fading Channel



- Time-varying channel or multiple receivers scenarios
- no pilot signal or explicit channel estimation!

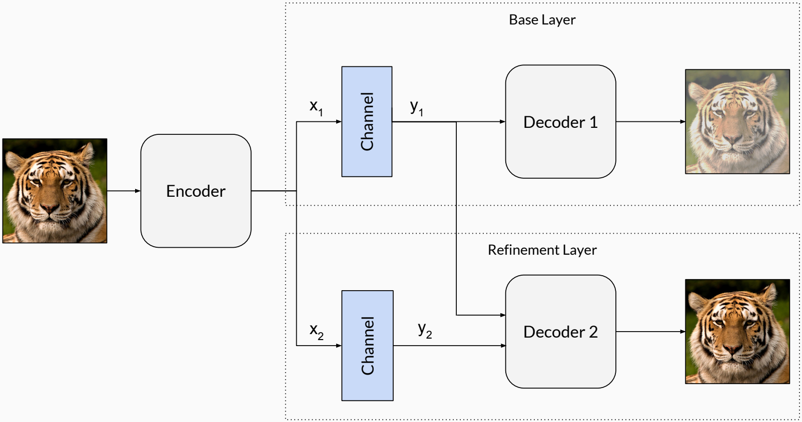
DeepJSCC - Perceived Image Quality



The more specific the domain, better the performance!

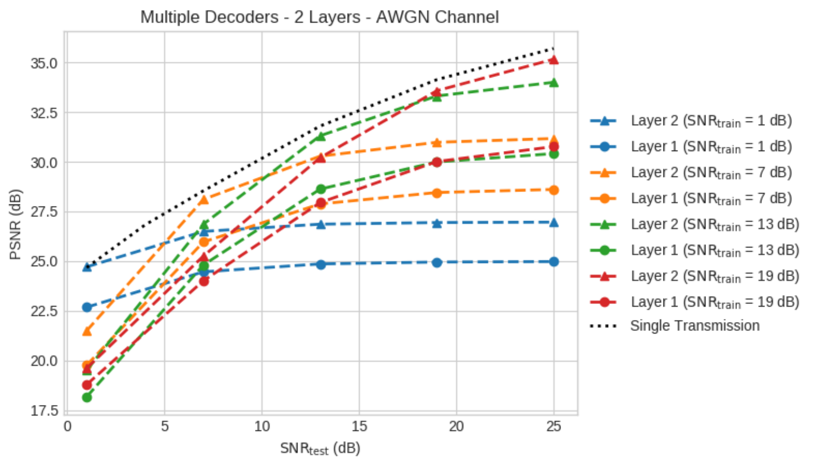
Original	BPG + LDPC	DeepJSCC
		
	PSNR: 24.096 MS-SSIM: 0.794	PSNR: 25.619 MS-SSIM: 0.882
		
	PSNR: 24.449 MS-SSIM: 0.872	PSNR: 26.095 MS-SSIM: 0.924
		
	PSNR: 22.279 MS-SSIM: 0.779	PSNR: 24.408 MS-SSIM: 0.907

Deep Wireless Successive Refinement



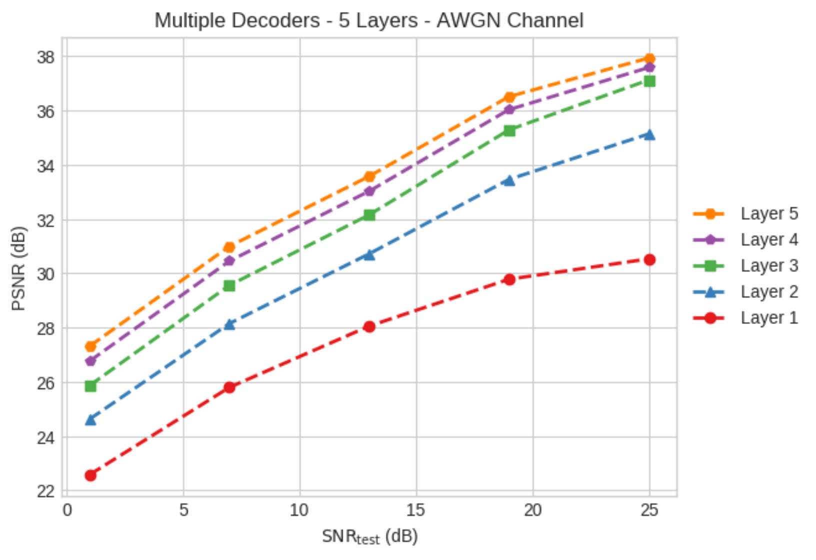
D. Burth Kurka, D. Gunduz, **Successive Refinement of Images with Deep Joint Source-Channel Coding**, IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Cannes, France, Jul 2019.

2-Layer Deep JSCC



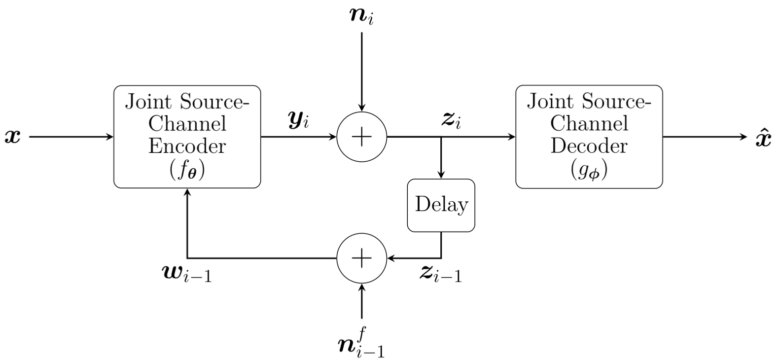
- Keep the same properties of single transmission (graceful degradation, analog behaviour)
- Negligible loss due to layering

5-Layer Deep JSCC



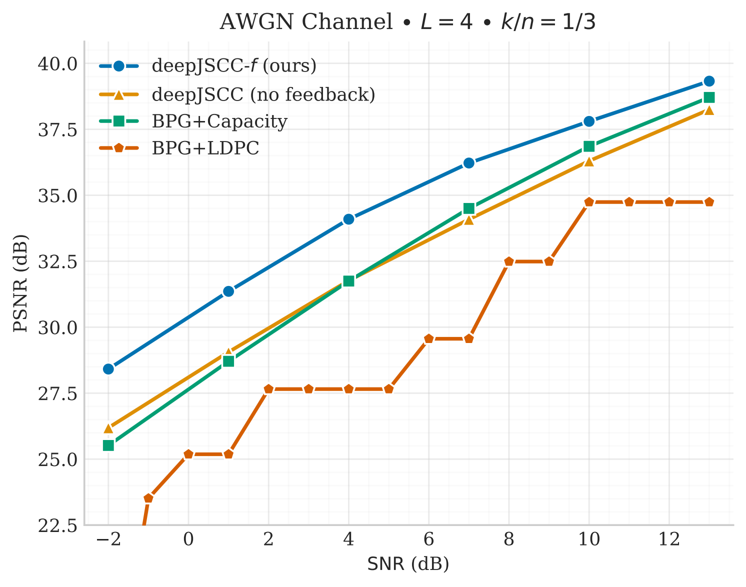
The addition of extra layers has small impact on previous layers

DeepJSCC-f: Exploiting Channel output feedback

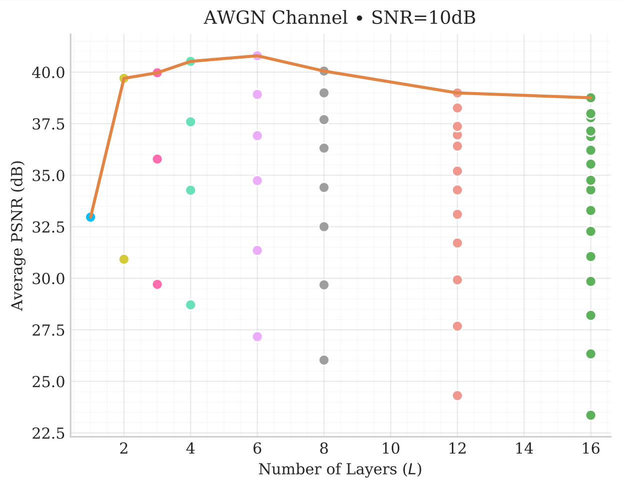


- First practical coding scheme that truly exploits feedback!

D. Burth Kurka and D. Gunduz, **Deep joint-source channel coding of images with feedback**, submitted to IEEE JSAIT, Oct. 2019.

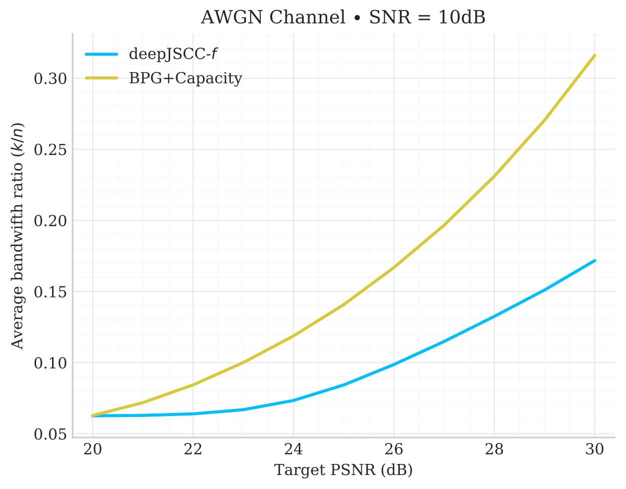

AWGN Channel • $L = 4$ • $k/n = 1/3$
deepJSCC-f (ours)
deepJSCC (no feedback)
BPG+Capacity
BPG+LDPC
PSNR (dB)
40.0
37.5
35.0
32.5
30.0
27.5
25.0
22.5
−2
0
2
4
6
8
10
12
SNR (dB)

DeepJSCC-f - impact of channel uses on performance



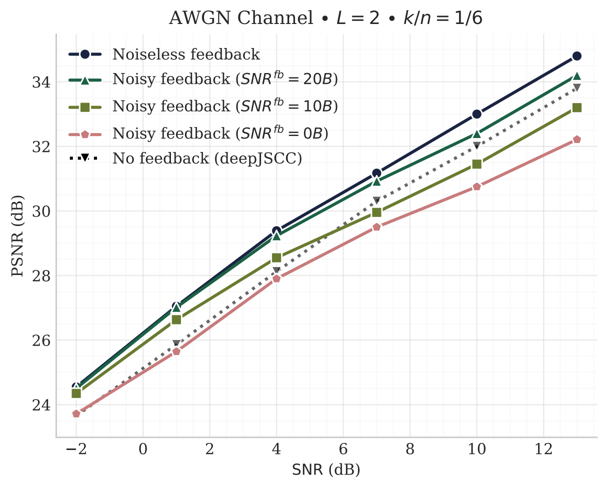
- Increasing the number of layers helps up to a certain point

DeepJSCC-f - Variable-length coding with feedback



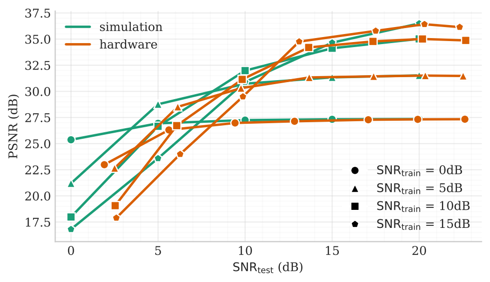
$\tilde{5}0\%$ bandwidth reduction!

DeepJSCC-f - Noisy feedback



Most theoretical results break down with noisy feedback!

Practical implementation



Average execution time (encoding+decoding):

JPEG2000+LDPC	BPG+LDPC	DeepJSCC (GPU)	DeepJSCC (CPU)
4.5ms	69.6ms	6.4ms	15.4ms

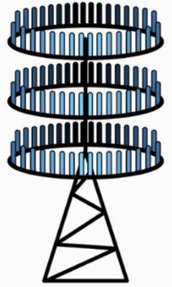
Summary so far...

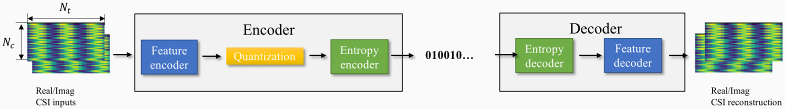
- Novel deep JSCC architecture for image transmission over wireless channels under low latency, bandwidth and energy
- Proposed model outperforms state-of-the-art separation-based transmission schemes, especially for limited channel bandwidth and SNR
- Absence of cliff effect and graceful degradation of the reconstruction quality with channel SNR
- Ability to communicate without explicit channel estimation
- Successive refinement achievable through multiple transmissions
- First practical model to exploit channel output feedback

Related research from Imperial IPC-Lab

Deep Learning based channel state information compression

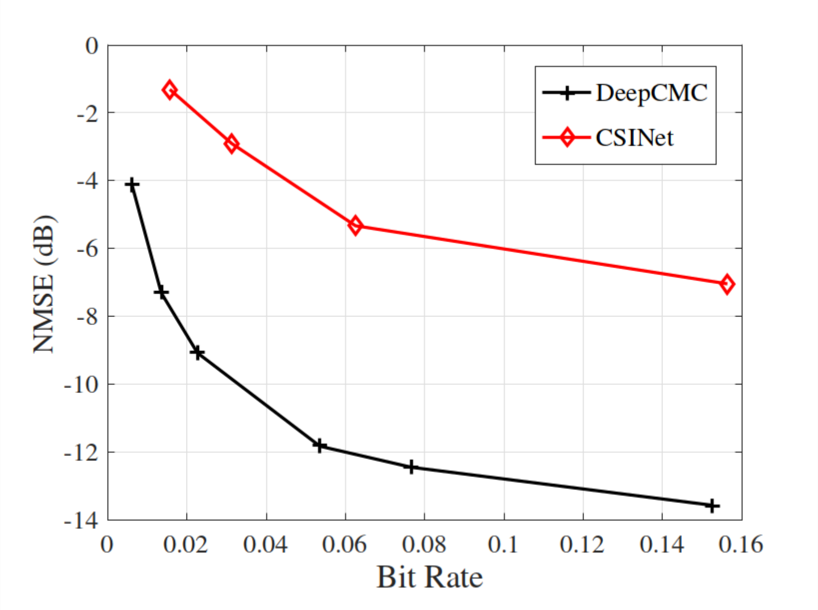
- Massive MIMO provides significant spectral efficiency, but require accurate channel state information (CSI)
- Excessive overhead required for massive MIMO training (i.e. pilot transmission, channel estimation and feedback)
- Training overhead even more critical for frequency division duplex (FDD) mode





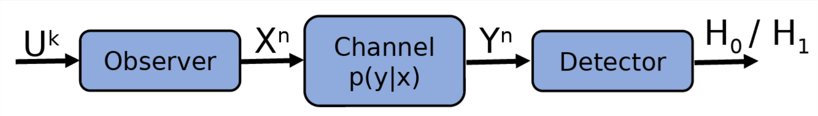
Q. Yang, M. Boloursaz Mashhadi, and D. Gunduz, **Deep convolutional compression for massive MIMO CSI feedback**, IEEE MLSP, Oct. 2019.

Deep Learning based channel state information compression



Hypothesis Testing Over Noisy Channels

- Distributed statistical inference problems in which data needs to be communicated to a remote decision maker.
  (e.g. Sensor networks, cloud computing, machine learning...)



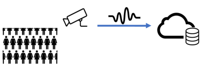
$$H_0 : U^k \sim \prod_{i-1}^{k} P_U \qquad\qquad H_1 : U^k \sim \prod_{i-1}^{k} Q_U$$

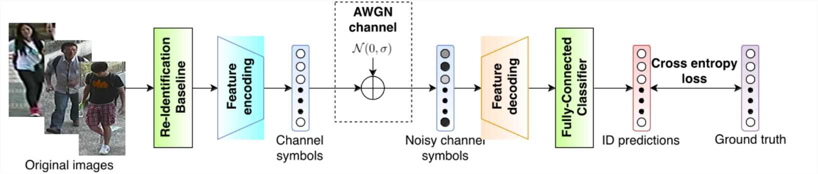
S. Sreekumar, D. Gunduz, **Hypothesis testing over a noisy channel**, IEEE ISIT, 2019.
S. Sreekumar and D. Gunduz, **Distributed hypothesis testing over discrete memoryless channels**, IEEE Transactions on Information Theory, to appear.
S Sreekumar, A Cohen, D Gunduz, **Distributed hypothesis testing with a privacy constraint**, Submitted.

Person re-identification over noisy channels

Standard approach:

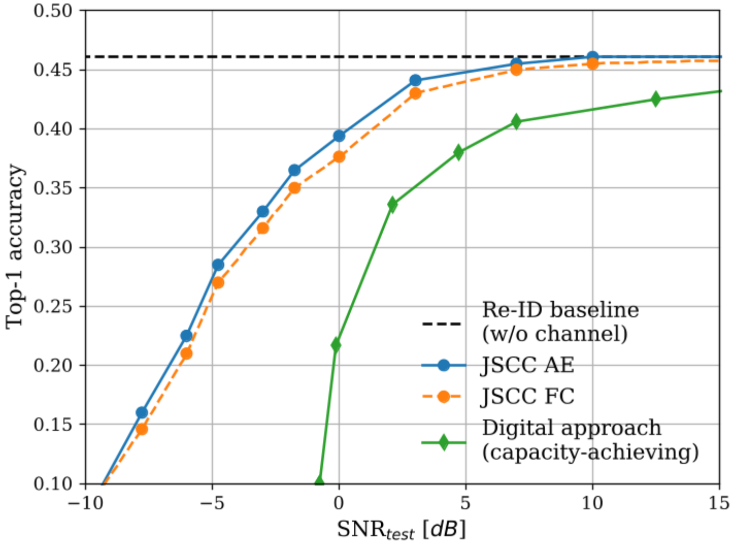
- Transmit images to the cloud
- Determine features most relevant for reidentification over image database
- Most of image's elements are useless for reId





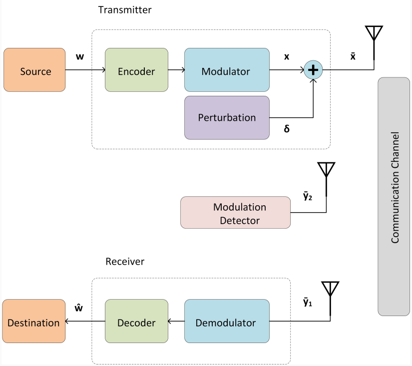
M. Jankowski, D. Gunduz, and K. Mikolajczyk, **Deep joint source-channel coding for wireless image retrieval**, arXiv:1910.12703v1, Oct. 2019.

Person re-identification over noisy channels



Adversarial ML for Secure Communications

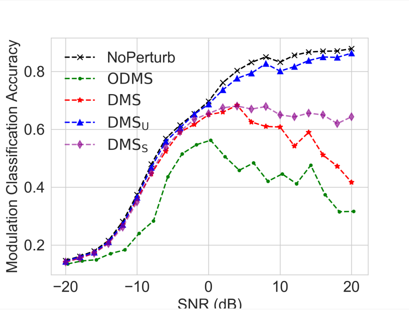
- Secure communication by preventing intruder form detecting the modulation scheme
- Small perturbations so that intruder is unable to identify modulation; receiver decodes message with acceptable BER (without knowing about the modifications)
- Use adversarial learning to solve trade-off



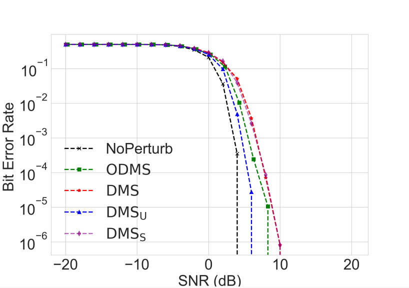
M. Z. Hameed, A. Gyorgy and D. Gunduz, **Communication without interception: Defense against modulation detection**, IEEE Global Conference on Signal and Information Processing (GlobalSIP), Ottawa, Canada, Nov. 2019. (Best paper award!)

Adversarial ML for Secure Communications - Results

Modulation Classification Accuracy Intruder



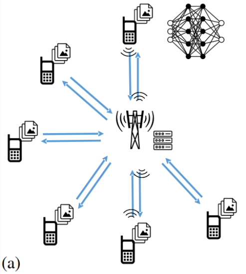
BER of legitimate Receiver



M. Z. Hameed, A. Gyorgy and D. Gunduz, **Communication without interception: Defense against modulation detection**, IEEE Global Conference on Signal and Information Processing (GlobalSIP), Ottawa, Canada, Nov. 2019.

Federated Learning and Distributed Computation Learning

- Terms like petabytes, exabytes and zetabytes are replacing megabytes, gigabytes and terabytes
- Handling such data volumes in a single machine is neither efficient nor possible DeepMind’s AlphaGo ran on 1920 CPUs and 280 GPUs
- In IoT, users or devices may not want to, or cannot have the bandwidth/ energy to share their data with cloud. Alternative is distributed / federated learning

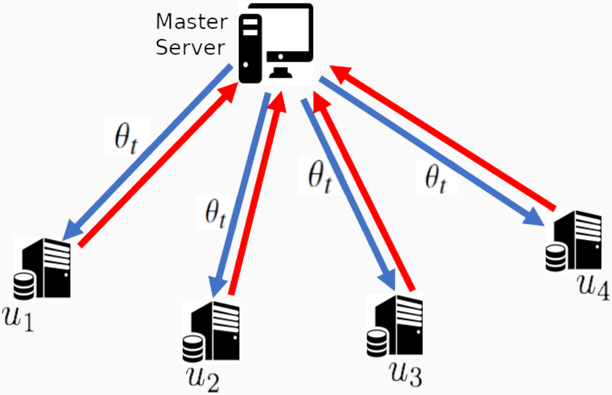


Federated Learning

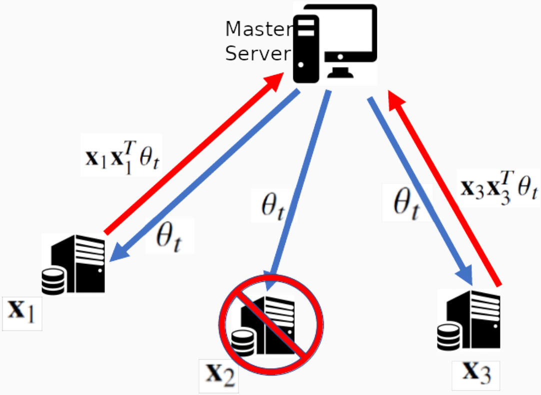
-

Distributed Learning Framework

$$F(\theta) = \frac{1}{M}\sum_{m=1}^{M} f(\theta, \mathcal{U}_m) \qquad \theta_{t+1} = \theta_t - \eta_t \frac{1}{M}\sum_{m=1}^{M} \nabla f(\theta_t, \mathcal{U}_m)$$



Straggling Servers



Treat system as a packet erasure communication channel, in which transmitted data packets are randomly erased.

Dealing with Straggling Servers

- Gradient Coding

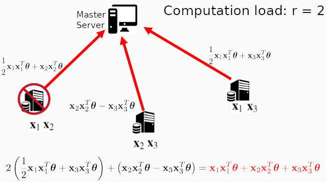
  R. Tandon, Q. Lei, A. G. Dimakis, and N. Karampatziakis, **Gradient coding**, arXiv preprint: arXiv:1612.03301, 2016.

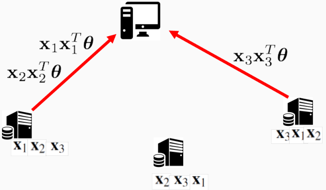
- Multi-message Distributed Computing (non-persistent stragglers)

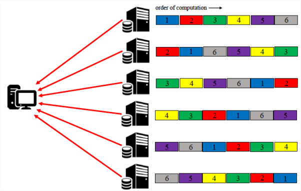
  E. Ozfatura, D. Gunduz and S. Ulukus, **Speeding up distributed gradient descent by utilizing non-persistent stragglers**, submitted to SysML, Stanford, CA, Mar. 2019.

- Multi-message Uncoded Computation

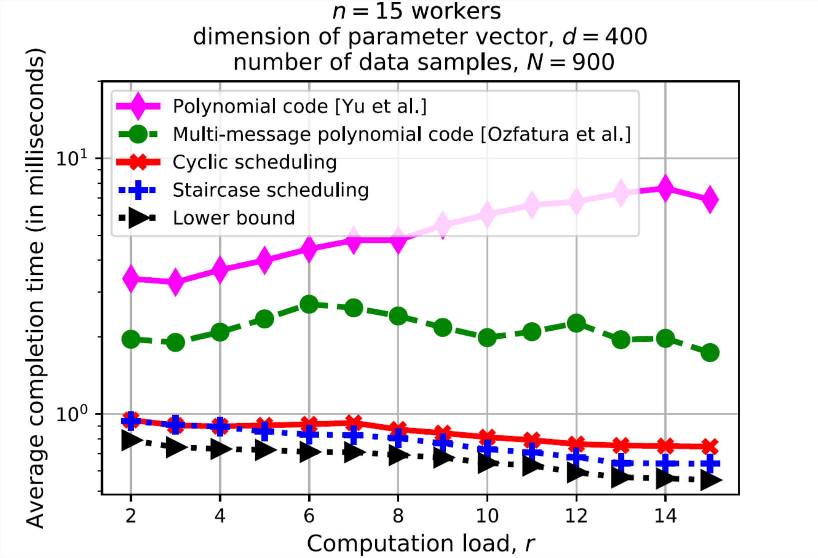
  M. Mohammadi Amiri and D. Gunduz, **Distributed uncoded computation**, submitted to IEEE Trans. on Signal Processing, 2018.





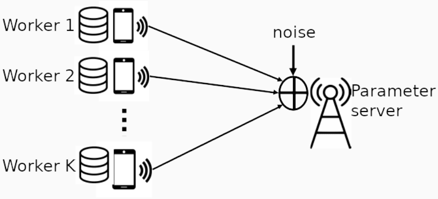


Straggling Servers



Wireless Edge Learning

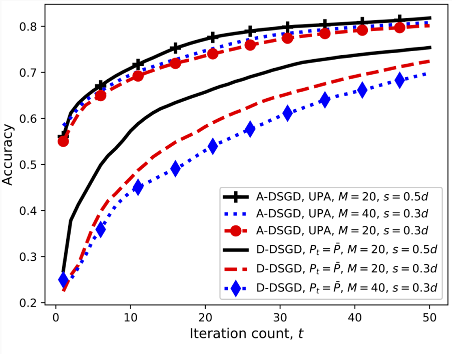
- Wireless devices with own data
- a parameter server (PS) enables learning
- Devices (workers) connected to PS through wireles links: bandlimited multiple access channel (MAC)
- Digital approach: separate learning an communications
- Analog approach: let channel do gradient averaging



M. Mohammadi Amiri and D. Gunduz, **Federated learning over wireless fading channels**, arXiv:1907.09769 [cs.IT], Jul. 2019.

M. Mohammadi Amiri and D. Gunduz, **Over-the-air machine learning at the wireless edge**, in Proc. IEEE SPAWC, Jul. 2019.

Wireless Edge Learning - accuracy vs iteration



- Distributed MNIST classification (single layer with 10 neurons)
- Parameter vector size $d = 28 \times 28 \times 10 + 10 = 7850$
- d: dimension of parameter vector
- s: symbols per iteration
- M: number of devices (i.e., workers)

Conclusion

Conclusions

- Machine Learning (ML) can improve communication networks, and learning also requires efficient communications
- ML brings along many new problems for communications researchers

Some other problems we look into, not mentioned in this talk:

- Deep RL for wireless resource optimization
- GANs for channel modeling
- Privacy-aware learning

Further References and Acknowledgements

Feel free to contact me at:

- d.kurka@imperial.ac.uk

To retrieve all papers mentioned in this presentation:

- **`https://www.imperial.ac.uk/information-processing-and-communications-lab/`**
- D. Gunduz, P. de Kerret, N. Sidiroupoulos, D. Gesbert, C. Murthy, M. van der Schaar, Machine learning in the air IEEE Journal on Selected Areas in Communications, Oct. 2019.
- "**Deniz Gunduz**" on Google Scholar

Sponsors:

CNPq; Fapesp; European Research Council (ERC), Starting Grant BEACON; Marie Skłodowska-Curie fellowship

Thank you!