From Conventional to Semantic Communications based on Deep Learning

Professor Geoffrey Li
Imperial College London
London, UK

ITP Lab Intelligent Transmission & Processing Laboratory
Imperial College London

Outline

- Overview
- DL-based Conventional Communications

  * Block-Wise

  * End-to-End
- DL-based Semantic Communications
- Conclusions

ITP Lab
Imperial

Motivation

➢ **Challenges in current/conventional communication systems**

- Mathematical models versus practical imperfection
- Block structures versus global optimality
- Complexity and performance of optimization
- Spectrum efficiency limited by Shannon capacity

➢ **Why deep learning?**

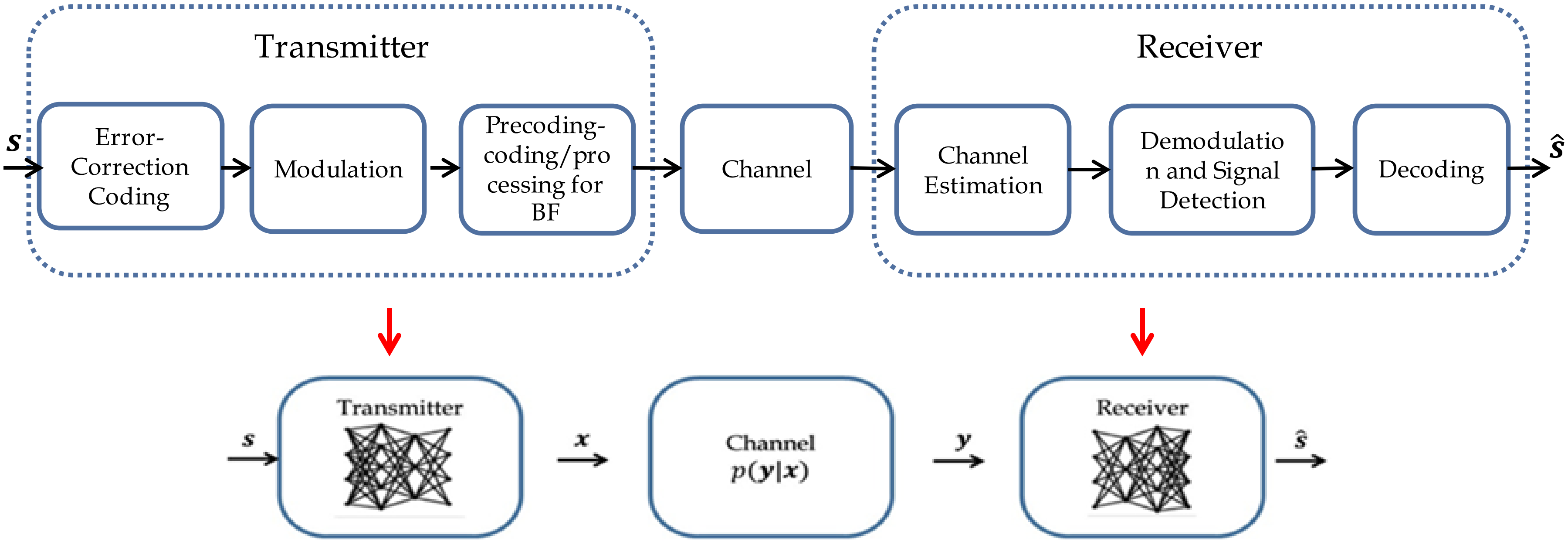
- No need for models for data-driven method
- End-to-end loss optimization for global optimality
- Deep learning enabled end-to-end and semantic communications

W. Tong and G. Y. Li "Nine critical issues in AI and wireless communications to ensure successful 6G," in *IEEE Wireless Commun.*, also at https://arxiv.org/abs/2109.11320, Aug. 2021.

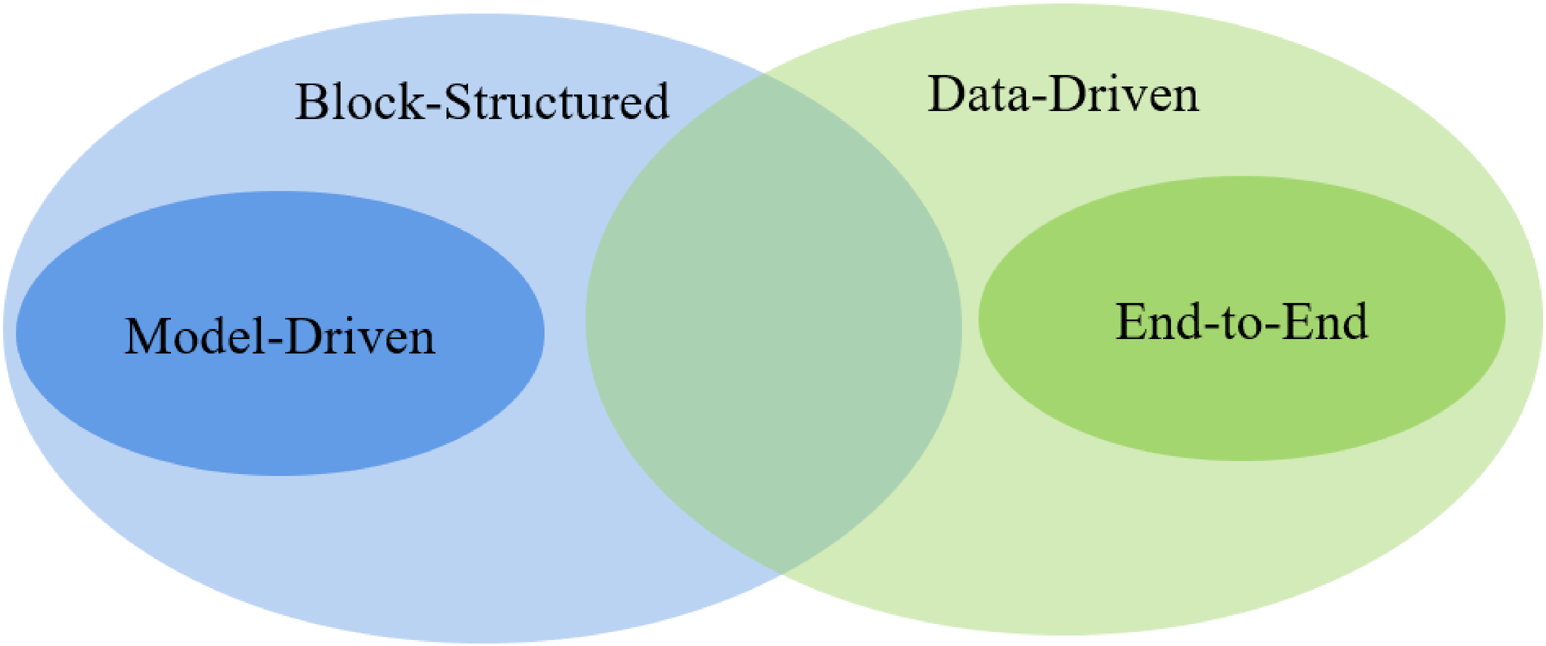
Block Structure or End-to-End for Conventional Communications

Conventional Communications:

Transmit symbols or bits, following Shannon Limit



DL in Physical Layer Conventional Communications



- Z.-J. Qin, H. Ye, G. Y. Li, and B.-H. Juang, "Deep learning in physical layer communications," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 93-98, April 2019. (2022 IEEE ComSoc Fred W. Ellersick Prize Paper Award)
- H.-T. He, S. Jin, C.-K. Wen, F.-F. Gao, G. Y. Li, and Z.-B. Xu, "Model-driven deep learning for physical layer communications," *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 77- 83, Oct. 2019
- H. Ye, G. Y. Li, and B.-H. F. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114 – 117, Feb. 2018.

Information Content of English and Semantic Encoding

Encoding English Words Letter-by-Letter

- In English, on average there are 4.5 letters per word
- 5.5 characters per word if including space
- 5 bits to encode each letter (26 letters)
- 27.5 bits/word (5X5.5=27.5) **Need a codebook of 26 letters**

Encoding English Words Word-by-Word

- 171,476 English words (from Google)
- 18 bits/word ($2^{17} < 171{,}476 < 2^{18}$) **Need a codebook of 171,476 words**

Encoding English Semantically

- For example, only 1 bit if answering a YES or NO question
- …. More Efficient! **Need an extremely huge codebook**

ITP Lab Imperial

From Symbol to Semantic Transmission

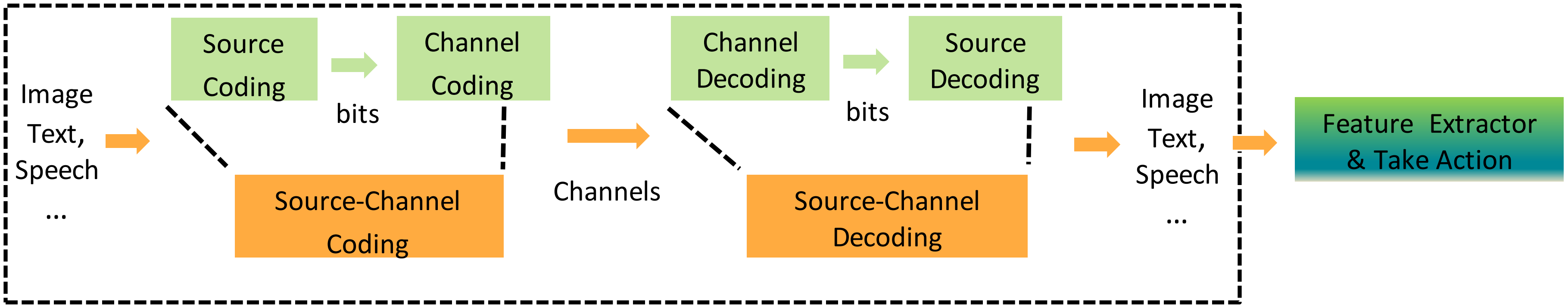
- **Three Levels of Communications: Shannon and Weaver**
  - **Transmission of symbols (Shannon Paradigm)**
    following Shannon limit & well-developed near limit
  - **Semantic exchange of source information**
    semantic communications (transmission of intelligence)
  - **Effects of semantic information exchange**

- **Semantic Communications: <u>Significantly improved efficiency</u>!**

C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. The University of Illinois Press, 1949.

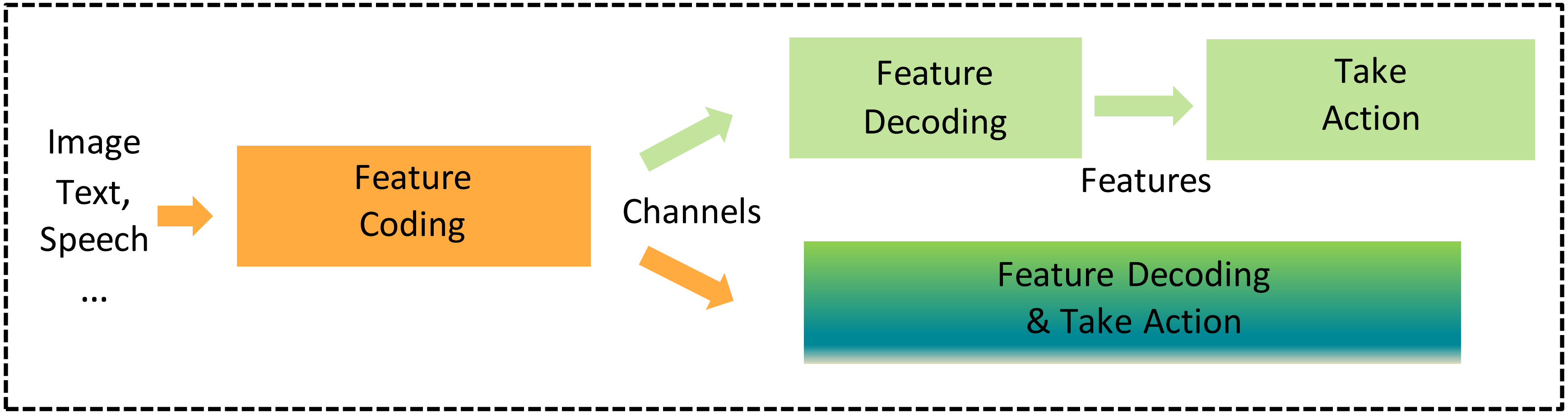
Conventional Communications

- Only consider the **data recovery accurately**
- Information redundancy are removed in **entropy-domain**
- All information (including **useless and irrelevant**) is transmitted to the receiver, part is useless for the target network



Semantic Communications

- **Feature** networks and **action** networks considered
- Information redundancy removed in **semantic domain**
- Only **useful and relevant** information transmitted to the receiver
- The features can serve **different** action networks



Z.-J. Qin, X.-M. Tao, J.-H. Lu, and G. Y. Li, "Semantic communications: Principles and challenges," https://arxiv.org/abs/2201.01389.

ITP Lab
Imperial

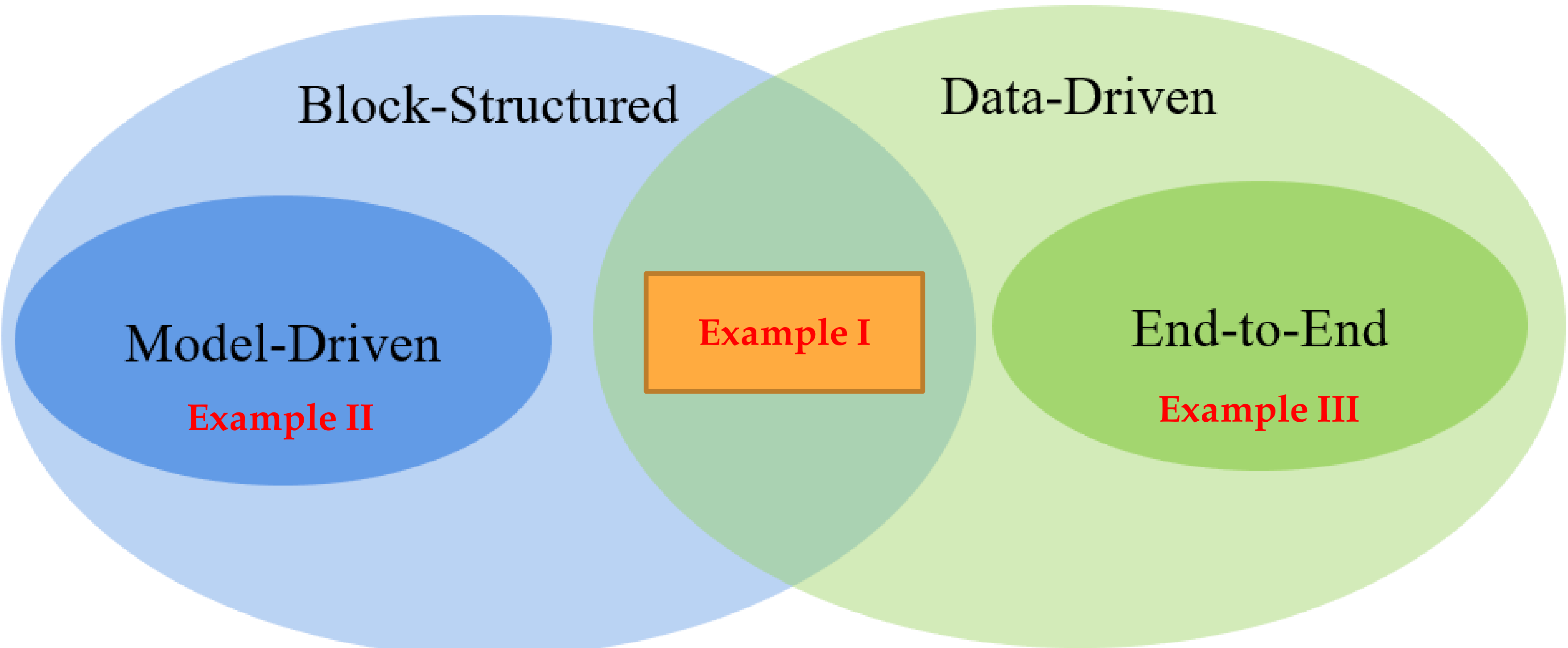
Outline

- Overview
- DL-based Conventional Communications

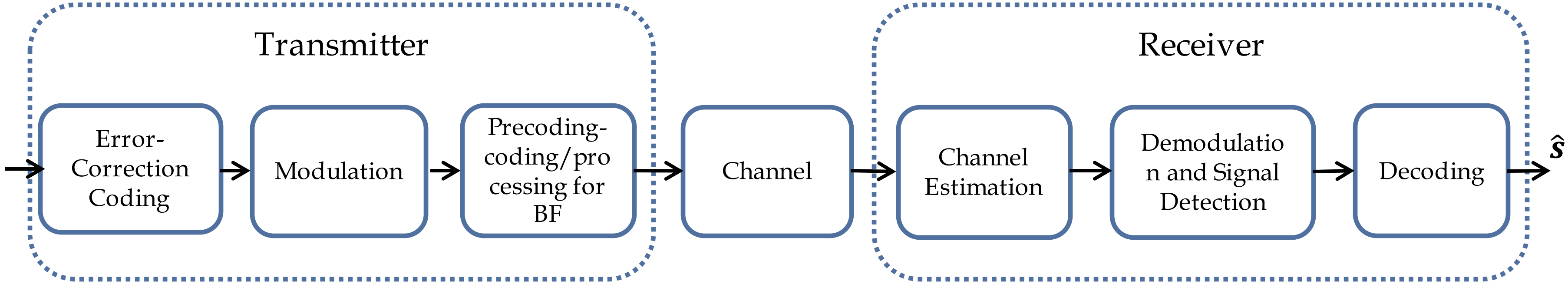
  * Block-Wise

  * End-to-End
- DL-based Semantic Communications
- Conclusions

DL in for Conventional Communications



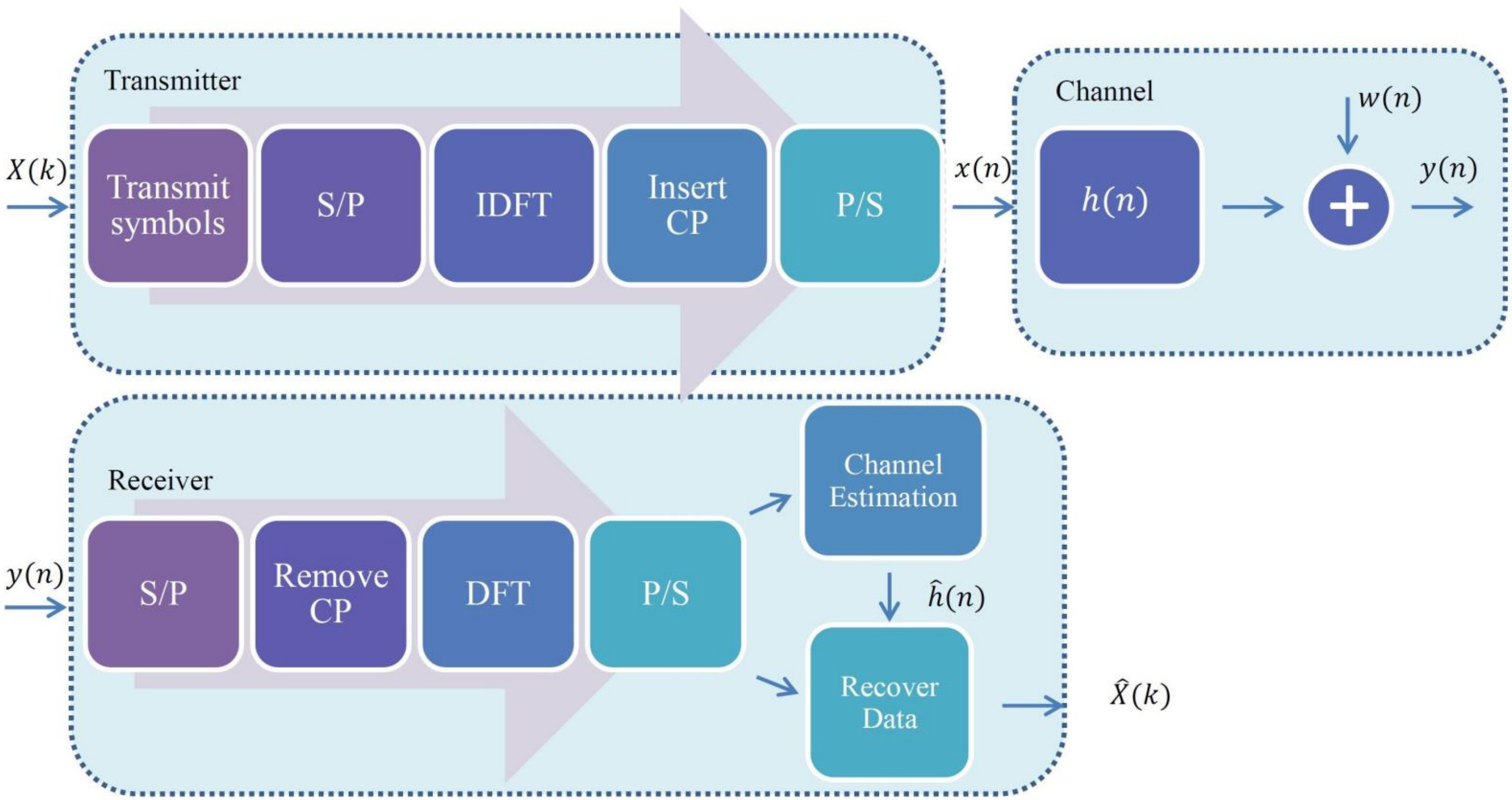
Channel Estimation (CE) and Signal Detection (SD)



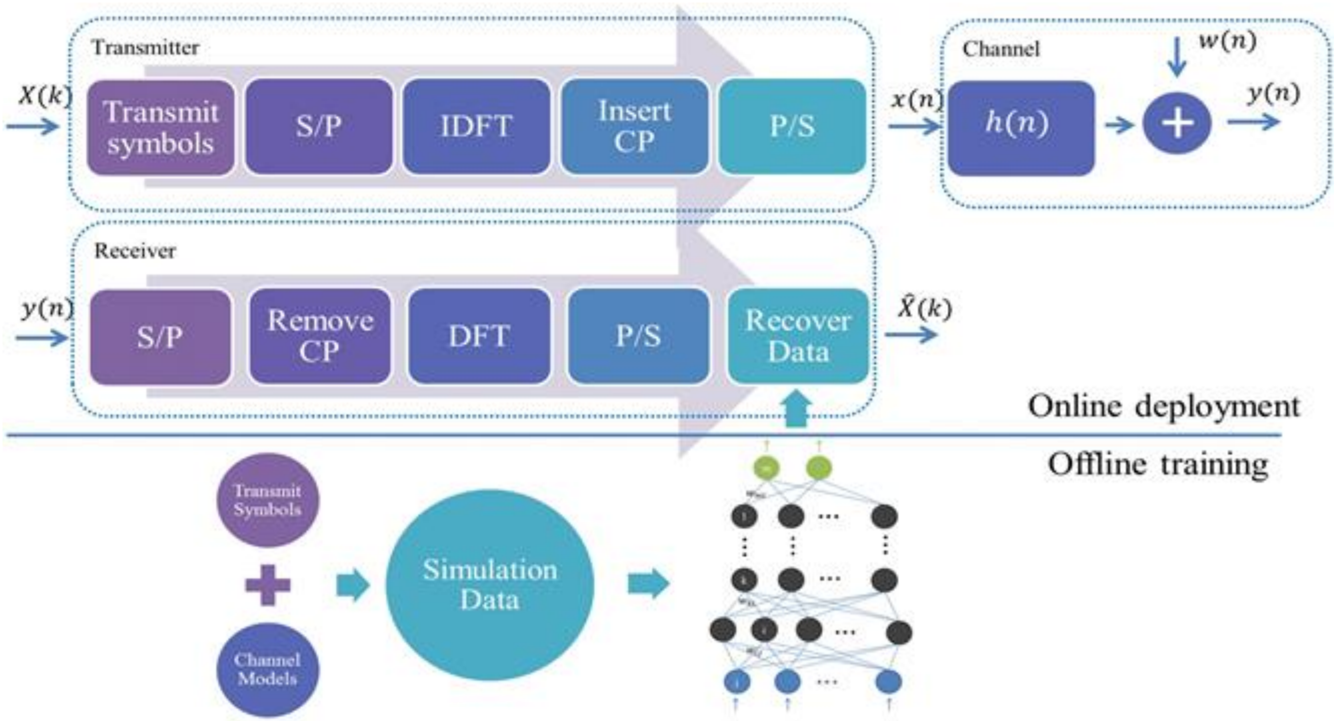
- **Related works:**
  - MMSE for channel estimation
  - Neural networks and DL in equalization and decoding
- **Challenges:**
  - Nonlinear distortion and interference
- **Innovations:**
  - DL for joint channel estimation and symbol detection
  - DL-based method: robust and insensitive to nonlinear distortion and interference

ITP Lab Imperial

Traditional CE and SD



DL-based CE and SD



- ➢ Input: received pilot OFDM block + received data OFDM block
- ➢ Output: recovered data

H. Ye, G. Y. Li, and B.-H. F. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114 – 117, Feb. 2018.

ITP Lab Imperial

DL-based CE and SD: DNN Model Training

- Training DNN to predict transmit data
- Training with received OFDM samples corresponding to pilots and data
- Generating label data under diverse channel conditions
- Optimizing model parameters to minimize L_2 loss function

$$L_2 = \frac{1}{N} \sum_{k} (\hat{X}(k) - X(k))^2$$

ITP Lab
Imperial

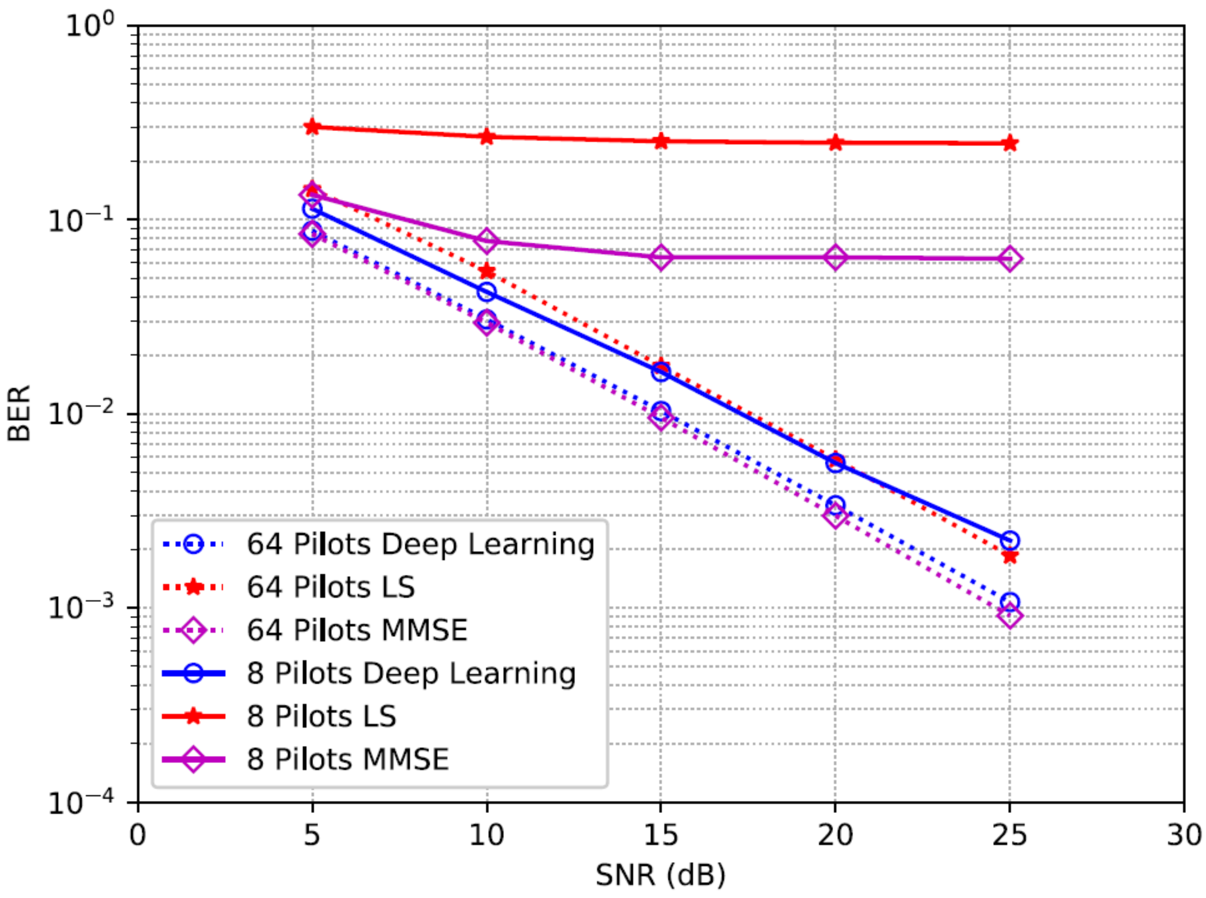
DL-based CE and SD: Impact of Pilot Number

64 pilots:

- ➢ Better than LS
- ➢ Comparable to LMMSE

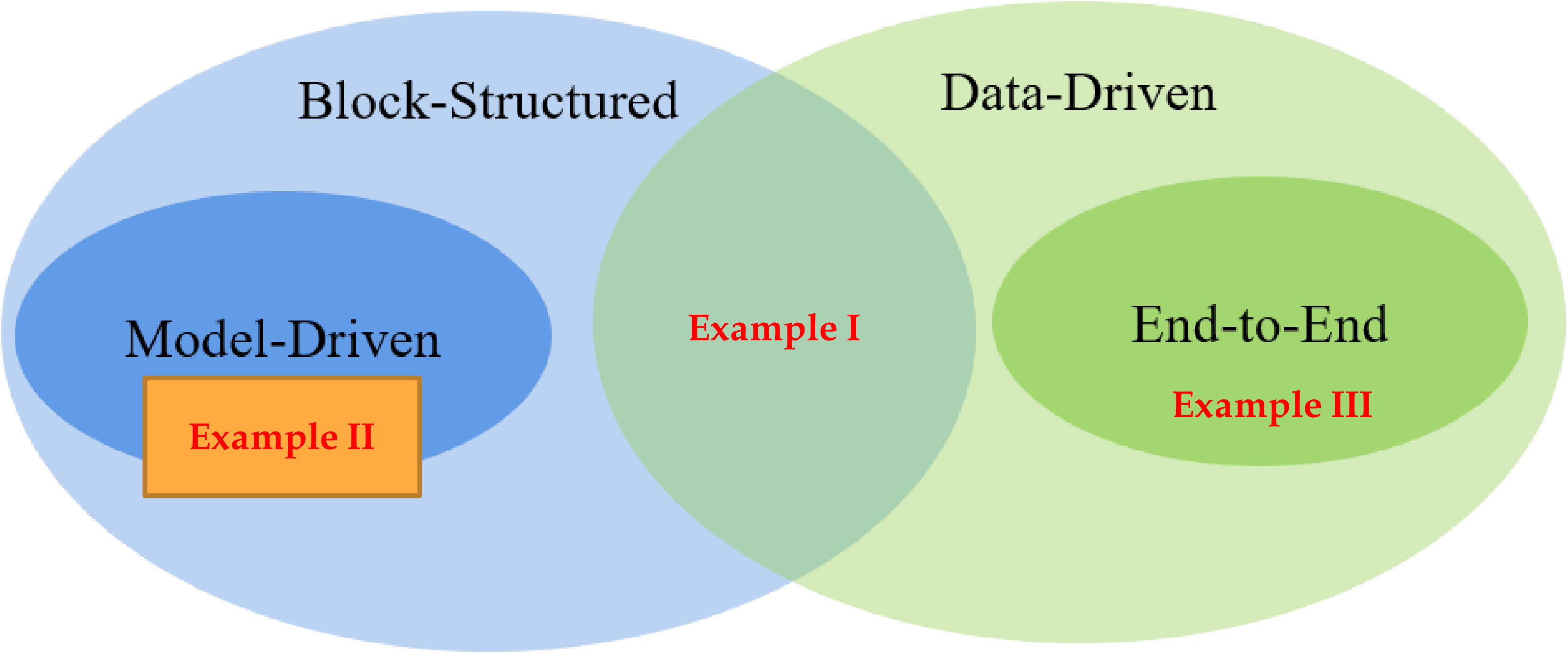
8 pilots:

- ➢ Better than LMMSE



ITP Lab
Imperial

DL in for Conventional Communications



Model-Driven DL

- Relying on relatively accurate model
- Exploiting rich domain/expert knowledge
- Easy to train with a small amount of data
- Explainable and predictable neural networks
- **Deep unfolding**: a popular model-driven approach

H.-T. He, S. Jin, C.-K. Wen, F.-F. Gao, G. Y. Li, and Z.-B. Xu, "Model-driven deep learning for physical layer communications," ***IEEE Wireless Commun,*** **vol. 26, no. 5, pp. 77- 83, Oct. 2019.**

Example: MIMO Detection

- **MIMO System**:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

- **Goal**: estimating $\mathbf{x}$ from received signal $\mathbf{y}$ and channel matrix $\mathbf{H}$
- **Conventional Detectors:**
  - Optimal detector: ML detector, high complexity
  - Linear detectors: ZF, LMMSE, low complexity but poor performance
  - Iterative detectors: AMP-based detector, EP-based detector, excellent performance, moderate complexity, performance degradation with ill-conditioned channel matrix
- **Motivation**: deep learning to perform iterative detection

H.-T. He, C.-K. Wen, S. Jin, and G. Y. Li, "Model-driven deep learning for MIMO detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 1702-1715, March 2020.

Example: OAMP for MIMO Detection

(Orthogonal Approximate Message Passing)

- **OAMP algorithm for MIMO detection:**

$$\hat{\mathbf{x}} = \int \mathbf{x}\mathcal{P}(\mathbf{x}|\mathbf{y}, \mathbf{H})d\mathbf{x}$$

Algorithm 1: OAMP algorithm for MIMO detection

Input: Received signal $\mathbf{y}$, channel matrix $\mathbf{H}$, noise level σ^2.

Output: Recovered signal $\mathbf{x}_t$.

Initialize: $\tau_t \leftarrow 1$, $\mathbf{x}_t \leftarrow \mathbf{0}$

$$\mathbf{r}_t = \hat{\mathbf{x}}_t + \mathbf{W}_t(\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_t), \tag{8}$$

$$\hat{\mathbf{x}}_{t+1} = \mathbb{E}\left\{\mathbf{x}|\mathbf{r}_t, \tau_t\right\}, \tag{9}$$

$$v_t^2 = \frac{\|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_t\|_2^2 - M\sigma^2}{\mathrm{tr}(\mathbf{H}^T\mathbf{H})} \tag{10}$$

$$\tau_t^2 = \frac{1}{N}\mathrm{tr}(\mathbf{B}_t\mathbf{B}_t^T)v_t^2 + \frac{1}{N}\mathrm{tr}(\mathbf{W}_t\mathbf{W}_t^T)\sigma^2. \tag{11}$$

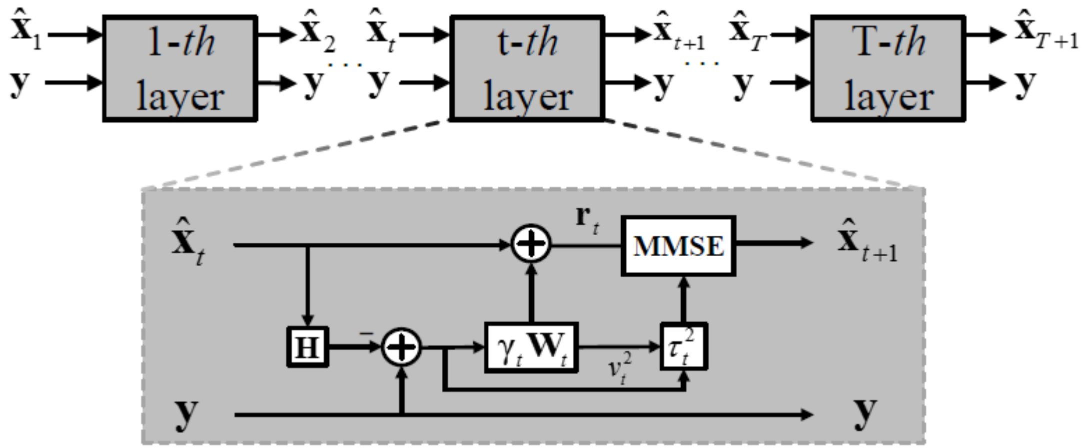
- **Obtained a Network by Unfolding OAMP Algorithm**

J. Ma and L. Ping, "Orthogonal OAMP," IEEE Access, vol. 5, no. 14, pp. 2020 – 2033, Jan. 2017

ITP Lab
Imperial

Example: Modified OAMP-Net

- **Architecture:**



- **Iterative Algorithm:**

$$v_t^2 = \frac{\|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_t\|_2^2 - M\sigma^2}{\mathrm{tr}(\mathbf{H}^T\mathbf{H})}$$

$$\mathbf{W}_t = \frac{2N}{\mathrm{tr}(\hat{\mathbf{W}}_t\mathbf{H})}\hat{\mathbf{W}}_t \quad \hat{\mathbf{W}}_t = v_t^2\mathbf{H}^T(v_t^2\mathbf{H}\mathbf{H}^T + \frac{\sigma^2}{2}\mathbf{I})^{-1}$$

$$\mathbf{r}_t = \hat{\mathbf{x}}_t + \gamma_t\mathbf{W}_t(\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_t)$$

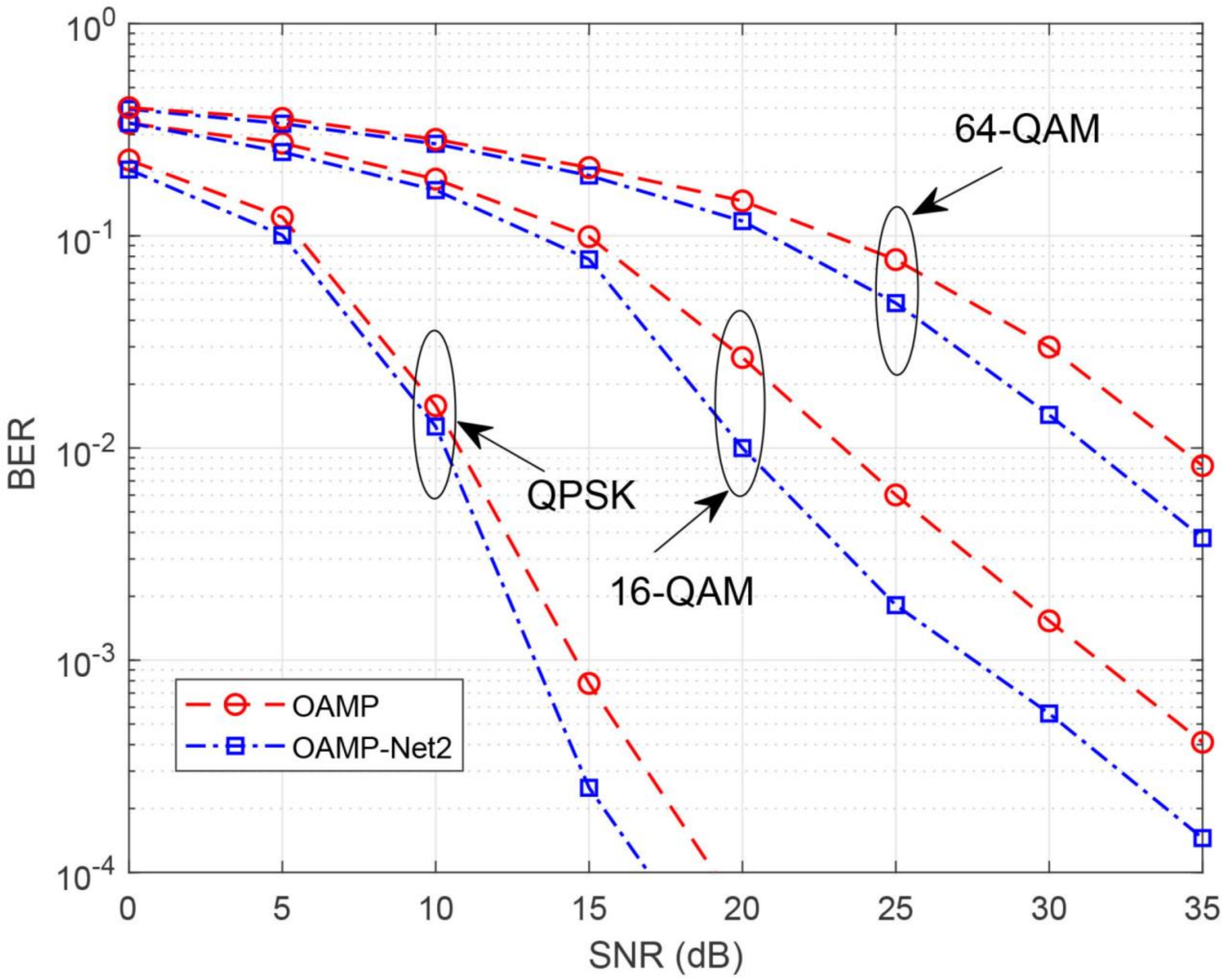
$$\tau_t^2 = \frac{1}{2N}\mathrm{tr}(\mathbf{C}_t\mathbf{C}_t^T)v_t^2 + \frac{\theta_t^2\sigma^2}{4N}\mathrm{tr}(\mathbf{W}_t\mathbf{W}_t^T) \quad \mathbf{C}_t = \mathbf{I} - \theta_t\mathbf{W}_t\mathbf{H}$$

$$\hat{\mathbf{x}}_{t+1} = \mathbb{E}\{\mathbf{x}|\mathbf{r}_t, \tau_t\}$$

- **Tainable Parameters:** **Only two parameters** (γ_t, θ_t) **for each iteration!**

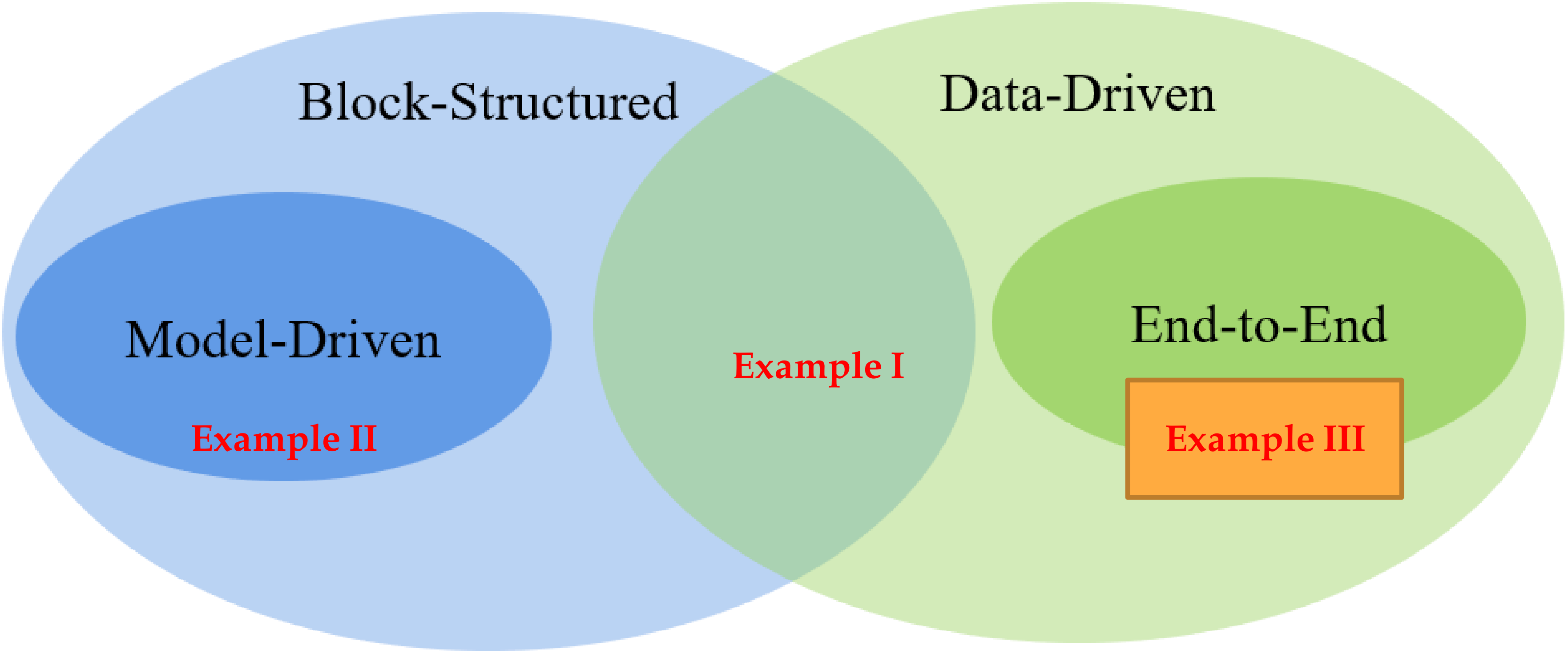
Performance of Modified OAMP-Net

- **16x16 MIMO**
- **Outperforming** original OAMP
- **No. trainable variables:** twice iteration no. Independ. of antenna no.

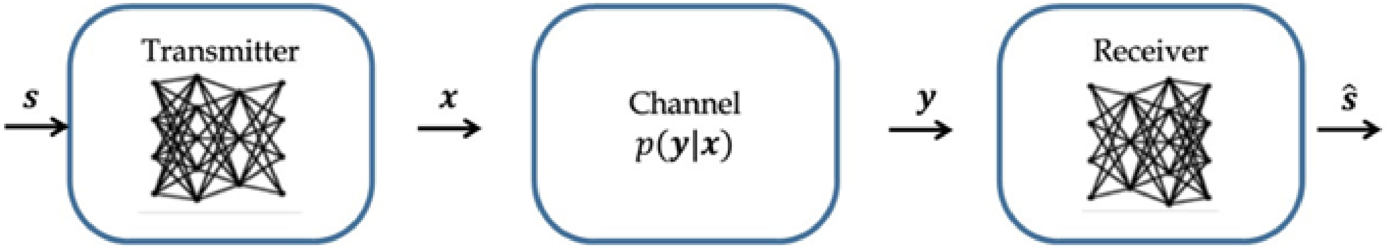


ITP Lab Imperial

DL in for Conventional Communications



Why End-to-End Learning?



- **Architecture:**
  - Representing both transmitter and receiver by DNNs
  - Leaning to encode transmit symbols at transmitter
  - Learning to recover transmit symbols at receiver
- **Merits:**
  - Achieving global optimum
  - Universal solution to different channels
  - Beating current state-of-arts

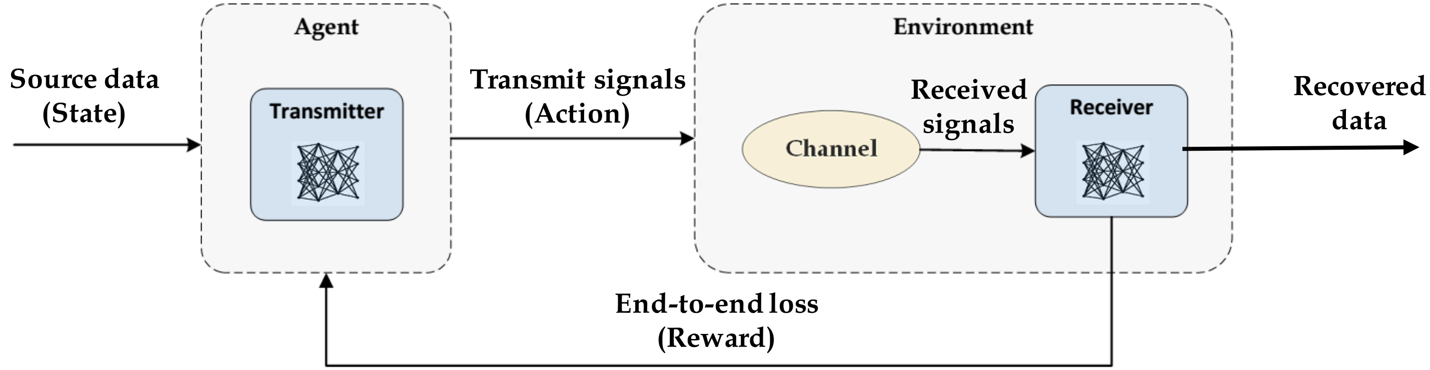
ITP Lab
Imperial

Channel Agnostic End-to-End Learning

- **Related Works:**
  - End-to-end communication for AWGN
  - End-to-end communication in OFDM
  - End-to-end communication with hardware impairment
- **Challenges:**
  - Back-propagation of the gradients is blocked by the unknown channel
  - Channel is time-varying
- **Approaches:**
  - Reinforcement Learning (following references)
  - Conditional *Generative Adversarial Net* (GAN) (our approach)

- T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. on Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563-575, Dec. 2017.
- S. Dorner, S. Cammerer, J. Hoydis, S. ten Brink, "Deep learning-based communication over the air", *IEEE J. Select. Topics Signal Process.*, vol.12, no. 1, pp. 132-143, Feb. 2018.
- A. Felix, S. Cammerer, S. Dorner, J. Hoydis, and S. ten Brink, "OFDM autoencoder for end-to-end learning of communications systems," in *Proc. IEEE Int. Workshop Signal Proc. Adv. Wireless Commun.(SPAWC)*, Jun. 2018.
- F. Aoudia, and J. Hoydis. "End-to-end learning of communications systems without a channel model," *arXiv preprint arXiv: 1804.02276*

E2E based on Reinforcement Learning

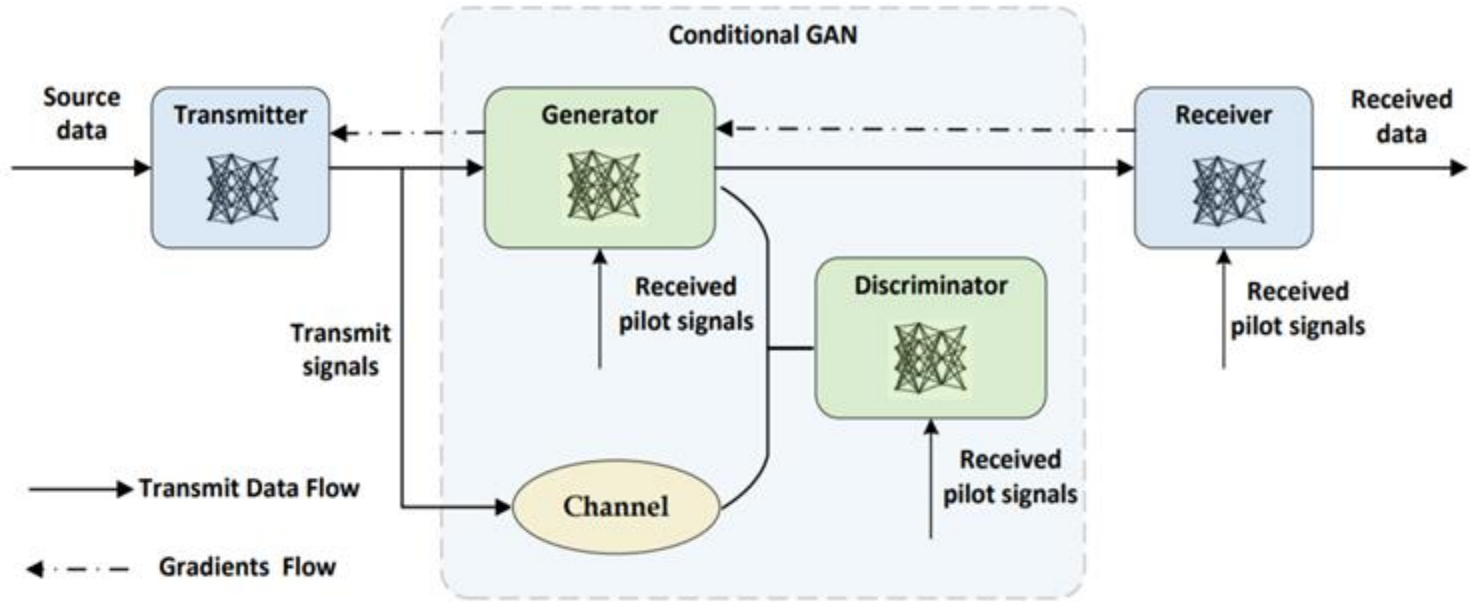


- **Reinforcement Learning Formation:**
  - ❑ Agent: transmitter
  - ❑ Environment: channel + receiver
  - ❑ States: source data
  - ❑ Actions: transmit signals
- **Advantage and Disadvantage:**
  - ❑ Unnecessary for channel modeling
  - ❑ Hard for continuous action in reinforcement learning

F. Aoudia, and J. Hoydis. "End-to-end learning of communications systems without a channel model," *arXiv preprint arXiv: 1804.02276*

ITP Lab
Imperial

E2E based on Conditional GAN

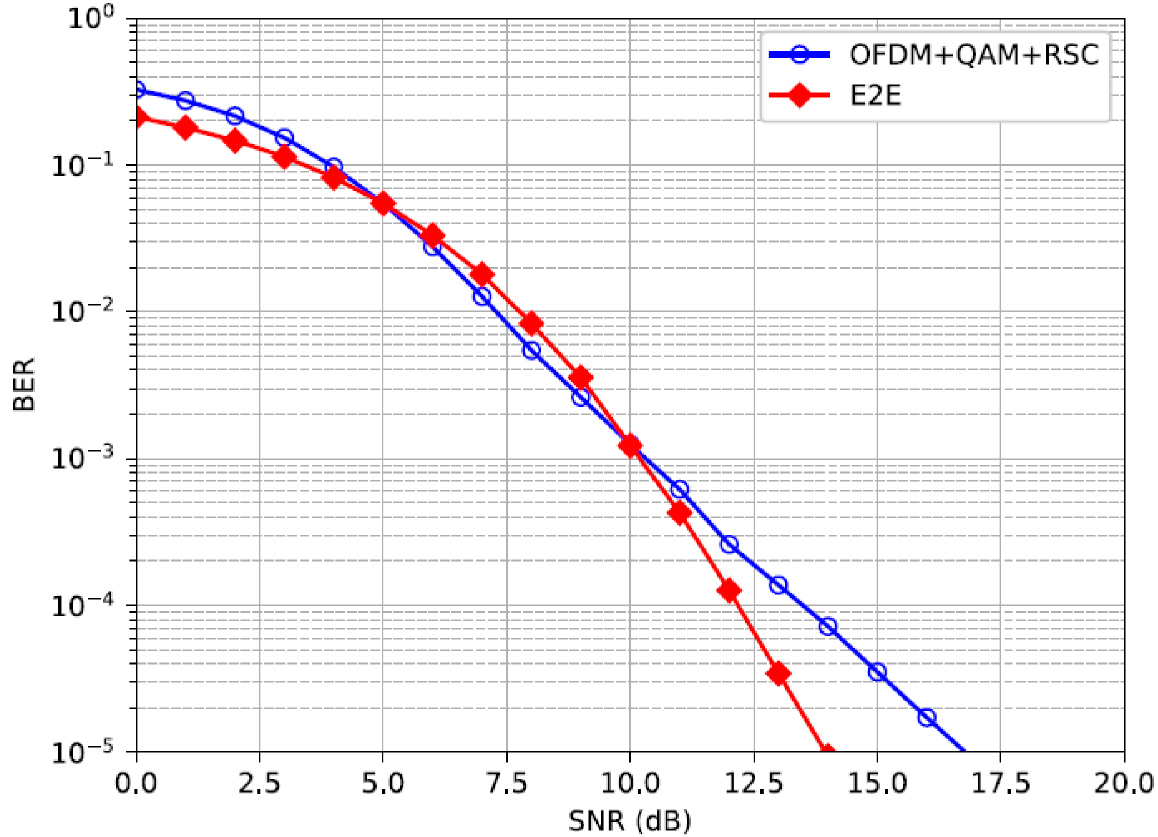


- Using CNN to address curse of dimentionity
- Conditional GAN: modelling the channel output distribution
- Surrogate of real channel when training the transmitter
- Received pilots as a part of conditioning for unknown channel

H. Ye, L. Liang, G. Y. Li, and B.-H. F. Juang, "Deep learning based end-to-end wireless communication systems with GAN as unknown channel," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3133-3143, May 2020.

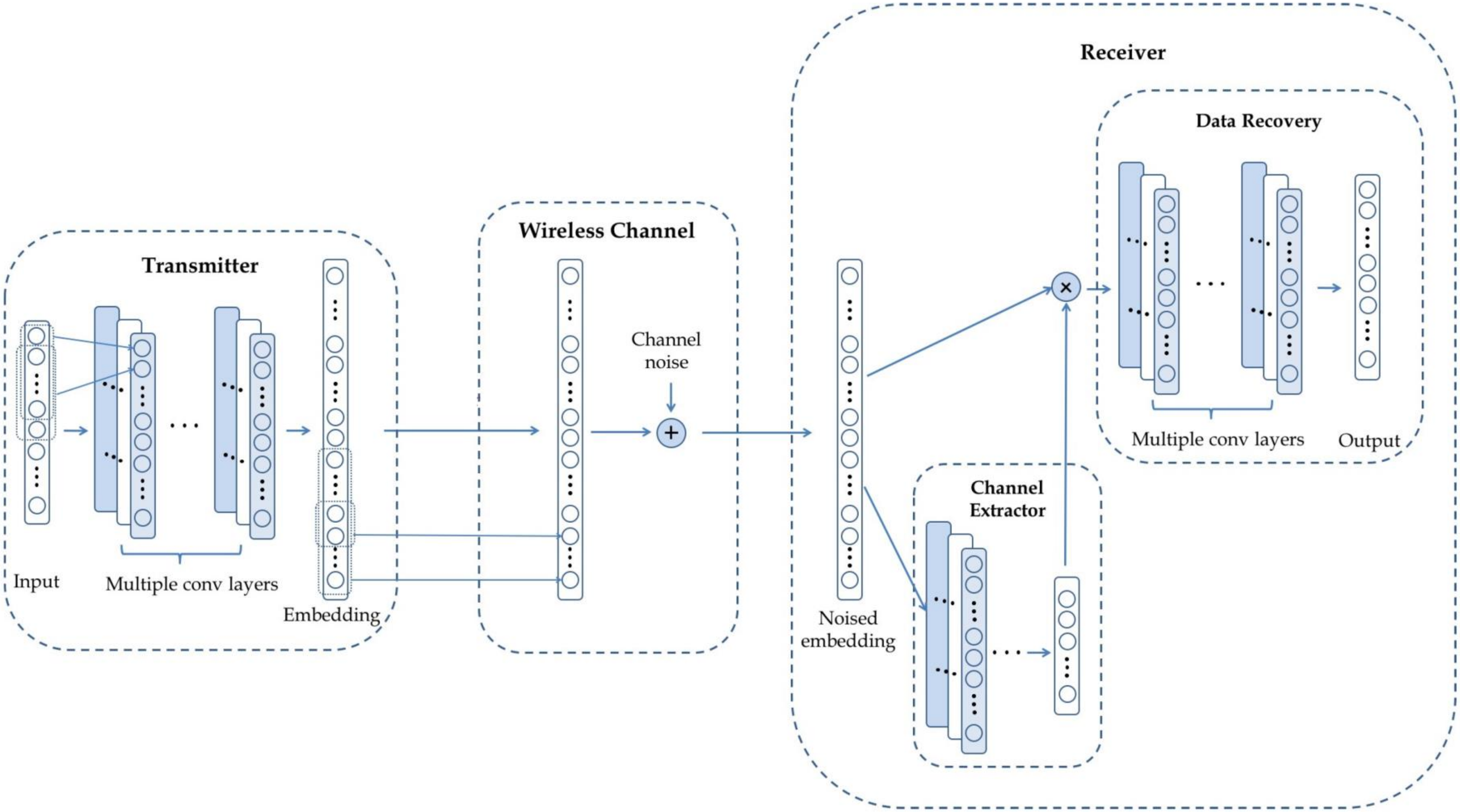
Performance for WINNER II Channels

- Similar BER at low SNR
- Better at high SNR



ITP Lab Imperial

E2E without Explicit Pilot



H. Ye, L. Liang, G. Y. Li, and B.-H. F. Juang, "Deep learning based end-to-end wireless communication systems without pilots," *IEEE Trans. Cognitive Commun. and Netw.*, vol. 7, no. 3, pp. 702 – 714, September 2021.

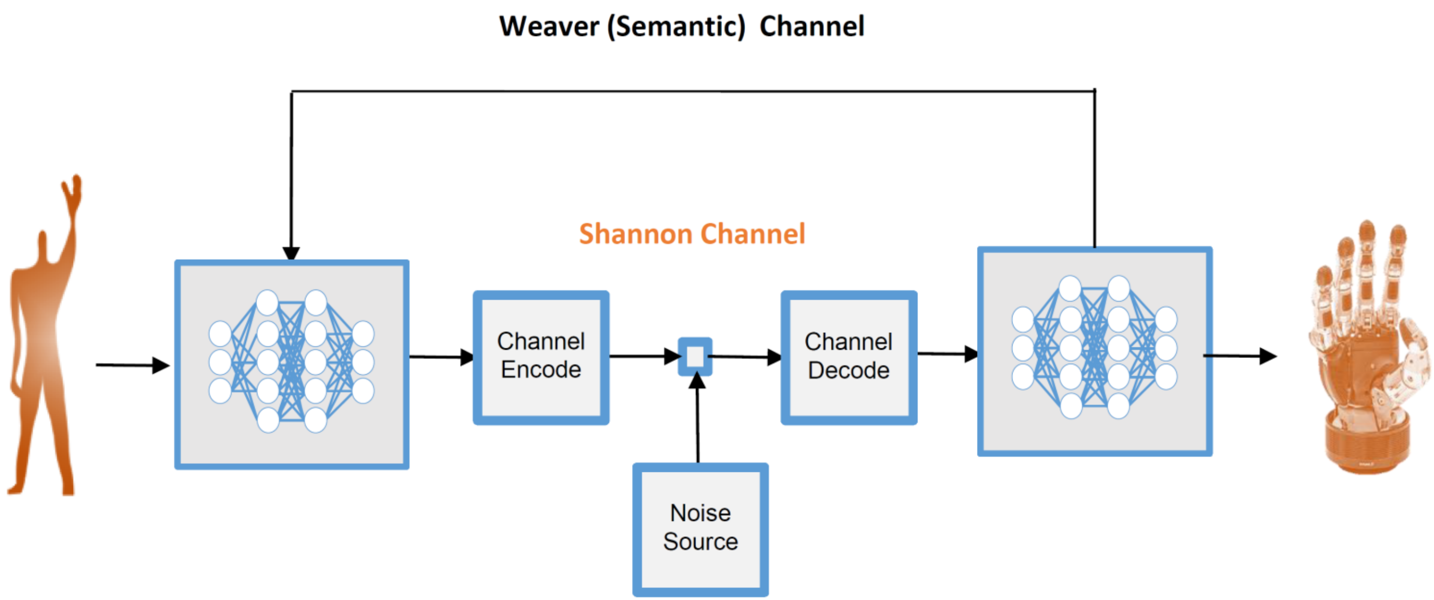
Outline

- Overview
- DL-based Conventional Communications

  * Block-Wise

  * End-to-End
- DL-based Semantic Communications
- Conclusions

Example on Semantic Communications



W. Tong and G. Y. Li "Nine critical issues in AI and wireless communications to ensure successful 6G," in *IEEE Wireless Commun.*, also at https://arxiv.org/abs/2109.11320, Aug. 2021.

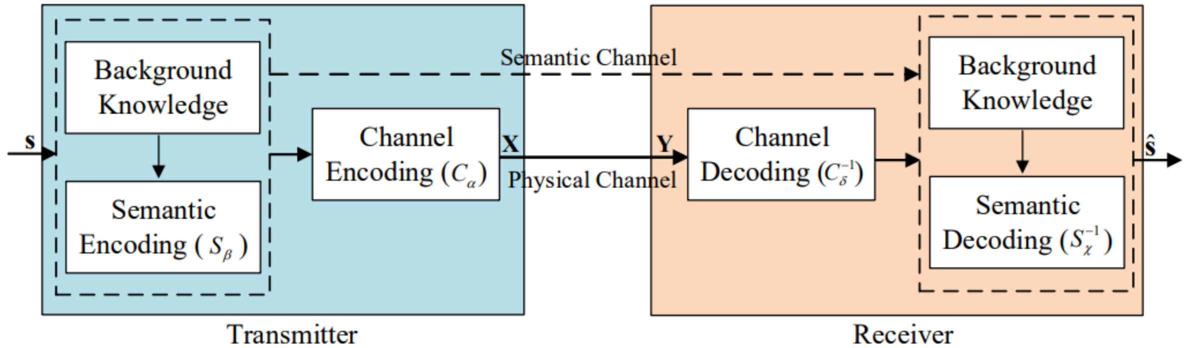
Semantic Transceiver

- **Transceiver**
  - Transmitter

    $\mathbf{X} = C_{\alpha}\left(S_{\beta}\left(\mathbf{s}\right)\right),$
  - Receiver

    $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N},$

    $\hat{\mathbf{s}} = S_{\chi}^{-1}\left(C_{\delta}^{-1}\left(\mathbf{Y}\right)\right)$



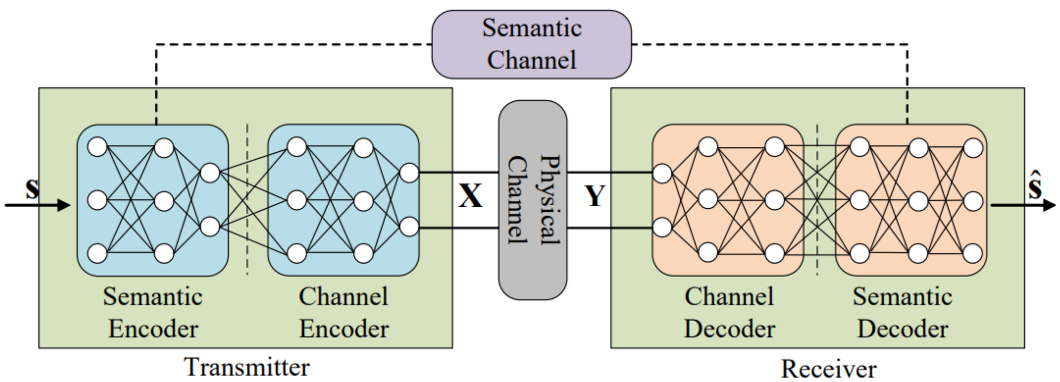
- **Channels**
  - Physical channel noise is caused by the physical channel impairment
    - AWGN, fading channels…
  - Semantic channel noise refers to misunderstanding
    - Caused by interpretation error and disturbance in estimated information.

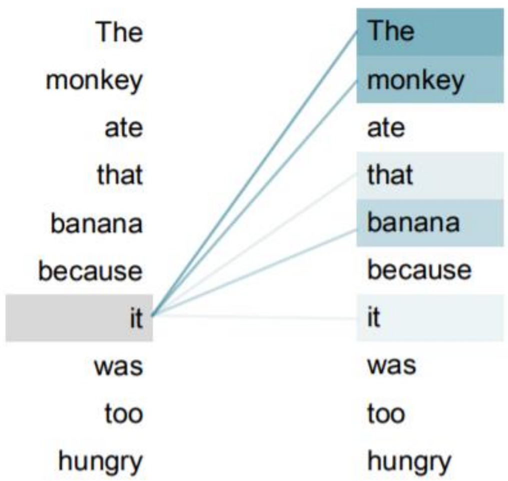
H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Trans. Signal Process.* vol. 69, pp. 2663-2675, 2021, Apr. 2021.

Transceiver Structure

Transformer based semantic communication

- Merge the traditional communication and semantic into DNNs
- Transformer can learn the semantic in text
  - e.g., “it” completes pronoun reference “the animal”





A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances Neural Info. Process. Systems (NIPS’17)*, Long Beach, CA, USA. Dec. 2017, pp. 5998–6008.

Loss Function

- Loss function used to train the transceiver

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}}(\mathbf{s}, \hat{\mathbf{s}}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\chi}, \boldsymbol{\delta}) - \lambda \mathcal{L}_{\text{MI}}(\mathbf{x}, \mathbf{y}; T, \boldsymbol{\alpha}, \boldsymbol{\beta})$$

- **Cross-Entropy**: Through reducing the loss value of channel encoder, the network can learn the syntax, phrase, the meaning of words

$$\mathcal{L}_{\text{CE}}(\mathbf{s}, \hat{\mathbf{s}}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\chi}, \boldsymbol{\delta}) = -\sum_{i=1} q(w_i) \log(p(w_i)) + (1 - q(w_i)) \log(1 - p(w_i))$$

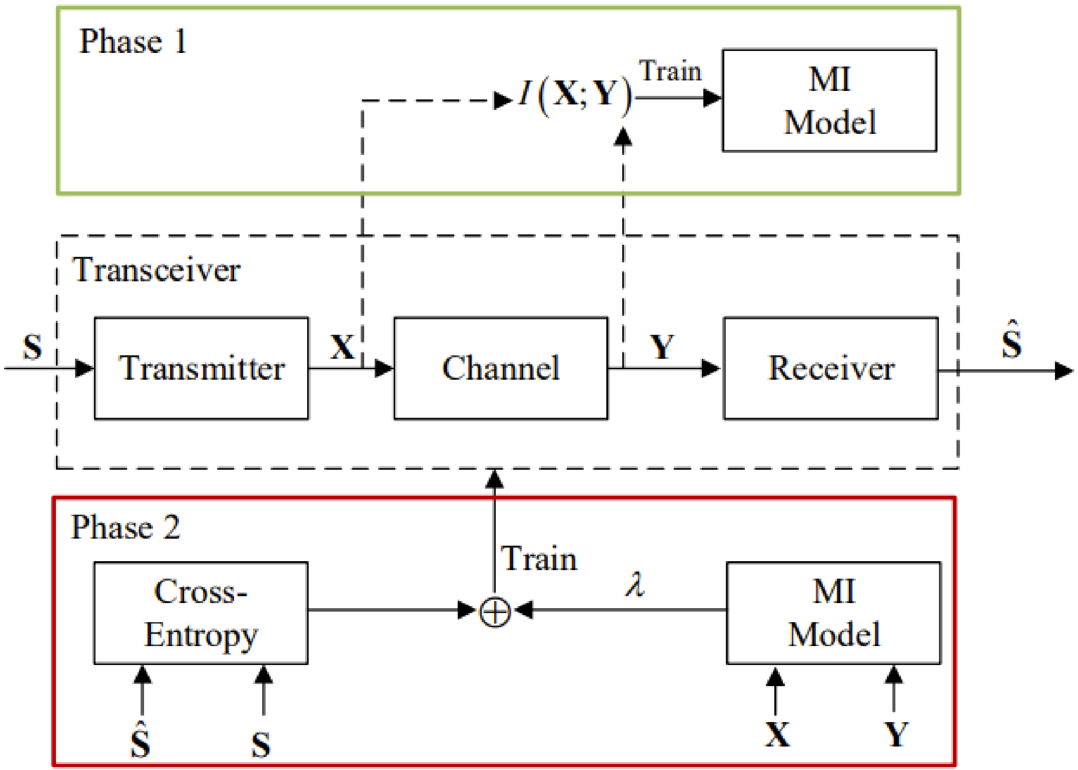
- **Mutual Information**: maximizing achieved data rate

$$\mathcal{L}_{\text{MI}}(\mathbf{X}, \mathbf{Y}; T) = \mathbb{E}_{p(x,y)}[f_T] - \log\left(\mathbb{E}_{p(x)p(y)}\left[e^{f_T}\right]\right)$$

ITP Lab Imperial

Two-Step Training

- Maximizing mutual information
- Train the whole model



ITP Lab Imperial

Performance Metrics

- **BLEU score**
  - Compare the difference between words in two sentences

$$\log \mathrm{BLEU} = \min\left(1 - \frac{l_{\hat{\mathbf{s}}}}{l_{\mathbf{s}}}, 0\right) + \sum_{n=1}^{N} u_n \log p_n,$$

  - l_{s} is the length of sentence $\mathbf{s}$, $l_{\hat{\mathrm{s}}}$ is the length of sentence $\hat{\mathbf{s}}$
  - p_n is the n-grams score, u_n is the weights of n-grams

- **Sentence Similarity**
  - Use siamese network to compute the semantic similarity

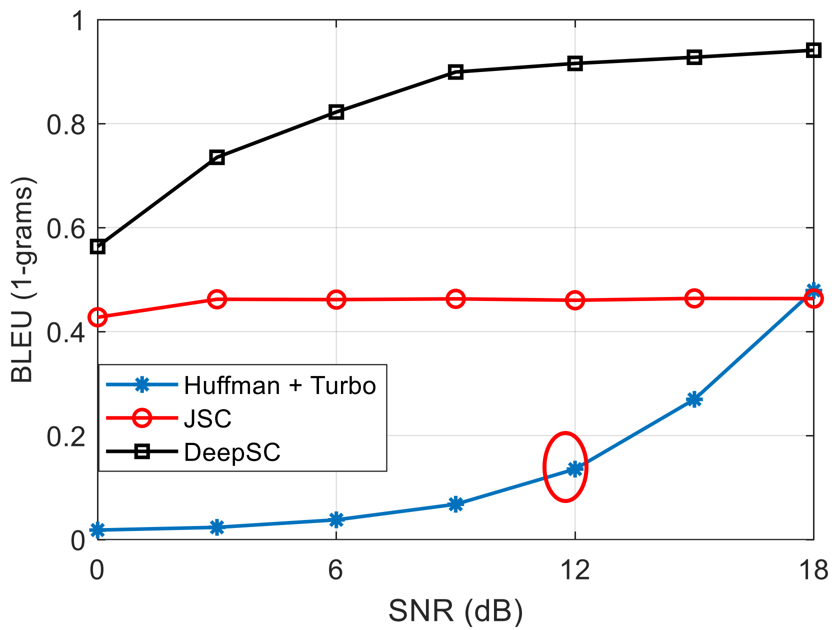
$$\mathrm{match}(\hat{\mathbf{s}}, \mathbf{s}) = \frac{\boldsymbol{B}_{\boldsymbol{\Phi}}(\mathbf{s}) \cdot \boldsymbol{B}_{\boldsymbol{\Phi}}(\hat{\mathbf{s}})^T}{\|\boldsymbol{B}_{\boldsymbol{\Phi}}(\mathbf{s})\| \, \|\boldsymbol{B}_{\boldsymbol{\Phi}}(\hat{\mathbf{s}})\|}$$

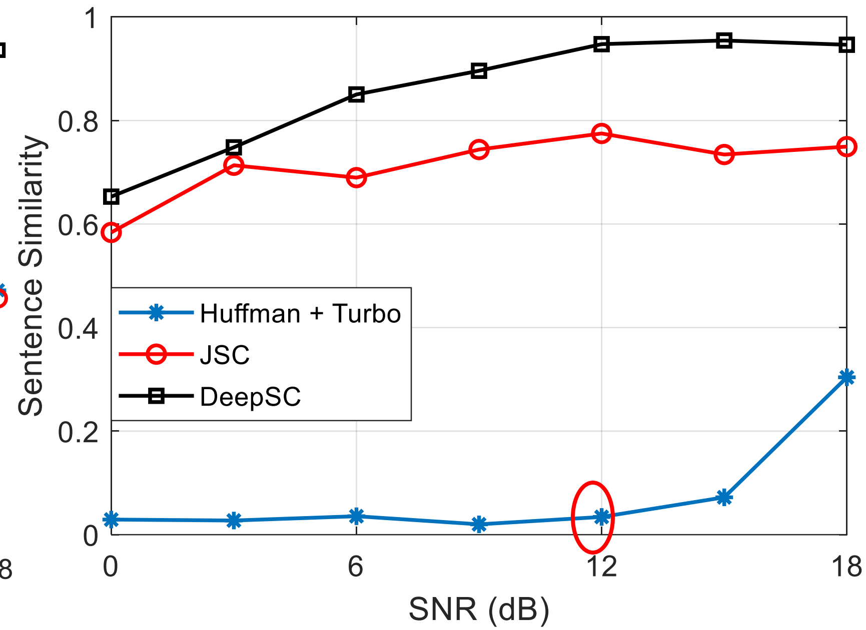
  - $\mathbf{B}_{\boldsymbol{\Phi}}(\mathrm{g})$ is the BERT model
  - Mapping sentence, $\mathbf{s}$, into **semantic vector space**, $\mathbf{B}_{\boldsymbol{\Phi}}(\mathbf{s})$, by BERT model
  - Computing similarity by measuring **distance** between $\mathbf{B}_{\boldsymbol{\Phi}}(\mathbf{s})$ and $\mathbf{B}_{\boldsymbol{\Phi}}(\hat{\mathbf{s}})$

Simulation Setting

- **Dataset**
  - The proceedings of the European Parliament
- **Proposed network architecture**
  - Transmitter:
    - 3 layers of Transformer encoder and 2 dense layers
  - Receiver:
    - 2 dense layers and 3 layers of Transformer decoder
- **Benchmark**
  - Deep Learning based joint source-channel coding (DL based JSC coding)
  - Traditional methods
    - Source coding: Huffman coding
    - Channel coding: Turbo code
    - Modulation: 64-QAM

Simulation Results





- All deep learning approaches are more competitive in the low SNR regime.
- The tendency in sentence similarity is much closer to human judgment.
  - In SNR = 12 dB, 20% BLEU score = approximate 0 sentence similarity
  - People are usually unable to understand the meaning of texts full of errors

ITP Lab
Imperial

Outline

- Overview
- DL-based Conventional Communications

  * Block-Wise

  * End-to-End
- DL-based Semantic Communications
- Conclusions

ITP Lab
Imperial

Conclusions

❑ For Conventional Communications

*Robust to nonlinear distortion, interference, & frequency selectivity

*Improving performance of iterative detectors and adapt to complicated channels

❑ End-to-end Communication Architecture

* Enabling global optimization of transceiver

* Potentially reducing the complexity

❑ Semantic Communications

*Significantly improving transmission efficiency

*Future of wireless communications