A Model-Driven Deep Learning Network for MIMO Detection

GlobalSIP, November 26-29, 2018

Hengtao He
National Mobile Communications Research Laboratory
Southeat University, Nanjing, China

Contributors:
- C.-K. Wen, National Sun Yat-sen University, Taiwan
- Shi Jin, Southeast University, Nanjing, China
- G. Y. Li, Georgia Institute of Technology, Atlanta, USA

Outline

- Introduction
- Model-Driven Deep Learning
- MIMO Detection
- OAMP-Net for MIMO Detection
- Simulation Results
- Future Works

Motivation

➢ **Challenges in current communication systems**

- ❑ Difficult channel modeling in complex scenarios:
- ❑ Demand for effective and fast signal processing
- ❑ Not adpat to communication environments dynamically

➢ **Why deep learning?**

- ❑ Without need for accurate channel models
- ❑ Distributed and parallel computing architectures
- ❑ Adapt to communication environments dynamically

T. Wang, C.-K. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, “Deep learning for wireless physical layer: Opportunities and challenges,” *China Communications*, vol. 14, no. 11, pp. 92-111, Nov. 2017.

Z.-J. Qin, H. Ye, G. Y. Li, and B.-H. Juang, “Deep learning in physical layer communications,” submitted to *IEEE Commun. Mag.*, July 2018/ also at https://arxiv.org/abs/1807.11713.

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,” submitted to *IEEE Wireless Communications*, Sept. 2018/ also at https://arxiv.org/abs/1809.06059. 2018.

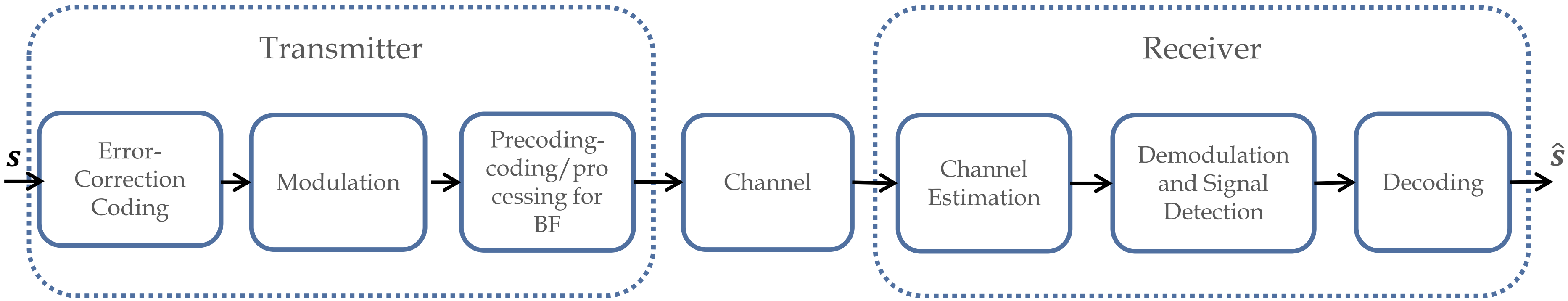
Data-Driven or Model-Driven DL?

- **Data-Driven DL**
  - Using standard neural networks (DNN, RNN, CNN,…)
  - Requiring little domain knowledge
  - Unexplainable and unpredictable neural network
  - Requiring a large amount of data
- **Model-Driven DL**
  - Relying on relatively accurate model
  - Exploiting rich domain/expert knowledge in physical layer communications
  - Explainable and predictable networks
  - Easy to train with a small amount of data

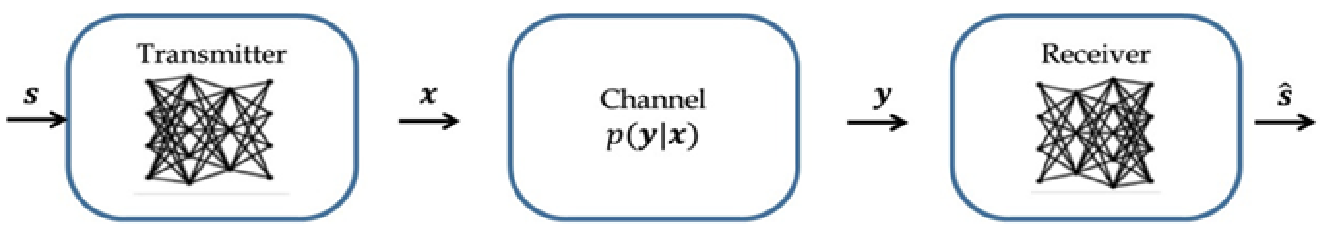
H.-T. He, S. Jin, C.-K. Wen, F.-F. Gao, G. Y. Li, and Z.-B. Xu, “Model-driven deep learning for physical layer communications,” submitted to *IEEE Wireless Communications*, Sept. 2018/ also at https://arxiv.org/abs/1809.06059. 2018.

Block Structure or End-to-End?

Block Structure

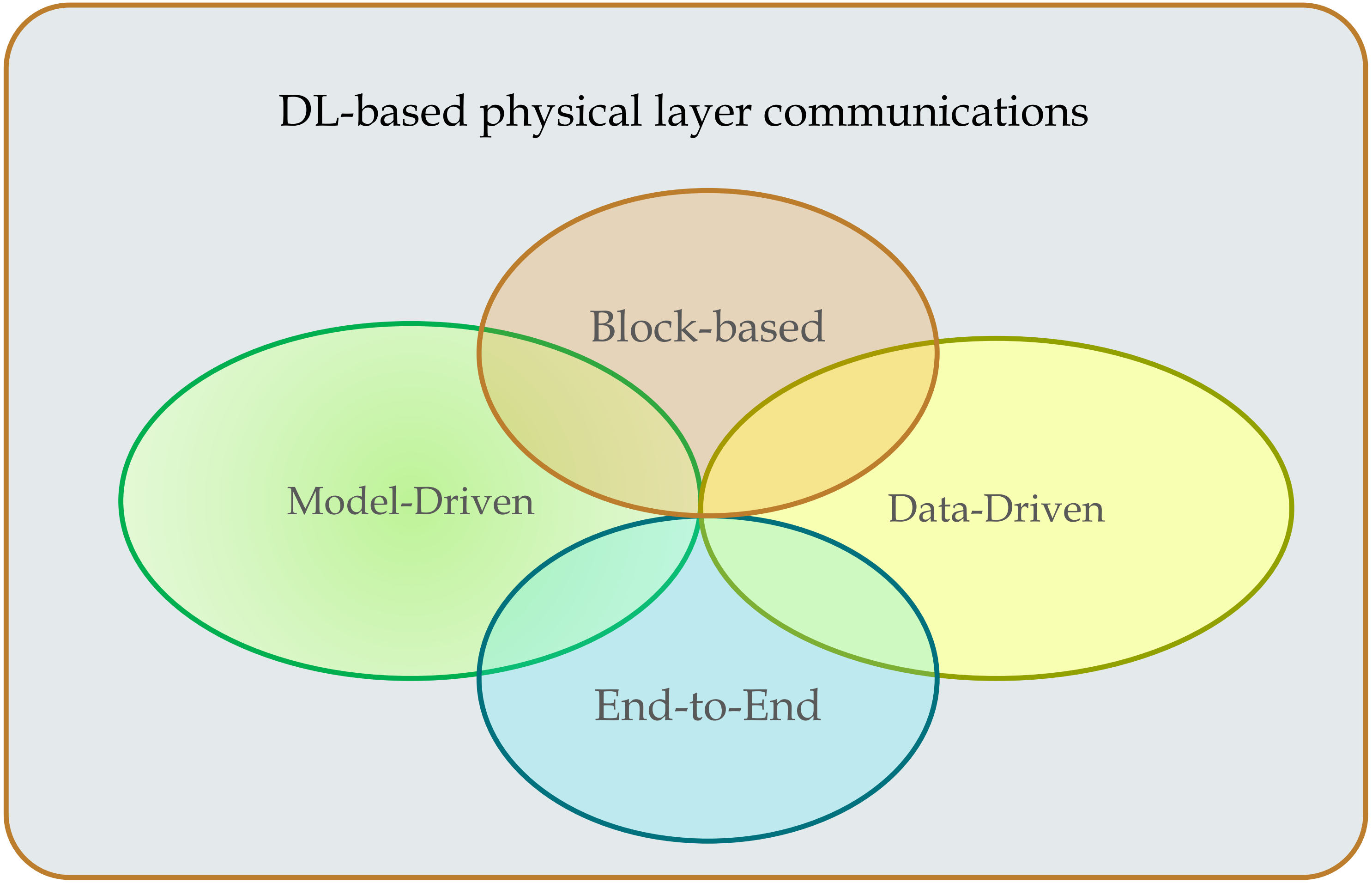


End-to-End



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.

Future Physical Layer Communications



Outline

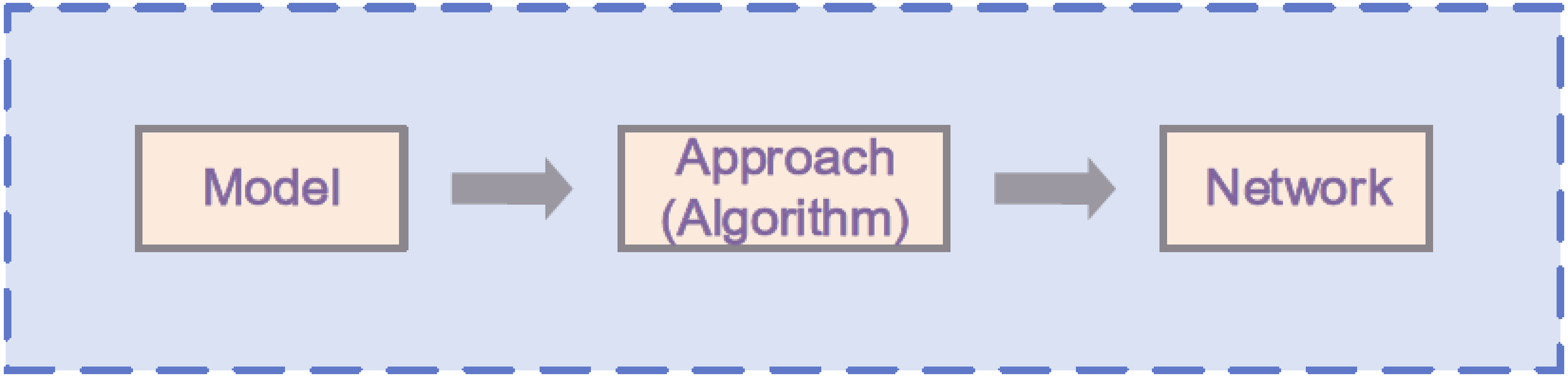
- Introduction
- Model-Driven Deep Learning
- MIMO Detection
- OAMP-Net for MIMO Detection
- Simulation Results
- Future Works

Motivation

- **Challenges of data-driven method in communications**
  - Difficulties in network design and its interpretation
  - Lack of understanding in its generation ability
  - Require a large number of data
- **Why model-driven deep learning?**
  - Design network topology with theoretical foundations
  - Communication expert knowledge can be utilized
  - Easy to train the network with a small amount of data

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," submitted to *IEEE Wireless Communications*, Sept. 2018/ also at https://arxiv.org/abs/1809.06059. 2018.

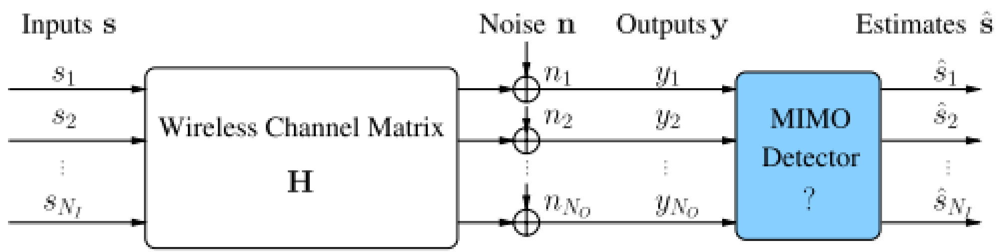
Components



- Model: based on domain knowledge, not required accurate
- Approach (Algorithm): based on the model
- Network: deep unfolding, using the algorithm as initialization, mimic conventional architecture

- Introduction
- Model-Driven Deep Learning
- MIMO Detection
- OAMP-Net for MIMO Detection
- Simulation Results
- Future Works

MIMO Detection



- **MIMO System**:

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

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

Orthogonal Approximate Message Passing (OAMP) algorithm

- Standard linear regression:

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

- Bayesian MMSE estimator:

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

- Decoupling principle:

$$\tilde{\mathcal{P}}(x_i|\mathbf{y},\mathbf{H})(i = 1, 2, \ldots, N)$$

- Equivalent AWGN channel:

$$\mathbf{r}_t = \mathbf{x} + \mathbf{w}_t$$

With $\mathbf{w}_t \sim \mathcal{N}(\mathbf{x}; \mathbf{0}, \tau_t^2\mathbf{I})$

- MMSE estimator:

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

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

Orthogonal Approximate Message Passing (OAMP)-based detector

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

$$\mathbf{r}_t = \hat{\mathbf{x}}_t + \mathbf{W}_t(\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_t), \tag{6}$$

$$\hat{\mathbf{x}}_{t+1} = \mathbb{E}\left\{\mathbf{x}|\mathbf{r}_t, \tau_t\right\}, \tag{7}$$

$$v_t^2 = \frac{\|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_t\|_2^2 - M\sigma^2}{\mathrm{tr}(\mathbf{H}^T\mathbf{H})}, \tag{8}$$

$$\tau_t^2 = \frac{1}{2N}\mathrm{tr}(\mathbf{B}_t\mathbf{B}_t^T)v_t^2 + \frac{1}{4N}\mathrm{tr}(\mathbf{W}_t\mathbf{W}_t^T)\sigma^2. \tag{9}$$

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

$$\mathbf{B}_t = \mathbf{I} - \mathbf{W}_t\mathbf{H}.$$

Real-valued channel matrix

$$\mathbf{H} = \begin{bmatrix} \Re(\bar{\mathbf{H}}) & -\Im(\bar{\mathbf{H}}) \\ \Im(\bar{\mathbf{H}}) & \Re(\bar{\mathbf{H}}) \end{bmatrix}$$

Decouple the posterior probability

$$\mathcal{P}(x_i|\mathbf{y}, \mathbf{H})(i = 1, 2, \ldots, 2N)$$

Equivalent AWGN channel:

$$\mathbf{r}_t = \mathbf{x} + \mathbf{w}_t$$

MMSE estimator: $\mathbb{E}\left\{\mathbf{x}|\mathbf{r}_t, \tau_t\right\}$

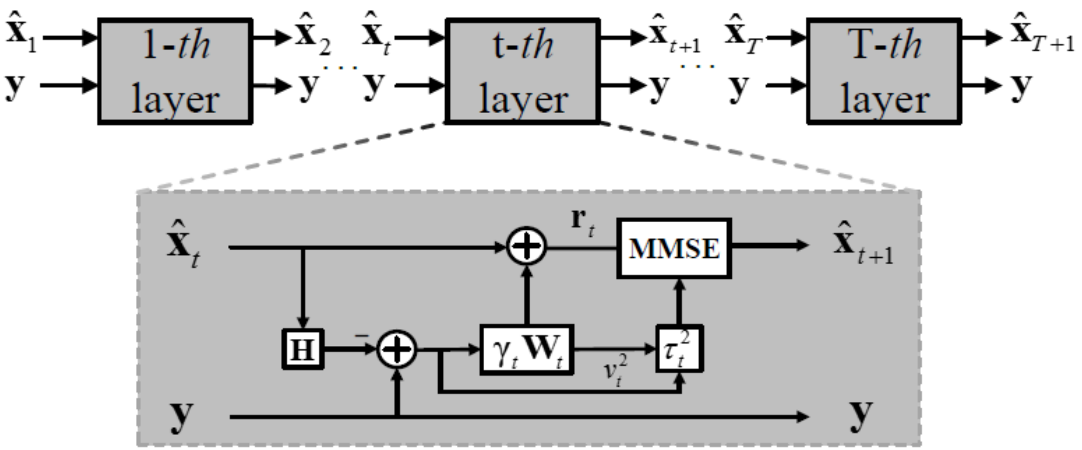
$$\mathbb{E}\left\{x_i|r_i, \tau_t\right\} = \frac{\sum_{s_i} s_i \mathcal{N}(s_i; r_i, \tau_t^2)p(s_i)}{\sum_{s_i} \mathcal{N}(s_i; r_i, \tau_t^2)p(s_i)}.$$

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

- Introduction
- Model-Driven Deep Learning
- MIMO Detection
- OAMP-Net for MIMO Detection
- Simulation Results
- Future Works

OAMP-Net for MIMO Detection

- **Architecture:**



- **Iterative Algorithms: :**

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

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

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

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

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

$$\mathbf{C}_t = \mathbf{I} - \theta_t\mathbf{W}_t\mathbf{H}$$

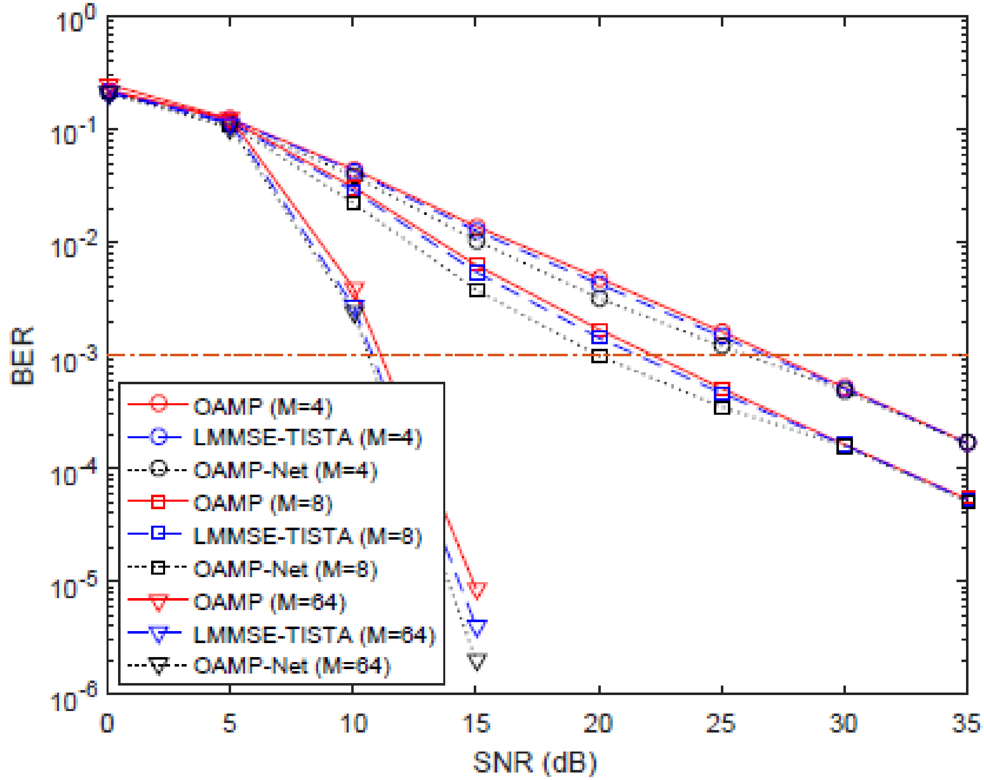
- **Tainable Parameters:** **Only two parameters (γ_t, θ_t) for each iteration!**

Why OAMP-Net work ?

- **Bayesian-optimal performances**
  - State evolution analysis
  - For large systems not for for small-size MIMO systems
  - For Rayleigh not for correlated MIMO channel
- **The effects of learned parameters** (γ_t, θ_t)
  - Provides appropriate step sizes for the update of mean and variance in the MMSE denoiser
  - Compensate for non-orthogonality of the two error(v_t^2 , τ_t^2)

- Introduction
- Model-Driven Deep Learning
- MIMO Detection
- OAMP-Net for MIMO Detection
- **Simulation Results**
- Future Works

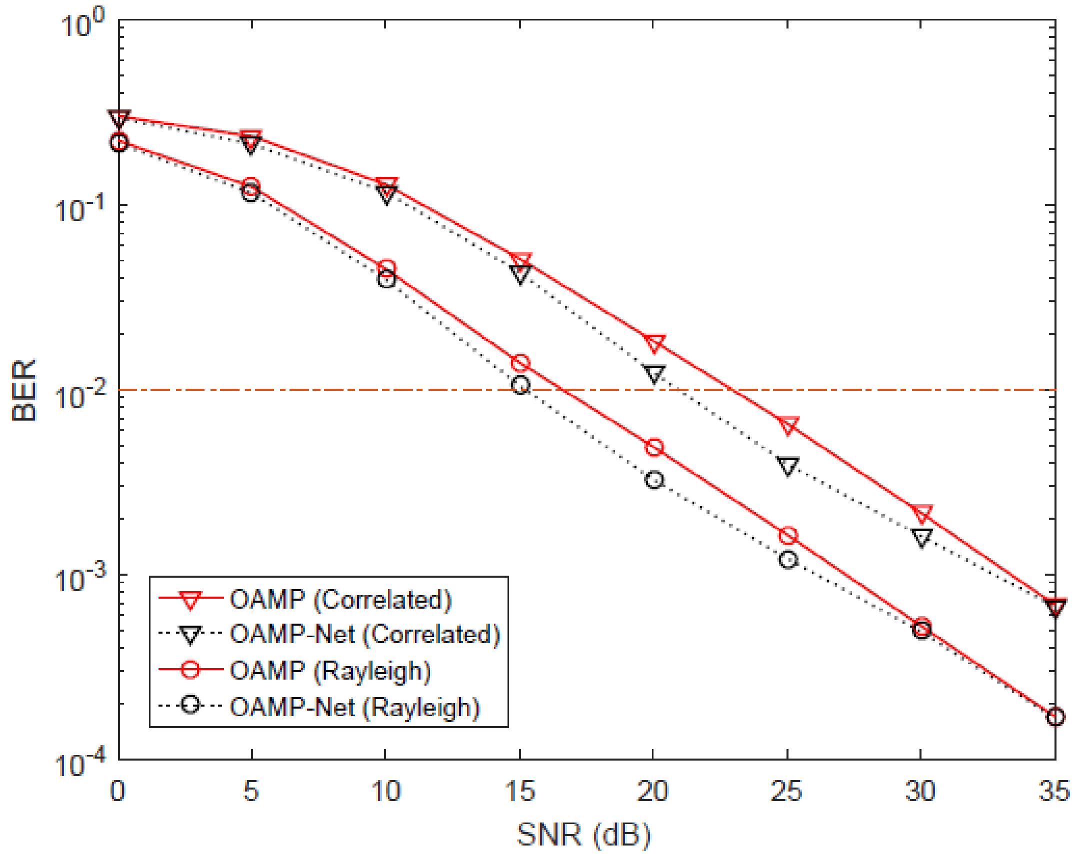
Rayleigh MIMO Channel



- OAMP-Net outperforms the OAMP algorithm and LMMSE-TISTA network
- Number of trainable variables is 2 *times of iteration number* and independent of the number of antennas N and M

D. Ito, S. Takabe, and T. Wadayama, "Trainable ISTA for sparse signal recovery," *arxiv preprint* arXiv:1801.01978, 2018.

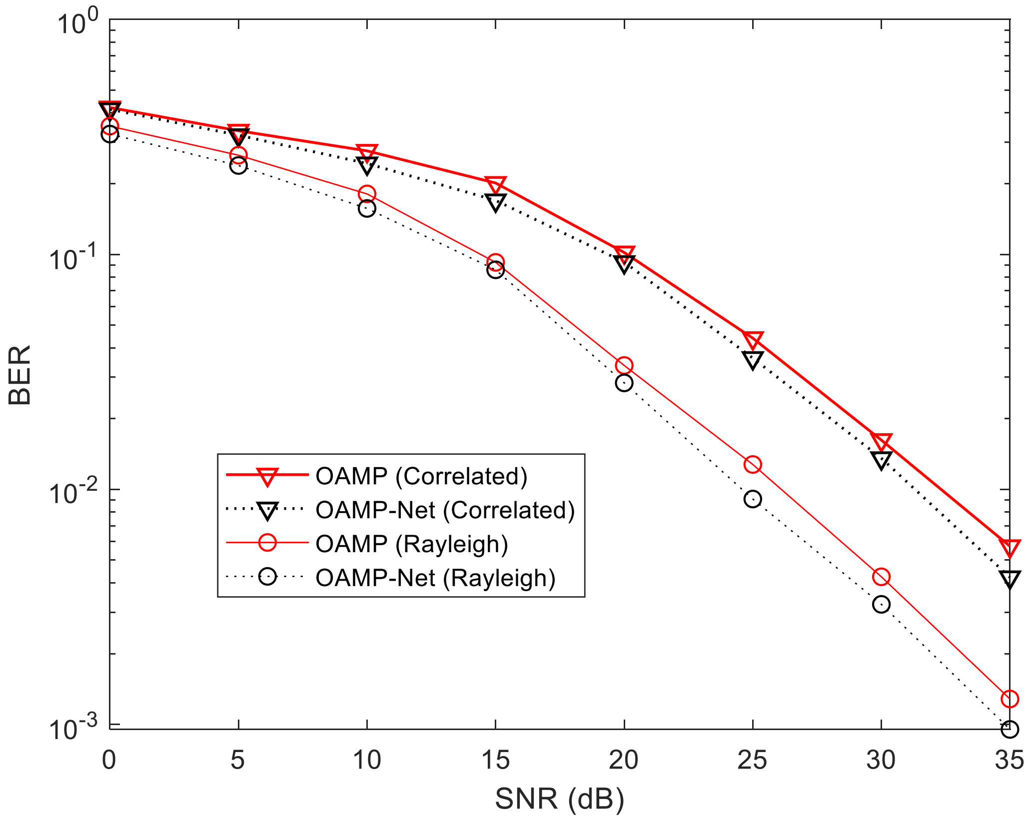
Correlated MIMO Channel



QPSK
M=N=4

- OAMP-Net outperforms the OAMP algorithm when M=N=4
- Obtain more gains under correlated MIMO channel

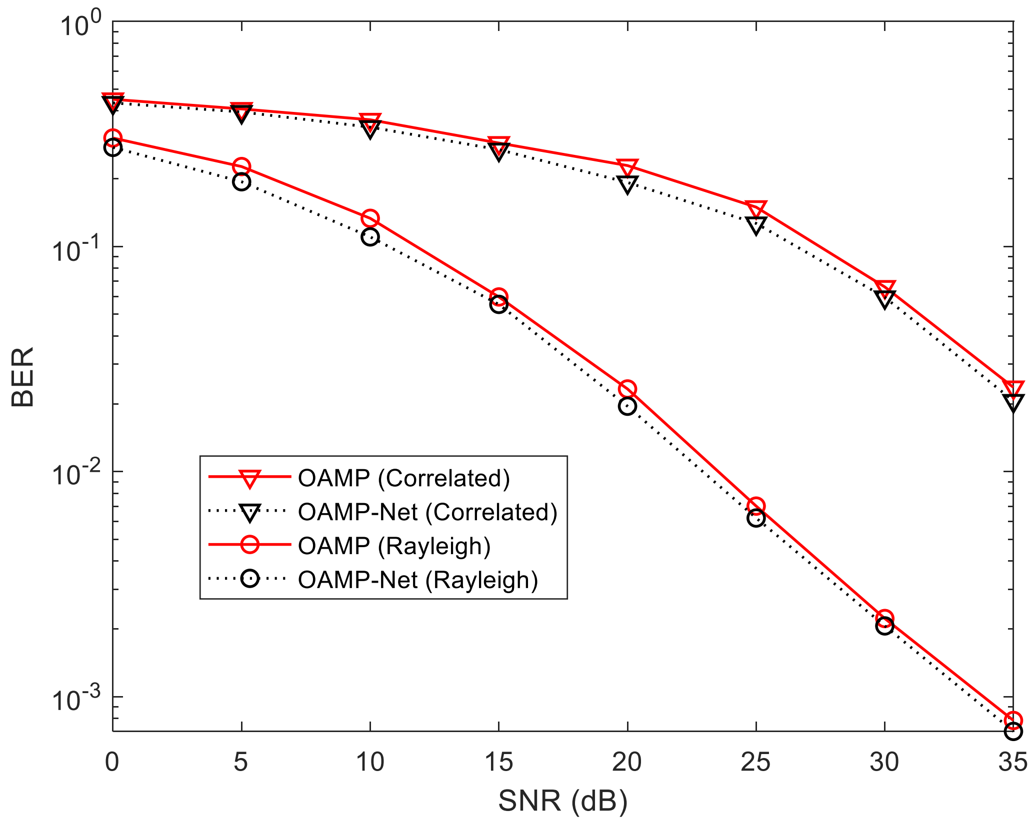
High-order Modulation



16QAM
M=N=4

- OAMP-Net outperforms the OAMP algorithm when M=N=4

High-order Modulation



64QAM
M=N=4

- OAMP-Net outperforms the OAMP algorithm when M=N=4

Future Works

- ❑ OAMP-Net for imperfect CSI

- ❑ Learn to damp

- ❑ Expectation propagation-based Network for MIMO detection:

- ❑ Low-resolution ADC architectures:

J. C´espedes, P. M. Olmos, M. S´anchez-Fernandez, and F. P´erez-Cruz, “Expectation propagation detection for high-order high-dimensional MIMO systems,” *IEEE Trans. Commun*., vol. 62, no. 8, pp. 2840-2849, Aug. 2014.

I. Santos and J. Murillo-Fuentes, “EP-based turbo detection for MIMO receivers and large-scale systems,” *arxiv preprint* arXiv:1805.05065, 2018.

H.-T. He, C.- K. Wen, and S. Jin, “Generalized expectation consistent signal recovery for nonlinear measurements,” in Proc. *IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 2333–2337.

THANKS !