MITSUBISHI ELECTRIC RESEARCH LABORATORIES (MERL)
Cambridge, Massachusetts

Learning to Modulate for Non-coherent MIMO

Ye Wang, Toshiaki Koike-Akino

IEEE ICC 2020
International Conference on Communications

This document does not contain Technology as defined in EAR Part 772.

Overview

- Apply machine learning to design modulation and detection
  - Target: space-time constellations for non-coherent MIMO channel
  - Recent trend: encode/decode with Deep Neural Networks (DNN)
- Simulation-driven, end-to-end, encoder/decoder optimization
  - Minimizing cross-entropy loss $\Leftrightarrow$ maximizing mutual info
  - We compare DNN-based vs DNN-free systems
- Learned schemes can outperform traditional designs at some SNRs
  - DNNs can be avoided altogether while keeping similar performance
  - Feasibility of non-coherent MIMO with only two time slots

Non-Coherent MIMO Channel

Using m Tx and n Rx antennas over L time slots

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z}$$

- Signal $\mathbf{X} \in \mathbb{C}^{m \times L}$ sent over m Tx antennas and L time slots
- Unknown block fading channel $\mathbf{H} \in \mathbb{C}^{n \times m} \overset{\text{iid}}{\sim} \mathcal{CN}(0, 1/m)$
- Gaussian noise $\mathbf{Z} \in \mathbb{C}^{n \times L} \overset{\text{iid}}{\sim} \mathcal{CN}(0, \sigma^2)$
- Power constraint: $\mathbb{E}[\|\mathbf{X}\|^2/(mL)] = 1$, average SNR $= 1/\sigma^2$
- Receive $\mathbf{Y} \in \mathbb{C}^{n \times L}$ on n Rx antennas

Goal: design k-bit modulation and non-coherent detection scheme

$$\mathbf{m} \in \{0,1\}^k \rightarrow \boxed{\text{ENC}} \rightarrow \mathbf{X} \rightarrow \boxed{\text{MIMO}} \rightarrow \mathbf{Y} \rightarrow \boxed{\text{DEC}} \rightarrow \hat{\mathbf{m}}$$

Encoding Signal Constellation via a Lookup Table

For small k (#bits), lookup table most effective and efficient

- Encoder : $\{0,1\}^k \to \mathbb{C}^{m\times L}$ fully captured by table $\mathbf{C} \in \mathbb{C}^{2^k \times m \times L}$
- Subtract centroid (mean across first axis) to get centered $\overline{\mathbf{C}}$
- Normalize average power of codebook $\widetilde{\mathbf{C}} := \overline{\mathbf{C}}\sqrt{2^k mL}/\|\overline{\mathbf{C}}\|$
- For message $\mathbf{m} \in \{0,1\}^k$, signal $\mathbf{X}_\mathbf{m}$ is selected from $\widetilde{\mathbf{C}}$

Encoding Signal Constellation via a Lookup Table

For small k (#bits), lookup table most effective and efficient

- Encoder : $\{0,1\}^k \to \mathbb{C}^{m \times L}$ fully captured by table $\mathbf{C} \in \mathbb{C}^{2^k \times m \times L}$
- Subtract centroid (mean across first axis) to get centered $\overline{\mathbf{C}}$
- Normalize average power of codebook $\widetilde{\mathbf{C}} := \overline{\mathbf{C}}\sqrt{2^k mL}/\|\overline{\mathbf{C}}\|$
- For message $\mathbf{m} \in \{0,1\}^k$, signal $\mathbf{X_m}$ is selected from $\widetilde{\mathbf{C}}$

Common alternative: DNN-encoder with one-hot message input

- Vastly over-parameterizes the codebook with extraneous layers
- Binary input encoding would scale better for large k
- Our work uses lookup table to avoid encoder DNN

Decoding with or without DNN

We optimize two soft-decision decoders: DNN-based vs DNN-free

- Both output unnormalized, log-likelihoods for each message
- Apply softmax to yield approximate posterior $P^{\theta}_{\mathbf{m}|\mathbf{Y}}$

Decoding with or without DNN

We optimize two soft-decision decoders: DNN-based vs DNN-free

- Both output unnormalized, log-likelihoods for each message
- Apply softmax to yield approximate posterior $P^{\theta}_{\mathbf{m}|\mathbf{Y}}$

Neural Network (NN) Decoder: θ is network parameters

- Used Multi-Layer Perceptron (MLP) or Residual MLP (ResMLP)
- Blind decoder trained end-to-end with cross-entropy loss

Decoding with or without DNN

We optimize two soft-decision decoders: DNN-based vs DNN-free

- Both output unnormalized, log-likelihoods for each message
- Apply softmax to yield approximate posterior $P^{\theta}_{\mathbf{m}|\mathbf{Y}}$

Neural Network (NN) Decoder: θ is network parameters

- Used Multi-Layer Perceptron (MLP) or Residual MLP (ResMLP)
- Blind decoder trained end-to-end with cross-entropy loss

Pseudo-ML (pML) Decoder: based on orthonormal code ML decoding

- If codeword orthonormal, i.e., $\forall \mathbf{m}, \mathbf{X}_{\mathbf{m}} \mathbf{X}_{\mathbf{m}}^{\dagger} = L \cdot \mathbf{I}_{m}$, then $\|\mathbf{Y}\mathbf{X}_{\mathbf{m}}^{\dagger}\|^2$ is proportional to unnormalized, log-likelihood $\log \alpha P(\mathbf{Y}|\mathbf{m})$
- Outputs $\{\theta \|\mathbf{Y}\mathbf{X}_{\mathbf{m}}^{\dagger}\|^2\}_{\mathbf{m}\in\{0,1\}^k}$ where $\theta > 0$ captures confidence
- Requires additional codebook orthonormality constraint

Cross-Entropy Loss Training Maximizes Mutual Info

$$\mathbf{m} \rightarrow \boxed{\mathsf{ENC}_{\mathbf{C}}} \rightarrow \mathbf{X}_{\mathbf{m}} \rightarrow \boxed{\mathsf{MIMO}} \rightarrow \mathbf{Y} \rightarrow \boxed{\mathsf{DEC}_{\theta}} \rightarrow P^{\theta}_{\mathbf{m}|\mathbf{Y}}$$

End-to-end optimization with cross-entropy loss

$$\min_{\mathbf{C},\theta} \mathbb{E}\left[-\log P^{\theta}_{\mathbf{m}|\mathbf{Y}}(\mathbf{m}|\mathbf{Y})\right]$$

Cross-Entropy Loss Training Maximizes Mutual Info

$$\mathbf{m} \to \boxed{\mathsf{ENC}_{\mathbf{C}}} \to \mathbf{X}_{\mathbf{m}} \to \boxed{\mathsf{MIMO}} \to \mathbf{Y} \to \boxed{\mathsf{DEC}_{\theta}} \to P^{\theta}_{\mathbf{m}|\mathbf{Y}}$$

End-to-end optimization with cross-entropy loss

$$\min_{\mathbf{C},\theta} \mathbb{E}\left[-\log P^{\theta}_{\mathbf{m}|\mathbf{Y}}(\mathbf{m}|\mathbf{Y})\right]$$

Equivalent to maximizing mutual info $\mathcal{I}(\mathbf{m};\mathbf{Y})$ since

- $\mathbb{E}\left[-\log P^{\theta}_{\mathbf{m}|\mathbf{Y}}(\mathbf{m}|\mathbf{Y})\right] = \mathcal{H}(\mathbf{m}|\mathbf{Y}) + \mathrm{KL}(P_{\mathbf{m}|\mathbf{Y}} \| P^{\theta}_{\mathbf{m}|\mathbf{Y}})$
- $\mathcal{I}(\mathbf{m};\mathbf{Y}) = \mathcal{H}(\mathbf{m}) - \mathcal{H}(\mathbf{m}|\mathbf{Y})$, with constant $\mathcal{H}(\mathbf{m}) = k$

Cross-Entropy Loss Training Maximizes Mutual Info

$$\mathbf{m} \to \boxed{\mathsf{ENC}_{\mathbf{C}}} \to \mathbf{X}_{\mathbf{m}} \to \boxed{\mathsf{MIMO}} \to \mathbf{Y} \to \boxed{\mathsf{DEC}_{\theta}} \to P^{\theta}_{\mathbf{m}|\mathbf{Y}}$$

End-to-end optimization with cross-entropy loss

$$\min_{\mathbf{C},\theta} \mathbb{E}\left[-\log P^{\theta}_{\mathbf{m}|\mathbf{Y}}(\mathbf{m}|\mathbf{Y})\right]$$

Equivalent to maximizing mutual info $\mathcal{I}(\mathbf{m};\mathbf{Y})$ since

- $\mathbb{E}\left[-\log P^{\theta}_{\mathbf{m}|\mathbf{Y}}(\mathbf{m}|\mathbf{Y})\right] = \mathcal{H}(\mathbf{m}|\mathbf{Y}) + \mathrm{KL}(P_{\mathbf{m}|\mathbf{Y}} \| P^{\theta}_{\mathbf{m}|\mathbf{Y}})$
- $\mathcal{I}(\mathbf{m};\mathbf{Y}) = \mathcal{H}(\mathbf{m}) - \mathcal{H}(\mathbf{m}|\mathbf{Y})$, with constant $\mathcal{H}(\mathbf{m}) = k$

For pML decoder, we also enforce orthonormality with soft penalty

$$\min_{\mathbf{C},\theta} \mathbb{E}\left[-\log P^{\theta}_{\mathbf{m}|\mathbf{Y}}(\mathbf{m}|\mathbf{Y})\right]\left(1 + \lambda \ell(\mathbf{C})\right), \quad \lambda > 0$$

$$\ell(\mathbf{C}) := \frac{1}{2^k m^2} \sum_{\mathbf{m}\in\{0,1\}^k} \left\| \mathbf{X}_{\mathbf{m}} \mathbf{X}_{\mathbf{m}}^{\dagger} / L - \mathbf{I}_m \right\|^2$$

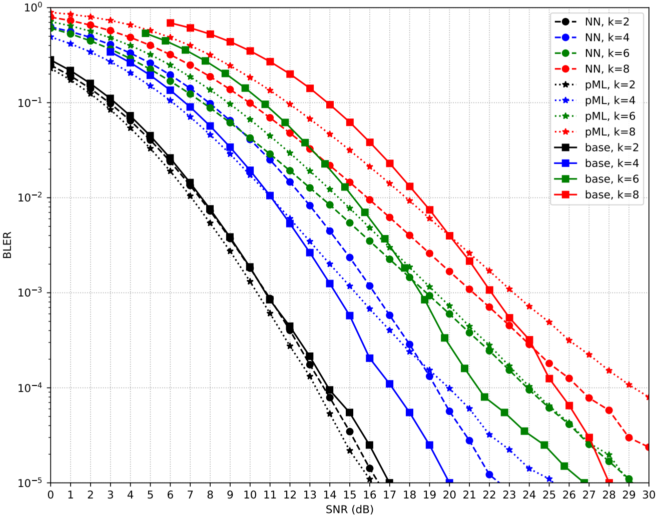
Experimental Evaluation

- Non-coherent MIMO parameters
  - Bits $k \in \{2, 4, 6, 8\}$, time slots $L \in \{2, 4\}$
  - RX $n \in \{2, 3, 4\}$, TX $m = n$ for $L = 4$, TX $m = 2$ for $L = 2$
- Encoders/decoders optimized over a range of hyperparameters
  - MLP: varied depth and width of fully-connected layers
  - ResMLP: MLP with additional skip connections and batch-norm
  - pML: varied λ parameter controlling code orthonormality
- Compare performance vs codes constructed by [Liang, Xia '02]
  - Existing Grassmann code designs require $L > m$
  - We demonstrate novel feasibility of learning for $L = m = 2$
- Cross-entropy loss gives approximate lower-bound on throughput:

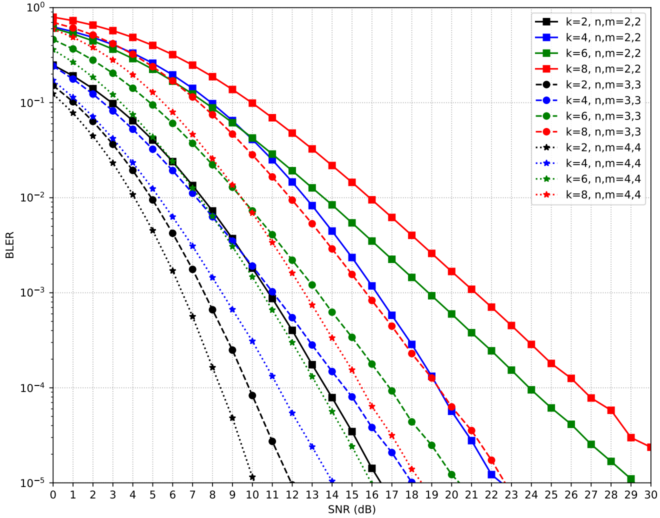
$$\frac{k - \mathbb{E}\left[-\log P^{\theta}_{\mathbf{m}|\mathbf{Y}}(\mathbf{m}|\mathbf{Y})\right]}{L} \lessapprox \frac{I(\mathbf{m};\mathbf{Y})}{L}$$

  - Compare with capacity lower-bounds of [Yang, Durisi, Riegler '13]

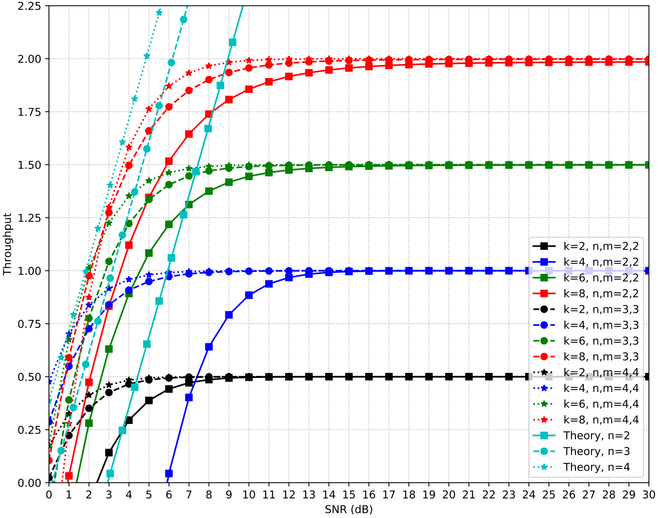
Comparison for $L = 4$ Time Slots, TX-RX $(m, n) = (2, 2)$



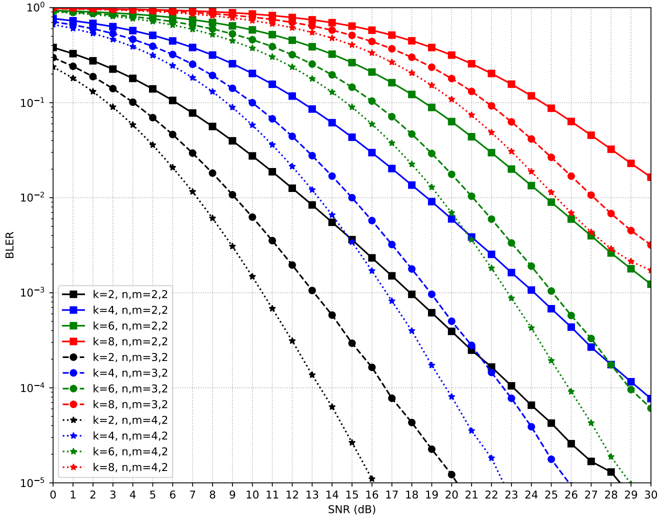
NN Decoder Performance for $L = 4$



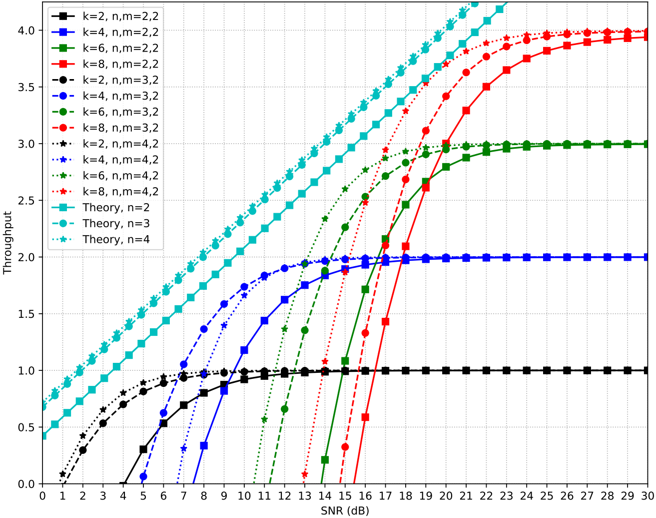
NN Decoder Throughput for $L = 4$



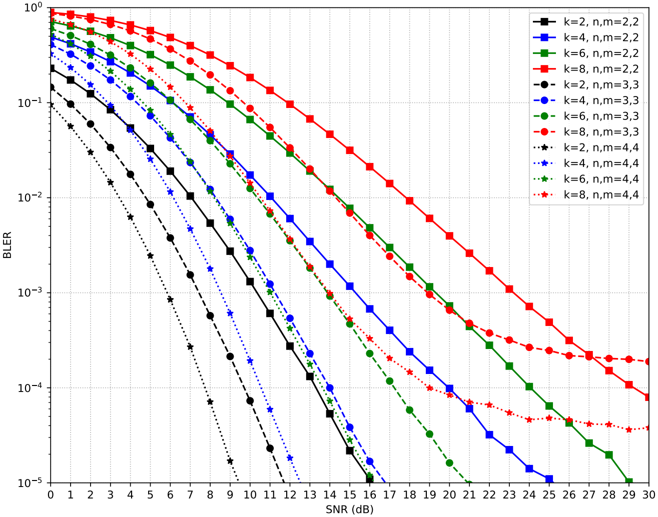
NN Decoder Performance for $L = 2$



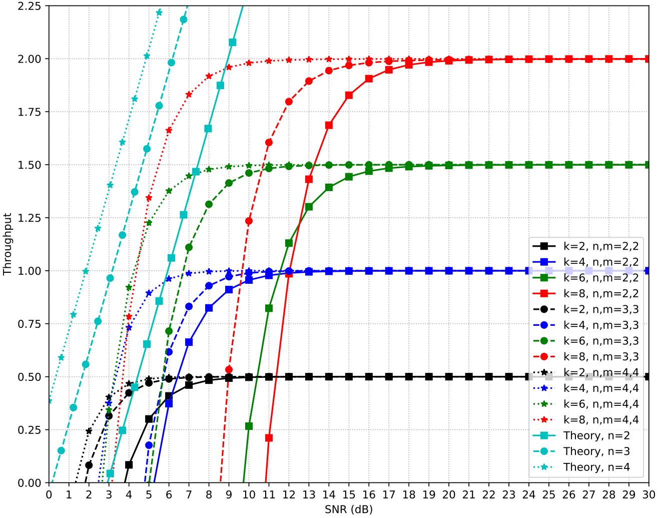
NN Decoder Throughput for $L = 2$



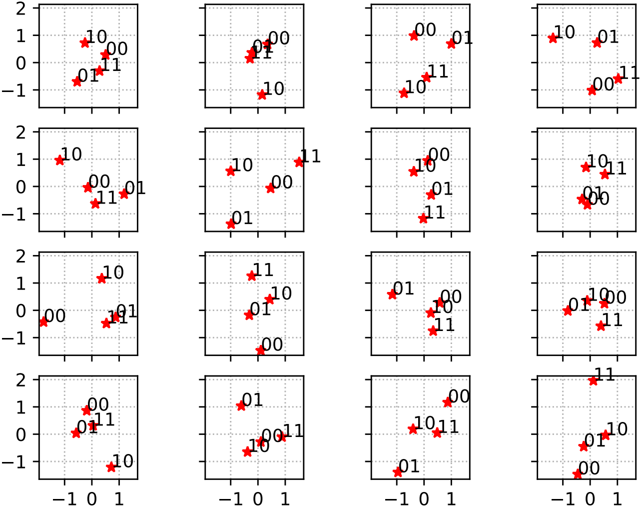
pML Decoder Performance for $L = 4$



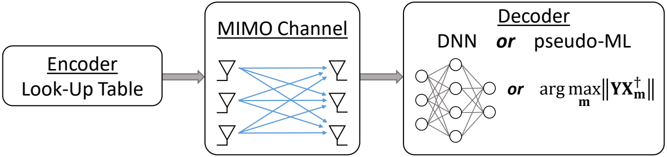
pML Decoder Throughput for $L = 4$



Codebook via pML Decoder for $k = 2$, $L = 4$, $(m, n) = (4, 4)$



Summary



- Applied learning to non-coherent MIMO modulation and detection
  - Simulation-driven, end-to-end, space-time constellation optimization
  - Minimizing cross-entropy loss $\Leftrightarrow$ maximizing mutual info
  - We offer a DNN-free system as alternative to using DNN
- Learned schemes can outperform traditional designs at some SNRs
  - DNNs can be avoided altogether while keeping similar performance
  - Non-coherent MIMO feasible with only two time slots