ON EQUIVALENCE OF NEURAL NETWORK RECEIVERS

Maqsood Careem*, Prof. Aveek Dutta*, Dr. Ngwe Thawdar**

*University at Albany SUNY, Albany, NY

**US Air Force Research Laboratory, Rome, NY

Motivation & Problem Statement

- Classical (MAP) Rx:
  - Optimal in BER if channel model known & stationary
  - Efficient & optimal if channel is AWGN: **Algorithm Deficit** [1]
  - Not optimal if channel model is unknown: **Model Deficit** [1]
- Neural Network (NN) Rx typically as Black Box Systems:
  - NN transceivers are more robust
  - Lacks bounds on BER→ Approach MAP performance for given data-set [2,3]
  - Lacks efficient model design → High hardware & training complexity [4,5]

Overview of Wireless Transceivers

Options for wireless transceivers	Explainable	Robust to any channel	BER Guarantees
Classical approaches: SISO: [10,11,12] MIMO: [13,14,15]	✓		(✓)

(✓) *Holds for certain channel conditions*

Problem Statement: *To mathematically bridge the gap in explainability of a NN Rx → To achieve guarantees on BER & Structure and training complexity*

[1] O. Simeone, "A very brief introduction to machine learning with applications to communication systems"
[2] H.Kim, "Communication Algorithms via Deep Learning "
[3] T. Gruber, "On Deep Learning-Based Channel Decoding"
[4] T. Wang, "Deep Learning for Wireless Physical Layer: Opportunities and Challenges"
[5] F. Restuccia, "Physical-Layer Deep Learning: Challenges and Applications to 5G and Beyond"

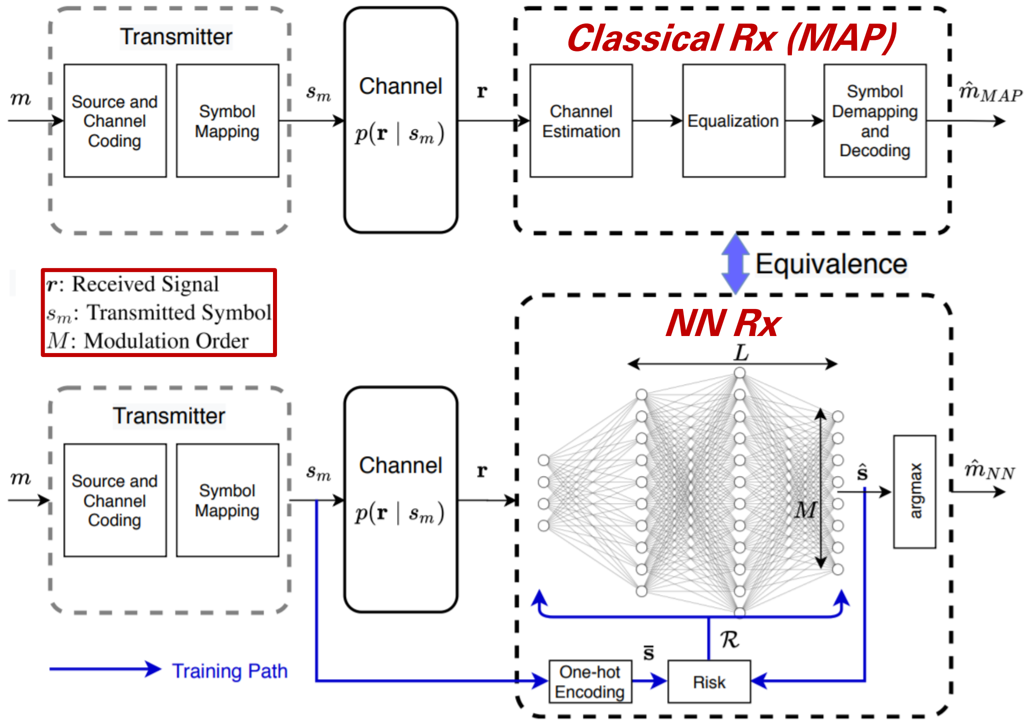
Towards Explainable NN-based Receivers

- ***Explainability:*** Math of how a NN classifies symbols → Guarantees on BER & complexity
- Challenging for channels[6] → Derive ***Equivalent*** receiver
- Equivalence is empirically understood[7] → Needs proof
- Challenging Task:

    (i) Data-driven (empirical) vs Model-driven (Bayesian)

    (ii) Parameter optimization vs Output optimization

Allows us to understand disparity of NN & Classical Rx
→ Derive tight BER bounds based on NN structure and training



[6] Guo, “Explainable Artificial Intelligence (XAI) for 6G: Improving Trust between Human and Machine”
[7] Gruber et al., “On Deep Learning-Based Channel Decoding”

Equivalence Under Complete Training

Classical Rx:

MAP Rx function

$$\hat{m}_{MAP} = \arg\max_{m} f_{MAP}(\boldsymbol{r}); \quad f_{MAP}(\boldsymbol{r}) := \mathrm{P}[s_m|\boldsymbol{r}]$$

Posterior

MAP estimate is the symbol with maximum Posterior Probability

NN-Rx with Complete Training ($n \rightarrow \infty$):

Assumption: Stationary channel

(i) Law of Large Numbers (ii) Universal Approximation[8]

MSE Risk

$$\mathcal{R}(f_{NN}(\boldsymbol{r};\delta_n), \bar{\boldsymbol{s}}) = \frac{1}{n}\left\| f_{NN}(\boldsymbol{r};\delta_n) - \bar{\boldsymbol{s}} \right\|^2 \rightarrow [\mathbb{E}\left\{\sum_{m=1}^{M}[f_{NN}(\boldsymbol{r};\delta_n) - \mathbb{E}\{\bar{s}_m|\boldsymbol{r}\}]^2\right\} + \mathbb{E}\left\{\sum_{m=1}^{M}\mathrm{var}\{\bar{s}_m|\boldsymbol{r}\}\right\}]$$

 # Training samples ***NN Function*** ***Symbol Labels***

Posterior Mean

Posterior Probability $\xrightarrow{\arg\max}$ ***Symbol Estimate***

Design Criteria:

a. Outputs: Mx1 vector
b. Outputs add to 1
c. One-hot encoded labels

Lemma 1: NN and Classical Rx are equivalent under above design criteria for known and stationary channels

Lemma 2: The BER of NN is related to BER of Classical Rx for known and stationary channels as,

$$BER_{NN} \geqslant BER_{MAP}$$

[8] Cybenko, "Approximation by superpositions of a sigmoidal function,"

Risk Statistics Under Incomplete Training

Incomplete Training: *Lemmas 3 & 4: Risk is Gaussian: Due to randomness in data, NN architecture, NN complexity*

$$\bar{\mathcal{R}}(f_{NN}(\boldsymbol{r};\delta_n^*))\sim\mathcal{N}(\mu_n,\sigma_n^2+\sigma_h^2)$$

NN Statistics ***Channel Statistic***

Channel Statistics:

Calculated from data (or channel if known)

$$\sigma_h^2=\mathbb{E}\left[\sum_{m=1}^{M}\mathrm{var}\left\{\bar{s}_m|\boldsymbol{r}\right\}\right]$$

If Channel Model is known (E.g., AWGN): Proposition 1

$$\sigma_h^2=\frac{M\lambda A_M}{\gamma_b\log_2 M}\quad where,\quad \lambda=\left(1-\frac{2}{\pi}\right),$$

SNR *Modulation Parameter*

If Channel Model is unknown

Calculate from Data: Average value of the conditional variance

NN Statistics:

Empirically determined for dif. NN architecture & complexity

$$\mu_n=\mathbb{E}\left[\left(\bar{f}_{NN}(\boldsymbol{r})-f_{MAP}(\boldsymbol{r})\right)^2\right],$$
$$\sigma_n^2=\mathbb{E}\left[\mathbb{E}_{\mathcal{D}}\left[\left(f_{NN}(\boldsymbol{r};\delta_n)-\bar{f}_{NN}(\boldsymbol{r})\right)^2\right]\right]$$

Expected NN Function

$$\bar{f}_{NN}(\boldsymbol{r})=\mathbb{E}_{\mathcal{D}}[f_{NN}(\boldsymbol{r};\delta_n)]$$

Data-dependent BER bounds (SISO)

Theorem 1 (SISO): Assumption: Equally Likely Prior

$$P_{LB} \leqslant BER_{NN} \leqslant P_{UB}$$

$P_{LB} = BER_{MAP}$ (Known Channel)

$P_{LB} = \mathbf{Q}\left(\frac{M-1}{\sqrt{M^2\sigma_h^2}}\right)$ (Unknown Channel)

Modulation Order

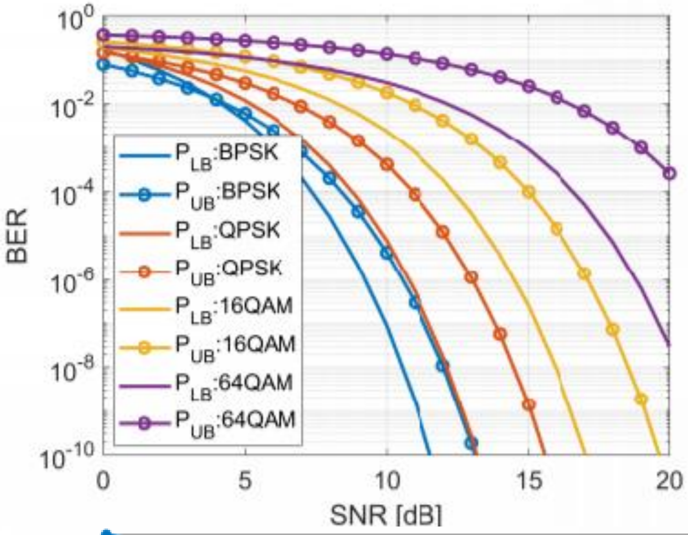
Depends on:
Data (channel) modulation

$$P_{UB} = \left(\frac{1}{2}\right) e^{-\left(\frac{M(1-\mu_n)-1}{\sqrt{2M^2(\sigma_n^2+\sigma_h^2)}}\right)}$$

Depends on:
Data (channel), mod. NN arch. & complexity

Take-Aways

- BER of NN-Rx lies within bounds (fixed size, arch. & data)
- Gap increases with M
- Equivalence under sufficient samples

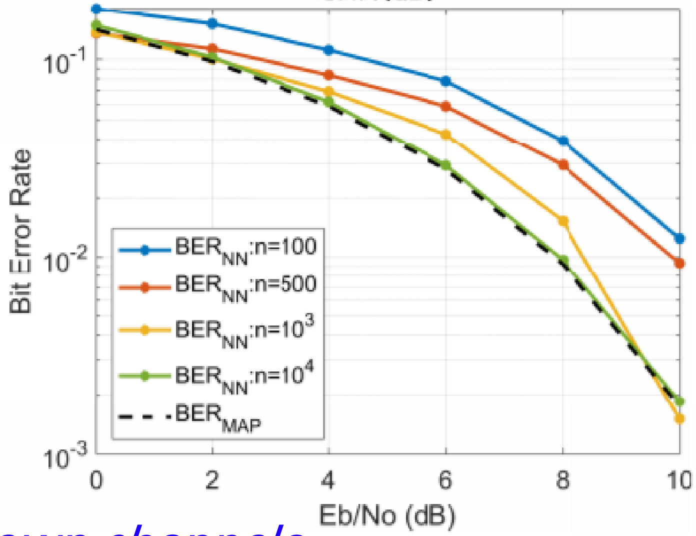


Theoretical

AWGN Channel

Feed-forward NN (FNN)

$n = 10^3$



Experimental

16-QAM

Tight data-dependent bounds on BER for known or unknown channels

Data-dependent BER bounds (MIMO)

Theorem 2 (MIMO): Assumption: Independent Fading

$$P_{LB} \leqslant BER_{NN} \leqslant P_{UB}$$

$$P_{LB} = BER_{MAP}$$ (Known Channel)

$$P_{LB}=\frac{M-1}{M}\prod_{k=1}^{K}\frac{1}{1+(M-1)\bar{\gamma}_k/M^2A_M}$$ (Unknown Channel)

 # Antennas *SNR* *Modulation Parameter*

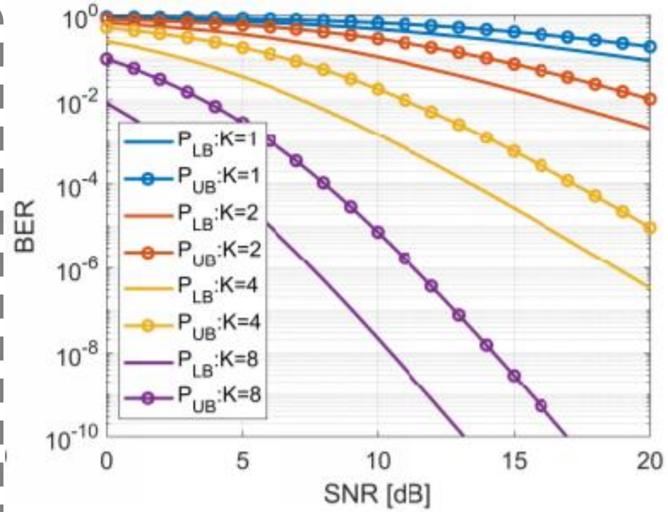
$$P_{UB}=\frac{M-1}{M}\prod_{k=1}^{K}\frac{1}{(1+g_{NN}\bar{\gamma}_k)},\quad g_{NN}=\frac{M(1-\mu_n)-1}{M^2A_M}$$

Depends on:
 # Antennas, SNR
Data (channel), mod.
NN arch.& complex

Rayleigh Fading

Take-Aways

- Gap between bounds increases with K
- Equivalence for MIMO fading channels (Assumption: frequency-flat fading)



Theoretical

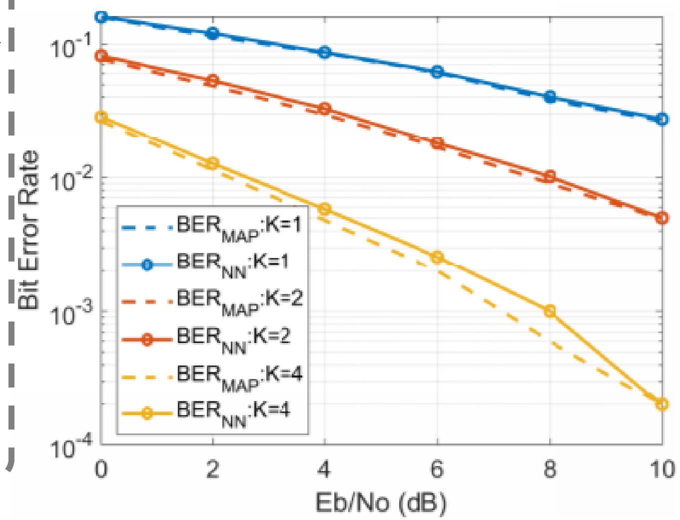
Rayleigh Fading

Maximum Ratio Combine

Feed-forward NN

QPSK

$n = 10^3$



Experimental

QPSK

Insights on Model & Training Complexity

- BER upper bound is a monotonically increasing function of the Risk statistics:

$$P_{UB}(\mu_n(N), \sigma_n^2(N), \sigma_h^2)$$

- σ_h^2 is independent of NN model and training and is fixed for a given dataset
- $\mu_n(N), \sigma_n^2(N)$ for FNN depend on NN structure (# parameters: N) and training set quality (# samples: n)
- Minimizing $\mu_n(N), \sigma_n^2(N)$ ensures minimum worst case BER
- ***Inferences:***
  - *For given data-set, find least complex NN structure with min. worst case BER*
  - *For given NN structure, find min. required training with min. worst case BER*

TABLE I: Simulation Parameters

Parameters/Hyperparameters	Value/Model
Neural Network Type	Deep Feed-Forward NN
Empirical Risk ($\mathcal{R}$)	Mean-Squared Error (MSE)
Training Algorithm	Scaled Conjugate Gradient Descent(SCGD)
Activation Function (ϕ_l)	Hidden:Tan-Sig [24], Output:Softmax
Channel Models	AWGN, Rayleigh
Antenna Configuration	SISO, MIMO
Number of Hidden Layers	[1, 3, 5]
Number of Input Neurons	2 (I,Q streams)
Number of Output Neurons	$M = [2, 4, 16, 64]$
Number of Samples (n)	10^2 - 10^5
Number of Parameters (N)	10^0 - 10^5
Training,Validation,Testing	[60, 20, 20] %

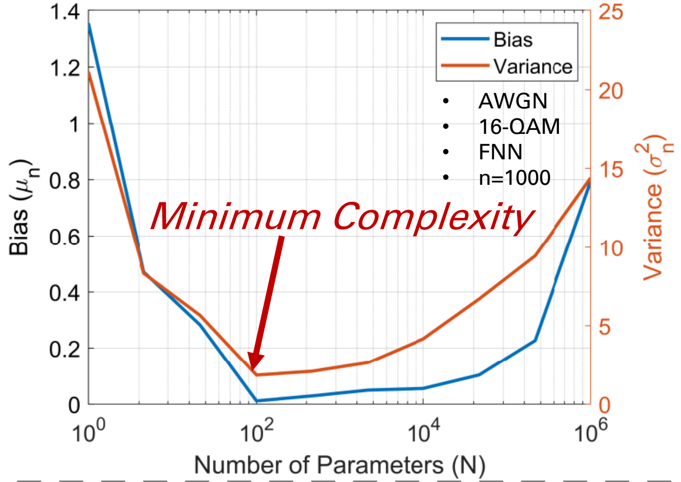
Impact of Model Complexity

Empirical Approach for Least Complex Receiver Structure

Evaluate NN Statistics for Validation-Set

- **Depends on**: Quality of training, NN Architecture & Structure, Parameter Initialization
- Least complex NN structure with min. worst BER is found for fixed training dataset ($n = 1000$)

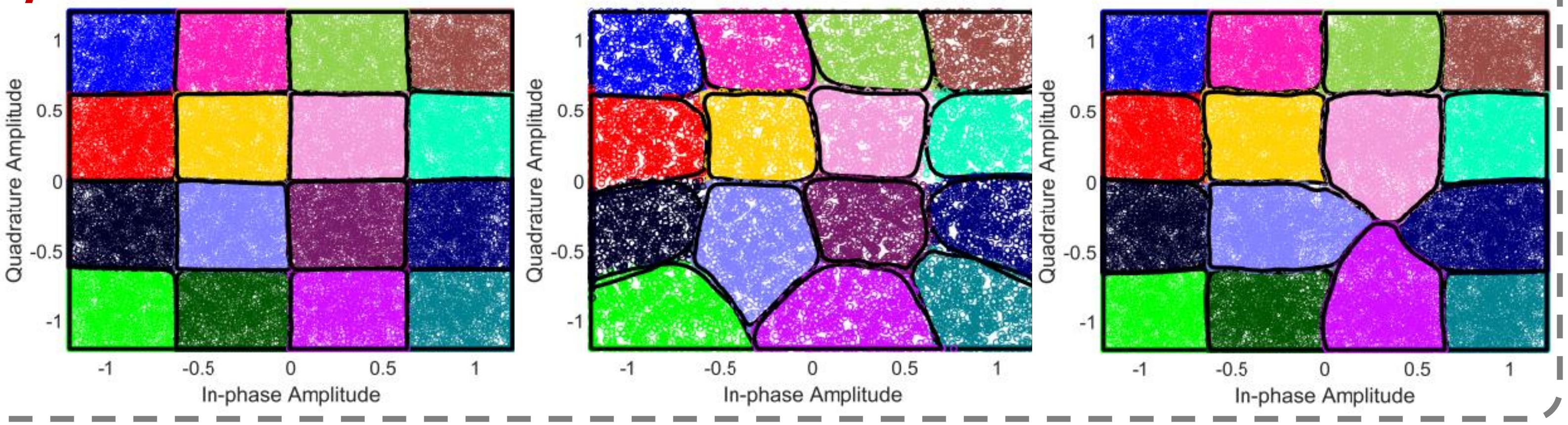
Impact of Model Complexity



Boundaries & Model Complexity

- Optimal Model is Determined
- Less Complex → Underfits
- More Complex → Overfits

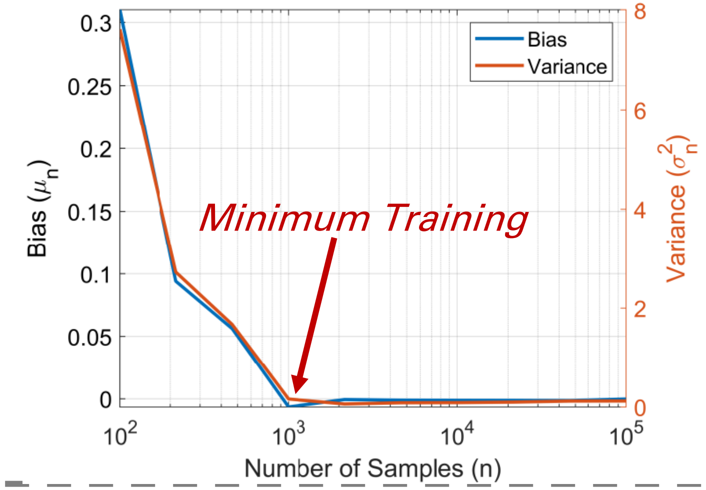
N: *Number of parameters*
$n = 1000$



Impact of Training Quality

Impact of # Training Samples

- Min. training is found for a given NN structure
- Simple FNN is sufficient for equivalence
- n samples are drawn from entire channel distribution

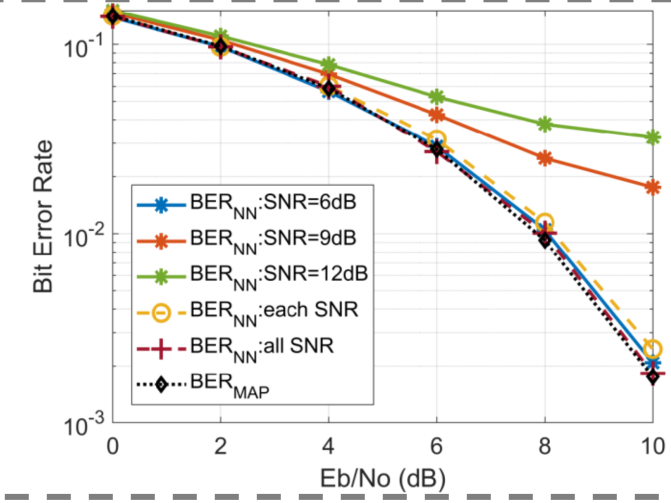


- AWGN channel
- 16-QAM
- FNN

Impact of Training Schemes

Achieves min. error under:

- Training one model at all SNR
- Training one model per SNR
- Training one model at low SNR ($\leq 6dB$)



- AWGN channel
- 16-QAM
- FNN
- $n = 10^3$

Conclusions

- Through theoretical and empirical analysis we show that:
  - NN Rx with MSE risk is equivalent to MAP Rx, under complete training
  - Under incomplete training, BER of NN Rx lies within the derived tight data & model dependent BER upper bound for NN receivers
  - Empirically derived least complex NN Rx structure and min. training to ensure min. worst BER

THANK YOU. QUESTIONS?

Paper contains all theorems and proofs presented

mabdulcareem@albany.edu

https://www.albany.edu/~ma952922