Machine learning-based solutions for channel decoding in M2M-type communications

PhD defense - December 13th, 2024

PhD candidate: Gastón DE BONI ROVELLA

Jury:
M. Raphaël LE BIDAN, Rapporteur (IMT Atlantique, Brest)
M. Maël LE TREUST, Rapporteur (Université de Rennes/CNRS)
Mme Inbar FIJALKOW, Examinatrice (ENSEA/CNRS)
M. Charly POULLIAT, Examinateur (ENSEEIHT)
M. Jérôme LACAN, Directeur de thèse (ISAE-SUPAERO)
Mme Meryem BENAMMAR, Co-directrice de thèse (ISAE-SUPAERO)

Membres invités:
M. Hugo MERIC (CNES)
M. Tarik BENADDI (TAS)

Table of contents

Table of contents

Context

Meaning of the title

"Machine learning-based solutions for channel decoding in M2M-type communications"

⟵ ⟵ ⟵

- **M2M communications**: direct exchange of data between devices.
- **Channel coding**: detect and correct errors caused by the channel.
- **Machine learning**: learn from data without explicit programming.

Machine learning for communications - why?

Several advantages over classical solutions:

- End-to-end design.
- Do not require an accurate model of the communication setting.
- Adaptability to varying communication conditions.
- **For channel decoding:** Online complexity (real-time) can be traded for offline complexity (training process).

Figure: ChatGPT-generated image that symbolizes *Intelligent communications*.

Machine learning for communications - why?

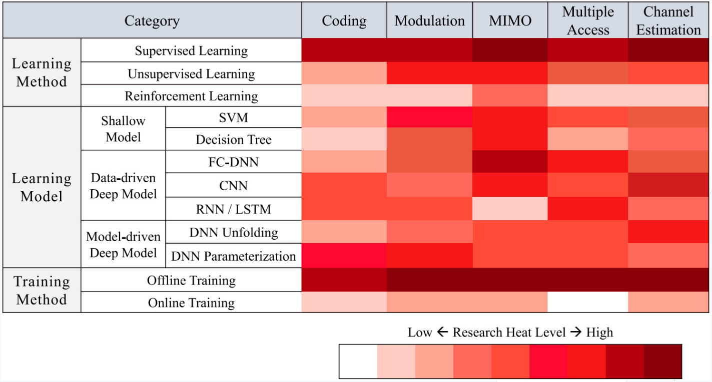


Figure: Research heatmap of Artificial Intelligence for Communications[1].

[1] Neng Ye et al. "Artificial Intelligence for Wireless Physical-Layer Technologies (AI4PHY): A Comprehensive Survey". In: *IEEE Transactions on Cognitive Communications and Networking* (2024), pp. 1–1

Machine learning for channel decoding

- Decoders in M2M require:
  - Low latency.
  - Short packet length.
  - Low complexity.
- However, optimal decoders for short codes are usually very complex and with considerable latency (e.g. SCL for Polar codes).
- Machine learning appears as a potential solution for optimal and low-complexity decoding of **short codes**.

Work in progress...

$\Uparrow$ **performance** $\Downarrow$ **complexity** $\Uparrow$ **applicability**

compared to previous machine learning-based decoders.

Channel decoding problem

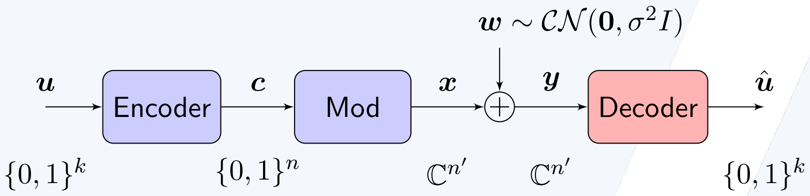


Figure: General system model.

The Bit Error Probability is defined as follows:

$$P_e \triangleq \frac{1}{k}\sum_{i=1}^{k} \mathbb{P}\{U_i \neq \hat{U}_i\}. \qquad (1)$$

The optimal decoder is the bit-MAP decoder, defined for every $i \in [1:k]$ as:

$$g^{(i)}(\boldsymbol{y}) \triangleq \underset{u\in\{0,1\}}{\operatorname{argmax}} \; P_{U_i|\boldsymbol{Y}}(u|\boldsymbol{y}). \qquad (2)$$

First machine learning-based solutions and limitations[2]

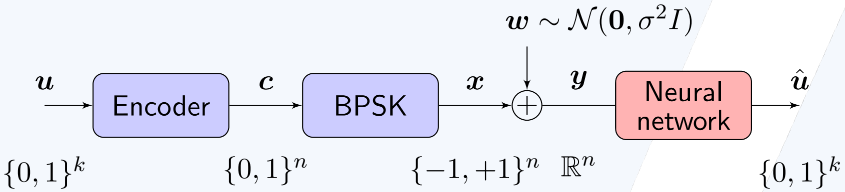


Figure: General system model.

We can build the training dataset as follows:

$$\textbf{Inputs} \rightarrow \boldsymbol{y}_i, \text{ for } i \in [1 : N_{\text{data}}]$$
$$\textbf{Outputs} \rightarrow \boldsymbol{u}_i, \text{ for } i \in [1 : N_{\text{data}}]$$

[2] Tobias Gruber et al. "On deep learning-based channel decoding". In: *2017 51st Annual Conference on Information Sciences and Systems (CISS)*. IEEE, Mar. 2017

Curse of dimensionality (CoD)

message size $= k$	codeword space size $= 2^k$
4	16
16	65536
64	18446744073709551616
200	1.6×10^{60} $\approx$ atoms in 100 Suns

Table: Number of valid codewords vs. message size.

Curse of dimensionality (CoD)

For a code of size (n, k) and error-correction capability of t bits:

Codeword space

2^k

Noise realizations

at least $\sum_{i=0}^{t} \binom{n}{i}$

per codeword

Size of the neural network

Increases significantly with the code dimensions to maintain performances

Summary of contributions

To train without noise realizations (Chapter 2):

1. Propose a new SVM-based approach that trains on only noiseless codewords.
2. Under AWGN, prove its equivalence to the bit-MAP decoder.

To reduce the training codeword space & size of the network (Chapters 3 and 4):

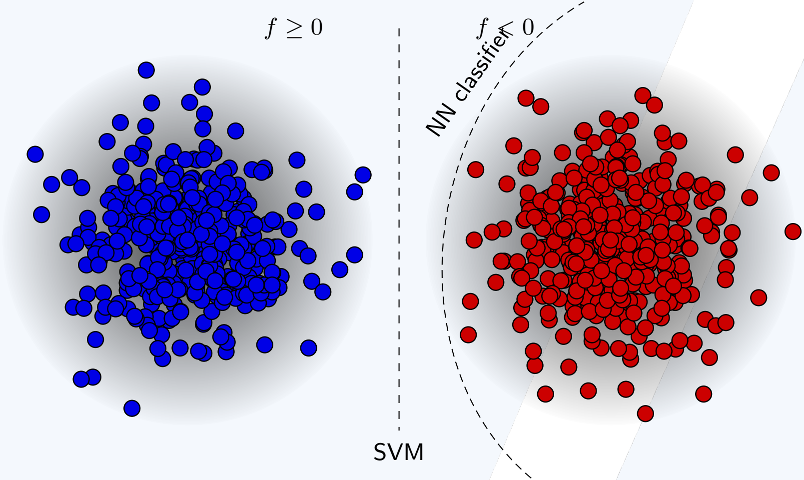
1. Employ the SBND approach –which is trained using a single codeword– and propose a message-oriented approach that improves performances.
2. Analyze the impact of the parity-check matrix and propose an algorithm to optimize it.
3. Introduce a reduced-complexity neural architecture with competitive performances.
4. Extend the SBND approach to higher-order modulations and discuss the changes in the training dataset.

github.com/gastondeboni/SVM_for_Channel_Decoding
github.com/gastondeboni/Syndrome_Based_Neural_Decoding

Table of contents

Why Support Vector Machine (SVM) for decoding?

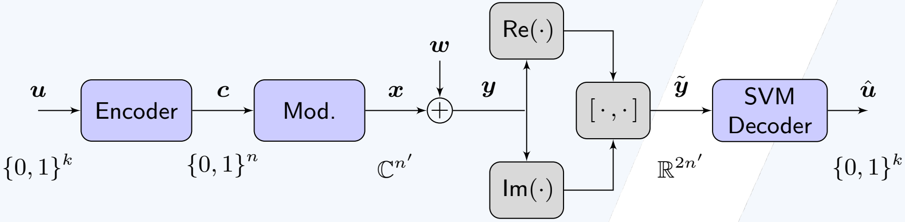
The maximum margin property



SVM for decoding

System model

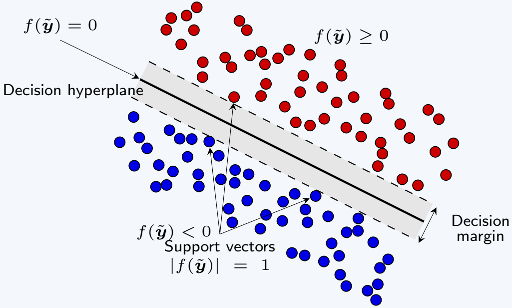
Because SVMs only work with real numbers...



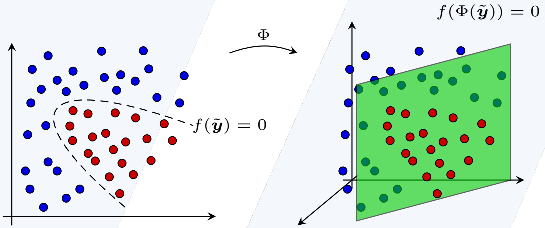
where $\boldsymbol{w} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 I)$.

Support Vector Machines - theory

With a linearly separable dataset $\{\tilde{\boldsymbol{y}}_i, l_i\}_{1\leq i\leq N}$, we compute the hyperplane $f(\tilde{\boldsymbol{y}}) = 0$ such that:



With linearly non-separable data, we employ the **kernel method**, where a function Φ projects the data into a high-dimensional space.



Support Vector Machines - theory

Mathematical foundation

Suppose a dataset $\{\tilde{\boldsymbol{y}}_i, l_i\}_{1 \leq i \leq N}$. We must compute the solution $\boldsymbol{\alpha}^\star, \nu^\star$ to the following opt. problem:

$$\begin{aligned} &\underset{\boldsymbol{\alpha}}{\operatorname{argmax}} \quad \mathcal{L}(\boldsymbol{\alpha}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} l_i l_j \alpha_i \alpha_j K(\tilde{\boldsymbol{y}}_i, \tilde{\boldsymbol{y}}_j), \quad \boldsymbol{\alpha} \in \mathbb{R}^{2n'} \\ &\text{subject to} \quad \alpha_i \geq 0 \quad \forall i \in [1:N] \quad \text{and} \quad \sum_{i=1}^{N} \alpha_i l_i = 0, \end{aligned} \tag{3}$$

where $\boxed{K(\tilde{\boldsymbol{y}}, \tilde{\boldsymbol{y}}') \triangleq e^{-\gamma ||\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{y}}'||^2}, \ \gamma \in \mathbb{R}^+}$. The final SVM classifier is given by:

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} l_i \alpha_i^\star K(\boldsymbol{x}, \tilde{\boldsymbol{y}}_i) + \nu^\star. \tag{4}$$

SVM for decoding

How is SVM applied to channel decoding in the literature?

SVM for decoding

Previous works - multi-class classification[3][4]

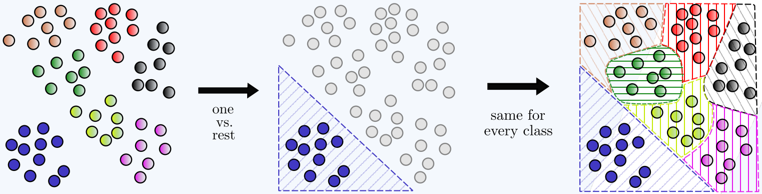


Figure: One vs. rest approach.

This solution implies the evaluation of 2^k functions:

$$\hat{\boldsymbol{u}} = \boldsymbol{u}_{j^*}, \text{ where } j^* = \underset{j \in [1:2^k]}{\operatorname{argmax}} f^{(j)}(\tilde{\boldsymbol{y}}). \quad (5)$$

[3] V. Sudharsan and B. Yamuna. "Support Vector Machine based Decoding Algorithm for BCH Codes". In: *Journal of Telecommunications and Information Technology* 2.2016 (June 2016), pp. 108–112

[4] R. Ramanathan et al. "Generalised and Channel Independent SVM Based Robust Decoders for Wireless Applications". In: *2009 International Conference on Advances in Recent Technologies in Communication and Computing*. IEEE, Oct. 2009

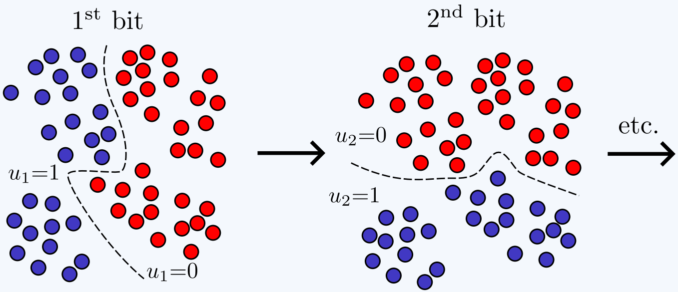
SVM for decoding

Limitation	**Contribution**
Need to evaluate 2^k functions	**Bit-wise approach**
Several noise realizations per codeword	**Noiseless training**
Lack of theoretical analysis	**Optimality study and closed-form solution**

Contributions

Contributions

1) Bit-wise SVM decoder



Outline of the decoding process:

1. we receive $\tilde{\boldsymbol{y}} \in \mathbb{R}^{2n'}$;
2. if $f^{(1)}(\tilde{\boldsymbol{y}}) \geq 0$ then $\hat{u}_1 = 1$;
3. else $\hat{u}_1 = 0$;
4. continue for $f^{(2)}, ..., f^{(k)}$.

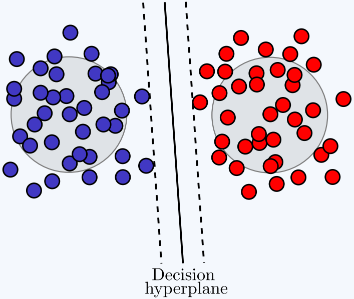
Obs: The k decison functions can be evaluated in parallel.

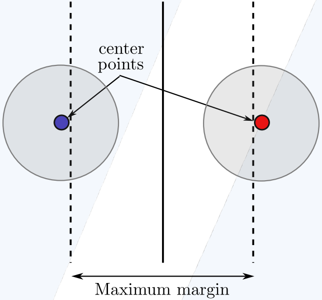
This solution reduces the number of functions from 2^k to k.

Contributions

2) Noiseless training

Because we are dealing with an AWGN channel...





The dataset is reduced to only 1 sample per class (i.e., per codeword).

Contributions

1+2) New optimization problem[5]

Bitwise approach + noiseless training = k **opt. problems** for $j \in [1:k]$:

$$\underset{\boldsymbol{\alpha}}{\text{argmin}}\ \frac{1}{2}\sum_{i,m=1}^{2^k} \alpha_i \alpha_m l_i^{(j)} l_m^{(j)} K(\tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{x}}_m) - \sum_{i=1}^{2^k} \alpha_i, \quad \text{where } K(\tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{x}}_m) = e^{-\gamma||\tilde{\boldsymbol{x}}_i - \tilde{\boldsymbol{x}}_m||^2}$$

$$\text{subject to: } \alpha_i \geqslant 0 \text{ and } \sum_{i=1}^{2^k} l_i^{(j)} \alpha_i = 0,$$

where $l_i^{(j)} = +1$ if the jth bit of the ith message is a 1, and $l_i^{(j)} = -1$ otherwise.

$$f^{(j)}(\tilde{\boldsymbol{y}}) = \sum_{i=1}^{2^k} l_i^{(j)} \alpha_i^{\star(j)} e^{-\gamma||\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{x}}_i||^2} + \nu^{\star(j)}. \tag{6}$$

[5] Gastón De Boni Rovella et al. "On the Optimality of Support Vector Machines for Channel Decoding". In: *2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*. IEEE, June 2024

Contributions

3) Optimality analysis

Bitwise approach + noiseless training = k **decision functions** for $j \in [1:k]$:

$$f^{(j)}(\tilde{\boldsymbol{y}}) = \sum_{i=1}^{2^k} l_i^{(j)} \alpha_i^{\star(j)} e^{-\gamma||\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{x}}_i||^2} + \nu^{\star(j)}. \quad (7)$$

Theorem (Optimal solution and equivalence to bit-MAP)

1. *For $\gamma \gg 1$, $\boldsymbol{\alpha}^\star = (1, 1, ..., 1)$, and $\nu^\star = 0$, for all k opt. problems.*
2. *With the previous solution $(\boldsymbol{\alpha}^\star, \nu^\star)$, if $\gamma = 1/\sigma^2$, the proposed decoding rule is equal to the bit-MAP decoder.*

Results

Results

Convergence to optimal solution

$\gamma_s \triangleq 1/\sigma_s^2$, where σ_s^2 is such that $E_b/N_0 = s$dB:

Theorem

1. *For* $\gamma \gg 1$, $\boldsymbol{\alpha}^\star = (1, 1, ..., 1)$, *and* $\nu^\star = 0$, *for all* k *opt. problems.*
2. *With the previous solution* $(\boldsymbol{\alpha}^\star, \nu^\star)$, *if* $\gamma = 1/\sigma^2$, *the proposed decoding rule is equal to the bit-MAP decoder.*

$$f(\tilde{\boldsymbol{y}}) = \sum_{i=1}^{2^k} l_i \alpha_i^\star e^{-\gamma||\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{x}}_i||^2} + \nu^\star.$$

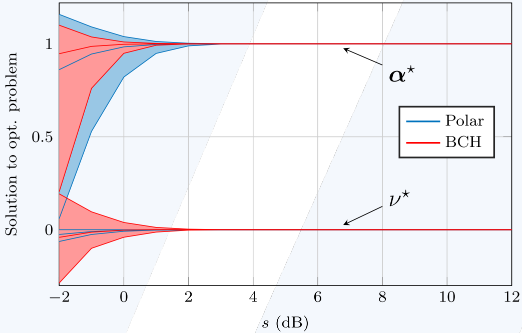


Figure: Solutions to the opt. problem vs. value of γ, for Polar and BCH codes of size (32,11) under 16QAM.

Results

Bit error rate studies

$\gamma_s \triangleq 1/\sigma_s^2$, where σ_s^2 is such that $E_b/N_0 = s$dB:

Theorem

1. *For* $\gamma \gg 1,\ \boldsymbol{\alpha}^\star = (1, 1, ..., 1)$, *and* $\nu^\star = 0$, *for all* k *opt. problems.*
2. *With the previous solution* $(\boldsymbol{\alpha}^\star, \nu^\star)$, *if* $\gamma = 1/\sigma^2$, *the proposed decoding rule is equal to the bit-MAP decoder.*

$$f(\boldsymbol{x}) = \sum_{i=1}^{2^k} l_i \alpha_i^\star e^{-\gamma||\boldsymbol{x}-\tilde{\boldsymbol{x}}_i||^2} + \nu^\star.$$

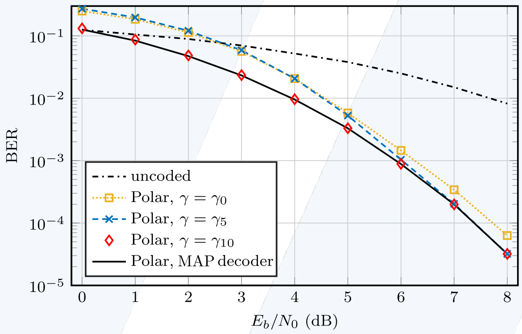


Figure: Polar (32,11), 16QAM.

Conclusion

Take-away points:

1. The proposed approach (bitwise + noiseless training) reduces the number of SVM classifiers from 2^k to k and the dataset to only one sample per class.
2. However, the theoretical analysis shows equivalence to MAP for AWGN.

Perspectives:

1. Applying the system in a more complex channel where the MAP decoding rule is not available in closed form? (frequency or time selective, fading, unknown, etc.).
2. Training on a subset of valid codewords?

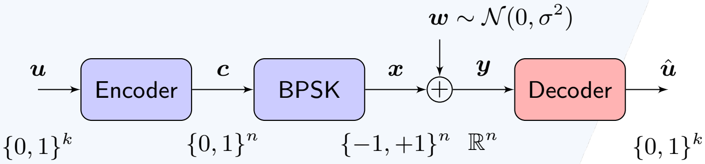
	One vs. rest	**Bitwise + noiseless opt. (ours)**
# of SVM functions	2^k	$\boldsymbol{k}$
dataset size	$N_{data} \gg 2^k$	$\boldsymbol{2^k}$

Table: Complexity comparison between methods.

$$f(\tilde{\boldsymbol{y}}) = \sum_{i=1}^{2^k} l_i \alpha_i e^{-\gamma ||\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{x}}_i||^2} + \nu.$$

Table of contents

Deep Neural Networks for decoding



Recall the **curse of dimensionality**: for message of length $k \Rightarrow 2^k$ possible codewords.

There are two approaches that employ single-codeword training:

- Model-based (*deep unfolding* of Belief Propagation)[6].
- Model-free (syndrome-based neural decoding)[7].

[6] Eliya Nachmani, Yair Be'ery, and David Burshtein. "Learning to Decode Linear Codes Using Deep Learning". In: *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, Sept. 2016

[7] Amir Bennatan, Yoni Choukroun, and Pavel Kisilev. "Deep Learning for Decoding of Linear Codes - A Syndrome-Based Approach". In: *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, June 2018

Previous works: model-based[6]

For the parity-check matrix

$$H = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

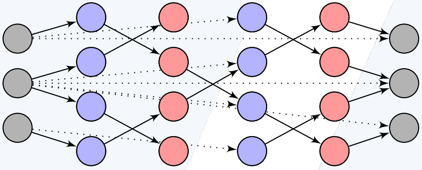
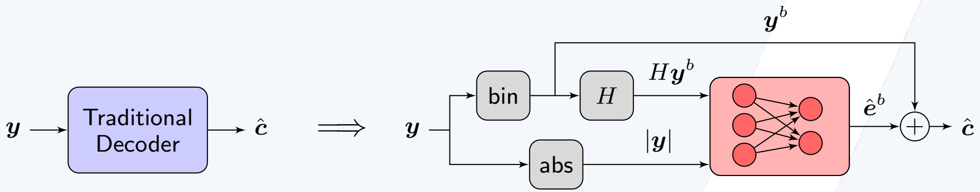


Figure: Neural Belief Propagation (2 iter).

- The structure of Tanner graph defines the network architecture.
- Improves on the BP algorithm for specific codes (short and/or dense).
- The performances are often worse than the model-free method.

[6] Eliya Nachmani, Yair Be'ery, and David Burshtein. "Learning to Decode Linear Codes Using Deep Learning". In: *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, Sept. 2016

Model-free approach: Syndrome-Based Neural Decoding (SBND)[7]

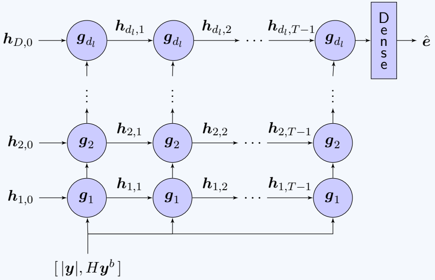


Why adopt this approach?

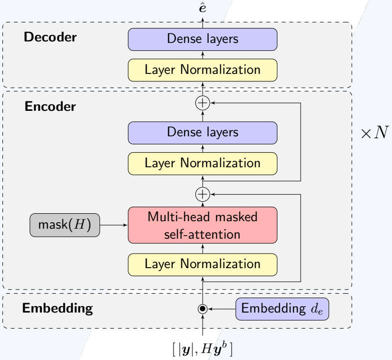
- No intrinsic loss of optimality: $P_{C_i|\boldsymbol{Y}}(c_i|\boldsymbol{y}) = P_{E_i^b||\boldsymbol{Y}|,H\boldsymbol{Y}^b}(c_i \oplus y_i^b||\boldsymbol{y}|, H\boldsymbol{y}^b)$, $\forall i \in [1:n]$.
- The inputs $(H\boldsymbol{y}^b, |\boldsymbol{y}|)$ are independent of $\boldsymbol{c}$ under BPSK, which enables the *single-codeword training property*.
- This *bypasses* the codeword space aspect of the CoD (train on 1 codeword instead of 2^k).

[7] Amir Bennatan, Yoni Choukroun, and Pavel Kisilev. "Deep Learning for Decoding of Linear Codes - A Syndrome-Based Approach". In: *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, June 2018

SBND: architectures in the literature[7][8]



(a) Recurrent Neural Network (RNN)[3].



(b) Transformer (ECCT)[4].

[7] Amir Bennatan, Yoni Choukroun, and Pavel Kisilev. "Deep Learning for Decoding of Linear Codes - A Syndrome-Based Approach". In: *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, June 2018

[8] Yoni Choukroun and Lior Wolf. "Error Correction Code Transformer". In: *Advances in Neural Information Processing Systems*. Ed. by S. Koyejo et al. Vol. 35. Curran Associates, Inc., 2022, pp. 38695–38705

SBND approach

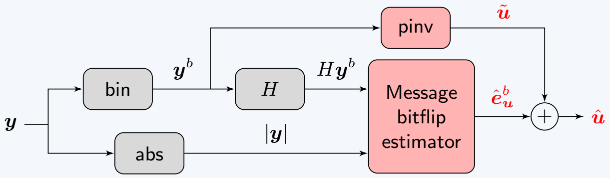
Contributions

Limitation	**Contribution**
The codeword c is estimated instead of the message u	**message-oriented decoder**
The network is often very large (1s to 10s million parameters)	**recurrent ECCT**
Large impact of the PC matrix H on performance	**PC matrix study and optimization**

Contributions

Contributions

1) Proposed message-oriented framework[9]



Theorem (Sufficient statistics)

The following equation holds, for all $i \in [1:k]$:

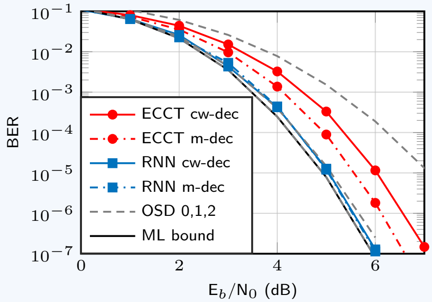
$$P_{U_i|\mathbf{Y}}(u_i|\boldsymbol{y}) = P_{E^b_{u,i}|\,|\mathbf{Y}|,H\mathbf{Y}^b}(u_i \oplus \tilde{u}_i \,|\, |\boldsymbol{y}|, H\boldsymbol{y}^b). \quad (8)$$

- Maintains the single-codeword training property.
- It allows for a deeper focus on the information bits during training (sacrificing redundant bits).
- Network complexity is reduced (only k outputs).
- It is directly applicable to non-systematic codes.

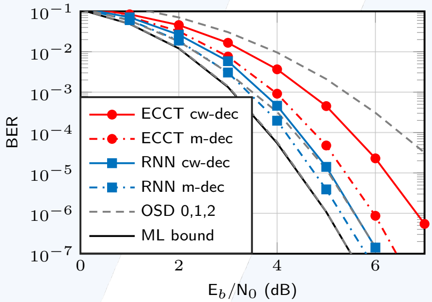
[9] Gastón De Boni Rovella and Meryem Benammar. "Improved Syndrome-based Neural Decoder for Linear Block Codes". In: *GLOBECOM 2023 - 2023 IEEE Global Communications Conference*. IEEE, Dec. 2023

Results

1) Message-oriented vs. codeword-oriented



(a) BCH (63,45).

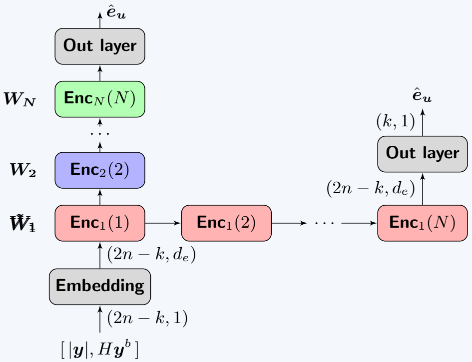


(b) BCH (63,39).

Obs: RNN has 4M weights, ECCT has 2M weights.

Contributions

2) Recurrent transformer-based (r-ECCT) architecture[10]



Results:

- Number of weights divided by N (≈ 10).
- Decoding performances globally maintained (even slightly improved).

[10] Gastón De Boni Rovella et al. "Syndrome-Based Neural Decoding for Higher-Order Modulations (**submitted**)". In: *IEEE Transactions on Communications* (2024)

Results

2) Recurrent Error Correction Code Transformer (r-ECCT) and complexity analysis

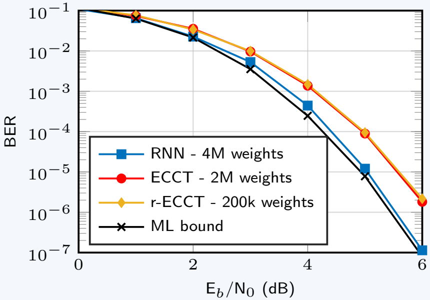


Figure: BCH (63,45).

$$W_{\text{RNN}} = 3\left((2d_l - 1)\alpha^2 + \alpha\right) r^2 + (3d_l + k)\alpha r + k \approx \boldsymbol{\mathcal{O}((2n-k)^2)}$$

$$W_{\text{ECCT}} = 12Nd_e^2 + (13N + r + 3)d_e + (r+1)k + 1 \approx \boldsymbol{\mathcal{O}(Nd_e^2)}$$

$$\boxed{W_{\text{r-ECCT}} = 12d_e^2 + (16 + r)d_e + (r+1)k + 1 \approx \boldsymbol{\mathcal{O}(d_e^2)}}$$

where:

- $r \triangleq 2n - k$, where (n, k) are the code parameters;
- N the number of encoders in the ECCT architecture;
- and d_e the embedding dimension of the ECCT and r-ECCT.

Contributions

3) Influence of the Parity-Check Matrix (PCM)

1. As a metric, we propose the *pairwise* Mutual Information (MI) between a single input and output of the decoder, and we compute the formula for our case:

$$\boxed{MI(E_i^b; S_j) = \left[\mathcal{H}_b(\mathcal{E}(N_j)) - \mathcal{H}_b(\mathcal{E}(N_j - 1))\right] \mathbb{I}\{H_{ij} = 1\},} \tag{9}$$

where

- $\mathcal{H}_b(a) \triangleq -a\log_2 a - (1-a)\log_2(1-a)$;
- N_j Hamming weight of the jth row of H;
- and $\mathcal{E}(N_j) \triangleq \frac{1}{2} + \frac{1}{2}(1-2p)^{N_j}$, p denoting the bitflip probability for a given E_b/N_0.

Contributions

3) Influence of the Parity-Check Matrix (PCM)

2. We notice that this pairwise MI decreases with the respective row's weight:

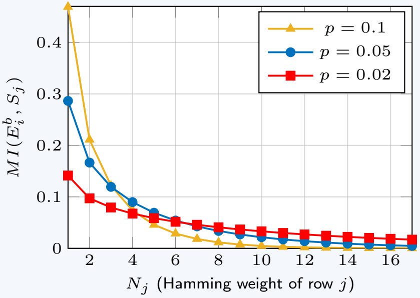
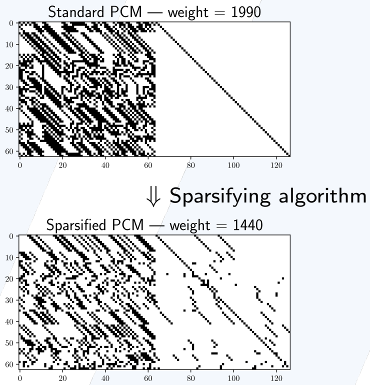


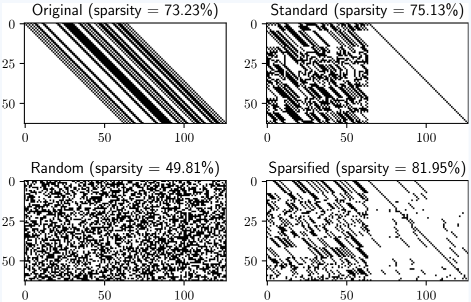
Figure: Pairwise MI vs. row's weight.

4. We propose a **sparsifying algorithm**.

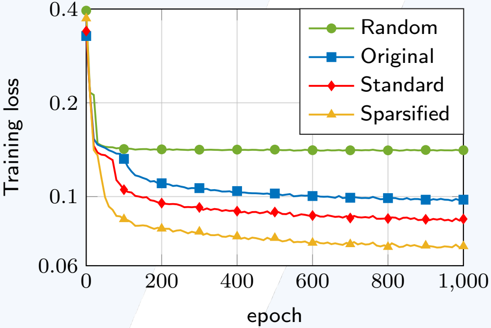


Results

3) Influence of the PCM



(a) Four considered PCMs for a BCH (127,64).



(b) Loss function with different sparsities.

Results

3) Influence of the PCM

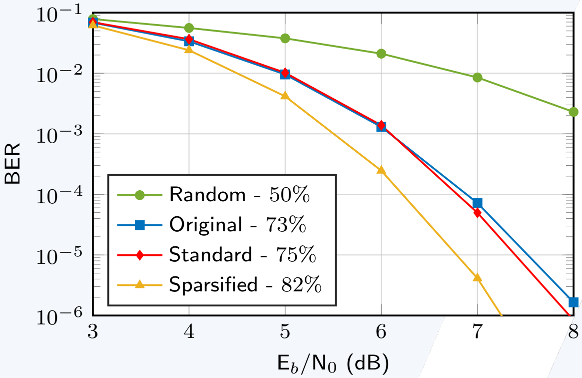


Figure: BCH (127,64), different sparsities.

Results (all combined)

Results

Every element put together vs. best neural solutions in the literature: BCH (127,64)

Message-oriented approach + r-ECCT (reduced complexity) + sparsified matrix:

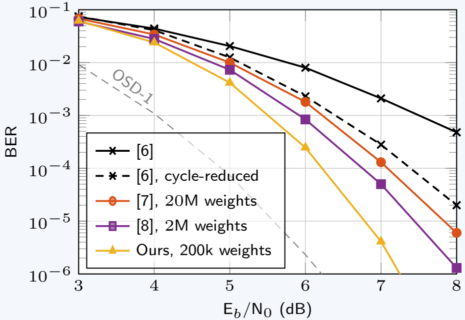


Figure: BCH (127,64), comparison with SOTA.

[6] Eliya Nachmani et al. "Deep Learning Methods for Improved Decoding of Linear Codes". In: IEEE Journal of Selected Topics in Signal Processing 12.1 (Feb. 2018).

[7] Amir Bennatan, Yoni Choukroun, and Pavel Kisilev. "Deep Learning for Decoding of Linear Codes - A Syndrome-Based Approach". In: 2018 IEEE International Symposium on Information Theory (ISIT). IEEE, June 2018.

[8] Yoni Choukroun and Lior Wolf. "Error Correction Code Transformer". In: Advances in Neural Information Processing Systems. Ed. by S. Koyejo et al. Vol. 35. Curran Associates, Inc., 2022.

Conclusion

Summary of contributions

Summary of contributions:

1. Message-oriented approach (increased performance).
2. r-ECCT architecture (reduced complexity).
3. Optimization of PCM (increased performance).

Future works:

1. Can we extend this to **higher-order modulations**?
2. To be continued...

Table of contents

System model for Higher-Order Modulations (HOM)

Bit-Interleaved Coded Modualtions

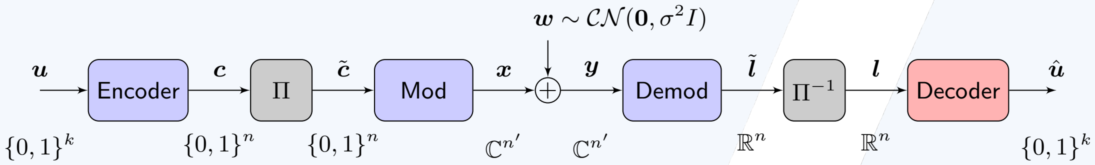


Figure: Proposed system with Bit-Interleaved Coded Modulations (BICM)[11].

1. The bit-interleaver Π **shuffles** the codewords before modulation.
2. The deinterleaver Π^{-1} retrieves the **original bit order**.

[11] G. Caire, G. Taricco, and E. Biglieri. "Bit-Interleaved Coded Modulation". In: *IEEE Transactions on Information Theory* 44.3 (May 1998), pp. 927–946

Contributions

SBND for HOM

1. SBND decoder for HOM construction.
2. Optimality analysis.
3. Training dataset design.

Contributions

Proposed decoder and optimality

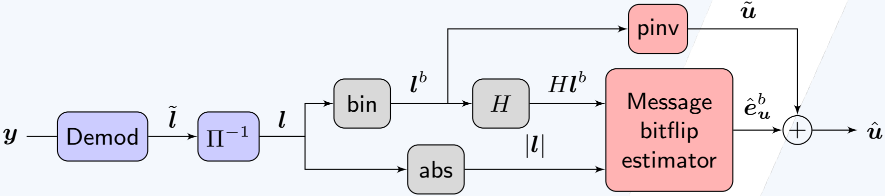


Figure: Proposed decoder with Bit-Interleaved Coded Modulations (BICM).

Theorem (Sufficient statistics for HOM)

The following equation holds, for all $i \in [1:k]$*:*

$$P_{\boldsymbol{E}_{\boldsymbol{u}}^b|\boldsymbol{L}}(\boldsymbol{e}_{\boldsymbol{u}}^b|\boldsymbol{l}) = P_{\boldsymbol{E}_{\boldsymbol{u}}^b||\boldsymbol{L}|,H\boldsymbol{L}^b}(\boldsymbol{e}_{\boldsymbol{u}}^b|\,|\boldsymbol{l}|, H\boldsymbol{l}^b). \quad (10)$$

Contributions

Proposed decoder and optimality - proof outline

$$P_{\boldsymbol{E}_{\boldsymbol{u}}^b|\boldsymbol{L}}(\boldsymbol{e}_{\boldsymbol{u}}^b|\boldsymbol{l}) = P_{\boldsymbol{E}_{\boldsymbol{u}}^b||\boldsymbol{L}|,\boldsymbol{L}^b}(\boldsymbol{e}_{\boldsymbol{u}}^b|\,|\boldsymbol{l}|,\boldsymbol{l}^b) \quad (11)$$

$$= P_{\boldsymbol{E}_{\boldsymbol{u}}^b||\boldsymbol{L}|,H\boldsymbol{L}^b,A\boldsymbol{L}^b}(\boldsymbol{e}_{\boldsymbol{u}}^b|\,|\boldsymbol{l}|,H\boldsymbol{l}^b,A\boldsymbol{l}^b) \quad ([H^T,A^T] \text{ is invertible}) \quad (12)$$

$$= P_{\boldsymbol{E}_{\boldsymbol{u}}^b||\boldsymbol{L}|,H\boldsymbol{L}^b,A\boldsymbol{L}^b}(\boldsymbol{e}_{\boldsymbol{u}}^b|\,|\boldsymbol{l}|,H\boldsymbol{l}^b,A(\boldsymbol{c}\oplus\boldsymbol{e}^b)) \quad (13)$$

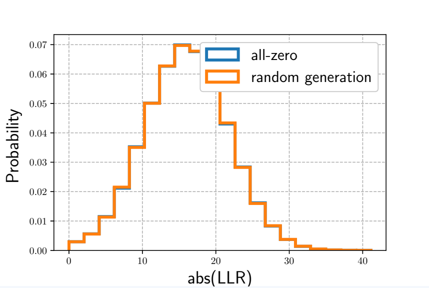
$$= P_{\boldsymbol{E}_{\boldsymbol{u}}^b||\boldsymbol{L}|,H\boldsymbol{L}^b,A\boldsymbol{L}^b}(\boldsymbol{e}_{\boldsymbol{u}}^b|\,|\boldsymbol{l}|,H\boldsymbol{l}^b,\boldsymbol{u}\oplus\boldsymbol{e}_{\boldsymbol{u}}^b)$$

$\boldsymbol{U} \sim$ Bern(0.5) and $\boldsymbol{E}_{\boldsymbol{u}}^b$ is independent of $\boldsymbol{U}$ (thanks to BICM):

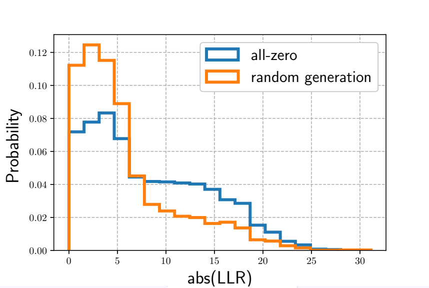
$$P_{\boldsymbol{E}_{\boldsymbol{u}}^b|\boldsymbol{L}}(\boldsymbol{e}_{\boldsymbol{u}}^b|\boldsymbol{l}) = P_{\boldsymbol{E}_{\boldsymbol{u}}^b||\boldsymbol{L}|,H\boldsymbol{L}^b}(\boldsymbol{e}_{\boldsymbol{u}}^b|\,|\boldsymbol{l}|,H\boldsymbol{l}^b) \quad (14)$$

Contributions

3) Training set design



(a) BPSK.



(b) 16-QAM - 1st position.

Figure: Distribution of $|\boldsymbol{l}|$ for BPSK modulation and 16-QAM.

We lose the single-codeword training property.

Results

Results

SBND + BICM employing different architectures[12]

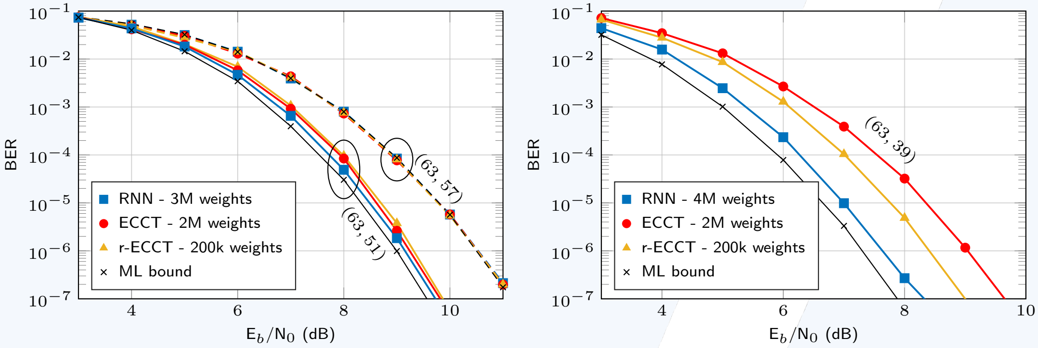


Figure: BCH codes, 8-PSK, different architectures.

[12] Gastón De Boni Rovella et al. "Scalable Syndrome-based Neural Decoders for Bit-Interleaved Coded Modulations". In: *2024 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*. IEEE, May 2024, pp. 341–346

Results

SBND + BICM employing different architectures[12]

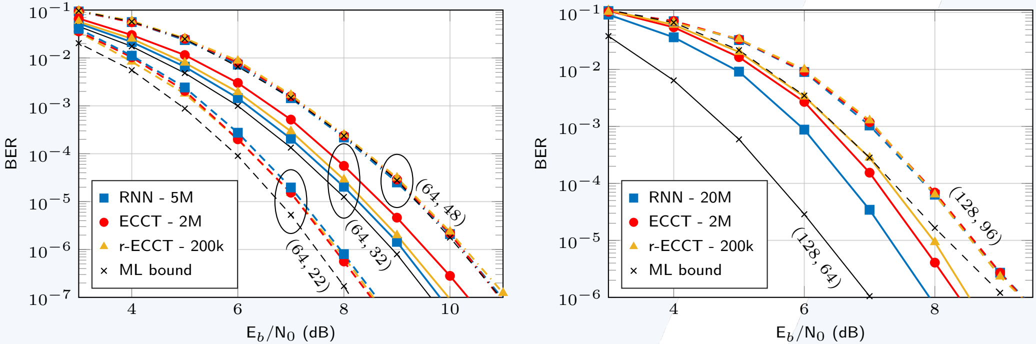


Figure: Polar codes, 16-QAM, different architectures.

[12] Gastón De Boni Rovella et al. "Scalable Syndrome-based Neural Decoders for Bit-Interleaved Coded Modulations". In: *2024 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*. IEEE, May 2024, pp. 341–346

Conclusion of SBND for HOM

Summary of contributions targeting applicability to HOM:

1. Extension of SBND to higher-order modulations (employing BICM).
2. Proof of optimality in this scenario.
3. Training set discussion.

Take-away points:

1. The SBND approach is successfully extended to higher-order modulations without loss of optimality.
2. The single-codeword training property is *mathematically lost* in the way. However, training was carried out on less than 3% of the codeword space for $n = 64$ and less than $10^{-9}\%$ for $n = 128$, with good performances depending on code size and rate.

Table of contents

Conclusion and final remarks

Final thoughts:

1. SVM: the complexity was reduced significantly, but still equivalent to the MAP decoder.
2. SBND: the different contributions resulted in **increased performance** (message-oriented approach, PCM study), **reduced complexity** (r-ECCT), and **wider applicability** (non-systematic codes and higher-order modulations) compared with existing solutions.

Future works:

1. Establish the basis for a comparative analysis against classical decoders.
2. Unified architecture[13].
3. Combine model-based and model-free approaches.

[13] Yongli Yan et al. *Error Correction Code Transformer: From Non-Unified to Unified.* 2024

Publications

National conferences:

[1] G. De Boni Rovella, M. Benammar, D. Gourmel and M. Djelloul. "SVM pour la démodulation et le décodage conjoints," colloque GRETSI, Grenoble, France, 2023.

International conferences:

[2] G. De Boni Rovella and M. Benammar, "Improved Syndrome-based Neural Decoder for Linear Block Codes," GLOBECOM 2023 - IEEE Global Communications Conference, Kuala Lumpur, Malaysia, 2023.

[3] G. De Boni Rovella, M. Benammar, H. Méric and T. Benaddi, "On the Optimality of Support Vector Machines for Channel Decoding," 2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Antwerp, Belgium, 2024.

[4] G. De Boni Rovella, M. Benammar, T. Benaddi and H. Meric, "Scalable Syndrome-based Neural Decoders for Bit-Interleaved Coded Modulations," 2024 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN), Stockholm, Sweden, 2024.

Journal articles:

[5] G. De Boni Rovella, M. Benammar, T. Benaddi and H. Meric, "Syndrome-Based Neural Decoding for Higher-Order Modulations." In: IEEE Transactions on Communications (2024). **(in review)**

[6] G. De Boni Rovella, M. Benammar, T. Benaddi and H. Meric, "Support Vector Machines for Optimal Channel Decoding." In: EURASIP Journal on Wireless Communications and Networking - Special Issue (2024). **(in review)**

The end

Thank you!

Questions?

All codes are available in the following repositories:
github.com/gastondeboni/SVM_for_Channel_Decoding
github.com/gastondeboni/Syndrome_Based_Neural_Decoding