Decoding of Polar Code by Using Deep Feed-Forward Neural Networks

PEDAVEGI ADITYA EE18BTECH11034
KOIDALA SURYA PRAKASH EE18BTECH11026

Supervised by : Prof Dr Shashank Vatedka
Dept of Electrical Engineering,
Indian Institute of Technology, Hyderabad

April 2022

Contents

- In this paper , the authors implemented a deep feed-forward neural network to decode a polar code and compare its decoding performance with the conventional list-style successive decoder performance.
- Additionally investigated the DNNs performance by varying model network parameters like hidden layers , number of nodes , activation functions .
- The Metric used for comparison is BER (Bit Error Rate).[1]

[1] https://ieeexplore.ieee.org/document/8390279

In channel coding a k-bit message word is converted into a N length codeword before transmission to improve the error performance. Polar Coding is one of the linear block channel coding framework which is used in general practical applications like 5G NR standards,3GPP etc.

Polar Transform :

- The main important mathematical formulation used in Polar code is Polar transform which is defined by Polar code generator matrix G_N

$$G_N = F^{\otimes n} \tag{1}$$

where F = $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $F^{\otimes n}$ denotes the nth Kronecker power (Tensor power) of F [2]. Hence G_N is a NxN matrix and where $N = 2^n$.

[2] https://www.youtube.com/watch?v=rB0rhQKyV34

Consider a vector u of dimension 1xN constructed from message word (m_K) of length k,the encoded codeword is given as $\boldsymbol{x} = \boldsymbol{u}\boldsymbol{G_N}$.[3]

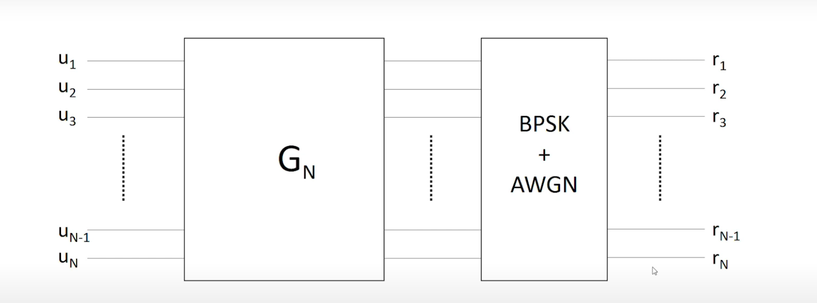

Figure: Polar Transform

[3] https://arxiv.org/abs/0807.3917

The transmission of the vector u across N different channels is observed as transmission across N different bit channels where each bit channel is characterised as follows :

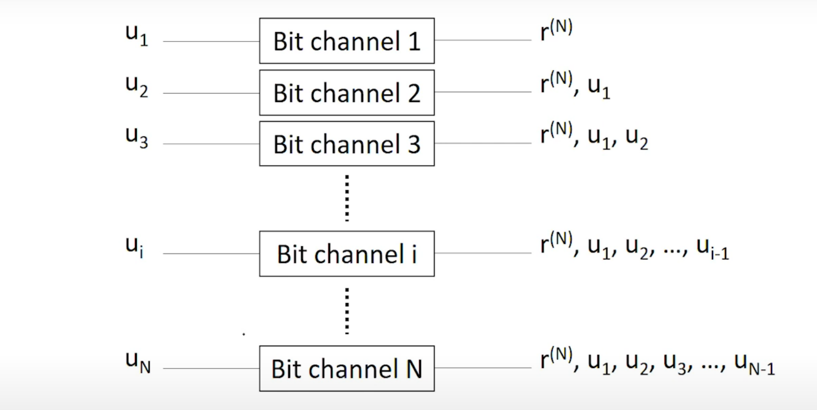

Figure: Bit Channel Representation

The Polar code generator matrix G_N is defined in such a way such that the channel gets polarized termed as channel polarization and the bit channels are ordered based on quality. The quality ranges from very good to very bad. The quality is given from worst to best based on the pre-defined reliability sequences in 5G standards for a given N.[4]

(N,k) Polar Code :

- Take message word (m_k) of length k bits.
- Form a vector u of length N bits as :
  - Find N-k least (worst) reliable channels from the reliability sequence.
  - Setting u_i for those N-k channels to zero. (Termed as frozen positions).
  - Remaining k positions of u_i are occupied by k bits of message word (m_k) (Termed as message positions).
- Forming the encoded codeword $x = uG_N$.

[4] https://www.youtube.com/watch?v=1uYEq4ue0ok

Example

For N = 16 the standard 5G reliability sequence is given as :

1 2 3 5 9 4 6 10 7 11 13 8 12 14 15 16

In the research paper the author had implemented a (16,8) Polar code where $N = 16$, $k = 8$ and rate $r = \frac{k}{N} = 0.5$.

Consider $\boldsymbol{m_k} = [\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4, \boldsymbol{m}_5, \boldsymbol{m}_6, \boldsymbol{m}_7, \boldsymbol{m}_8]$.

The vector u is given as

$\boldsymbol{u} = [0, 0, 0, 0, 0, 0, \boldsymbol{m}_1, \boldsymbol{m}_2, 0, 0, \boldsymbol{m}_3, \boldsymbol{m}_4, \boldsymbol{m}_5, \boldsymbol{m}_6, \boldsymbol{m}_7, \boldsymbol{m}_8]$.

The encoded codeword is given as $\boldsymbol{x} = \boldsymbol{u}\boldsymbol{G}_{16}$.[5]

[5] https://ieeexplore.ieee.org/document/7055304

Decoding

In the research paper the author has compared the decoding performance of neural networks with conventional Successive Cancellation (SC) list decoding algorithm.

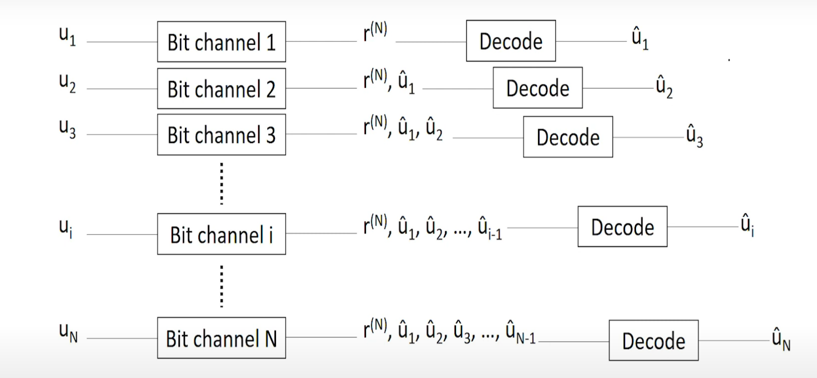

Figure: SC Decoding

In SC decoding the previously estimated $\hat{u}_{i-1}$ is used successively for estimating $\hat{u}_i$.[6]

In SC list decoding instead of giving one possible codeword it gives a list (L) possible codewords. (The decoder proceeds with the both the decisions of the estimate for successive decoding till it reaches giving L possible codewords). [7]

[6] https://www.youtube.com/watch?v=O3JWkvEY8Lc

[7] https://www.youtube.com/watch?v=WbC5Ux5Pjp8

- As in Fig. a message of k information bits is encoded with a regular polar coding scheme to yield a N-bit codeword.
- The encoded code word is BPSK modulated while adding a white gaussian noise of variance : σ^2
- For evaluating , a simple (16, 8) polar code is considered.

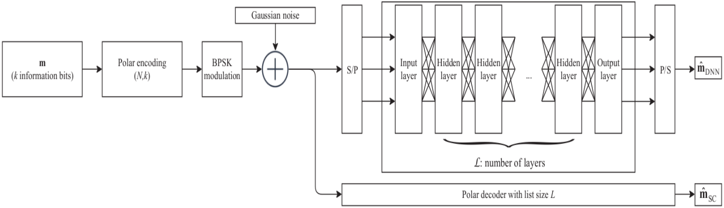

Figure: Overall block diagram of decoding of polar codes with DNN and list decoding

Setup

- DNN Specifications
  - Input layer size : 16 ; Output layer size : 8.
  - Size of training set : x : $(2^8, 16)$, y:$(2^8, 8)$.
  - Since this can be formualated as a classification problem , sigmoid activation is used at the end to generate bit streams .
- Compared BER performances over different SNR with list-style successive decoding algorithms like "SC with L=8", "SC with L=1".

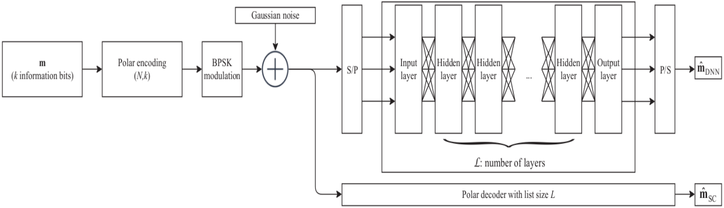

Figure: Overall block diagram of decoding of polar codes with DNN and list decoding

Results : Effect of Activation in hidden layers

- 5 different activation functions were explored namely, the linear, the sigmoid, the softplus, the tanh and the ReLU. A DNN with hidden layer configuration of 128-64-32 has been used.
- ReLU outperforms "SC with L=1", while linear and sigmoid performs the poorest.

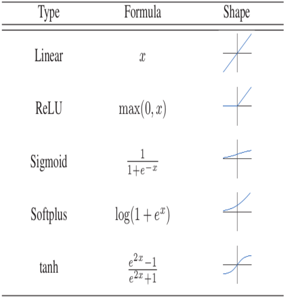

Figure: Example of Activation Functions explored

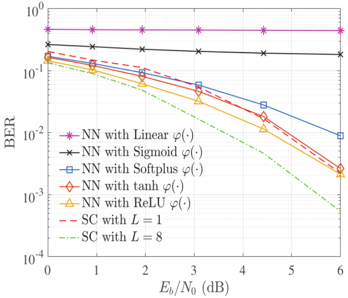

Figure: BER with changing activation functions.

- Hidden layers : (while no. of nodes per layer is fixed to 64) : BER improves as number of layers increases. A four layered network (64-64-64-64) performs closer to SC with L=8 performance.
- No. of nodes : (Fixed number of layers to 2) : As number of nodes per layer increases BER performance improves.

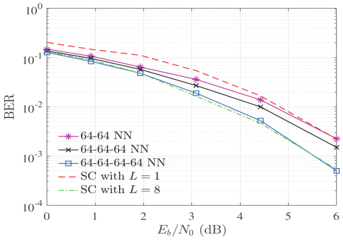

Figure: BER with changing no. of hidden layers

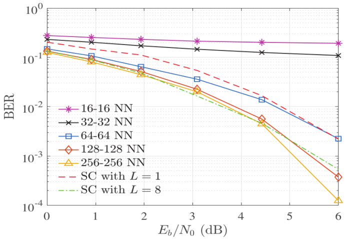

Figure: BER with changing no. of nodes

Resutls : Effect of learning

- Epoch : Authors suggest to train a DNN with **M** epochs , such that $M >= 2^N$ where N denotes the length of codewords.
- Dropout : As dropout ratio increases , BER gets worse. The reason could be since stochasticity is being introduced by noisy modulation , adding dropout is making the model much worse.

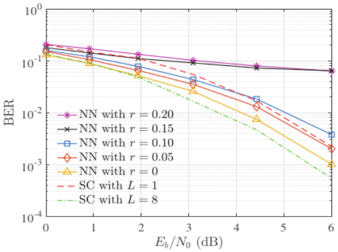

Figure: BER with changing dropout ratios

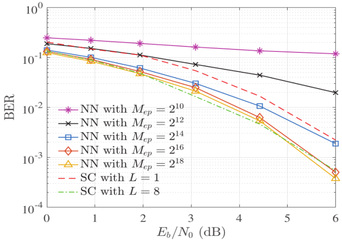

Figure: BER with changing no. of epochs

Effect of learning

- Weight distribution : Shapes of the normalised weight distribution curves are similar regardless of DNN configurations.
- They tend to get sharper as number of nodes increase. [8]

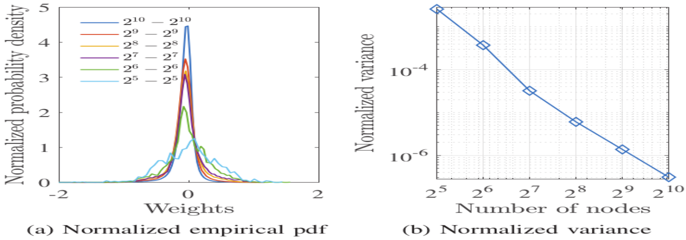

Figure: Statistical behaviour of normalised weights of the network

[8] https://arxiv.org/abs/1701.07738

Conclusion

Deep Learning techniques are upon a keen interest in the area of communication engineering and deep learning frameworks are proposed for channel decoding problems as they belong to category of classification problem. In the research paper such a DNN has been modelled for decoding a polar code.