

An Analytical Study of the Min-Sum Approximation for Polar Codes

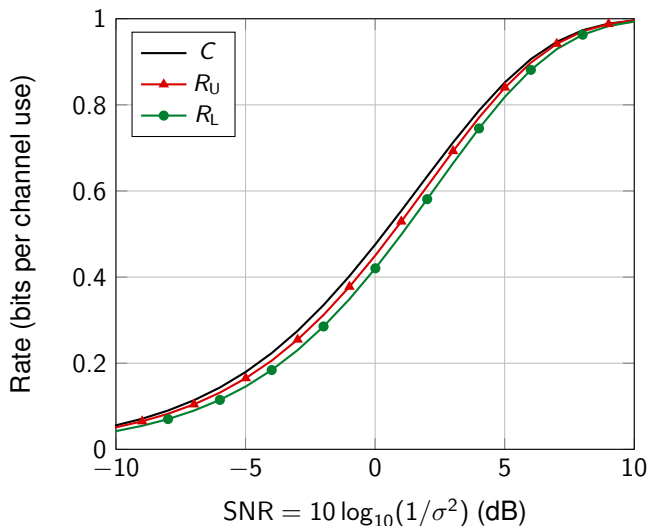
Nir Chisnevski, Ido Tal, Shlomo Shamai (Shitz)

The Andrew and Erna Viterbi Faculty of
Electrical and Computer Engineering
Technion, Israel

June 2025

IEEE ISIT 2025

A graph is worth a thousand words



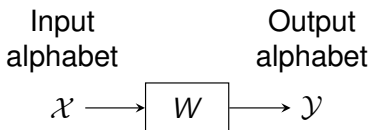
Setting: Channel $W : \mathcal{X} \rightarrow \mathcal{Y}$

- Binary input: $\mathcal{X} = \{0, 1\}$
- Memoryless
- Symmetric:
exists $\pi : \mathcal{Y} \rightarrow \mathcal{Y}$ such that:
 - $W(y|0) = W(\pi(y)|1)$
 - $\pi(\pi(y)) = y$
- Uniform input distribution:

$$p(x = 0) = p(x = 1) = 1/2$$

- Joint distribution:

$$W(y; x) = p(x) \cdot W(y|x)$$



Transition
distribution

$$W(y|x) \forall y \in \mathcal{Y}, x \in \mathcal{X}$$

Successive cancellation decoding

- Notation:

$$W_N^{(i)}(y_0^{N-1}, u_0^{i-1}; u_i) = \Pr(Y_0^{N-1} = y_0^{N-1}, U_0^{i-1} = u_0^{i-1}, U_i = u_i)$$

- Log-likelihood ratio (LLR):

$$L_N^{(i)}(y_0^{N-1}, u_0^{i-1}) = \log_2 \left(\frac{W_N^{(i)}(y_0^{N-1}, u_0^{i-1}; u_i = 0)}{W_N^{(i)}(y_0^{N-1}, u_0^{i-1}; u_i = 1)} \right)$$

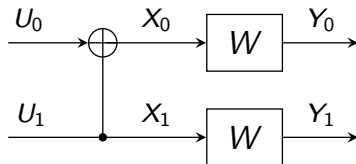
- Sequentially for $i = 0, 1, \dots, N-1$:

- if $i \in \mathcal{A}$ (information bit):

$$\hat{u}_i = \operatorname{argmax}_{u_i} W_N^{(i)}(y_0^{N-1}, \hat{u}_0^{i-1}; u_i) = \begin{cases} 0 & L_N^{(i)}(y_0^{N-1}, \hat{u}_0^{i-1}) \geq 0 \\ 1 & L_N^{(i)}(y_0^{N-1}, \hat{u}_0^{i-1}) < 0 \end{cases}$$

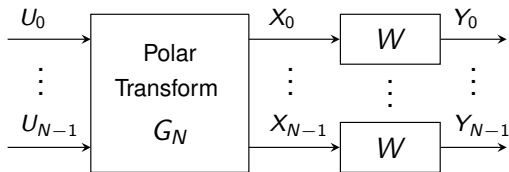
- if $i \notin \mathcal{A}$ (frozen bit): $\hat{u}_i = 0$

Decoding with LLRs for $N = 2$



- Recall, for $i \in \mathcal{A}$: $\hat{u}_i$ is set according to the sign of $L_N^{(i)}(y_0^{N-1}, \hat{u}_0^{i-1})$
- We have:
 - $L_a \triangleq L_1^{(0)}(y_0) = \log_2 \frac{W(y_0; x_0=0)}{W(y_0; x_0=1)}$
 - $L_b \triangleq L_1^{(0)}(y_1) = \log_2 \frac{W(y_1; x_1=0)}{W(y_1; x_1=1)}$
- We calculate:
 - $u_0 : L_2^{(0)}(y_0, y_1) = 2 \tanh^{-1} \left(\tanh \left(\frac{L_a}{2} \right) \cdot \tanh \left(\frac{L_b}{2} \right) \right) \triangleq f(L_a, L_b)$
 - $u_1 : L_2^{(1)}(y_0, y_1, \hat{u}_0) = (-1)^{\hat{u}_0} \cdot L_a + L_b \triangleq g_{\hat{u}_0}(L_a, L_b)$

Decoding for $N > 2$



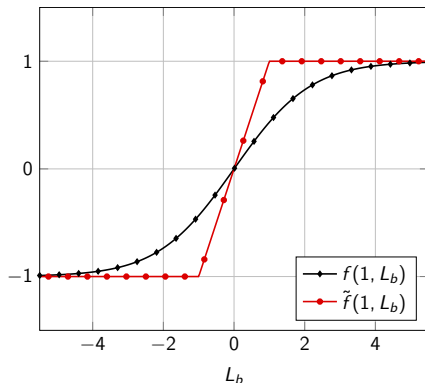
- Start with $\left\{ L_1^{(0)}(y_i) \right\}_{i=0}^{N-1}$
- Produce $\left\{ L_N^{(i)}(y_0^{N-1}, \hat{u}_0^{i-1}) \right\}_{i=0}^{N-1}$
- By recursively using f and g (applying “-” and “+” transforms):

$$f(L_a, L_b) = 2 \tanh^{-1} \left(\tanh \left(\frac{L_a}{2} \right) \cdot \tanh \left(\frac{L_b}{2} \right) \right)$$

$$g_u(L_a, L_b) = (-1)^u \cdot L_a + L_b$$

The min-sum approximation

- Practically:
 - g is simple
 - f is complicated
- Replace f with $\tilde{f}$



$$g_u(L_a, L_b) = (-1)^u \cdot L_a + L_b$$

$$f(L_a, L_b) = 2 \tanh^{-1} \left(\tanh \left(\frac{L_a}{2} \right) \cdot \tanh \left(\frac{L_b}{2} \right) \right)$$

$$\tilde{f}(L_a, L_b) = \text{sgn}(L_a) \cdot \text{sgn}(L_b) \cdot \min \{|L_a|, |L_b|\}$$

Non-approximated Min-Sum + Labeler

Exact calculation of all $P_e(U_i U_0^{i-1}, Y_0^{N-1})$	Hard	$O(N^{1.585})$
$P_e(\text{codeword}) < 2^{-N^\beta}$ for $\beta \in (0, 1/2)$	$R < C$	$R < R_L$
$P_e(\text{codeword}) \rightarrow 1$	$R > C$	$R > R_U$

Main ideas

- Use **posynomial** $Q(\xi)$ in indeterminate ξ to encapsulate all labels (approximated LLRs) and their respective probabilities:

$$Q(\xi) = \sum_t Q(t; 0) \xi^t$$

- Apply $\tilde{f}/g$ ($-/+$ transforms) in posynomial domain
- Use $Q(\xi)$ to obtain **Bhattacharyya-like** bound, satisfying

$$Z^- \leq 2Z, \quad Z^+ = Z^2$$

$\tilde{f}$ and g are positive homogeneous

- For $\alpha > 0$:
 - $\tilde{f}(\alpha \cdot L_a, \alpha \cdot L_b) = \alpha \cdot \tilde{f}(L_a, L_b)$
 - $g_u(\alpha \cdot L_a, \alpha \cdot L_b) = \alpha \cdot g_u(L_a, L_b)$
- Scaling initial LLRs doesn't change the estimations $\{\hat{u}_i\}_{i=0}^{N-1}$
 - decisions are made by the sign of $L_N^{(i)}(y_0^{N-1}, \hat{u}_0^{i-1})$
- Replace initial LLRs with generalized **labels**
 - $L_1^{(0)} = \text{LLR}(y) \implies \tilde{L}_1^{(0)} = \lambda(y)$
 - $L_N^{(i)}$ obtained using f and $g \implies \tilde{L}_N^{(i)}$ obtained using $\tilde{f}$ and g

$$\tilde{f}(L_a, L_b) = \text{sgn}(L_a) \cdot \text{sgn}(L_b) \cdot \min\{|L_a|, |L_b|\}$$

$$g_u(L_a, L_b) = (-1)^u \cdot L_a + L_b$$

Definition of good labeler $\lambda : \mathcal{Y} \rightarrow \mathbb{R}$

1 Symmetry preservation:

- Since $W(y|0) = W(\pi(y)|1) \implies L_1^{(0)}(\pi(y)) = -L_1^{(0)}(y)$
- Require $\lambda(\pi(y)) = -\lambda(y)$

2 Sign consistency: for all positive label t , $x = 0$ is more likely.

$$t > 0 \implies \sum_{y:\lambda(y)=t} W(y|0) \geq \sum_{y:\lambda(y)=t} W(y|1)$$

The inequality is strict for at least one label (not the “all erasures” channel).

3 Finite integer range: the range of λ is contained in

$$\{-\gamma, -\gamma + 1, \dots, \gamma - 1, \gamma\}$$

for some positive integer γ .

Min-sum joint distribution

- Define joint-distribution between input u_i and output label t :

$$\tilde{Q}_N^{(i)}(t; u_i) \triangleq \sum_{\substack{y_0^{N-1}, u_0^{i-1}: \\ \tilde{L}_N^{(i)}(y_0^{N-1}, u_0^{i-1})=t}} W_N^{(i)}(y_0^{N-1}, u_0^{i-1}; u_i)$$

- $\tilde{\mathcal{T}}_N^{(i)}$: the support of $\tilde{Q}_N^{(i)}(t; u_i)$ with respect to t
- Under a good labeler:

$$\begin{aligned} \tilde{\mathcal{T}}_1^{(0)} &\subseteq \{-\gamma, -\gamma + 1, \dots, \gamma - 1, \gamma\} \\ \implies \tilde{\mathcal{T}}_N^{(i)} &\subseteq \{-2^{\text{wt}(i)} \cdot \gamma, \dots, 2^{\text{wt}(i)} \cdot \gamma\} \end{aligned}$$

where $\text{wt}(i)$ is the Hamming weight of i , i.e. the number of “+”.

- The size of $\tilde{\mathcal{T}}_N^{(i)}$ grows **linearly** (not exponentially) with N

$$\tilde{Q}_N^{(i)}(\xi) \triangleq \sum_{t \in \tilde{T}_N^{(i)}} \tilde{Q}_N^{(i)}(t; 0) \cdot \xi^t$$

- Coefficients - probabilities (non-negative)
- Exponents - labels (real, can be negative)
- Contains all information about $\tilde{Q}_N^{(i)}(t; u)$ (joint-distribution)
 - Information on $\tilde{Q}_N^{(i)}(t; 1)$ is available due to symmetry

Posynomial (explicit, low-complexity) transforms

Plus transform (applying g)

$$\tilde{Q}_N^{(2j+1)}(\xi) = 2 \cdot \left(\tilde{Q}_{N/2}^{(j)}(\xi) \right)^2$$

- Polynomial multiplication
- Complexity $\mathcal{O} \left(|\tilde{T}_{N/2}^{(j)}| \cdot \log(|\tilde{T}_{N/2}^{(j)}|) \right)$

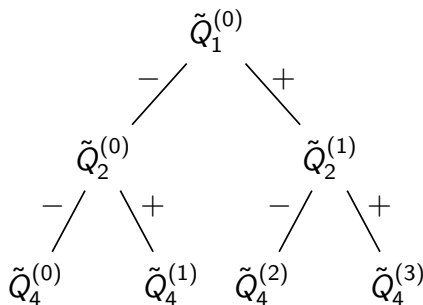
Minus transform (applying $\tilde{f}$)

$$\begin{aligned} \tilde{Q}_N^{(2j)}(\xi) = & 2 \left(\tilde{Q}_{N/2}^{(j)}(\xi) \odot \left(2 \cdot \text{pos} \left\langle \tilde{A}_{N/2}^{(j)}(\xi) \right\rangle + \tilde{Q}_{N/2}^{(j)}(\xi) \right) \right) \\ & + 2 \left(\tilde{Q}_{N/2}^{(j)}(1/\xi) \odot \left(2 \cdot \text{neg} \left\langle \tilde{B}_{N/2}^{(j)}(\xi) \right\rangle + \tilde{Q}_{N/2}^{(j)}(1/\xi) \right) \right) \\ & - 2 \left([\xi^0] \tilde{Q}_{N/2}^{(j)}(\xi) \right)^2 \end{aligned}$$

- Linear complexity $\mathcal{O} \left(|\tilde{T}_{N/2}^{(j)}| \right)$

Calculating probability of error (finite-length N)

- Calculate all posynomials $\{\tilde{Q}_N^{(i)}(\xi)\}_{i=0}^{N-1}$ in time $\mathcal{O}(N^{1.585} \cdot \gamma \log \gamma)$



- Calculate all error probabilities (in the above complexity):

$$P_e \left(\tilde{Q}_N^{(i)} \right) = \tilde{Q}_N^{(i)}(0; 0) + 2 \cdot \sum_{t < 0} \tilde{Q}_N^{(i)}(t; 0)$$

- Yields a **construction algorithm** (finding the information set $\mathcal{A}$) with the above complexity

Asymptotic case ($N \rightarrow \infty$)

- Define a Bhattacharyya-like parameter

$$Z\left(\tilde{Q}_N^{(i)}, \xi_0\right) \triangleq 2 \cdot \tilde{Q}_N^{(i)}(\xi_0) \quad \text{for } \xi_0 \in (0, 1]$$

- Upper bound:

$$P_e\left(\tilde{Q}_N^{(i)}\right) \leq Z\left(\tilde{Q}_N^{(i)}, \xi_0\right)$$

- Simple evolutions:

$$\begin{aligned} Z\left(\tilde{Q}_N^{(2j)}, \xi_0\right) &\leq 2 \cdot Z\left(\tilde{Q}_{N/2}^{(j)}, \xi_0\right) \\ Z\left(\tilde{Q}_N^{(2j+1)}, \xi_0\right) &= \left(Z\left(\tilde{Q}_{N/2}^{(j)}, \xi_0\right)\right)^2 \end{aligned}$$

same as Bhattacharyya
parameter for non-
approximated decoding

- Want tight bound on $P_e \implies$ use smallest Z (pick best ξ_0)
- Define:

$$Z^* \left(\tilde{Q}_N^{(i)} \right) \triangleq \min_{0 < \xi_0 \leq 1} Z \left(\tilde{Q}_N^{(i)}, \xi_0 \right)$$

- Instance of geometric programming
 - Solved using convex optimization methods

Key proposition

Polarization

Let $Z_0, Z_1, \dots$ be a $[0, 1]$ -valued random process that satisfies for $n \geq 0$

$$Z_{n+1} \leq 2 \cdot \begin{cases} Z_n, & \text{w.p. } \frac{1}{2}, \\ Z_n^2, & \text{w.p. } \frac{1}{2}. \end{cases}$$

Fix $\beta \in (0, 1/2)$, $\eta > 0$, and $\delta > \delta'(\eta)$ where

$$\delta'(\eta) \triangleq 2 \cdot (8\eta)^{\log_2 \varphi} \quad \text{and} \quad \varphi = \frac{1 + \sqrt{5}}{2}.$$

Then, there exists $n_0 = n_0(\beta, \delta - \delta'(\eta))$ such that if $Z_0 \leq \eta$ then

$$\Pr\left(Z_n \leq 2^{-2^{n\beta}} \text{ for all } n \geq n_0\right) \geq 1 - \delta.$$

Concisely

$$Z_0 \ll 1 \implies \Pr(Z_n \leq 2^{-2^{n\beta}}) \approx 1$$

Main theorem

Formally

Fix $\beta \in (0, 1/2)$. There exist thresholds $R_L = R_L(\lambda)$ and $R_U = R_U(\lambda)$, such that $0 < R_L \leq R_U$. For any code rate $R < R_L$ there exists a family of polar codes with growing lengths such that their rates are at least R and their word error probabilities are at most 2^{-N^β} , where N is the codeword length. Conversely, for code rates exceeding R_U , the word error probability approaches 1 as the code-length increases, irrespective of which bits are frozen.

Concisely

- Below R_L : $P_e(\text{codeword}) < 2^{-N^\beta}$
- Above R_U : $P_e(\text{codeword}) \rightarrow 1$

Upper rate threshold R_U

- Trivial threshold is $R_U = I(W) = I(X; Y)$
- Calculate posynomials up to depth $d = v$ ($V = 2^v$)
- Calculate all mutual informations $I(\tilde{Q}_V^{(j)})$
- Define:

$$R_U(V) = \frac{1}{V} \sum_{j=0}^{V-1} I(\tilde{Q}_V^{(j)})$$

Lower rate threshold R_L

- Calculate posynomials up to depth $d = v$ ($V = 2^v$)
- Calculate all Bhattacharyya-like bounds $Z^*(\tilde{Q}_V^{(j)})$
- Apply previous “key proposition” and define:

$$R_L(V) = \frac{1}{V} \sum_{j=0}^{V-1} \max \left\{ 1 - \delta' \left(Z^*(\tilde{Q}_V^{(j)}) \right), 0 \right\}$$