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, Haifa 32000, Israel.
Email: {nir.ch@campus, idotal@ee, sshlomo@ee}.technion.ac.il

***Abstract*— The min-sum approximation is widely used in the decoding of polar codes. Although it is a numerical approximation, hardly any penalties are incurred in practice. We give a theoretical justification for this. We consider the common case of a binary-input, memoryless, and symmetric channel, decoded using successive cancellation and the min-sum approximation. Under mild assumptions, we show the following. For the finite length case, we show how to exactly calculate the error probabilities of all synthetic (bit) channels in time $\mathcal{O}(N^{1.585})$, where N is the codeword length. This implies a code construction algorithm with the above complexity. For the asymptotic case, we develop two rate thresholds, denoted $R_{\mathrm{L}} = R_{\mathrm{L}}(\lambda)$ and $R_{\mathrm{U}} = R_{\mathrm{U}}(\lambda)$, where $\lambda(\cdot)$ is the labeler of the channel outputs (essentially, a quantizer). For any $0 < \beta < \frac{1}{2}$ and any code rate $R < R_{\mathrm{L}}$, there exists a family of polar codes with growing lengths such that their rates are at least R and their error probabilities are at most $2^{-N^{\beta}}$. That is, strong polarization continues to hold under the min-sum approximation. Conversely, for code rates exceeding R_{U}, the error probability approaches 1 as the code-length increases, irrespective of which bits are frozen. We show that $0 < R_{\mathrm{L}} \leq R_{\mathrm{U}} \leq C$, where C is the channel capacity. The last inequality is often strict, in which case the ramification of using the min-sum approximation is that we can no longer achieve capacity.**

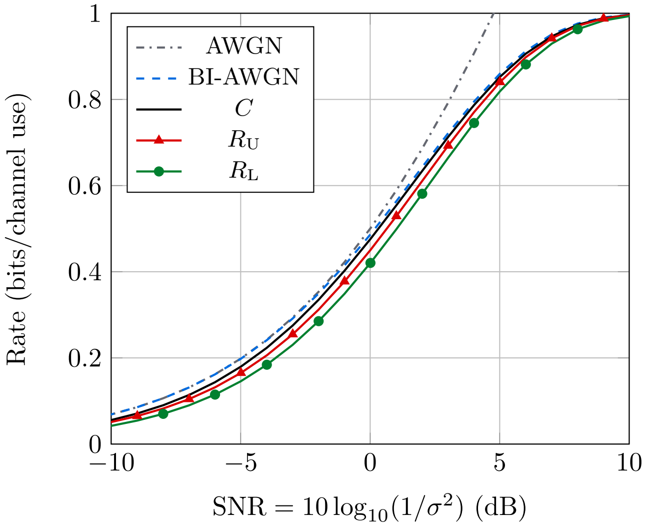

Fig. 1. The capacity C and the thresholds R_{U} and R_{L} of a BI-AWGN with 3-bit quantized output, using the labeling function λ given in (48). For reference, the capacities of the corresponding non-quantized BI-AWGN and AWGN are also given.

I. Introduction

Polar codes are a family of capacity-achieving error correcting codes with efficient encoding and decoding algorithms, introduced by Arıkan [1]. In this paper, we study the setting of a binary-input, memoryless and symmetric channel. Although many generalizations to this case exist [2]–[16], it is arguably the most basic and common one. Moreover, it affords a very efficient hardware implementation using the numerical min-sum approximation (MSA) in the decoder.

The seminal decoding algorithm of polar codes is called successive-cancellation (SC) decoding. It is a recursive algorithm that makes repeated use of the following two functions:

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

$$g_u(L_a, L_b) = (-1)^u \cdot L_a + L_b\,. \quad (2)$$

The functions g_0 and g_1 are simple to implement, since addition and subtraction are hardware-friendly operations. However, the f function is somewhat complicated, since hyperbolic functions are expensive in terms of calculation time and power consumption. Therefore, in many practical implementations the MSA is used [17]. That is, similar to what is done in LDPC decoder implementation [18], the f function is replaced with a simpler function $\tilde{f}$ given by

$$\tilde{f}(L_a, L_b) = \mathrm{sgn}(L_a)\cdot\mathrm{sgn}(L_b)\cdot\min\{|L_a|, |L_b|\}\,, \quad (3)$$

where $\mathrm{sgn}(\cdot)$ is the sign function defined as

$$\mathrm{sgn}(x) \triangleq \begin{cases} 1 & \text{if } x > 0\,, \\ -1 & \text{if } x < 0\,, \\ 0 & \text{if } x = 0\,. \end{cases}$$

For the non-approximated setting, L_a, L_b, and the outputs of f and g are log-likelihood ratios (LLRs) corresponding to certain channel outputs. For the approximated setting, we use the generalized term ‘labels’ for the corresponding quantities. At the base of the recursion the labels L_a and L_b are obtained by applying a labeling function $\lambda(\cdot)$ on the channel outputs. The full definition of $\lambda(\cdot)$ is given in Section II. Informally, $\lambda(y)$ is a quantized version of the LLR corresponding to the channel output y, up to a positive scaling constant.

The MSA is also used in decoders that are derivatives of the SC decoder, such as the SC list decoder [19] and the SC stack decoder [20]. Often, the MSA incurs only a small penalty in

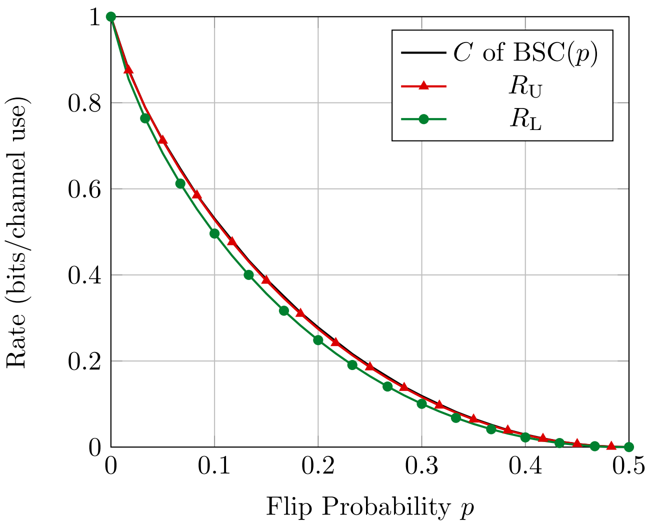

Fig. 2. The capacity C and the thresholds R_{U} and R_{L} for the BSC(p).

error rate [21, Figure 7], [22, Figures 7-8], and [23, Figure 4]. In this paper, we analyze this phenomenon.

The following theorem is our main result for the asymptotic case. The theorem promises two rate thresholds, R_{L} and R_{U}, when employing the MSA in SC decoding. Below R_{L}, strong polarization is guaranteed, while above R_{U} the error probability approaches 1. Figures 1 and 2 plot these thresholds and the channel capacity C for the binary-input additive white Gaussian noise (BI-AWGN) channel with quantized output, and for the binary symmetric channel (BSC), respectively. As can be seen in these figures, R_{L}, R_{U}, and C are all rather close. However, note that R_{U} is strictly smaller than C. That is, in these cases using the MSA means that we can no longer achieve capacity. The theorem assumes that a "fair labeler" is used, as defined in Definition 2 below.

Theorem 1. *Let W be a binary-input, memoryless and symmetric channel. Fix $0 < \beta < \frac{1}{2}$. Let $\lambda(\cdot)$ be a fair labeler. Then, there exist thresholds $R_{\mathrm{L}} = R_{\mathrm{L}}(\lambda)$ and $R_{\mathrm{U}} = R_{\mathrm{U}}(\lambda)$, such that $0 < R_{\mathrm{L}} \leq R_{\mathrm{U}}$. When using SC decoding and the MSA, the following holds. For any code rate $R < R_{\mathrm{L}}$ 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 $2^{-N^{\beta}}$, 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.*

If we only assume a fair labeler, R_{L} is weak but still positive, and R_{U} is trivial. For a significant subclass of fair labelers, "good labelers" (Definition 1), both bounds can be significantly strengthened. A good labeler is often the case in practice.

For the finite-length case and the good labeler setting, we develop an algorithm for calculating the exact error probability of each min-sum synthetic channel, defined in (13). The running time of our algorithm is $\mathcal{O}(N^{1.585})$. Note that in the non-approximated setting, no such algorithm exists, only a method to calculate bounds on the error probabilities [24].

II. Notation

Denote by $W : \mathcal{X} \to \mathcal{Y}$ a general binary-input, memoryless, and symmetric channel with input alphabet $\mathcal{X} = \{0, 1\}$ and output alphabet $\mathcal{Y}$. For each pair $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ the input probability is $p(x)$, and the transition probability is $W(y|x)$. Hence, the joint probability is given by $W(y; x) = p(x) \cdot W(y|x)$. We will assume that $p(x)$ is symmetric, i.e. $p(0) = p(1) = 1/2$.

For $n \in \mathbb{N}$ denote $N = 2^n$ and let $(X_i, Y_i)_{i=0}^{N-1}$ be N i.i.d. pairs, each distributed according to $W(y; x)$. Denote by U_0^{N-1} the polar transform of X_0^{N-1}. For $0 \leq i < N$, define the following synthetic joint distribution[1]:

$$\begin{aligned} W_N^{(i)}(y_0^{N-1}, u_0^{i-1}; u_i) &= \\ &\Pr(Y_0^{N-1} = y_0^{N-1}, U_0^{i-1} = u_0^{i-1}, U_i = u_i) \,. \end{aligned} \tag{4}$$

By [1, Proposition 3],

$$\begin{aligned} &W_N^{(2j)}\left(y_0^{N-1}, u_0^{2j-1}; u_{2j}\right) = \\ &\sum_{u_{2j+1}} W_{N/2}^{(j)}\left(y_0^{\frac{N}{2}-1}, u_{0,e}^{2j-1} \oplus u_{0,o}^{2j-1}; u_{2j} \oplus u_{2j+1}\right) \\ &\qquad \cdot W_{N/2}^{(j)}\left(y_{\frac{N}{2}}^{N-1}, u_{0,o}^{2j-1}; u_{2j+1}\right), \end{aligned} \tag{5}$$

and

$$\begin{aligned} &W_N^{(2j+1)}\left(y_0^{N-1}, u_0^{2j}; u_{2j+1}\right) = \\ &W_{N/2}^{(j)}\left(y_0^{\frac{N}{2}-1}, u_{0,e}^{2j-1} \oplus u_{0,o}^{2j-1}; u_{2j} \oplus u_{2j+1}\right) \\ &\qquad \cdot W_{N/2}^{(j)}\left(y_{\frac{N}{2}}^{N-1}, u_{0,o}^{2j-1}; u_{2j+1}\right), \end{aligned} \tag{6}$$

where $W_1^{(0)}(y; x) = W(y; x)$ and "$\oplus$" is addition over GF(2). In the above, $u_{0,e}^{2j-1}$ and $u_{0,o}^{2j-1}$ are the even and odd entries of u_0^{2j-1}, respectively. As shown in [1], $W_N^{(2j)}$ and $W_N^{(2j+1)}$ are the result of applying the "$-$" and "+" transforms, respectively, on $W_{N/2}^{(j)}$, up to a relabeling of the output.

For each joint distribution $W_N^{(i)}$ we define the LLR $L_N^{(i)}$ as

$$L_N^{(i)}(y_0^{N-1}, u_0^{i-1}) \triangleq \log_2\left(\frac{W_N^{(i)}(y_0^{N-1}, u_0^{i-1}; u_i = 0)}{W_N^{(i)}(y_0^{N-1}, u_0^{i-1}; u_i = 1)}\right) . \tag{7}$$

Using the relations described in (5) and (6) we obtain the following recursive transforms for the LLRs:

$$\begin{aligned} &L_N^{(2j)}\left(y_0^{N-1}, u_0^{2j-1}\right) = \\ &f\left(L_{N/2}^{(j)}\left(y_0^{N/2-1}, u_{0,e}^{2j-1} \oplus u_{0,o}^{2j-1}\right), L_{N/2}^{(j)}\left(y_{N/2}^{N-1}, u_{0,o}^{2j-1}\right)\right), \end{aligned} \tag{8a}$$

$$\begin{aligned} &L_N^{(2j+1)}\left(y_0^{N-1}, u_0^{2j}\right) = \\ &g_{u_{2j}}\left(L_{N/2}^{(j)}\left(y_0^{N/2-1}, u_{0,e}^{2j-1} \oplus u_{0,o}^{2j-1}\right), L_{N/2}^{(j)}\left(y_{N/2}^{N-1}, u_{0,o}^{2j-1}\right)\right), \end{aligned} \tag{8b}$$

[1] We find it notationally easier to track joint distributions instead of channels. The latter is simply obtained from the former by multiplying by 2.

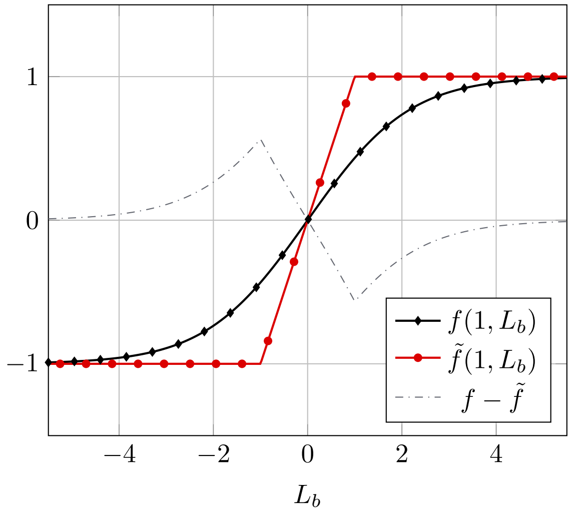

Fig. 3. A comparison between the non-approximated function $f(L_a, L_b)$ and the approximated function $\tilde{f}(L_a, L_b)$ for $L_a = 1$.

where f and g are defined in (1) and (2), respectively. The starting condition for this recursion is

$$L_1^{(0)}(y) = \mathrm{LLR}(y) \triangleq \log_2\left(W(y;0)/W(y;1)\right) \ . \tag{9}$$

The SC decoder uses f and g to recursively calculate the LLRs of all synthetic joint distributions, yielding a decoding algorithm with running time $\mathcal{O}(N \log N)$.

The min-sum SC decoder is a simplified version of the original SC decoder, as it uses $\tilde{f}$ (see (3)) instead of the computationally heavier f during the recursion. A graphical comparison between these two functions if given in Figure 3. Unlike f, both $\tilde{f}$ and g are positive homogeneous (i.e. multiplying both inputs by a positive constant multiplies the output by the same constant). This implies that the min-sum decoder is not affected by scaling. Therefore, we further extend the approximation and allow the initial labels at the base of the recursion not to be LLRs, but some values obtained by applying a labeling function λ on the channel outputs. We now list 3 properties required of a labeler λ to be called a "good labeler".

Definition 1 (Good labeler)**.** A labeler $\lambda : \mathcal{Y} \to \mathbb{R}$ is a good labeler with respect to a binary-input memoryless symmetric channel $W : \mathcal{X} \to \mathcal{Y}$ if the following holds:

1) Symmetry preservation: since W is symmetric, there exists a permutation $\pi : \mathcal{Y} \to \mathcal{Y}$ such that for all $y \in \mathcal{Y}$, $W(y|1) = W(\pi(y)|0)$ and $\pi(\pi(y)) = y$. We require that $\lambda(\pi(y)) = -\lambda(y)$ for all $y \in \mathcal{Y}$.
2) Sign consistency: for all positive t we have $\alpha_t \geq \alpha_{-t}$, where $\alpha_t = \sum_{y:\lambda(y)=t} W(y|0)$, and the inequality is strict for at least one t.
3) Finite integer range: the range of λ is contained in $\{-\gamma, -\gamma+1, \ldots, \gamma-1, \gamma\}$, for some positive integer γ.

Note that the requirement of a strict inequality in the second property rules out channels with capacity zero. We also define a "fair labeler" as follows.

Definition 2 (Fair labeler)**.** A labeler λ is a fair labeler if the first two requirements of a good labeler are met.

Note that if we were to take $\lambda(y) = \mathrm{LLR}(y)$, we would have a fair labeler, for any channel with positive capacity. The last property of the good labeler is required only for computational reasons, and is often the case due to quantization. The justification for it is by the homogeneous property of $\tilde{f}$ and g and its implications, as described above.

Under the MSA, labels are calculated recursively by

$$\tilde{L}_N^{(2j)}\left(y_0^{N-1}, u_0^{2j-1}\right) = \tag{10a}$$
$$\tilde{f}\left(\tilde{L}_{N/2}^{(j)}\left(y_0^{N/2-1}, u_{0,e}^{2j-1} \oplus u_{0,o}^{2j-1}\right), \tilde{L}_{N/2}^{(j)}\left(y_{N/2}^{N-1}, u_{0,o}^{2j-1}\right)\right) ,$$
$$\tilde{L}_N^{(2j+1)}\left(y_0^{N-1}, u_0^{2j}\right) = \tag{10b}$$
$$g_{u_{2j}}\left(\tilde{L}_{N/2}^{(j)}\left(y_0^{N/2-1}, u_{0,e}^{2j-1} \oplus u_{0,o}^{2j-1}\right), \tilde{L}_{N/2}^{(j)}\left(y_{N/2}^{N-1}, u_{0,o}^{2j-1}\right)\right) ,$$

where $\tilde{f}$ and g are defined in (3) and (2), respectively. Note the similarity between (8) and (10). As opposed to the starting condition (9) for the non-approximated setting, the starting condition under the MSA is

$$\tilde{L}_1^{(0)}(y) = \lambda(y) \ . \tag{11}$$

The use of a labeling function λ is beneficial in practice, since it allows us to avoid the estimation of unknown channel parameters. For example, consider the case of an AWGN channel with unknown noise level σ^2. Thus, the LLR of output symbol y is given by $2y/\sigma^2$. However, using $\lambda(y) = \mathrm{LLR}(y) = 2y/\sigma^2$ will give exactly the same results as using $\lambda(y) = \alpha \cdot y$, where $\alpha > 0$ is some fixed positive constant. The utility of the latter fair labeling function is that σ^2 need not be estimated. In practice, we use the following good labeler, which is a quantized version of the previous fair labeler,

$$\lambda(y) = \begin{cases} \mathrm{sgn}(y) \cdot \lfloor \alpha \cdot |y| \rfloor & \text{if } |y| < \gamma/\alpha \ , \\ \mathrm{sgn}(y) \cdot \gamma & \text{if } |y| \geq \gamma/\alpha \ . \end{cases}$$

We optimize α and γ to work well over the range σ^2 is likely to belong to.

III. Posynomial representation

For a fair labeler, we now define the synthetic joint distributions (on the label t and input u_i) at stage i of the SC decoder and min-sum SC decoder. These are, respectively,

$$Q_N^{(i)}(t; u_i) \triangleq \sum_{\substack{y_0^{N-1}, u_0^{i-1}: \\ L_N^{(i)}(y_0^{N-1}, u_0^{i-1}) = t}} W_N^{(i)}(y_0^{N-1}, u_0^{i-1}; u_i) \ , \tag{12}$$

$$\tilde{Q}_N^{(i)}(t; u_i) \triangleq \sum_{\substack{y_0^{N-1}, u_0^{i-1}: \\ \tilde{L}_N^{(i)}(y_0^{N-1}, u_0^{i-1}) = t}} W_N^{(i)}(y_0^{N-1}, u_0^{i-1}; u_i) \ . \tag{13}$$

Denote by $\mathcal{T}_N^{(i)}$ and $\tilde{\mathcal{T}}_N^{(i)}$ the support of $Q_N^{(i)}(t; u_i)$ and $\tilde{Q}_N^{(i)}(t; u_i)$, respectively, with respect to t.

Using the relations in (5)–(11), we obtain the following minus and plus transforms of synthetic joint distributions.

Lemma 2 (Transforms of synthetic joint distributions)**.**

$$\tilde{Q}_N^{(2j)}(t;u_{2j}) = \sum_{\substack{t_a,t_b,u_{2j+1}: \\ \tilde{f}(t_a,t_b)=t}} \tilde{Q}_{N/2}^{(j)}(t_a;u_{2j}\oplus u_{2j+1}) \cdot \tilde{Q}_{N/2}^{(j)}(t_b;u_{2j+1}) \,, \tag{14a}$$

$$\tilde{Q}_N^{(2j+1)}(t;u_{2j+1}) = \sum_{\substack{t_a,t_b,u_{2j}: \\ g_{u_{2j}}(t_a,t_b)=t}} \tilde{Q}_{N/2}^{(j)}(t_a;u_{2j}\oplus u_{2j+1}) \cdot \tilde{Q}_{N/2}^{(j)}(t_b;u_{2j+1}) \,. \tag{14b}$$

The above continues to hold if we remove all tildes.

The following lemma ensures that symmetry holds for the min-sum synthetic distributions.

Lemma 3 (Symmetry of synthetic joint distribution)**.**

$$\tilde{Q}_N^{(i)}(t;u_i) = \tilde{Q}_N^{(i)}(-t;u_i\oplus 1) \,. \tag{15}$$

The above continues to hold if we remove all tildes.

We now give intuition as to why a setting in which the MSA and a good labeler are used is much easier in terms of exactly calculating quantities of interest, as opposed to the non-approximated setting.

In the non-approximated setting, for the minus transform we have that

$$|\mathcal{T}_N^{(2j)}| \leq |\mathcal{T}_{N/2}^{(j)}|^2 \,. \tag{16}$$

Indeed, this follows since in (14a) with the tildes removed we sum over all $(t_a,t_b) \in \mathcal{T}_{N/2}^{(j)} \times \mathcal{T}_{N/2}^{(j)}$ which determine $t = f(t_a,t_b)$, and also over $u_{2j+1} \in \{0,1\}$, which does not appear in $f(t_a,t_b)$. For the plus transform, we have by inspection of (14b) with the tildes removed that $|\mathcal{T}_N^{(2j+1)}| \leq 2 \cdot |\mathcal{T}_{N/2}^{(j)}|^2$. In fact, for a symmetric channel, this can be strengthened to

$$|\mathcal{T}_N^{(2j+1)}| \leq |\mathcal{T}_{N/2}^{(j)}|^2 \,. \tag{17}$$

Indeed, by Lemma 3, $t \in \mathcal{T}_{N/2}^{(j)}$ iff $-t \in \mathcal{T}_{N/2}^{(j)}$. Thus, (17) follows by inspection of g in (2). Hence, by (16), (17), and a straightforward induction,

$$|\mathcal{T}_N^{(i)}| \leq |\mathcal{T}_1^{(0)}|^{2^{\mathrm{wt}(i)}} \,, \tag{18}$$

where $\mathrm{wt}(i)$ is the Hamming weight of the vector whose entries are the binary representation of i.

In contrast, consider the case of the MSA and a good labeler. By definition, $\tilde{\mathcal{T}}_1^{(0)} \subseteq \{-\gamma,\ldots,\gamma\}$. By inspection of (2), (3), (14), and a straightforward induction, it follows that

$$\tilde{\mathcal{T}}_N^{(i)} \subseteq \{-2^{\mathrm{wt}(i)} \cdot \gamma, \ldots, 2^{\mathrm{wt}(i)} \cdot \gamma\} \,. \tag{19}$$

This is the reason we can carry out our calculations efficiently in this case: as opposed to (18), the size of $\tilde{\mathcal{T}}_N^{(i)}$ grows linearly with N, since $\mathrm{wt}(i)$ is at most n and $N = 2^n$.

A further consequence of the symmetry is Lemma 3 is the following lemma. It gives a simple expression for the probability of error at the i-th stage of the min-sum SC decoder when aided by a genie that reveals the correct values of u_0^{i-1}.

Lemma 4.

$$P_{\mathrm{e}}\left(\tilde{Q}_N^{(i)}\right) = \tilde{Q}_N^{(i)}(0;0) + 2 \cdot \sum_{t<0} \tilde{Q}_N^{(i)}(t;0) \,. \tag{20}$$

To derive a Bhattacharyya-like upper bound on P_{e}, and to aid in notation in general, we abuse notation and define the following posynomial, in the indeterminate ξ:

$$\tilde{Q}_N^{(i)}(\xi) \triangleq \sum_t \tilde{Q}_N^{(i)}(t;0) \cdot \xi^t \,. \tag{21}$$

The above is indeed a posynomial: all the coefficients are non-negative as they are probabilities, while t is not restricted to non-negative numbers.

The following corollary justifies why in (21) we define the posynomial $\tilde{Q}_N^{(i)}(\xi)$ without taking into account terms of the form $\tilde{Q}_N^{(i)}(t;1)$.

Corollary 5. *$\tilde{Q}_N^{(i)}(t;1)$ equals the coefficient of ξ^t in $\tilde{Q}_N^{(i)}(1/\xi)$.*

We further define the following:

$$Z\left(\tilde{Q}_N^{(i)},\xi\right) \triangleq 2 \cdot \tilde{Q}_N^{(i)}(\xi) \,. \tag{22}$$

Our upper bound on P_{e} is presented in the following lemma.

Lemma 6 (Bhattacharyya-like bound)**.** *For $0 < \xi_0 \leq 1$,*

$$P_{\mathrm{e}}\left(\tilde{Q}_N^{(i)}\right) \leq Z\left(\tilde{Q}_N^{(i)},\xi_0\right) \,. \tag{23}$$

We remark that setting $\xi_0 = 1/\sqrt{2}$ and removing the tildes yields the Bhattacharyya bound on the error probability at the i-th stage of the non-approximated genie-aided SC decoder. Also, we may optimize over ξ_0 to yield the tightest upper bound, denoted

$$Z^\star\left(\tilde{Q}_N^{(i)}\right) \triangleq \min_{0<\xi_0\leq 1} Z\left(\tilde{Q}_N^{(i)},\xi_0\right) \,. \tag{24}$$

The above optimization is an instance of geometric programming, and can thus be efficiently computed [25, Section 4.5].

The following shows that the evolution of Z and $Z^\star$ is similar to the evolution of the Bhattacharyya parameter in the non-approximated setting.

Lemma 7 (Bhattacharyya-like evolutions)**.** *For $0 < \xi_0 \leq 1$ and $0 \leq j < N/2$ we have*

$$Z\left(\tilde{Q}_N^{(2j)},\xi_0\right) \leq 2 \cdot Z\left(\tilde{Q}_{N/2}^{(j)},\xi_0\right) \,, \tag{25a}$$

$$Z\left(\tilde{Q}_N^{(2j+1)},\xi_0\right) = \left(Z\left(\tilde{Q}_{N/2}^{(j)},\xi_0\right)\right)^2 \,. \tag{25b}$$

Furthermore,

$$Z^\star\left(\tilde{Q}_N^{(2j)}\right) \leq 2 \cdot Z^\star\left(\tilde{Q}_{N/2}^{(j)}\right) \,, \tag{26a}$$

$$Z^\star\left(\tilde{Q}_N^{(2j+1)}\right) = \left(Z^\star\left(\tilde{Q}_{N/2}^{(j)}\right)\right)^2 \,. \tag{26b}$$

To prove the above, we state the following two lemmas.

Lemma 8 (Bound on posynomial minus transform)**.** *For all* $0 < \xi_0 \le 1$ *we have*

$$\tilde{Q}_N^{(2j)}(\xi_0) \le 2 \cdot \tilde{Q}_{N/2}^{(j)}(\xi_0) \; . \tag{27}$$

Lemma 9 (Posynomial plus transform)**.**

$$\tilde{Q}_N^{(2j+1)}(\xi) = 2 \cdot \left(\tilde{Q}_{N/2}^{(j)}(\xi)\right)^2 \; . \tag{28}$$

The previous lemma implies that the coefficients of $\tilde{Q}_N^{(2j+1)}(\xi)$ can be calculated efficiently from those of $\tilde{Q}_{N/2}^{(j)}(\xi)$. We now show an analogous result for $\tilde{Q}_N^{(2j)}(\xi)$. In aid of this, we define the "above" and "below" posynomials:

$$\tilde{A}_N^{(i)}(\xi) \triangleq \sum_{t \in \tilde{\mathcal{T}}_N^{(i)}} \left(\sum_{t' > t} \tilde{Q}_N^{(i)}(t';0)\right) \cdot \xi^t \; , \tag{29}$$

$$\tilde{B}_N^{(i)}(\xi) \triangleq \sum_{t \in \tilde{\mathcal{T}}_N^{(i)}} \left(\sum_{t' < t} \tilde{Q}_N^{(i)}(t';0)\right) \cdot \xi^t \; . \tag{30}$$

Namely, if we write out $\tilde{Q}_N^{(i)}(\xi)$ in ascending order of powers of ξ, then the coefficient of ξ^t in $\tilde{A}_N^{(i)}(\xi)$ (resp. $\tilde{B}_N^{(i)}(\xi)$) is the sum of the coefficients strictly above (resp. below) the monomial $\tilde{Q}_N^{(i)}(t;0)\xi^t$.

Let $\Gamma(\xi)$ and $\Lambda(\xi)$ be two posynomials. Denote by $[\xi^t]\,\Gamma(\xi)$ the coefficient of ξ^t in $\Gamma(\xi)$. Define the "positive" and "negative" operators, and Hadamard (element-wise) product: these operators return posynomials, where for all t,

$$[\xi^t]\ \mathrm{pos}\left\langle \Gamma(\xi)\right\rangle = \begin{cases} [\xi^t]\,\Gamma(\xi) & t \ge 0 \; , \\ [\xi^{-t}]\,\Gamma(\xi) & t < 0 \; . \end{cases} \tag{31}$$

$$[\xi^t]\ \mathrm{neg}\left\langle \Gamma(\xi)\right\rangle = \begin{cases} [\xi^t]\,\Gamma(\xi) & t \le 0 \; , \\ [\xi^{-t}]\,\Gamma(\xi) & t > 0 \; . \end{cases} \tag{32}$$

$$[\xi^t]\,\left(\Gamma(\xi) \odot \Lambda(\xi)\right) = \left([\xi^t]\,\Gamma(\xi)\right) \cdot \left([\xi^t]\,\Lambda(\xi)\right) \; . \tag{33}$$

Lemma 10 (Posynomial minus transform)**.**

$$\begin{aligned} \tilde{Q}_N^{(2j)}(\xi) = \\ & 2\left(\tilde{Q}_{N/2}^{(j)}(\xi) \odot \left(2 \cdot \mathrm{pos}\left\langle \tilde{A}_{N/2}^{(j)}(\xi)\right\rangle + \tilde{Q}_{N/2}^{(j)}(\xi)\right)\right) \\ & + 2\left(\tilde{Q}_{N/2}^{(j)}(1/\xi) \odot \left(2 \cdot \mathrm{neg}\left\langle \tilde{B}_{N/2}^{(j)}(\xi)\right\rangle + \tilde{Q}_{N/2}^{(j)}(1/\xi)\right)\right) \\ & - 2\left([\xi^0]\,\tilde{Q}_{N/2}^{(j)}(\xi)\right)^2 \; . \end{aligned} \tag{34}$$

IV. Finite-Length Case

In this section, we assume a good labeler. For the finite length case, our aim is to calculate $P_{\mathrm{e}}\left(\tilde{Q}_N^{(i)}\right)$ for all $0 \le i < N$, where the codeword length is $N = 2^n$. The expression for this is given in (20), which we can recast using (32) as

$$P_{\mathrm{e}}\left(\tilde{Q}_N^{(i)}\right) = \mathrm{neg}\left\langle \tilde{Q}_N^{(i)}(\xi)\right\rangle\Big|_{\xi=1} \; . \tag{35}$$

We use (28) and (34) to calculate $\tilde{Q}_N^{(i)}(\xi)$ for all i, and then apply (35) to yield the error probability. The following two lemmas specify the complexity of calculating $\tilde{Q}_N^{(2j)}(\xi)$ and $\tilde{Q}_N^{(2j+1)}(\xi)$ from $\tilde{Q}_{N/2}^{(j)}(\xi)$. Namely, the complexity of calculating all the coefficients of the former, given all the coefficients of the latter. Recall that $\tilde{\mathcal{T}}_N^{(i)}$ is defined in (19).

Lemma 11 (Complexity of posynomial minus transform)**.** *The complexity of calculating* $\tilde{Q}_N^{(2j)}(\xi)$ *from* $\tilde{Q}_{N/2}^{(j)}(\xi)$ *is* $\mathcal{O}\left(|\tilde{\mathcal{T}}_{N/2}^{(j)}|\right)$*.*

Lemma 12 (Complexity of posynomial plus transform)**.** *The complexity of calculating* $\tilde{Q}_N^{(2j+1)}(\xi)$ *from* $\tilde{Q}_{N/2}^{(j)}(\xi)$ *is* $\mathcal{O}\left(|\tilde{\mathcal{T}}_{N/2}^{(j)}| \cdot \log(|\tilde{\mathcal{T}}_{N/2}^{(j)}|)\right)$*.*

The following theorem is our main result for this section. It shows that the complexity of calculating all the probabilities of error $P_{\mathrm{e}}\left(\tilde{Q}_N^{(i)}\right)$ is polynomial in the codeword length N and in γ (recall Definition 1).

Theorem 13 (Total complexity of evaluating P_{e})**.** *When using a good labeler* λ*, the complexity of calculating* $P_{\mathrm{e}}\left(\tilde{Q}_N^{(i)}\right)$ *for all* $0 \le i < N$ *is* $\mathcal{O}(N^{\log_2 3}\log N \cdot \gamma \log \gamma)$*. We simplify this to* $\mathcal{O}(N^{1.585} \cdot \gamma \log \gamma)$*.*

V. Asymptotic Case

In this section we prove Theorem 1. We first do so assuming a fair labeler, and then show how to significantly improve the thresholds R_{L} and R_{U} for the case of a good labeler. The following three results are required for deriving R_{L}.

Proposition 14. *Let* $B_1, B_2, \ldots$ *be i.i.d. random variables such that* $\Pr(B_i = 0) = \Pr(B_i = 1) = 1/2$*. Let* $S_0, S_1, \ldots$ *be a* $[0,1]$*-valued random process that satisfies*

$$S_{n+1} \le \kappa \cdot \begin{cases} S_n, & B_{n+1} = 0 \; , \\ S_n^2, & B_{n+1} = 1 \; , \end{cases} \qquad n \ge 0 \; . \tag{36}$$

Then, for every $\epsilon' > 0$ *and* $\delta' > 0$ *there exist* $n' = n'(\epsilon', \delta', \kappa)$ *and* $\eta = \eta(\epsilon', \delta', \kappa) > 0$ *such that if* $S_0 \le \eta$ *then*

$$\Pr\left(S_n \le \epsilon' \text{ for all } n \ge n'\right) \ge 1 - \delta' \; . \tag{37}$$

This is [26, Equation 171], and is the crux of proving [26, Proposition 49].

The expression for η is given in the penultimate displayed equation in [26, Appendix A], where r is defined slightly before as the largest positive solution of $\kappa^r + (2\kappa)^{-r} = 2$. In our setting, S_n will be related to $Z^\star$. Thus, by (26), we specialize to $\kappa = 2$. Plugging $x = 2^r$ into $2^r + 4^{-r} = 2$ yields $x + 1/x^2 = 2$. The three roots of this equation are 1, φ, and $-\varphi^{-1}$, where $\varphi = \frac{1}{2} \cdot \left(1 + \sqrt{5}\right)$ is the golden ratio. Thus, $r = \log_2(\varphi)$ and

$$\eta(\delta') = \frac{1}{8} \cdot (\delta'/2)^{1/\log_2 \varphi} \; . \tag{38}$$

The following result is an immediate corollary.

Corollary 15. *Let* $S_0, S_1, \ldots$ *be as in Proposition 14, with* $\kappa = 2$*. Fix* $\epsilon' > 0$ *and* $\eta > 0$*. Then there exists* $n' = n'(\epsilon', \eta)$ *such that if* $S_0 \le \eta$ *then*

$$\Pr\left(S_n \le \epsilon' \text{ for all } n \ge n'\right) \ge 1 - \delta'(\eta) \; , \tag{39}$$

where

$$\delta'(\eta) \triangleq 2 \cdot (8\eta)^{\log_2 \varphi} \quad \textit{and} \quad \varphi = (1+\sqrt{5})/2\,. \tag{40}$$

The following result is of primary importance and will be used directly to prove Theorem 1.

Proposition 16. *Let $S_0, S_1, \ldots$ be as in Proposition 14, with $\kappa = 2$. Fix $0 < \beta < 1/2$, $\eta > 0$, and $\delta > \delta'(\eta)$, where $\delta'(\eta)$ is given in (40). Then, there exists $n_0 = n_0(\beta, \delta - \delta'(\eta))$ such that if $S_0 \leq \eta$ then*

$$\Pr\left(S_n \leq 2^{-2^{n\beta}} \text{ for all } n \geq n_0\right) \geq 1 - \delta\,. \tag{41}$$

A. Fair Labeler

Proof of Theorem 1: For R_{U}, we first recall that any decoder operates on the output of W, after it has been labeled by λ. Thus, it effectively sees the channel $\tilde{Q}(t|x) = 2 \cdot \tilde{Q}_1^{(0)}(t;x)$, as defined in (13). We take R_{U} as the capacity of this channel, which is valid by the strong converse to the coding theorem, see [27, Theorem 5.8.5].

We now work towards deriving R_{L}. Consider a polar code of length $N = 2^n$ with non-frozen index set $\mathcal{A} = \{0 \leq i < N : Z^\star(\tilde{Q}_N^{(i)}) < 2^{-N^{\beta'}}\}$, where $\beta' = \frac{\beta+1/2}{2}$. By the "genie-aided decoder" argument in [1], the union bound, and Lemma 6, the error probability of such a code is at most $|\mathcal{A}| \cdot 2^{-N^{\beta'}} \leq N \cdot 2^{-N^{\beta'}} < 2^{-N^{\beta}}$, where the last inequality holds for N large enough. Thus, we must find an R_{L} such that for $R < R_{\mathrm{L}}$ fixed and all N large enough, $|\mathcal{A}| \geq N \cdot R$. Consider the set $\mathcal{A}' = \{0 \leq i < N : \zeta_N^{(i)} < 2^{-N^{\beta'}}\}$, where $\zeta_1^{(0)} = Z^\star(\tilde{Q}_1^{(0)})$ and

$$\zeta_N^{(i)} = \begin{cases} 2 \cdot \zeta_{N/2}^{(i/2)} & i \text{ is even,} \\ \left(\zeta_{N/2}^{((i-1)/2)}\right)^2 & i \text{ is odd.} \end{cases} \tag{42}$$

By (26), we have for all i that $\zeta_N^{(i)} \geq Z^\star(\tilde{Q}_N^{(i)})$. Namely, $\mathcal{A}' \subseteq \mathcal{A}$. Thus, it suffices to find an R_{L} such that for $R < R_{\mathrm{L}}$ fixed and all N large enough, $|\mathcal{A}'| \geq N \cdot R$. For any $M = 2^m$, we use the definition of $\delta'(\cdot)$ in (40) and define the following:

$$R_{\mathrm{L}}(M) \triangleq \frac{1}{M} \sum_{j=0}^{M-1} \max\left\{1 - \delta'(\zeta_M^{(j)}), 0\right\}\,. \tag{43}$$

Proving the following two items will complete the proof:

1) For a given M and $R < R_{\mathrm{L}}(M)$ there exists n_0 such that for all $n \geq n_0$ we have $|\mathcal{A}'| \geq N \cdot R$.
2) There exists an M such that $R_{\mathrm{L}}(M) > 0$.

To prove the first item, assume that R, and therefore $R_{\mathrm{L}}(M)$, are positive, otherwise the claim is trivial. For each one of the M indices $0 \leq j < M$, we invoke Proposition 16 with $\delta = \delta'(\zeta_M^{(j)}) + (R_{\mathrm{L}}(M) - R)$, $\eta = S_0 = \zeta_M^{(j)}$, and $\beta'' = \frac{\beta'+1/2}{2}$ in place of β. Denote the n_0 promised by the proposition as $n_0^{(j)}$. Now define $n_0^{\max} = \max_j n_0^{(j)}$. By (41), for $n \geq m + n_0^{\max}$ the fraction of indices $0 \leq i < N$ such that $\zeta_N^{(i)} \leq 2^{-2^{(n-m)\beta''}}$ is at least

$$\begin{aligned}
&\frac{1}{M} \sum_{j=0}^{M} \max\left\{1 - \left(\delta'(\zeta_M^{(j)}) + R_{\mathrm{L}}(M) - R\right), 0\right\} \\
&= \frac{1}{M} \sum_{j=0}^{M} \max\left\{1 - \delta'(\zeta_M^{(j)}) + R - R_{\mathrm{L}}(M), 0\right\} \\
&\geq \frac{1}{M} \sum_{j=0}^{M} \max\left\{1 - \delta'(\zeta_M^{(j)}) + R - R_{\mathrm{L}}(M), R - R_{\mathrm{L}}(M)\right\} \\
&= \frac{1}{M} \sum_{j=0}^{M} \max\left\{1 - \delta'(\zeta_M^{(j)}), 0\right\} + R - R_{\mathrm{L}}(M) \\
&= R\,.
\end{aligned}$$

For the first item to hold, we take $n_0 \geq m + n_0^{\max}$ large enough such that for all $n \geq n_0$ we have $2^{-2^{(n-m)\beta''}} \leq 2^{-2^{n\beta'}} = 2^{-N^{\beta'}}$ (ensuring $|\mathcal{A}'| \geq N \cdot R$) and $N \cdot 2^{-N^{\beta'}} < 2^{-N^{\beta}}$.

We now prove the second item. That is, it is always possible to find an M such that $R_{\mathrm{L}}(M) > 0$. We first show that $Z^\star(\tilde{Q}_1^{(0)}) < 1$. Indeed,

$$Z(\tilde{Q}_1^{(0)}, \xi)\Big|_{\xi=1} = 1 \quad \text{and} \quad \frac{d}{d\xi} Z(\tilde{Q}_1^{(0)}, \xi)\Big|_{\xi=1} > 1\,,$$

where the inequality follows by item 2 in Definition 1. Hence, for $\xi_0 < 1$ sufficiently close to 1 it must hold that $Z(\tilde{Q}_1^{(0)}, \xi_0) < 1$. Thus, $\zeta_1^{(0)} = Z^\star(\tilde{Q}_1^{(0)}) < 1$. Next, note that $\zeta_M^{(M-1)} = \left(\zeta_1^{(0)}\right)^M$. Take M as the smallest power of 2 that is at least $\log_a b$ where $a = \zeta_1^{(0)}$ and $b = \eta(1/2) \approx 0.327254$. For this choice, $R_{\mathrm{L}}(M) \geq \frac{1}{2M} > 0$, by considering the last term in (43). ■

B. Good Labeler

We now show how both thresholds R_{L} and R_{U} can be strengthened in the case of a good labeler. We give a simplified description here. We give a full and more nuanced description in the expanded version Section VI. For R_{L}, we observe the following regarding the proof of Theorem 1. Any definition of $\zeta_N^{(i)}$ that satisfies $\zeta_N^{(i)} \geq Z^\star(\tilde{Q}_N^{(i)})$ for all $0 \leq i < N$ is valid. Thus, for a parameter $V = 2^v \leq M$, and all indices $0 \leq k < V$, define $\zeta_V^{(k)} = Z^\star(\tilde{Q}_V^{(k)})$. For $N > V$, define $\zeta_N^{(i)}$ recursively according to (42). This improves $R_{\mathrm{L}}(M)$, which we now denote as $R_{\mathrm{L}}(V, M)$, since for polarization stage v we are calculating the exact values of $Z^\star(\tilde{Q}_V^{(k)})$, as opposed to bounds on them. By Lemmas 11 and 12, we can indeed calculate $\tilde{Q}_V^{(k)}$ efficiently.

To strengthen R_{U}, we now define $R_{\mathrm{U}}(V)$ as the average capacity of the channels corresponding to $\tilde{Q}_V^{(k)}$ over $0 \leq k < V$. The proof of this threshold being valid is given in Section VI. In essence, we employ a so called "block-genie" that corrects us after N/V decisions have been made. Each block of size N/V corresponds to N/V uses of one of the above channels, and hence we cannot code for this block at a rate exceeding the capacity of that channel.

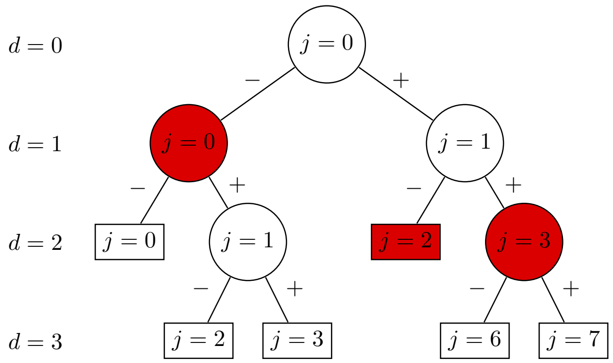

Fig. 4. A full binary tree, where the nodes in $\mathcal{G}$ are red and the nodes in $\mathcal{E}$ are rectangular (leaves). The node $(d = 2, j = 2)$ is both in $\mathcal{G}$ and $\mathcal{E}$.

VI. Improved Thresholds

In this section we give a full description of how R_{L} and R_{U} were calculated in Figures 1 and 2. These methods can be applied to any setting in which a good labeler is used.

A. Definition of $R_{\mathrm{L}}(\mathcal{G}, \mathcal{E})$ and $R_{\mathrm{U}}(\mathcal{G})$

We start by defining two sets: $\mathcal{G}$ and $\mathcal{E}$. Both sets contain depth-index pairs (d, j), where $d \geq 0$ and $0 \leq j < 2^d$. We think of each pair in these sets as a vertex of a full binary tree[2]. An example of such a tree is presented in Figure 4. The root of the tree is $(d = 0, j = 0)$. A vertex (d, j) has either no children, in which case it is a leaf, or two children: $(d+1, 2j)$ as the left child and $(d+1, 2j+1)$ as the right child. Hence, we can view j as representing a path of d edges labelled "$-$" and "$+$" starting at the root and ending at (d, j). The binary representation of $j = \sum_{i=0}^{d-1} b_i 2^i$ dictates the corresponding path. Namely, $b_i = 0$ means that the $(d-i)$-th edge is a "$-$" (left) edge, while $b_i = 1$ means that the $(d-i)$-th edge is a "$+$" (right) edge.

Given a full binary tree, the sets $\mathcal{E}$ and $\mathcal{G}$ are defined as follows. The set $\mathcal{E}$ contains the leaves of the tree. We call such a set valid. The set $\mathcal{G}$ is defined such that any path from the root to a leaf contains exactly one vertex in $\mathcal{G}$. Such a pair $(\mathcal{G}, \mathcal{E})$ is termed valid. Note that if we were to delete all descendants of vertices in $\mathcal{G}$ from the tree, we would again have a full binary tree whose leaves are $\mathcal{G}$. Hence, if $(\mathcal{G}, \mathcal{E})$ is a valid pair, then both $\mathcal{G}$ and $\mathcal{E}$ are valid.

For now, we assume that $\mathcal{G}$ and $\mathcal{E}$ are given (we will latter describe how to choose them). Our thresholds are now denoted $R_{\mathrm{L}}(\mathcal{G}, \mathcal{E})$ and $R_{\mathrm{U}}(\mathcal{G})$. For $R_{\mathrm{L}}(\mathcal{G}, \mathcal{E})$, we generalize (43) to

$$R_{\mathrm{L}}(\mathcal{G}, \mathcal{E}) = \sum_{(d,j) \in \mathcal{E}} \frac{1}{2^d} \cdot \max\left\{1 - \delta'(\zeta_{2^d}^{(j)}), 0\right\} \ , \tag{44}$$

where δ' is defined in (40) and $\zeta_{2^d}^{(j)}$ is defined recursively in (42), with the following starting conditions:

$$\zeta_{2^d}^{(j)} = Z^{\star}(\tilde{Q}_{2^d}^{(j)}) \ , \quad \text{for all } (d, j) \in \mathcal{G}. \tag{45}$$

Note that with respect to the simplified description in Section V-B, if we define (for $v \leq m$):

$$\begin{aligned} \mathcal{G}(V) &= \{(d, j) : d = v \quad \text{and} \quad 0 \leq j < V = 2^v\} \ , \qquad (46) \\ \mathcal{E}(M) &= \{(d, j) : d = m \quad \text{and} \quad 0 \leq j < M = 2^m\} \ , \end{aligned}$$

then $R_{\mathrm{L}}(V, M) = R_{\mathrm{L}}(\mathcal{G}(V), \mathcal{E}(M))$. For $R_{\mathrm{U}}(\mathcal{G})$, we have

$$R_{\mathrm{U}}(\mathcal{G}) = \sum_{(d,j) \in \mathcal{G}} \frac{1}{2^d} \cdot I\left(\tilde{Q}_{2^d}^{(j)}\right) \ , \tag{47}$$

where $I\left(\tilde{Q}_{2^d}^{(j)}\right)$ is the mutual information corresponding to the joint distribution $\tilde{Q}_{2^d}^{(j)}$. That is, the capacity of the channel corresponding to $\tilde{Q}_{2^d}^{(j)}$. Note that with respect to the simplified description in Section V-B, $R_{\mathrm{U}}(V) = R_{\mathrm{U}}(\mathcal{G}(V))$.

Computationally, given $\mathcal{G}$ and $\mathcal{E}$, the calculation of $R_{\mathrm{L}}(\mathcal{G}, \mathcal{E})$ and $R_{\mathrm{U}}(\mathcal{G})$ is implemented as follows. We carry out a pre-order scan of the tree starting from the root. That is, we scan the root, scan the subtree rooted at its left child recursively, and then scan the subtree rooted at its right child recursively. The first node scanned is thus the root, for which $\tilde{Q}_1^{(0)}(\xi)$ is given. Assume we are currently scanning a node (d, j) which is not the root. Hence, this node has a parent, $(d', j') = (d-1, \lfloor j/2 \rfloor)$.

- If the path from the root to (d, j) has not yet traversed a vertex in $\mathcal{G}$, then by induction we have already calculated $\tilde{Q}_{2^{d'}}^{(j')}(\xi)$, and now calculate $\tilde{Q}_{2^d}^{(j)}(\xi)$ according to either (28) or (34), depending on the parity of j.
  - If $(d, j) \in \mathcal{G}$, then we also calculate $Z^{\star}(\tilde{Q}_{2^d}^{(j)}(\xi))$ and set $\zeta_{2^d}^{(j)} = Z^{\star}(\tilde{Q}_{2^d}^{(j)})$, in accordance with the starting condition (45).
- If the path from the root to (d, j) has already traversed a vertex in $\mathcal{G}$, then by induction we have already calculated $\zeta_{2^{d'}}^{(j')}$ and now calculate $\zeta_{2^d}^{(j)}$ according to (42).
  - If $(d, j) \in \mathcal{E}$, we do not recursively continue the scan, since we have reached a leaf.

We now describe how we chose $\mathcal{G}$ and $\mathcal{E}$ and calculated R_{L} and R_{U} in Figures 1 and 2. We set parameters $d_{\mathcal{G}} = 12$ and $d_{\mathcal{E}} = 36$ as the maximal depth of a vertex in $\mathcal{G}$ and $\mathcal{E}$, respectively. We further set a numeric threshold $\epsilon = 10^{-3}$ that allows us to add vertices to $\mathcal{G}$ and $\mathcal{E}$ at a depth shallower than $d_{\mathcal{G}}$ and $d_{\mathcal{E}}$, respectively, in case sufficient polarization has already occurred. Conceptually, we carry out a pre-order scan of a perfect binary tree[3] of height $d_{\mathcal{E}}$, trimming it as we go along. That is, $\mathcal{G}$ and $\mathcal{E}$ are generated dynamically as the scan progresses. We initialize variables $R_{\mathrm{L}} = R_{\mathrm{U}} = 0$. Each time a vertex is added to $\mathcal{G}$, R_{U} is incremented according to (47). Each time a vertex is added to $\mathcal{E}$, R_{L} is incremented according to (44).

During the scan of vertex (d, j) as described in the itemed list above:

- If the path from the root to (d, j) has not yet traveresed a vertex in $\mathcal{G}$, we add (d, j) to both $\mathcal{G}$ and $\mathcal{E}$ and increment R_{U} and R_{L} if

[2] A binary tree in which each node has either two children or no children.

[3] A full binary tree in which all the leaves are at the same depth.

  - $I(\tilde{Q}_{2^d}^{(j)}(\xi)) < \epsilon$, or
  - $1 - \delta'(Z^\star(\tilde{Q}_{2^d}^{(j)}(\xi))) > 1 - \epsilon$.

  Otherwise, we add (d, j) to $\mathcal{G}$ and increment R_{U} if $d = d_{\mathcal{G}}$.

- If the path from the root to (d, j) has already traversed a vertex in $\mathcal{G}$, we add (d, j) to $\mathcal{E}$ and increment R_{L} if
  - $1 - \delta'(\zeta_{2^d}^{(j)}(\xi)) > 1 - \epsilon$, or
  - $\zeta_{2^d}^{(j)}(\xi) > 1$, or
  - $d = d_{\mathcal{E}}$.

The curves for $R_{\mathrm{L}}, R_{\mathrm{U}}$ and C in Figure 1 are plotted with respect to a BI-AWGN channel quantized by a labeler λ to have eight possible outputs. At the input we assume a normalized BPSK mapping from $\mathcal{X} = \{0, 1\}$ to $\mathcal{X}' = \{1, -1\}$ such that $x' = 1 - 2x$. At the output we assume that the labeler maps $\mathcal{Y} = \mathbb{R}$ to $\{-4, -3, -2, -1, 1, 2, 3, 4\}$. The channel is defined by the above two mappings and by the relation $y = x' + \nu$, where ν is the realization of a Gaussian random variable with zero mean and variance σ^2. The labeler is

$$\lambda(y) = \begin{cases} 4 & q_3 \le y \;, \\ 3 & q_2 \le y < q_3 \;, \\ 2 & q_1 \le y < q_2 \;, \\ 1 & 0 \le y < q_1 \;, \\ -1 & -q_1 \le y < 0 \;, \\ -2 & -q_2 \le y < -q_1 \;, \\ -3 & -q_3 \le y < -q_2 \;, \\ -4 & y < -q_3 \;, \end{cases} \tag{48}$$

where we used $q_1 = 0.2$, $q_2 = 0.6$, and $q_3 = 1.2$ to define the labeler regions. For Figure 2 we have a BSC with $\mathcal{Y} = \{0, 1\}$ and the labeler is $\lambda(y) = 1 - 2y$. Note that both labelers above are good labelers.

We now state and prove two propositions that justify $R_{\mathrm{L}}(\mathcal{G}, \mathcal{E})$ and $R_{\mathrm{U}}(\mathcal{G})$ as valid thresholds. These are generalizations of claims and proofs made in Section V for simpler choices of R_{L} and R_{U}.

B. Justification of $R_{\mathrm{L}}(\mathcal{G}, \mathcal{E})$

Proposition 17. *Setting $R_{\mathrm{L}} = R_{\mathrm{L}}(\mathcal{G}, \mathcal{E})$ in Theorem 1 is valid.*

Proof: Recall that in Theorem 1 we assume that $R < R_{\mathrm{L}}(\mathcal{G}, \mathcal{E})$, and our aim is to prove the existence of a family of polar codes with growing lengths such that their rates are at least R and their word error probabilities at most 2^{-N^β}, where N is the codeword length.

As in Section V-A, we use the recursive relation (42) to define $\zeta_N^{(i)}$, where now the starting conditions are $\zeta_{2^d}^{(j)} = Z^\star(\tilde{Q}_{2^d}^{(j)})$ for $(d, j) \in \mathcal{G}$. Note that by the definition of $\mathcal{E}$ and $\mathcal{G}$ and our description of the steps carried out when a node is scanned, the value of $\zeta_{2^d}^{(j)}$ calculated during the scan of $(d, j) \in \mathcal{E}$ is the same $\zeta_{2^d}^{(j)}$ defined by the above recursion.

Again by (26), we have for all N large enough and $0 \le i < N$ that $\zeta_N^{(i)} \ge Z^\star(\tilde{Q}_N^{(i)})$. Namely, $\mathcal{A}' \subseteq \mathcal{A}$, where $\mathcal{A}$ and $\mathcal{A}'$ are defined in Section V-A. Thus, it suffices to show that for $R < R_{\mathrm{L}}(\mathcal{G}, \mathcal{E})$ fixed and all N large enough, $|\mathcal{A}'| \ge N \cdot R$.

Assume that R, and therefore $R_{\mathrm{L}}(\mathcal{G}, \mathcal{E})$ are positive, otherwise the claim is trivial. As before, denote $\beta' = \frac{\beta + 1/2}{2}$. For each one of the pairs $(d, j) \in \mathcal{E}$, we invoke Proposition 16 with $\delta = \delta'(\zeta_{2^d}^{(j)}) + (R_{\mathrm{L}}(\mathcal{G}, \mathcal{E})) - R)$, $\eta = S_0 = \zeta_{2^d}^{(j)}$, and $\beta'' = \frac{\beta' + 1/2}{2}$ in place of β. Denote the n_0 promised by the proposition as $n_0^{(d,j)}$. Now define $n_0^{\max} = \max_{(d,j) \in \mathcal{E}} n_0^{(d,j)}$ and $d_{\mathcal{E}}^{\max} = \max_{(d,j) \in \mathcal{E}} d$. Thus, for any $(d, j) \in \mathcal{E}$, $2^{-2^{(n-d)\beta''}} \le 2^{-2^{(n - d_{\mathcal{E}}^{\max})\beta''}}$. Hence, by (41), for $n \ge d_{\mathcal{E}}^{\max} + n_0^{\max}$ the fraction of indices $0 \le i < N$ such that $\zeta_N^{(i)} \le 2^{-2^{(n - d_{\mathcal{E}}^{\max})\beta''}}$ is at least

$$\begin{aligned} & \sum_{(d,j) \in \mathcal{E}} \frac{1}{2^d} \max\left\{1 - \left(\delta'(\zeta_{2^d}^{(j)}) + R_{\mathrm{L}}(\mathcal{G}, \mathcal{E}) - R\right), 0\right\} \\ & = \sum_{(d,j) \in \mathcal{E}} \frac{1}{2^d} \max\left\{1 - \delta'(\zeta_{2^d}^{(j)}) + R - R_{\mathrm{L}}(\mathcal{G}, \mathcal{E}), 0\right\} \\ & \ge \sum_{(d,j) \in \mathcal{E}} \frac{1}{2^d} \max\left\{1 - \delta'(\zeta_{2^d}^{(j)}) + R - R_{\mathrm{L}}(\mathcal{G}, \mathcal{E}), R - R_{\mathrm{L}}(\mathcal{G}, \mathcal{E})\right\} \\ & = \sum_{(d,j) \in \mathcal{E}} \frac{1}{2^d} \left(\max\left\{1 - \delta'(\zeta_{2^d}^{(j)}), 0\right\} + R - R_{\mathrm{L}}(\mathcal{G}, \mathcal{E})\right) \\ & \overset{\text{(a)}}{=} \sum_{(d,j) \in \mathcal{E}} \frac{1}{2^d} \left(\max\left\{1 - \delta'(\zeta_{2^d}^{(j)}), 0\right\}\right) + R - R_{\mathrm{L}}(\mathcal{G}, \mathcal{E}) \\ & = R \,, \end{aligned}$$

where (a) follows since the Kraft inequality [28, Equation 5.8] is tight on full binary trees, as can easily be proven by induction.

We take n_0 in Theorem 1 such that $n_0 \ge d_{\mathcal{E}}^{\max} + n_0^{\max}$. We further require that n_0 is large enough so that for all $n \ge n_0$ we have $2^{-2^{(n - d_{\mathcal{E}}^{\max})\beta''}} \le 2^{-2^{n\beta'}} = 2^{-N^{\beta'}}$. By the above, this ensures that $|\mathcal{A}'| \ge N \cdot R$. Lastly, we require that n_0 is large enough such that for all $n \ge n_0$ we have $N \cdot 2^{-N^{\beta'}} < 2^{-N^\beta}$. This ensures that the word error rate is at most 2^{-N^β}. ■

C. Definition of block-genie and justification of $R_{\mathrm{U}}(\mathcal{G})$

Our aim now is to prove an analogous claim to Proposition 17 for R_{U}. This is Proposition 18 below. In the proof of Proposition 18 we use a "block-genie", a concept we now define. Recall that in the seminal paper [1], a genie-aided decoder is used. That is, a variant of SC decoding, in which at stage i the genie reveals u_1^{i-1}. Thus, at stage i, the relevant distribution is $W_N^{(i)}$, given in (4). The genie-aided decoder is used since it is easier to analyze than SC decoding, but still has exactly the same word error rate as the SC decoder. Our block-genie will have this property as well.

The block-genie-aided SC decoder is defined in Algorithms A to C.

For a code of length N and a received word y_0^{N-1}, decoding is preformed by calling Algorithm C with $(\lambda(y_0), \lambda(y_1), \ldots, \lambda(y_{N-1}))$ and $d = j = 0$. Note that the set $\mathcal{G}$ is used in Algorithm C. Conceptually, we break the task of decoding $u_0, u_1, \ldots, u_{N-1}$ into the decoding of $|\mathcal{G}|$ blocks. We assume a code of length $N = 2^n$, where $n \ge d_{\mathcal{G}}^{\max} = \max_{(d,j) \in \mathcal{G}} d$. For $(d, j) \in \mathcal{G}$, the corresponding

Algorithm A: Make Decision

```
MakeDecision($\lambda, i$)
if $i \in \mathcal{A}$ then
    $\hat{u}_i = \begin{cases} 0 & \lambda \geq 0 \\ 1 & \lambda < 0 \end{cases}$
else
    $\hat{u}_i = 0$
return $\hat{u}_i$
```

Algorithm B: Genie Correct

```
GenieCorrect($i, T$)
return $(u_i, u_{i+1}, \ldots, u_{i+T-1}) \cdot B_T \cdot F^{\otimes t}$
/* $B_T$ is the bit reversal matrix,
    $t = \log_2 T$, $F = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, and $\otimes$ is the
    Kronecker product                                   */
```

block is $u_i, u_{i+1}, \ldots, u_{i+T-1}$, where $i = j \cdot T$ and $T = 2^{n-d}$. When decoding this block, the genie has already revealed u_1^{i-1} and thus the relevant distribution under the MSA is $\tilde{Q}_{2^d}^{(j)}$ (applying $\tilde{f}$ in place of f in Algorithm C). Specifically, after this block has been decoded, the genie corrects any errors the decoder may have introduced. This is done by invoking the GenieCorrect function defined in Algorithm B and used at the bottom of Algorithm C.

Note that for $\mathcal{G} = \emptyset$, Algorithms A to C simply describe SC decoding, without any help from a genie. Moreover, for $\mathcal{G} = \mathcal{G}(N)$ as defined in (46), Algorithms A to C describe Arıkan's genie-aided SC decoding. For the above two choices of $\mathcal{G}$, as well as for any other valid choice, the word error probability is the same, since correction are made only after decisions on $\hat{u}_i$ have been made in Algorithm A.

Algorithm C: Decode

```
Decode($\lambda_0, \lambda_1, \ldots, \lambda_{T-1}; d, j$)
if $T = 1$ then
    c $=$ MakeDecision($\lambda_0, j$)              // c $= (c_0)$
else
    $\Lambda_f = \left(f(\lambda_0, \lambda_1), \ldots, f(\lambda_{T-2}, \lambda_{T-1})\right)$
    // In the MSA, $f$ is replaced by $\tilde{f}$
    a $=$ Decode($\Lambda_f; d+1, 2j$)           // a $= a_0^{T/2-1}$
    $\Lambda_g = \left(g_{a_0}(\lambda_0, \lambda_1), \ldots, g_{a_{T/2-1}}(\lambda_{T-2}, \lambda_{T-1})\right)$
    b $=$ Decode($\Lambda_g; d+1, 2j+1$)         // b $= b_0^{T/2-1}$
    c $= \left(a_0 \oplus b_0, b_0, \ldots, a_{T/2-1} \oplus b_{T/2-1}, b_{T/2-1}\right)$
                                                // c $= c_0^{T-1}$
if $(d, j) \in \mathcal{G}$ then
    $i = j \cdot T$
    /* Genie corrects decisions on
        $\hat{u}_i, \hat{u}_{i+1}, \ldots, \hat{u}_{i+T-1}$, after all these
        are made                                              */
    c $=$ GenieCorrect($i, T$)                    // c $= c_0^{T-1}$
return c
```

Proposition 18. *Setting $R_{\mathrm{U}} = R_{\mathrm{U}}(\mathcal{G})$ in Theorem 1 is valid.*

Proof: Fix $\Delta > 0$ and consider a code with rate $R \geq R_{\mathrm{U}}(\mathcal{G}) + \Delta$. Denote the information set of this code as $\mathcal{A}$ and its length as $N = 2^n$, Thus $|\mathcal{A}| = N \cdot R$. Assume that $n \geq d_{\mathcal{G}}^{\max} = \max_{(d,j) \in \mathcal{G}} d$.

Consider the block corresponding to $(d, j) \in \mathcal{G}$. The number of indices in this block is $T = 2^{n-d}$. Of these, denote the indices in $\mathcal{A}$ by

$$\mathcal{A}_{(d,j)} = \{j \cdot T \leq i < j \cdot (T+1) : i \in \mathcal{A}\} .$$

Thus, the rate at which this block is coded for is

$$R_{(d,j)} \triangleq |\mathcal{A}_{(d,j)}| / 2^{n-d} .$$

Since every index $0 \leq i < N$ is contained in exactly one block,

$$R = \sum_{(d,j) \in \mathcal{G}} \frac{1}{2^d} R_{(d,j)} .$$

Thus, by the above and (47),

$$\Delta \leq R - R_{\mathrm{U}}(\mathcal{G}) = \sum_{(d,j) \in \mathcal{G}} \frac{1}{2^d} \left(R_{(d,j)} - I\left(\tilde{Q}_{2^d}^{(j)}\right) \right) .$$

By the pigeon-hole principle and the Kraft inequality being tight for a full binary tree, there exists at least one $(d, j) \in \mathcal{G}$ such that

$$R_{(d,j)} - I\left(\tilde{Q}_{2^d}^{(j)}\right) \geq \Delta .$$

By the strong converse to the coding theorem [27, Theorem 5.8.5], the probability of misdecoding such a block converges to 1, as the block size tends to infinity. Thus, the word error rate must converge to 1 as N tends to infinity, since all blocks have lengths that tend to infinity with N. ■

We end this section by stating the following proposition. As will become apparent in the proof, this is a special case of Corollary 23, given in the following subsection.

Proposition 19. *$R_{\mathrm{U}}(\mathcal{G}) \leq C$, where C is the capacity of W.*

D. Monotonic Properties of $R_{\mathrm{L}}(\mathcal{G}, \mathcal{E})$ and $R_{\mathrm{U}}(\mathcal{G})$

In this section, we show that the deeper we carry out our calculations, the tighter our thresholds become. Specifically, a corollary of what we are about to prove is that increasing $d_{\mathcal{G}}$ or $d_{\mathcal{E}}$ (or decreasing ϵ) yields better results.

Recall from Section VI-A the definitions and properties of a valid $\mathcal{G}$, a valid $\mathcal{E}$, and a valid pair $(\mathcal{G}, \mathcal{E})$. The following defines a set $\mathcal{G}'$ obtained by replacing a vertex in $\mathcal{G}$ by its two sons.

Definition 3. For a valid $\mathcal{G}$ and a vertex $(d', j') \in \mathcal{G}$, let

$$\mathcal{G}'(d', j') \triangleq \mathcal{G} \cup \{(d'+1, 2j'), (d'+1, 2j'+1)\} \setminus \{(d', j')\} .$$

Note that since we assume that $\mathcal{G}$ is valid, then so is $\mathcal{G}'$.
The following defines the set $\mathcal{E}'$ similarly to the above.

Definition 4. For a valid $\mathcal{E}$ and a vertex $(d', j') \in \mathcal{E}$, let

$$\mathcal{E}'(d', j') \triangleq \mathcal{E} \cup \{(d'+1, 2j'), (d'+1, 2j'+1)\} \setminus \{(d', j')\} \ .$$

As before, since $\mathcal{E}$ is valid, so is $\mathcal{E}'$.

As an example of the above, consider the sets $\mathcal{G}$ and $\mathcal{E}$ depicted in Figure 4.

- The depiction of $\mathcal{E}'(d' = 3, j' = 3)$ would be to add two white rectangular sons to $(3, 3)$, and to change $(3, 3)$ from a white rectangle to a white circle. Note that the pair $(\mathcal{G}, \mathcal{E}')$ is valid.
- The depiction of $\mathcal{E}'(d' = 2, j' = 2)$ would be to add two white rectangular sons to $(2, 2)$, and to change $(2, 2)$ from a red rectangle to a red circle. Note that the pair $(\mathcal{G}, \mathcal{E}')$ is valid.
- The depiction of $\mathcal{G}'(d' = 2, j' = 3)$ would be to change the color of $(3, 6)$ and $(3, 7)$ from white to red, and to change the color of $(2, 3)$ from red to white. Note that the pair $(\mathcal{G}', \mathcal{E})$ is valid.
- Lastly, note that for $\mathcal{G}'(d' = 2, j' = 2)$ the pair $(\mathcal{G}', \mathcal{E})$ is not valid. In general, this happens if both $(d', j') \in \mathcal{G}$ and $(d', j') \in \mathcal{E}$. To keep validity for these cases, we enlarge both sets. That is, $(\mathcal{G}'(d', j'), \mathcal{E}'(d', j'))$ is valid. Thus, the depiction of $\mathcal{G}'(d' = 2, j' = 2)$ and $\mathcal{E}'(d' = 2, j' = 2)$ is to add two red rectangular sons to $(2, 2)$ and to change $(2, 2)$ from a red rectangle to a white circle.

The following propositions show that our thresholds become tighter when replacing either $\mathcal{G}$ by $\mathcal{G}'$ or $\mathcal{E}$ by $\mathcal{E}'$.

Proposition 20. *For a valid $\mathcal{G}$, and a vertex $(d', j') \in \mathcal{G}$,*

$$R_{\mathrm{U}}(\mathcal{G}') \leq R_{\mathrm{U}}(\mathcal{G}) \ . \tag{49}$$

Proposition 21. *For a valid pair $(\mathcal{G}, \mathcal{E})$, and a vertex $(d', j') \in \mathcal{E}$,*

$$R_{\mathrm{L}}(\mathcal{G}, \mathcal{E}') \geq R_{\mathrm{L}}(\mathcal{G}, \mathcal{E}) \ . \tag{50}$$

Proposition 22. *For a valid pair $(\mathcal{G}, \mathcal{E})$, and a vertex (d', j') such that $(d', j') \in \mathcal{G}$ and $(d', j') \notin \mathcal{E}$,*

$$R_{\mathrm{L}}(\mathcal{G}', \mathcal{E}) \geq R_{\mathrm{L}}(\mathcal{G}, \mathcal{E}) \ . \tag{51}$$

We denote $\mathcal{G}^* \geq \mathcal{G}$ if $\mathcal{G}^*$ is obtained from $\mathcal{G}$ by a finite series of operations as in Definition 3. Similarly, we denote $\mathcal{E}^* \geq \mathcal{E}$ if $\mathcal{E}^*$ is obtained from $\mathcal{E}$ by a finite series of operations as in Definition 4.

As we will show, repeated application of the above three propositions yield the following.

Corollary 23 (Monotonicity of rate thresholds)**.** *Let the pair $(\mathcal{G}, \mathcal{E})$ be valid. Let the pair $(\mathcal{G}^*, \mathcal{E}^*)$ be valid as well, where $\mathcal{G}^* \geq \mathcal{G}$ and $\mathcal{E}^* \geq \mathcal{E}$. Then,*

$$R_{\mathrm{U}}(\mathcal{G}^*) \leq R_{\mathrm{U}}(\mathcal{G}) \tag{52}$$

and

$$R_{\mathrm{L}}(\mathcal{G}^*, \mathcal{E}^*) \geq R_{\mathrm{L}}(\mathcal{G}, \mathcal{E}) \ . \tag{53}$$

Recall the definition of $\mathcal{G}(V)$ and $\mathcal{E}(M)$, given in (46) for $V = 2^v$ and $M = 2^m$. Note that for $v' \geq v$ and $m' \geq m$, we have $\mathcal{G}(V') \geq \mathcal{G}(V)$ and $\mathcal{E}(M') \geq \mathcal{E}(M)$, where $V' = 2^{v'}$ and $M' = 2^{m'}$. Thus, the above corollary implies that setting the depths v and m larger in Section V-B indeed yields tighter thresholds. Namely, $R_{\mathrm{U}}(V') \leq R_{\mathrm{U}}(V)$ and $R_{\mathrm{L}}(V', M') \geq R_{\mathrm{L}}(V, M)$. Similarly, increasing $d_{\mathcal{G}}$ or $d_{\mathcal{E}}$, or decreasing ϵ in Section VI-A also yields tighter thresholds.

Appendix A
Non-recursive interpretation of the MSA

Recall the definition of $L_N^{(i)}$ in (7). This definition is explicit (non-recursive). However, when implementing a decoder the recursive definition given in (8)–(9) is used. For the MSA, the corresponding recursive definition of $\tilde{L}_N^{(i)}$ is given in (10)–(11). In this appendix we give an explicit definition of $\tilde{L}_N^{(i)}$, under the assumption $\lambda(y) = \mathrm{LLR}(y) = \log_2 (W(y; 0)/W(y; 1))$. This is (55) below, and is the analog of (7), which is rephrased below as (54). To save space, we use standard shorthand. For example, $\Pr(Y_0^{N-1} = y_0^{N-1}, U_0^{i-1} = u_0^{i-1}, U_i = 0, U_{i+1}^{N-1} = u_{i+1}^{N-1})$ is shortened to $p(y_0^{N-1}, u_0^{i-1}, u_i = 0, u_{i+1}^{N-1})$.

Proposition 24. *For the non-approximated setting,*

$$L_N^{(i)}\left(y_0^{N-1}, u_0^{i-1}\right) = \log_2\left(\frac{\sum_{u_{i+1}^{N-1}} p(y_0^{N-1}, u_0^{i-1}, u_i = 0, u_{i+1}^{N-1})}{\sum_{u_{i+1}^{N-1}} p(y_0^{N-1}, u_0^{i-1}, u_i = 1, u_{i+1}^{N-1})}\right) \ . \tag{54}$$

Under the MSA with $\lambda(y) = \log_2 (W(y; 0)/W(y; 1))$,

$$\tilde{L}_N^{(i)}\left(y_0^{N-1}, u_0^{i-1}\right) = \log_2\left(\frac{\max_{u_{i+1}^{N-1}} p(y_0^{N-1}, u_0^{i-1}, u_i = 0, u_{i+1}^{N-1})}{\max_{u_{i+1}^{N-1}} p(y_0^{N-1}, u_0^{i-1}, u_i = 1, u_{i+1}^{N-1})}\right) \ . \tag{55}$$

Notice that the only difference between (54) and (55) is that in the former we use a "$\sum$" while in the latter we use a "max".

Although our paper would be self contained without this appendix, we feel that the explicit definition (55) gives intuition about the MSA. Specifically, consider stage i of the decoding, in which we have already decided on $\hat{u}_0^{i-1}$, and must now decide the value of $\hat{u}_i$. Define

$$\mathcal{C}_0^{(i)} \triangleq \left\{u_0^{N-1} \in \{0, 1\}^N : u_i = 0, u_0^{i-1} = \hat{u}_0^{i-1}\right\} \ ,$$
$$\mathcal{C}_1^{(i)} \triangleq \left\{u_0^{N-1} \in \{0, 1\}^N : u_i = 1, u_0^{i-1} = \hat{u}_0^{i-1}\right\} \ .$$

Recall from [1, Equation 4] the definition of the combined channel

$$W_N(y_0^{N-1} | u_0^{N-1}) = \prod_{i=0}^{N-1} W(y_i | x_i) \ ,$$

where $x_0^{N-1} = u_0^{N-1} B_N F^{\otimes n}$ is the codeword corresponding to u_0^{N-1}, B_N is the bit-reversal matrix, and $F \triangleq \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ is the Arıkan kernel. Recall also that in both the non-approximated and approximated settings the decision rule is based on the sign of $L_N^{(i)}$ and $\tilde{L}_N^{(i)}$, respectively. Therefore, an immediate corollary of Proposition 24 is the following decision rules. The non-approximated SC decoder sets $\hat{u}_i$ according to

$$\sum_{u_0^{N-1}\in\mathcal{C}_0^{(i)}} W_N(y_0^{N-1}|u_0^{N-1}) \underset{\hat{u}_i=1}{\overset{\hat{u}_i=0}{\gtrless}} \sum_{u_0^{N-1}\in\mathcal{C}_1^{(i)}} W_N(y_0^{N-1}|u_0^{N-1}) \, , \tag{56}$$

whereas the min-sum SC decoder sets $\hat{u}_i$ according to

$$\max_{u_0^{N-1}\in\mathcal{C}_0^{(i)}} W_N(y_0^{N-1}|u_0^{N-1}) \underset{\hat{u}_i=1}{\overset{\hat{u}_i=0}{\gtrless}} \max_{u_0^{N-1}\in\mathcal{C}_1^{(i)}} W_N(y_0^{N-1}|u_0^{N-1}) \, . \tag{57}$$

Indeed, the above follows by the assumption that the input distribution is i.i.d. symmetric, which implies that a-priori, all u_0^{N-1} are equally likely.

Informally, in both settings we must choose one of two cosets: $\mathcal{C}_0^{(i)}$ or $\mathcal{C}_1^{(i)}$. In the non-approximated setting we base our decision on a weighting of all the words in each coset, whereas in the min-sum setting we base our decision on only the most probable word in each coset.

As we will see, the proof of (54) is straightforward. To prove (55), we take an indirect but simple route. Namely, we define $L_N^{\star(i)}$ as the RHS of (55). That is,

$$L_N^{\star(i)}\left(y_0^{N-1}, u_0^{i-1}\right) \triangleq \log_2\left(\frac{\max\limits_{u_{i+1}^{N-1}} p(y_0^{N-1}, u_0^{i-1}, u_i = 0, u_{i+1}^{N-1})}{\max\limits_{u_{i+1}^{N-1}} p(y_0^{N-1}, u_0^{i-1}, u_i = 1, u_{i+1}^{N-1})}\right) \, . \tag{58}$$

Our proof will follow by showing that $L_N^{\star(i)}$ satisfies the same recursive relations as $\tilde{L}_N^{(i)}$ in (10)–(11). Indeed, by inspection of (58),

$$L_1^{\star(0)}(y) = \mathrm{LLR}(y) = \log_2\left(W(y;0)/W(y;1)\right) \, .$$

Recalling (11) and our assumption that $\lambda(y) = \mathrm{LLR}(y)$, we have that the starting condition is the same for $L_N^{\star(i)}$ and $\tilde{L}_N^{(i)}$. Thus, to prove (55) all that remains is to show that (10) holds with "$\star$" in place of "$\sim$".

In aid of the above we introduce the following notation:

$$\mu_N^{\star(i)}(y_0^{N-1}, u_0^{i-1}; u_i) \triangleq \max_{u_{i+1}^{N-1}} p(y_0^{N-1}, u_0^{i-1}, u_i, u_{i+1}^{N-1}) \tag{59}$$

and

$$\ell_N^{\star(i)}(y_0^{N-1}, u_0^{i-1}) = \left(\frac{\mu_N^{\star(i)}(y_0^{N-1}, u_0^{i-1}; u_i = 0)}{\mu_N^{\star(i)}(y_0^{N-1}, u_0^{i-1}; u_i = 1)}\right) . \tag{60}$$

Therefore, $L_N^{\star(i)}(y_0^{N-1}, u_0^{i-1}) = \log_2 \ell_N^{\star(i)}(y_0^{N-1}, u_0^{i-1})$.

To derive the required recursive relations of $L_N^{\star(i)}$ we first derive the following recursive relations of $\mu_N^{\star(i)}$.

Lemma 25 ($\mu^\star$ minus and plus transforms)**.** *For $N \geq 2$ and $0 \leq j < N/2$,*

$$\begin{aligned}\mu_N^{\star(2j)}(y_0^{N-1}, u_0^{2j-1}; u_{2j}) = \\ \max_{u_{2j+1}} \Big\{ \mu_{N/2}^{\star(j)}\Big(y_0^{N/2-1}, u_{0,e}^{2j-1} \oplus u_{0,o}^{2j-1}; u_{2j} \oplus u_{2j+1}\Big) \\ \cdot \mu_{N/2}^{\star(j)}\Big(y_{N/2}^{N-1}, u_{0,o}^{2j-1}; u_{2j+1}\Big)\Big\}\end{aligned} \tag{61}$$

and

$$\begin{aligned}\mu_N^{\star(2j+1)}(y_0^{N-1}, u_0^{2j-1}; u_{2j}) = \\ \mu_{N/2}^{\star(j)}\Big(y_0^{N/2-1}, u_{0,e}^{2j-1} \oplus u_{0,o}^{2j-1}; u_{2j} \oplus u_{2j+1}\Big) \\ \cdot \mu_{N/2}^{\star(j)}\Big(y_{N/2}^{N-1}, u_{0,o}^{2j-1}; u_{2j+1}\Big) \, .\end{aligned} \tag{62}$$

Proof: For (61) we have

$$\begin{aligned}
&\mu_N^{\star(2j)}(y_0^{N-1}, u_0^{2j-1}; u_{2j}) \\
&\overset{\text{(a)}}{=} \max_{u_{2j+1}^{N-1}} p\Big(y_0^{N-1}, u_0^{2j-1}, u_{2j}, u_{2j+1}^{N-1}\Big) \\
&\overset{\text{(b)}}{=} \max_{u_{2j+1}} \max_{u_{2j+2}^{N-1}} p\Big(y_0^{N/2-1}, y_{N/2}^{N-1}, u_{0,e}^{2j-1} \oplus u_{0,o}^{2j-1}, u_{0,o}^{2j-1} \\
&\qquad , u_{2j} \oplus u_{2j+1}, u_{2j+1}, u_{2j+2,e}^{N-1} \oplus u_{2j+2,o}^{N-1}, u_{2j+2,o}^{N-1}\Big) \\
&\overset{\text{(c)}}{=} \max_{u_{2j+1}} \max_{u_{2j+2}^{N-1}} p\Big(y_0^{N/2-1}, u_{0,e}^{2j-1} \oplus u_{0,o}^{2j-1} \\
&\qquad , u_{2j} \oplus u_{2j+1}, u_{2j+2,e}^{N-1} \oplus u_{2j+2,o}^{N-1}\Big) \\
&\qquad \cdot p\Big(y_{N/2}^{N-1}, u_{0,o}^{2j-1}, u_{2j+1}, u_{2j+2,o}^{N-1} | y_0^{N/2-1}, u_{0,e}^{2j-1} \oplus u_{0,o}^{2j-1} \\
&\qquad , u_{2j} \oplus, u_{2j+1}, u_{2j+2,e}^{N-1} \oplus u_{2j+2,o}^{N-1}\Big) \\
&\overset{\text{(d)}}{=} \max_{u_{2j+1}} \max_{u_{2j+2}^{N-1}} p\Big(y_0^{N/2-1}, u_{0,e}^{2j-1} \oplus u_{0,o}^{2j-1} \\
&\qquad , u_{2j} \oplus u_{2j+1}, u_{2j+2,e}^{N-1} \oplus u_{2j+2,o}^{N-1}\Big) \\
&\qquad \cdot p\Big(y_{N/2}^{N-1}, u_{0,o}^{2j-1}, u_{2j+1}, u_{2j+2,o}^{N-1}\Big) \\
&\overset{\text{(e)}}{=} \max_{u_{2j+1}} \max_{u_{2j+2,o}^{N-1}} \Big\{ p\Big(y_{N/2}^{N-1}, u_{0,o}^{2j-1}, u_{2j+1}, u_{2j+2,o}^{N-1}\Big) \\
&\qquad \cdot \max_{u_{2j+2,e}^{N-1}} p\Big(y_0^{N/2-1}, u_{0,e}^{2j-1} \oplus u_{0,o}^{2j-1} \\
&\qquad , u_{2j} \oplus u_{2j+1}, u_{2j+2,e}^{N-1} \oplus u_{2j+2,o}^{N-1}\Big)\Big\} \\
&\overset{\text{(f)}}{=} \max_{u_{2j+1}} \max_{u_{2j+2,o}^{N-1}} \Big\{ p\Big(y_{N/2}^{N-1}, u_{0,o}^{2j-1}, u_{2j+1}, u_{2j+2,o}^{N-1}\Big) \\
&\qquad \mu_{N/2}^{\star(j)}\Big(y_0^{N/2-1}, u_{0,e}^{2j-1} \oplus u_{0,o}^{2j-1}; u_{2j} \oplus u_{2j+1}\Big)\Big\} \\
&\overset{\text{(g)}}{=} \max_{u_{2j+1}} \mu_{N/2}^{\star(j)}\Big(y_0^{N/2-1}, u_{0,e}^{2j-1} \oplus u_{0,o}^{2j-1}; u_{2j} \oplus u_{2j+1}\Big) \\
&\qquad \cdot \mu_{N/2}^{\star(j)}\Big(y_{N/2}^{N-1}, u_{0,o}^{2j-1}; u_{2j+1}\Big) \, ,
\end{aligned}$$

where (a) is by (59), (b) follows since there is a one-to-one and onto mapping between the arguments of p on the LHS and the arguments of p on the RHS, (c) is by the definition of conditional probability, (d) is since we need not condition on independent random variables, (e) is since $\max_{x,y}\{f(x)\cdot g(x,y)\} = \max_x\{f(x)\cdot \max_y g(x,y)\}$, (f) is by (59) since for fixed $u_{2j+2,o}^{N-1}$ the maximization over $u_{2j+2,e}^{N-1}$ ranges over all possible values in $\mathcal{X}^{N/2-1-j}$, and (g) is again by (59).

For (62), we follow the same steps with maximization over u_{2j+2}^{N-1} instead of u_{2j+1}^{N-1}. Therefore, all the above equalities remain the same, apart from not containing the outer $\max_{u_{2j+1}}$. ■

proof of Proposition 24: For (54) we notice that

$$\begin{aligned} W_N^{(i)}(y_0^{N-1}, u_0^{i-1}; u_i) &= p(y_0^{N-1}, u_0^{i-1}, u_i) \\ &= \sum_{u_{i+1}^{N-1}} p(y_0^{N-1}, u_0^{i-1}, u_i, u_{i+1}^{N-1}) \; . \end{aligned}$$

Thus, substituting the above into (7) with $u_i = 0$ for the numerator and $u_i = 1$ for the denominator yields (54).

For (55) recall from the above discussion that the proof will be completed once we show that (10) holds with "$\star$" in place of "$\sim$". We begin by proving the minus case (10a), and then prove the plus case (10b). We define the following shorthand:

$$\begin{aligned} y_L &\triangleq y_0^{N/2-1} \; , & y_R &\triangleq y_{N/2-1}^{N-1} \; , \\ u_\oplus &\triangleq u_{0,e}^{2j-1} \oplus u_{0,o}^{2j-1} \; , & u_o &\triangleq u_{0,o}^{2j-1} \; , \\ u &\triangleq u_{2j} \; , & v &\triangleq u_{2j+1} \; , \\ \ell_a^\star &\triangleq \ell_{N/2}^{\star(j)}(y_L, u_\oplus) \; , & \ell_b^\star &\triangleq \ell_{N/2}^{\star(j)}(y_R, u_o) \; , \\ L_a^\star &\triangleq \log_2 \ell_a^\star \; , & L_b^\star &\triangleq \log_2 \ell_b^\star \; . \end{aligned}$$

For (10a) we have

$$\begin{aligned} &\ell_N^{\star(2j)}(y_0^{N-1}, u_0^{2j-1}) \\ &\overset{(a)}{=} \left(\frac{\mu_N^{\star(2j)}(y_0^{N-1}, u_0^{2j-1}; u_{2j}=0)}{\mu_N^{\star(2j)}(y_0^{N-1}, u_0^{2j-1}; u_{2j}=1)} \right) \\ &\overset{(b)}{=} \left(\frac{\max_v \left\{ \mu_{N/2}^{\star(j)}(y_L, u_\oplus; 0 \oplus v) \cdot \mu_{N/2}^{\star(j)}(y_R, u_o; v) \right\}}{\max_v \left\{ \mu_{N/2}^{\star(j)}(y_L, u_\oplus; 1 \oplus v) \cdot \mu_{N/2}^{\star(j)}(y_R, u_o; v) \right\}} \right) \\ &\overset{(c)}{=} \left(\frac{\max\{\alpha, \beta\}}{\max\{\gamma, \delta\}} \right) , \end{aligned}$$

where (a) is by (60), (b) is by (61), and (c) due to the following notation:

$$\begin{aligned} \alpha &= \mu_{N/2}^{\star(j)}(y_L, u_\oplus; 0) \cdot \mu_{N/2}^{\star(j)}(y_R, u_o; 0) \; , \\ \beta &= \mu_{N/2}^{\star(j)}(y_L, u_\oplus; 1) \cdot \mu_{N/2}^{\star(j)}(y_R, u_o; 1) \; , \\ \gamma &= \mu_{N/2}^{\star(j)}(y_L, u_\oplus; 1) \cdot \mu_{N/2}^{\star(j)}(y_R, u_o; 0) \; , \\ \delta &= \mu_{N/2}^{\star(j)}(y_L, u_\oplus; 0) \cdot \mu_{N/2}^{\star(j)}(y_R, u_o; 1) \; . \end{aligned}$$

Thus, $\ell_N^{\star(2j)}(y_0^{N-1}, u_0^{2j-1})$ can equal one of four possible values: α/γ, β/γ, α/δ, β/δ. We consider each case separately.

- Consider the case α/γ. We have

$$\begin{aligned} &\ell_N^{\star(2j)}(y_0^{N-1}, u_0^{2j-1}) \\ &\quad = \alpha/\gamma \\ &\quad = \frac{\mu_{N/2}^{\star(j)}(y_L, u_\oplus; 0) \cdot \mu_{N/2}^{\star(j)}(y_R, u_o; 0)}{\mu_{N/2}^{\star(j)}(y_L, u_\oplus; 1) \cdot \mu_{N/2}^{\star(j)}(y_R, u_o; 0)} \\ &\quad = \frac{\mu_{N/2}^{\star(j)}(y_L, u_\oplus; 0)}{\mu_{N/2}^{\star(j)}(y_L, u_\oplus; 1)} \\ &\quad = \ell_a^\star \end{aligned}$$

$\alpha \geq \beta$ yields $1/\ell_b^\star \leq \ell_a^\star$. That is $-L_b^\star \leq L_a^\star$.
$\gamma \geq \delta$ yields $\ell_a^\star \leq \ell_b^\star$. That is $L_a^\star \leq L_b^\star$.
Combining both we have $-L_b^\star \leq L_a^\star \leq L_b^\star$. Thus, $\mathrm{sgn}(L_b^\star) \geq 0$ and $\min\{|L_a^\star|, |L_b^\star|\} = |L_a^\star|$. Therefore,

$$\begin{aligned} &L_N^{\star(2j)}(y_0^{N-1}, u_0^{2j-1}) \\ &\quad = L_a^\star \\ &\quad = \mathrm{sgn}(L_a^\star) \cdot |L_a^\star| \\ &\quad = \mathrm{sgn}(L_a^\star) \cdot \mathrm{sgn}(L_b^\star) \cdot \min\{|L_a^\star|, |L_b^\star|\} \\ &\quad = \tilde{f}(L_a^\star, L_b^\star) \end{aligned}$$

- Consider the case α/δ. We have

$$\begin{aligned} &\ell_N^{\star(2j)}(y_0^{N-1}, u_0^{2j-1}) \\ &\quad = \alpha/\delta \\ &\quad = \frac{\mu_{N/2}^{\star(j)}(y_L, u_\oplus; 0) \cdot \mu_{N/2}^{\star(j)}(y_R, u_o; 0)}{\mu_{N/2}^{\star(j)}(y_L, u_\oplus; 0) \cdot \mu_{N/2}^{\star(j)}(y_R, u_o; 1)} \\ &\quad = \frac{\mu_{N/2}^{\star(j)}(y_R, u_o; 0)}{\mu_{N/2}^{\star(j)}(y_R, u_o; 1)} \\ &\quad = \ell_b^\star \end{aligned}$$

$\alpha \geq \beta$ yields $1/\ell_a^\star \leq \ell_b^\star$. That is $-L_a^\star \leq L_b^\star$.
$\delta \geq \gamma$ yields $\ell_b^\star \leq \ell_a^\star$. That is $L_b^\star \leq L_a^\star$.
Combining both we have $-L_a^\star \leq L_b^\star \leq L_a^\star$. Thus, $\mathrm{sgn}(L_a^\star) \geq 0$ and $\min\{|L_a^\star|, |L_b^\star|\} = |L_b^\star|$. Therefore,

$$\begin{aligned} &L_N^{\star(2j)}(y_0^{N-1}, u_0^{2j-1}) \\ &\quad = L_b^\star \\ &\quad = \mathrm{sgn}(L_b^\star) \cdot |L_b^\star| \\ &\quad = \mathrm{sgn}(L_a^\star) \cdot \mathrm{sgn}(L_b^\star) \cdot \min\{|L_a^\star|, |L_b^\star|\} \\ &\quad = \tilde{f}(L_a^\star, L_b^\star) \end{aligned}$$

- Consider the case β/γ. We have

$$\begin{aligned} &\ell_N^{\star(2j)}(y_0^{N-1}, u_0^{2j-1}) \\ &\quad = \beta/\gamma \\ &\quad = \frac{\mu_{N/2}^{\star(j)}(y_L, u_\oplus; 1) \cdot \mu_{N/2}^{\star(j)}(y_R, u_o; 1)}{\mu_{N/2}^{\star(j)}(y_L, u_\oplus; 1) \cdot \mu_{N/2}^{\star(j)}(y_R, u_o; 0)} \\ &\quad = \frac{\mu_{N/2}^{\star(j)}(y_R, u_o; 1)}{\mu_{N/2}^{\star(j)}(y_R, u_o; 0)} \\ &\quad = 1/\ell_b^\star \end{aligned}$$

$\beta \geq \alpha$ yields $\ell_b^\star \leq 1/\ell_a^\star$. That is $L_b^\star \leq -L_a^\star$.
$\gamma \geq \delta$ yields $\ell_a^\star \leq \ell_b^\star$. That is $L_a^\star \leq L_b^\star$.
Combining both we have $L_a^\star \leq L_b^\star \leq -L_a^\star$. Thus, $\operatorname{sgn}(L_a^\star) \leq 0$ and $\min\{|L_a^\star|, |L_b^\star|\} = |L_b^\star|$. Therefore,

$$\begin{aligned} & L_N^{\star(2j)}(y_0^{N-1}, u_0^{2j-1}) \\ &\quad = -L_b^\star \\ &\quad = -\operatorname{sgn}(L_b^\star) \cdot |L_b^\star| \\ &\quad = \operatorname{sgn}(L_a^\star) \cdot \operatorname{sgn}(L_b^\star) \cdot \min\{|L_a^\star|, |L_b^\star|\} \\ &\quad = \tilde{f}(L_a^\star, L_b^\star) \end{aligned}$$

- Consider the case β/δ. We have

$$\begin{aligned} & \ell_N^{\star(2j)}(y_0^{N-1}, u_0^{2j-1}) \\ &\quad = \beta/\delta \\ &\quad = \frac{\mu_{N/2}^{\star(j)}(y_L, u_\oplus; 1) \cdot \mu_{N/2}^{\star(j)}(y_R, u_o; 1)}{\mu_{N/2}^{\star(j)}(y_L, u_\oplus; 0) \cdot \mu_{N/2}^{\star(j)}(y_R, u_o; 1)} \\ &\quad = \frac{\mu_{N/2}^{\star(j)}(y_L, u_\oplus; 1)}{\mu_{N/2}^{\star(j)}(y_L, u_\oplus; 0)} \\ &\quad = 1/\ell_a^\star \end{aligned}$$

$\beta \geq \alpha$ yields $\ell_a^\star \leq 1/\ell_b^\star$. That is $L_a^\star \leq -L_b^\star$.
$\delta \geq \gamma$ yields $\ell_b^\star \leq \ell_a^\star$. That is $L_b^\star \leq L_a^\star$.
Combining both we have $L_b^\star \leq L_a^\star \leq -L_b^\star$. Thus, $\operatorname{sgn}(L_b^\star) \leq 0$ and $\min\{|L_a^\star|, |L_b^\star|\} = |L_a^\star|$. Therefore,

$$\begin{aligned} & L_N^{\star(2j)}(y_0^{N-1}, u_0^{2j-1}) \\ &\quad = -L_a^\star \\ &\quad = -\operatorname{sgn}(L_a^\star) \cdot |L_a^\star| \\ &\quad = \operatorname{sgn}(L_a^\star) \cdot \operatorname{sgn}(L_b^\star) \cdot \min\{|L_a^\star|, |L_b^\star|\} \\ &\quad = \tilde{f}(L_a^\star, L_b^\star) \end{aligned}$$

To summarize, in all four cases (10a) holds with "$\star$" in place of "$\sim$".

For (10b) we have

$$\begin{aligned} & \ell_N^{\star(2j+1)}(y_0^{N-1}, u_0^{2j}) \\ &\overset{(a)}{=} \left(\frac{\mu_N^{\star(2j+1)}(y_0^{N-1}, u_0^{2j}; u_{2j+1} = 0)}{\mu_N^{\star(2j+1)}(y_0^{N-1}, u_0^{2j}; u_{2j+1} = 1)} \right) \\ &\overset{(b)}{=} \frac{\mu_{N/2}^{\star(j)}(y_R, u_o; 0)}{\mu_{N/2}^{\star(j)}(y_R, u_o; 1)} \cdot \frac{\mu_{N/2}^{\star(j)}(y_L, u_\oplus; u \oplus 0)}{\mu_{N/2}^{\star(j)}(y_L, u_\oplus; u \oplus 1)} \\ &\overset{(c)}{=} \begin{cases} \ell_{N/2}^{\star(j)}(y_R, u_o) \cdot \ell_{N/2}^{\star(j)}(y_L, u_\oplus) & \text{if } u = 0\,, \\ \ell_{N/2}^{\star(j)}(y_{N/2}^{N-1}, u_{0,o}^{2j-1}) / \ell_{N/2}^{\star(j)}(y_L, u_\oplus) & \text{if } u = 1 \end{cases} \\ &= \begin{cases} \ell_b^\star \cdot \ell_a^\star & \text{if } u = 0\,, \\ \ell_b^\star / \ell_a^\star & \text{if } u = 1\,, \end{cases} \end{aligned}$$

where (a) is by (60), (b) is by (62), and (c) is again by (60). In log-domain the above is simply

$$\begin{aligned} L_N^{\star(2j+1)}(y_0^{N-1}, u_0^{2j}) &= \begin{cases} L_b^\star + L_a^\star & \text{if } u = 0 \\ L_b^\star - L_a^\star & \text{if } u = 1 \end{cases} \\ &= g_u(L_a^\star, L_b^\star) \end{aligned}$$

Therefore, (10b) also holds with "$\star$" in place of "$\sim$". ■

Appendix B
Proofs

A. Proofs for Section III

proof of Lemma 2: We prove (14a) and (14b). We begin with (14a). Define $\theta \triangleq (y_0^{N-1}, u_0^{2j-1})$, then by the definition of the synthetic min-sum joint distribution in (13) we have

$$\tilde{Q}_N^{(2j)}(t; u_{2j}) \;=\; \sum_{\theta : \tilde{L}_N^{(2j)}(\theta) = t} W_N^{(2j)}(\theta; u_{2j})\,. \tag{63}$$

Further define the following: $\alpha \triangleq (y_0^{N/2-1}, u_{0,e}^{2j-1} \oplus u_{0,o}^{2j-1})$, $\beta \triangleq (y_{N/2}^{N-1}, u_{0,o}^{2j-1})$, $\tau_a \triangleq \tilde{L}_{N/2}^{(j)}(\alpha)$, and $\tau_b \triangleq \tilde{L}_{N/2}^{(j)}(\beta)$. By (5) we have

$$W_N^{(2j)}(\theta; u_{2j}) = \sum_{u_{2j+1}} W_{N/2}^{(j)}(\alpha; u_{2j} \oplus u_{2j+1}) \cdot W_{N/2}^{(j)}(\beta; u_{2j+1}),$$

and by (10a) we have

$$\tilde{L}_N^{(2j)}(\theta) = \tilde{f}\left(\tilde{L}_{N/2}^{(j)}(\alpha), \tilde{L}_{N/2}^{(j)}(\beta)\right) = \tilde{f}(\tau_a, \tau_b)\,.$$

Therefore, by applying a change of variables from θ to (α, β), which by inspection iterate over the same set of possible values, we can rewrite the sum in (63) as follows:

$$\sum_{u_{2j+1}} \sum_{\substack{\alpha, \beta: \\ \tilde{f}(\tau_a, \tau_b) = t}} W_{N/2}^{(j)}(\alpha; u_{2j} \oplus u_{2j+1}) \cdot W_{N/2}^{(j)}(\beta; u_{2j+1})\,.$$

The sum over $\{\alpha, \beta : \tilde{f}(\tau_a, \tau_b) = t\}$ can be modified into two sums: an outer sum over $\{t_a, t_b : \tilde{f}(t_a, t_b) = t\}$ and an inner sum over $\{\alpha, \beta : \tilde{L}_{N/2}^{(j)}(\alpha) = t_a, \tilde{L}_{N/2}^{(j)}(\beta) = t_b\}$. By doing that, the innermost sum becomes

$$\begin{aligned} &\sum_{\substack{\alpha, \beta: \\ \tilde{L}_{N/2}^{(j)}(\alpha) = t_a \\ \tilde{L}_{N/2}^{(j)}(\beta) = t_b}} W_{N/2}^{(j)}(\alpha; u_{2j} \oplus u_{2j+1}) \cdot W_{N/2}^{(j)}(\beta; u_{2j+1}) \\ &\qquad = \tilde{Q}_{N/2}^{(j)}(t_a; u_{2j} \oplus u_{2j+1}) \cdot \tilde{Q}_{N/2}^{(j)}(t_b; u_{2j+1})\,. \end{aligned} \tag{64}$$

where the equality follows by the definition in (13). Now, the result in (64) is summed over $\{u_{2j+1}, t_a, t_b : \tilde{f}(t_a, t_b) = t\}$, which is (14a).

Similarly, (14b) can be obtained by following the same steps, using (6) instead of (5) and (10b) instead of (10a). That is, define $\theta' \triangleq (\theta, u_{2j})$. Then by (13) we have

$$\tilde{Q}_N^{(2j+1)}(t; u_{2j+1}) = \sum_{\theta' : \tilde{L}_N^{(2j+1)}(\theta') = t} W_N^{(2j+1)}(\theta'; u_{2j+1})\,. \tag{65}$$

By (6) we have

$$W_N^{(2j+1)}(\theta'; u_{2j+1}) = W_{N/2}^{(j)}(\alpha; u_{2j} \oplus u_{2j+1}) \cdot W_{N/2}^{(j)}(\beta; u_{2j+1}),$$

and by (10b) we have

$$\tilde{L}_N^{(2j)}(\theta') = g_{u_{2j}}\left(\tilde{L}_{N/2}^{(j)}(\alpha), \tilde{L}_{N/2}^{(j)}(\beta)\right) = g_{u_{2j}}(\tau_a, \tau_b)\,.$$

Therefore, by applying a change of variables from θ' to (α, β, u_{2j}), we can rewrite the sum in (65) as follows:

$$\sum_{u_{2j}} \sum_{\substack{\alpha,\beta: \\ g_{u_{2j}}(\tau_a,\tau_b)=t}} W_{N/2}^{(j)}(\alpha; u_{2j} \oplus u_{2j+1}) \cdot W_{N/2}^{(j)}(\beta; u_{2j+1})$$

The rest of the proof is proceeds as before, by modifying the inner sum and using the definition in (13). ■

proof of Lemma 3: We prove (15) by induction. For the base case $(N = 1, i = 0)$, we have $W_1^{(0)}(y; x) = W(y; x)$ and $\tilde{L}_1^{(0)}(y) = \lambda(y)$, see (11). By the first item of Definition 1 we have $\lambda(\pi(y)) = -\lambda(y)$. Therefore, by (13) we have

$$\begin{aligned} \tilde{Q}_1^{(0)}(-t; x_i \oplus 1) &= \sum_{y:\lambda(y)=-t} W(y; x_i \oplus 1) \\ &= \sum_{y:\lambda(\pi(y))=t} W(\pi(y); x_i) = \tilde{Q}_1^{(0)}(t; x_i) \ . \end{aligned}$$

We now assume that (15) holds for $(\frac{N}{2}, j)$ and show that it also holds for $(N, 2j)$ and $(N, 2j + 1)$. We define the shorthands: $u_{2j} \triangleq u$, and $u_{2j+1} \triangleq v$. Then, for $(N, 2j)$,

$$\begin{aligned} \tilde{Q}_N^{(2j)}(-t, u \oplus 1) &\overset{\text{(a)}}{=} \sum_{\substack{t_a,t_b,v: \\ \tilde{f}(t_a,t_b)=-t}} \tilde{Q}_{N/2}^{(j)}(t_a; u \oplus 1 \oplus v) \cdot \tilde{Q}_{N/2}^{(j)}(t_b; v) \\ &\overset{\text{(b)}}{=} \sum_{\substack{t_a,t_b,v: \\ \tilde{f}(t_a,t_b)=-t}} \tilde{Q}_{N/2}^{(j)}(-t_a; u \oplus v) \cdot \tilde{Q}_{N/2}^{(j)}(t_b; v) \\ &\overset{\text{(c)}}{=} \sum_{\substack{t_A,t_b,v: \\ \tilde{f}(-t_A,t_b)=-t}} \tilde{Q}_{N/2}^{(j)}(t_A; u \oplus v) \cdot \tilde{Q}_{N/2}^{(j)}(t_b; v) \\ &\overset{\text{(d)}}{=} \sum_{\substack{t_A,t_b,v: \\ \tilde{f}(t_A,t_b)=t}} \tilde{Q}_{N/2}^{(j)}(t_A; u \oplus v) \cdot \tilde{Q}_{N/2}^{(j)}(t_b; v) \\ &\overset{\text{(e)}}{=} \tilde{Q}_N^{(2j)}(t, u) \ , \end{aligned}$$

where (a) and (e) are by (14a), (b) is by the induction hypothesis, (c) is by changing variables $t_A = -t_a$, and (d) is by inspection of (3) which reveals that $\tilde{f}(-t_A, t_b) = -t$ is the same condition as $\tilde{f}(t_A, t_b) = t$.

Similarly for $(N, 2j + 1)$,

$$\begin{aligned} &\tilde{Q}_N^{(2j+1)}(-t, v \oplus 1) \\ &\overset{\text{(a)}}{=} \sum_{\substack{t_a,t_b,u: \\ g_u(t_a,t_b)=-t}} \tilde{Q}_{N/2}^{(j)}(t_a; u \oplus v \oplus 1) \cdot \tilde{Q}_{N/2}^{(j)}(t_b; v \oplus 1) \\ &\overset{\text{(b)}}{=} \sum_{\substack{t_a,t_b,u: \\ g_u(t_a,t_b)=-t}} \tilde{Q}_{N/2}^{(j)}(-t_a; u \oplus v) \cdot \tilde{Q}_{N/2}^{(j)}(-t_b; v) \\ &\overset{\text{(c)}}{=} \sum_{\substack{t_A,t_B,u: \\ g_u(-t_A,-t_B)=-t}} \tilde{Q}_{N/2}^{(j)}(t_A; u \oplus v) \cdot \tilde{Q}_{N/2}^{(j)}(t_B; v) \\ &\overset{\text{(d)}}{=} \sum_{\substack{t_A,t_B,u: \\ g_u(t_A,t_B)=t}} \tilde{Q}_{N/2}^{(j)}(t_A; u \oplus v) \cdot \tilde{Q}_{N/2}^{(j)}(t_B; v) \\ &\overset{\text{(e)}}{=} \tilde{Q}_N^{(2j)}(t, v) \ , \end{aligned}$$

where now (a) and (e) are by (14b), (b) is by the induction hypothesis, (c) is by changing variables $t_A = -t_a$ and $t_B = -t_b$, and (d) is by inspection of (2) which reveals that $g_v(-t_A, -t_b) = -t$ is the same condition as $g_v(t_A, t_B) = t$.

We now prove the claim in Lemma 3 with the tildes removed. For the base case, the symmetry of W implies that $L_1^{(0)}(\pi(y)) = -L_1^{(0)}(y)$, see (9). Thus, by (12), we have

$$\begin{aligned} Q_1^{(0)}(-t; x_i \oplus 1) &= \sum_{y:L_1^{(0)}(y)=-t} W(y; x_i \oplus 1) \\ &= \sum_{y:L_1^{(0)}(\pi(y))=t} W(\pi(y); x_i) = Q_1^{(0)}(t; x_i) \ . \end{aligned}$$

The induction is unchanged by removing the tildes. All that must be verified is that $f(-t_A, t_b) = -t$ is the same condition as $f(t_A, t_b) = t$. This follows by (1), and recalling that $\tanh$ is an odd function. ■

Proof of Lemma 4: We prove (20). Denote by $[A]$ an indicator of the event A.

$$\begin{aligned} P_{\text{e}}\left(\tilde{Q}_N^{(i)}\right) &= \Pr(\hat{u}_i \neq u_i) = \sum_{t,u_i} \tilde{Q}_N^{(i)}(t; u_i) \cdot [\hat{u}_i \neq u_i] \\ &= \sum_{t} \tilde{Q}_N^{(i)}(t; 0) \cdot [\hat{u}_i \neq 0] + \sum_{t} \tilde{Q}_N^{(i)}(t; 1) \cdot [\hat{u}_i \neq 1] \\ &\overset{\text{(a)}}{=} \sum_{t<0} \tilde{Q}_N^{(i)}(t; 0) + \sum_{t\geq 0} \tilde{Q}_N^{(i)}(t; 1) \\ &\overset{\text{(b)}}{=} \sum_{t<0} \tilde{Q}_N^{(i)}(t; 0) + \sum_{t\geq 0} \tilde{Q}_N^{(i)}(-t; 0) \\ &= \sum_{t<0} \tilde{Q}_N^{(i)}(t; 0) + \sum_{t\leq 0} \tilde{Q}_N^{(i)}(t; 0) \\ &= \tilde{Q}_N^{(i)}(0; 0) + 2 \cdot \sum_{t<0} \tilde{Q}_N^{(i)}(t; 0) \ , \end{aligned}$$

where (a) is by the decision rule of the decoder as described in the `MakeDecision` function given in Algorithm A, and (b) is by the symmetry property in (15). ■

proof of Corollary 5: The coefficient of ξ^t in $\tilde{Q}_N^{(i)}(1/\xi)$ is the coefficient of ξ^{-t} in $\tilde{Q}_N^{(i)}(\xi)$, which equals $\tilde{Q}_N^{(i)}(-t; 0)$. By (15) this is $\tilde{Q}_N^{(i)}(t; 1)$. ■

proof of Lemma 6: We recall (20)–(22) and prove (23). We have

$$\begin{aligned} Z\left(\tilde{Q}_N^{(i)}, \xi_0\right) &\overset{\text{(a)}}{=} 2 \cdot \sum_{t} \tilde{Q}_N^{(i)}(t; 0) \cdot \xi_0^t \\ &\overset{\text{(b)}}{\geq} \tilde{Q}_N^{(i)}(0; 0) + 2 \cdot \sum_{t<0} \tilde{Q}_N^{(i)}(t; 0) \cdot \xi_0^t \\ &\overset{\text{(c)}}{\geq} \tilde{Q}_N^{(i)}(0; 0) + 2 \cdot \sum_{t<0} \tilde{Q}_N^{(i)}(t; 0) \\ &\overset{\text{(d)}}{=} P_{\text{e}}\left(\tilde{Q}_N^{(i)}\right) \ , \end{aligned}$$

where (a) is by (21) and (22), (b) is since we have thrown away non-negative terms, (c) is since $\xi_0^t \geq 1$, for $t < 0$ and $0 < \xi_0 \leq 1$, and (d) is by (20). ■

We now prove Lemmas 8 and 9, which directly leads to the proof of Lemma 7.

proof of Lemma 8: We prove (27). Using the shorthand $u_{2j+1} \triangleq v$ we have

$$\begin{aligned}\tilde{Q}_N^{(2j)}(\xi) &\overset{\text{(a)}}{=} \sum_t \tilde{Q}_N^{(2j)}(t;0)\xi^t \\ &\overset{\text{(b)}}{=} \sum_t \left(\sum_{\substack{t_a,t_b,v: \\ \tilde{f}(t_a,t_b)=t}} \tilde{Q}_{N/2}^{(j)}(t_a;v) \cdot \tilde{Q}_{N/2}^{(j)}(t_b;v) \right) \cdot \xi^t\end{aligned}$$

where (a) is by (21) and (b) is by (14a). Evaluating the inner sum for the case $v=1$ yields

$$\begin{aligned}&\sum_{\substack{t_a,t_b: \\ \tilde{f}(t_a,t_b)=t}} \tilde{Q}_{N/2}^{(j)}(t_a;1) \cdot \tilde{Q}_{N/2}^{(j)}(t_b;1) \\ &\overset{\text{(a)}}{=} \sum_{\substack{t_a,t_b: \\ \tilde{f}(t_a,t_b)=t}} \tilde{Q}_{N/2}^{(j)}(-t_a;0) \cdot \tilde{Q}_{N/2}^{(j)}(-t_b;0) \\ &\overset{\text{(b)}}{=} \sum_{\substack{t_A,t_B: \\ \tilde{f}(-t_A,-t_B)=t}} \tilde{Q}_{N/2}^{(j)}(t_A;0) \cdot \tilde{Q}_{N/2}^{(j)}(t_B;0) \\ &\overset{\text{(c)}}{=} \sum_{\substack{t_A,t_B: \\ \tilde{f}(t_A,t_B)=t}} \tilde{Q}_{N/2}^{(j)}(t_A;0) \cdot \tilde{Q}_{N/2}^{(j)}(t_B;0)\ ,\end{aligned}$$

where (a) is by (15), (b) is by changing variables $t_A = -t_a$ and $t_B = -t_b$, and (c) is by inspection of (3) which reveals that $\tilde{f}(-t_A,-t_B) = \tilde{f}(t_A,t_B)$. Therefore the inner sum is the same for $v=1$ and for $v=0$. Therefore we have

$$\begin{aligned}\tilde{Q}_N^{(2j)}(\xi) &= 2 \cdot \sum_t \left(\sum_{\substack{t_a,t_b: \\ \tilde{f}(t_a,t_b)=t}} \tilde{Q}_{N/2}^{(j)}(t_a;0) \cdot \tilde{Q}_{N/2}^{(j)}(t_b;0) \right) \cdot \xi^t \\ &= 2 \cdot \sum_{t_a,t_b} \tilde{Q}_{N/2}^{(j)}(t_a;0) \cdot \tilde{Q}_{N/2}^{(j)}(t_b;0) \cdot \xi^{\tilde{f}(t_a,t_b)} \qquad (66)\end{aligned}$$

To upper bound the above expression for $0 < \xi_0 \leq 1$ in place of ξ, we divide all pairs $(t_a,t_b) \in \tilde{\mathcal{T}}_{N/2}^{(j)} \times \tilde{\mathcal{T}}_{N/2}^{(j)}$ into eight disjoint sets denoted $\{S_k\}_{k=1}^8$. For each set we evaluate $\xi_0^{\tilde{f}(t_a,t_b)}$ and upper bound this expression by either $\xi_0^{t_a}$ or $\xi_0^{t_b}$ as described in Table I. All the upper bounds are justified since for $0 < \xi_0 \leq 1$ the function ξ_0^t is non-increasing in t.

We now define two disjoint sets regarding the two possible values for upper bounds described in the table. That is,

$$\begin{aligned}\boxed{\mathcal{A}} &\triangleq \boxed{S_1 \sqcup S_3 \sqcup S_7 \sqcup S_8} \\ &\subseteq \left\{ (t_a,t_b) \in \tilde{\mathcal{T}}_{N/2}^{(j)} \times \tilde{\mathcal{T}}_{N/2}^{(j)} : \xi_0^{\tilde{f}(t_a,t_b)} \leq \xi_0^{t_a} \right\}\ , \\ \boxed{\mathcal{B}} &\triangleq \boxed{S_2 \sqcup S_4 \sqcup S_5 \sqcup S_6} \\ &\subseteq \left\{ (t_a,t_b) \in \tilde{\mathcal{T}}_{N/2}^{(j)} \times \tilde{\mathcal{T}}_{N/2}^{(j)} : \xi_0^{\tilde{f}(t_a,t_b)} \leq \xi_0^{t_b} \right\}\ ,\end{aligned}$$

where "$\sqcup$" denotes disjoint union, and the "$\subseteq$" relations follow from the last column of Table I. Note that by definition

$$\tilde{\mathcal{T}}_{N/2}^{(j)} \times \tilde{\mathcal{T}}_{N/2}^{(j)} = \mathcal{A} \sqcup \mathcal{B}\ . \qquad (67)$$

	sgn(t_a)	**sgn(t_b)**	**min{$\lvert t_a\rvert$, $\lvert t_b\rvert$}**	$\boldsymbol{\xi_0^{\tilde{f}(t_a,t_b)}}$
S_1	$+/0$	$+/0$	$\lvert t_a\rvert$	$\xi_0^{\lvert t_a\rvert} = \xi_0^{t_a}$
S_2	$+/0$	$-$	$\lvert t_a\rvert$	$\xi_0^{-\lvert t_a\rvert} = \xi_0^{-t_a} \leq \xi_0^{t_b}$
S_3	$-$	$+/0$	$\lvert t_a\rvert$	$\xi_0^{-\lvert t_a\rvert} = \xi_0^{t_a}$
S_4	$-$	$-$	$\lvert t_a\rvert$	$\xi_0^{\lvert t_a\rvert} = \xi_0^{-t_a} \leq \xi_0^{t_b}$
S_5	$+/0$	$+/0$	$\lvert t_b\rvert$	$\xi_0^{\lvert t_b\rvert} = \xi_0^{t_b}$
S_6	$+/0$	$-$	$\lvert t_b\rvert$	$\xi_0^{-\lvert t_b\rvert} = \xi_0^{t_b}$
S_7	$-$	$+/0$	$\lvert t_b\rvert$	$\xi_0^{-\lvert t_b\rvert} = \xi_0^{-t_b} \leq \xi_0^{t_a}$
S_8	$-$	$-$	$\lvert t_b\rvert$	$\xi_0^{\lvert t_b\rvert} = \xi_0^{-t_b} \leq \xi_0^{t_a}$

TABLE I
THE SETS S_1 TO S_8.

Denote the shorthands $q \triangleq \tilde{Q}_{N/2}^{(j)}(t_a;0) \cdot \tilde{Q}_{N/2}^{(j)}(t_b;0)$ and $\tilde{\mathcal{T}} \triangleq \tilde{\mathcal{T}}_{N/2}^{(j)} \times \tilde{\mathcal{T}}_{N/2}^{(j)}$. By (66) we have

$$\begin{aligned}\tilde{Q}_N^{(2j)}(\xi_0) &= 2 \cdot \sum_{t_a,t_b \in \tilde{\mathcal{T}}} q \cdot \xi_0^{\tilde{f}(t_a,t_b)} \\ &\overset{\text{(a)}}{=} 2 \cdot \sum_{t_a,t_b \in \mathcal{A}} q \cdot \xi_0^{\tilde{f}(t_a,t_b)} + 2 \cdot \sum_{t_a,t_b \in \mathcal{B}} q \cdot \xi_0^{\tilde{f}(t_a,t_b)} \\ &\overset{\text{(b)}}{\leq} 2 \cdot \sum_{t_a,t_b \in \mathcal{A}} q \cdot \xi_0^{t_a} + 2 \cdot \sum_{t_a,t_b \in \mathcal{B}} q \cdot \xi_0^{t_b} \\ &\overset{\text{(c)}}{\leq} 2 \cdot \sum_{t_a,t_b \in \tilde{\mathcal{T}}} q \cdot \xi_0^{t_a} + 2 \cdot \sum_{t_a,t_b \in \tilde{\mathcal{T}}} q \cdot \xi_0^{t_b} \\ &= 2 \cdot \sum_{t_a} \tilde{Q}_{N/2}^{(j)}(t_a;0) \cdot \xi_0^{t_a} \cdot \sum_{t_b} \tilde{Q}_{N/2}^{(j)}(t_b;0) \\ &\quad +2 \cdot \sum_{t_a} \tilde{Q}_{N/2}^{(j)}(t_a;0) \cdot \sum_{t_b} \tilde{Q}_{N/2}^{(j)}(t_b;0) \cdot \xi_0^{t_b} \\ &\overset{\text{(d)}}{=} 2 \cdot \frac{1}{2} \cdot \sum_{t_a} \tilde{Q}_{N/2}^{(j)}(t_a;0) \cdot \xi_0^{t_a} \\ &\qquad +2 \cdot \frac{1}{2} \cdot \sum_{t_b} \tilde{Q}_{N/2}^{(j)}(t_b;0) \cdot \xi_0^{t_b} \\ &\overset{\text{(e)}}{=} 2 \cdot \tilde{Q}_{N/2}^{(j)}(\xi_0)\ ,\end{aligned}$$

where (a) is by (67), (b) is by the last column in Table I, (c) is since we are adding non-negative terms, (d) is since $\tilde{Q}_{N/2}^{(j)}(t;v)$ is a joint distribution and summing over all t yields $\Pr(v=0) = 1/2$, and (e) is by (21). ■

proof of Lemma 9: We prove (28). Using the shorthand $u_{2j} \triangleq u$ we have

$$\begin{aligned}\tilde{Q}_N^{(2j+1)}(\xi) &\overset{\text{(a)}}{=} \sum_t \tilde{Q}_N^{(2j+1)}(t;0)\xi^t \\ &\overset{\text{(b)}}{=} \sum_t \left(\sum_{\substack{t_a,t_b,u: \\ g_u(t_a,t_b)=t}} \tilde{Q}_{N/2}^{(j)}(t_a;u) \cdot \tilde{Q}_{N/2}^{(j)}(t_b;0) \right) \cdot \xi^t \\ &= \sum_{t_a,t_b,u} \tilde{Q}_{N/2}^{(j)}(t_a;u) \cdot \tilde{Q}_{N/2}^{(j)}(t_b;0) \cdot \xi^{g_u(t_a,t_b)}\end{aligned}$$

$$
\begin{aligned}
&= \sum_{t_a,t_b} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\tilde{Q}^{(j)}_{N/2}(t_b;0)\cdot\xi^{g_0(t_a,t_b)} \\
&\qquad + \sum_{t_a,t_b} \tilde{Q}^{(j)}_{N/2}(t_a;1)\cdot\tilde{Q}^{(j)}_{N/2}(t_b;0)\cdot\xi^{g_1(t_a,t_b)} \\
&\overset{\text{(c)}}{=} \sum_{t_a,t_b} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\tilde{Q}^{(j)}_{N/2}(t_b;0)\cdot\xi^{t_a+t_b} \\
&\qquad + \sum_{t_a,t_b} \tilde{Q}^{(j)}_{N/2}(t_a;1)\cdot\tilde{Q}^{(j)}_{N/2}(t_b;0)\cdot\xi^{t_b-t_a} \\
&= \sum_{t_a} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\xi^{t_a}\cdot\sum_{t_b} \tilde{Q}^{(j)}_{N/2}(t_b;0)\cdot\xi^{t_b} \\
&\qquad + \sum_{t_a} \tilde{Q}^{(j)}_{N/2}(t_a;1)\cdot\xi^{-t_a}\cdot\sum_{t_b} \tilde{Q}^{(j)}_{N/2}(t_b;0)\cdot\xi^{t_b} \\
&\overset{\text{(d)}}{=} 2\cdot\left(\sum_{t_a} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\xi^{t_a}\right)^2 = 2\cdot\left(\tilde{Q}^{(j)}_{N/2}(\xi)\right)^2 ,
\end{aligned}
$$

where (a) is by (21), (b) is by (14b), (c) is by (2), and (d) is by using change of variables $t_A = -t_a$ and the symmetry property in (15). That is, (d) follows since

$$
\begin{aligned}
\sum_{t_a} \tilde{Q}^{(j)}_{N/2}(t_a;1)\cdot\xi^{-t_a} &= \sum_{t_A} \tilde{Q}^{(j)}_{N/2}(-t_A;1)\cdot\xi^{t_A} \\
&= \sum_{t_A} \tilde{Q}^{(j)}_{N/2}(t_A;0)\cdot\xi^{t_A} .
\end{aligned}
$$

■

proof of Lemma 7: We first consider the "-" case, and prove (25a) and (26a). To prove (25a), we have by (27) that for all $0 < \xi_0 \leq 1$ that $\tilde{Q}^{(2j)}_N(\xi_0) \leq 2\cdot\tilde{Q}^{(j)}_{N/2}(\xi_0)$. Plugging the definition of Z from (22), which is $\tilde{Q}^{(2j)}_N(\xi_0) = \frac{1}{2}\cdot Z\left(\tilde{Q}^{(2j)}_N,\xi_0\right)$, yields (25a). To obtain (26a) we optimize both sides of (25a) separately, such that $Z^\star\left(\tilde{Q}^{(j)}_{N/2}\right) = Z\left(\tilde{Q}^{(j)}_{N/2},\xi_{opt1}\right)$, and $Z^\star\left(\tilde{Q}^{(2j)}_N\right) = Z\left(\tilde{Q}^{(2j)}_N,\xi_{opt2}\right)$. Therefore,

$$
\begin{aligned}
Z^\star\left(\tilde{Q}^{(2j)}_N\right) &= Z\left(\tilde{Q}^{(2j)}_N,\xi_{opt2}\right) \leq Z\left(\tilde{Q}^{(2j)}_N,\xi_{opt1}\right) \\
&\leq 2\cdot Z\left(\tilde{Q}^{(j)}_{N/2},\xi_{opt1}\right) = 2\cdot Z^\star\left(\tilde{Q}^{(j)}_{N/2}\right) ,
\end{aligned}
$$

where the first inequality is by the optimization and the second inequality is by (25a).

We now consider the "+" case, and prove (25b) and (26b). To prove (25b) we have by (28) (for all ξ and specifically for $0 < \xi_0 \leq 1$) that $\tilde{Q}^{(2j+1)}_N(\xi_0) = 2\cdot\tilde{Q}^{(j)}_{N/2}(\xi_0)$. Again, plugging $\tilde{Q}^{(2j+1)}_N(\xi_0) = \frac{1}{2}\cdot Z\left(\tilde{Q}^{(2j+1)}_N,\xi_0\right)$, yields (25b). To obtain (26b) we use the same optimization argument as before. This concludes the proof. We remark that the equality in (25b) implies that now $\xi_{opt1} = \xi_{opt2}$. This has a practical computational advantage: it implies that roughly half of the optimizations in Section VI need not be carried out. ■

proof of Lemma 10: We prove (34) by considering each coefficient of $\tilde{Q}^{(2j)}_N(\xi)$ separately.

$$
\begin{aligned}
[\xi^t]\ \tilde{Q}^{(2j)}_N(\xi) &\overset{\text{(a)}}{=} \tilde{Q}^{(2j)}_N(t;0) \\
&\overset{\text{(b)}}{=} \sum_{\substack{t_a,t_b,u:\\ \tilde{f}(t_a,t_b)=t}} \tilde{Q}^{(j)}_{N/2}(t_a;u)\cdot\tilde{Q}^{(j)}_{N/2}(t_b;u) \\
&= \sum_{\substack{t_a,t_b:\\ \tilde{f}(t_a,t_b)=t}} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\tilde{Q}^{(j)}_{N/2}(t_b;0) \\
&\qquad + \sum_{\substack{t_a,t_b:\\ \tilde{f}(t_a,t_b)=t}} \tilde{Q}^{(j)}_{N/2}(t_a;1)\cdot\tilde{Q}^{(j)}_{N/2}(t_b;1) \\
&\overset{\text{(c)}}{=} 2\cdot\sum_{\substack{t_a,t_b:\\ \tilde{f}(t_a,t_b)=t}} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\tilde{Q}^{(j)}_{N/2}(t_b;0) . \qquad (68)
\end{aligned}
$$

where (a) is by (21), (b) is by (14a), and (c) is since by (15) we have

$$\tilde{Q}^{(j)}_{N/2}(t_a;1)\cdot\tilde{Q}^{(j)}_{N/2}(t_b;1) = \tilde{Q}^{(j)}_{N/2}(-t_a;0)\cdot\tilde{Q}^{(j)}_{N/2}(-t_b;0) ,$$

and by (3) we have $\tilde{f}(-t_a,-t_b) = f(t_a,t_b)$.

We define the set $\mathcal{C}_t$ as the set of pairs (t_a,t_b) which contribute to the sum in (68),

$$\mathcal{C}_t = \left\{(t_a,t_b)\in\tilde{\mathcal{T}}^{(j)}_{N/2}\times\tilde{\mathcal{T}}^{(j)}_{N/2} : \tilde{f}(t_a,t_b) = t\right\} . \qquad (69)$$

We first consider the case $t > 0$. By inspection of $\tilde{f}$ in (3), the set $\mathcal{C}_t$ can be partitioned into six disjoint sets denoted $S_{\text{I}}, S_{\text{II}}, \ldots S_{\text{VI}}$ and defined as follows:

$$
\begin{aligned}
S_{\text{I}} &\triangleq \{(t_a,t_b)\in\mathcal{C}_t : t_a = t, t_b > t\} , && \text{(70a)}\\
S_{\text{II}} &\triangleq \{(t_a,t_b)\in\mathcal{C}_t : t_a > t, t_b = t\} , && \text{(70b)}\\
S_{\text{III}} &\triangleq \{(t_a,t_b)\in\mathcal{C}_t : t_a = -t, t_b < -t\} , && \text{(70c)}\\
S_{\text{IV}} &\triangleq \{(t_a,t_b)\in\mathcal{C}_t : t_a < -t, t_b = -t\} , && \text{(70d)}\\
S_{\text{V}} &\triangleq \{(t_a,t_b)\in\mathcal{C}_t : t_a = t, t_b = t\} , && \text{(70e)}\\
S_{\text{VI}} &\triangleq \{(t_a,t_b)\in\mathcal{C}_t : t_a = -t, t_b = -t\} . && \text{(70f)}
\end{aligned}
$$

Thus, the sum in (68) can be broken into six sums. By the symmetry between t_a and t_b in (68), both (70a) and (70b) have the same contribution, as well as both (70c) and (70d). We now consider the contribution of each sum.

For S_{I} (and also for S_{II}, as explained above) we have

$$
\begin{aligned}
&\sum_{(t_a,t_b)\in S_{\text{I}}} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\tilde{Q}^{(j)}_{N/2}(t_b;0) \\
&\overset{\text{(a)}}{=} \sum_{t_a=t} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\sum_{t_b>t} \tilde{Q}^{(j)}_{N/2}(t_b;0) \\
&\overset{\text{(b)}}{=} [\xi^t]\ \tilde{Q}^{(j)}_{N/2}(\xi)\cdot[\xi^t]\ \tilde{A}^{(j)}_{N/2}(\xi) \\
&\overset{\text{(c)}}{=} [\xi^t]\ \tilde{Q}^{(j)}_{N/2}(\xi)\cdot[\xi^t]\ \operatorname{pos}\left\langle\tilde{A}^{(j)}_{N/2}(\xi)\right\rangle ,
\end{aligned}
$$

where (a) is by (70a), (b) is by (21) and (29), and (c) is by (31) since $t > 0$.

For S_{III} (and also for S_{IV}, as explained above) we have

$$\begin{aligned}
&\sum_{(t_a,t_b)\in S_{\mathrm{III}}} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\tilde{Q}^{(j)}_{N/2}(t_b;0)\\
&\overset{\text{(a)}}{=} \sum_{t_a=-t} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\sum_{t_b<-t} \tilde{Q}^{(j)}_{N/2}(t_b;0)\\
&\overset{\text{(b)}}{=} [\xi^{-t}]\,\tilde{Q}^{(j)}_{N/2}(\xi)\cdot[\xi^{-t}]\,\tilde{B}^{(j)}_{N/2}(\xi)\\
&\overset{\text{(c)}}{=} [\xi^{t}]\,\tilde{Q}^{(j)}_{N/2}(1/\xi)\cdot[\xi^{t}]\,\mathrm{neg}\left\langle\tilde{B}^{(j)}_{N/2}(\xi)\right\rangle\,,
\end{aligned}$$

where (a) is by (70c), (b) is by (21) and (30), and (c) is by (32) since $t>0$.

For S_{V} we have

$$\begin{aligned}
&\sum_{(t_a,t_b)\in S_{\mathrm{V}}} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\tilde{Q}^{(j)}_{N/2}(t_b;0)\\
&\overset{\text{(a)}}{=} \sum_{t_a=t} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\sum_{t_b=t} \tilde{Q}^{(j)}_{N/2}(t_b;0)\\
&\overset{\text{(b)}}{=} [\xi^{t}]\,\tilde{Q}^{(j)}_{N/2}(\xi)\cdot[\xi^{t}]\,\tilde{Q}^{(j)}_{N/2}(\xi)\,,
\end{aligned}\tag{71}$$

where (a) is by (70e) and (b) is by (21).

For S_{VI} we have

$$\begin{aligned}
&\sum_{(t_a,t_b)\in S_{\mathrm{VI}}} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\tilde{Q}^{(j)}_{N/2}(t_b;0)\\
&\overset{\text{(a)}}{=} \sum_{t_a=-t} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\sum_{t_b=-t} \tilde{Q}^{(j)}_{N/2}(t_b;0)\\
&\overset{\text{(b)}}{=} [\xi^{-t}]\,\tilde{Q}^{(j)}_{N/2}(\xi)\cdot[\xi^{-t}]\,\tilde{Q}^{(j)}_{N/2}(\xi)\\
&= [\xi^{t}]\,\tilde{Q}^{(j)}_{N/2}(1/\xi)\cdot[\xi^{t}]\,\tilde{Q}^{(j)}_{N/2}(1/\xi)\,,
\end{aligned}\tag{72}$$

where (a) is by (70f) and (b) is by (21).

Plugging all six sums into (68) yields (for $t>0$)

$$\begin{aligned}
&[\xi^t]\,\tilde{Q}^{(2j)}_{N}(\xi) =\\
&\quad[\xi^t]\,2\left(\tilde{Q}^{(j)}_{N/2}(\xi)\odot\left(2\cdot\mathrm{pos}\left\langle\tilde{A}^{(j)}_{N/2}(\xi)\right\rangle+\tilde{Q}^{(j)}_{N/2}(\xi)\right)\right)\\
&+[\xi^t]\,2\left(\tilde{Q}^{(j)}_{N/2}(1/\xi)\odot\left(2\cdot\mathrm{neg}\left\langle\tilde{B}^{(j)}_{N/2}(\xi)\right\rangle+\tilde{Q}^{(j)}_{N/2}(1/\xi)\right)\right)
\end{aligned}\tag{73}$$

where we used the definition of "$\odot$" in (33).

We now consider the case $t<0$. Similarly to what we did before for the case $t>0$, we partition the set $\mathcal{C}_t$ defined in (69) into six disjoined sets denoted $\tilde{S}_{\mathrm{I}},\tilde{S}_{\mathrm{II}},\ldots\tilde{S}_{\mathrm{VI}}$ as follows:

$$\tilde{S}_{\mathrm{I}} \triangleq \{(t_a,t_b)\in\mathcal{C}_t : t_a = t, t_b > -t\}\,,\tag{74a}$$
$$\tilde{S}_{\mathrm{II}} \triangleq \{(t_a,t_b)\in\mathcal{C}_t : t_a > -t, t_b = t\}\,,\tag{74b}$$
$$\tilde{S}_{\mathrm{III}} \triangleq \{(t_a,t_b)\in\mathcal{C}_t : t_a = -t, t_b < t\}\,,\tag{74c}$$
$$\tilde{S}_{\mathrm{IV}} \triangleq \{(t_a,t_b)\in\mathcal{C}_t : t_a < t, t_b = -t\}\,,\tag{74d}$$
$$\tilde{S}_{\mathrm{V}} \triangleq \{(t_a,t_b)\in\mathcal{C}_t : t_a = t, t_b = t\}\,,\tag{74e}$$
$$\tilde{S}_{\mathrm{VI}} \triangleq \{(t_a,t_b)\in\mathcal{C}_t : t_a = -t, t_b = -t\}\,.\tag{74f}$$

We now consider the contribution of each sum.

For $\tilde{S}_{\mathrm{I}}$ (and also for $\tilde{S}_{\mathrm{II}}$) we have

$$\begin{aligned}
&\sum_{(t_a,t_b)\in \tilde{S}_{\mathrm{I}}} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\tilde{Q}^{(j)}_{N/2}(t_b;0)\\
&\overset{\text{(a)}}{=} \sum_{t_a=t} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\sum_{t_b>-t} \tilde{Q}^{(j)}_{N/2}(t_b;0)\\
&\overset{\text{(b)}}{=} [\xi^{t}]\,\tilde{Q}^{(j)}_{N/2}(\xi)\cdot[\xi^{-t}]\,\tilde{A}^{(j)}_{N/2}(\xi)\\
&\overset{\text{(c)}}{=} [\xi^{t}]\,\tilde{Q}^{(j)}_{N/2}(\xi)\cdot[\xi^{t}]\,\mathrm{pos}\left\langle\tilde{A}^{(j)}_{N/2}(\xi)\right\rangle\,,
\end{aligned}$$

where (a) is by (74a), (b) is by (21) and (29), and (c) is by (31) since $t<0$.

For $\tilde{S}_{\mathrm{III}}$ (and also for $\tilde{S}_{\mathrm{IV}}$) we have

$$\begin{aligned}
&\sum_{(t_a,t_b)\in \tilde{S}_{\mathrm{III}}} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\tilde{Q}^{(j)}_{N/2}(t_b;0)\\
&\overset{\text{(a)}}{=} \sum_{t_a=-t} \tilde{Q}^{(j)}_{N/2}(t_a;0)\cdot\sum_{t_b<t} \tilde{Q}^{(j)}_{N/2}(t_b;0)\\
&\overset{\text{(b)}}{=} [\xi^{-t}]\,\tilde{Q}^{(j)}_{N/2}(\xi)\cdot[\xi^{t}]\,\tilde{B}^{(j)}_{N/2}(\xi)\\
&\overset{\text{(c)}}{=} [\xi^{t}]\,\tilde{Q}^{(j)}_{N/2}(1/\xi)\cdot[\xi^{t}]\,\mathrm{neg}\left\langle\tilde{B}^{(j)}_{N/2}(\xi)\right\rangle\,,
\end{aligned}$$

where (a) is by (74c), (b) is by (21) and (30), and (c) is by (32) since $t<0$.

For $\tilde{S}_{\mathrm{V}}$ and for $\tilde{S}_{\mathrm{VI}}$ we have exactly the same expressions as for S_{V} and S_{VI} given in (71) and (72), respectively. Indeed, this is because in deriving these we did not use the assumption that $t>0$, and by comparing (70e) and (70f) to (74e) and (74f), respectively. Plugging all six sums into (68), reveals that (73) also holds for $t<0$.

For the case $t=0$ we follow the same steps as those for the case $t>0$, but note that now $S_{\mathrm{V}}=S_{\mathrm{VI}}$. That is, the contribution of $S_{\mathrm{V}}\cup S_{\mathrm{VI}}$ is now

$$[\xi^0]\,\tilde{Q}^{(j)}_{N/2}(\xi)\cdot[\xi^0]\,\tilde{Q}^{(j)}_{N/2}(\xi) = [\xi^0]\,\tilde{Q}^{(j)}_{N/2}(1/\xi)\cdot[\xi^0]\,\tilde{Q}^{(j)}_{N/2}(1/\xi)\,,$$

as opposed to the contribution for $t>0$:

$$[\xi^t]\,\tilde{Q}^{(j)}_{N/2}(\xi)\cdot[\xi^t]\,\tilde{Q}^{(j)}_{N/2}(\xi)+[\xi^t]\,\tilde{Q}^{(j)}_{N/2}(1/\xi)\cdot[\xi^t]\,\tilde{Q}^{(j)}_{N/2}(1/\xi)\,.$$

Thus, (73) will hold also for $t=0$, once we subtract $2\left([\xi^0]\,\tilde{Q}^{(j)}_{N/2}\right)^2$, which yields (34). ■

B. Proofs for Section IV

This subsection is devoted to proving the results in Section IV regarding the complexity of calculations for the finite-length case. That is, we show how the posynomials $\tilde{Q}^{(i)}_N(\xi)$ are efficiently calculated. Recall that $\tilde{Q}^{(i)}_N(\xi)$ is defined in (21), where the summation index t ranges over $\tilde{\mathcal{T}}^{(i)}_N$ given in (19).

Recall that $\tilde{\mathcal{T}}^{(i)}_N$ contains all integers from $-\gamma\cdot 2^{\mathrm{wt}(i)}$ to $\gamma\cdot 2^{\mathrm{wt}(i)}$. Thus, we represent $\tilde{Q}^{(i)}_N(\xi)$ by an array indexed over this integer range, where entry t contains the coefficient of ξ^t. Namely, if we denote this array as $q[\cdot]$, then $q[t]=[\xi^t]\,\tilde{Q}^{(i)}_N(\xi)$. The same representation is used for all posynomials arising from intermediate steps in the calculation.

proof of Lemma 11: We first note by inspection that all intermediate posynomials taking part in the calculation have the same range of indices. That is, all these posynomials have indices ranging over $\tilde{\mathcal{T}}^{(j)}_{N/2} = \tilde{\mathcal{T}}^{(2j)}_{N}$, where the equality is also in accordance with (19) since $\mathrm{wt}(2j) = \mathrm{wt}(j)$. The result will follow by showing that all the intermediate calculations in (34) can be carried out in linear time. That is, in time $\mathcal{O}\left(|\tilde{\mathcal{T}}^{(j)}_{N/2}|\right)$.

Denote the array of $\tilde{Q}^{(j)}_{N/2}(\xi)$ as $q[\cdot]$. First consider the calculation of $\tilde{A}^{(j)}_{N/2}(\xi)$, defined in (29). This is done by allocating an array $a[\cdot]$, indexed over $\tilde{\mathcal{T}}^{(j)}_{N/2}$, and populating its entries from highest to lowest. That is, we set $a[\gamma \cdot 2^{\mathrm{wt}(j)}] = 0$ and for all smaller t we set $a[t] = a[t+1] + q[t+1]$. Clearly, this calculation is linear in the size of $a[\cdot]$. Similarly, $\tilde{B}^{(j)}_{N/2}(\xi)$, defined in (30), is calculated by allocating an array $b[\cdot]$, and populating its entries from lowest to highest such that $b[-\gamma \cdot 2^{\mathrm{wt}(j)}] = 0$ and for all larger t, $b[t] = b[t-1] + q[t-1]$. This operation is linear as well. By inspection, all the other operations involved in (34) are also linear. Note that $\tilde{Q}^{(j)}_{N/2}(1/\xi)$ is simply the posynomial $\tilde{Q}^{(j)}_{N/2}(\xi)$, reversed. That is, $[\xi^t]\ \tilde{Q}^{(j)}_{N/2}(1/\xi) = [\xi^{-t}]\ \tilde{Q}^{(j)}_{N/2}(\xi)$. ■

proof of Lemma 12: By (19), the largest and smallest powers of $\tilde{Q}^{(j)}_{N/2}(\xi)$ are $\gamma' \triangleq 2^{\mathrm{wt}(j)}\gamma$ and $-\gamma'$, respectively. We recast (28) as follows:

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

Notice that $\xi^{\gamma'} \cdot \tilde{Q}^{(j)}_{N/2}(\xi)$ is a polynomial. Therefore, the complexity of calculating $\tilde{Q}^{(2j+1)}_{N}(\xi)$ is that of squaring a polynomial of degree $2\gamma'$. By [29, Chapter 30], this can be done using fast Fourier transform in time $\mathcal{O}(2\gamma' \cdot \log(2\gamma')) = \mathcal{O}\left(|\tilde{\mathcal{T}}^{(j)}_{N/2}| \cdot \log(|\tilde{\mathcal{T}}^{(j)}_{N/2}|)\right)$, where the equality follows by (19). ■

proof of Theorem 13: We first note that both $|\tilde{T}^{(2j)}_{N}|$ and $|\tilde{T}^{(2j+1)}_{N}|$ are at least $|\tilde{T}^{(j)}_{N/2}|$, by (19). Hence, by Lemmas 11 and 12 we may bound the computational complexity of calculating $\tilde{Q}^{(i)}(\xi)$ from $\tilde{Q}^{(\lfloor i/2 \rfloor)}(\xi)$ by $\mathcal{O}(|\tilde{T}^{(i)}_{N}| \log |\tilde{T}^{(i)}_{N}|)$. That is, using (19), by $\mathcal{O}(2^{\mathrm{wt}(i)} \cdot \gamma \cdot \log(2^{\mathrm{wt}(i)} \cdot \gamma))$.

For $1 \leq m \leq n$, consider the complexity of the last step of calculating all $\tilde{Q}^{(j)}_{M}(\xi)$, where $0 \leq j < M = 2^m$. That is, calculating the $\tilde{Q}^{(j)}_{M}(\xi)$, when we have already calculated all $\tilde{Q}^{(k)}_{M/2}(\xi)$, $0 \leq k < M/2 = 2^{m-1}$. Since the number of indices j of weight w is $\binom{m}{w}$, the complexity is of order

$$\sum_{w=0}^{m} \binom{m}{w} 2^w \gamma \log_2(2^w \cdot \gamma) .$$

Thus, the overall complexity is of order

$$\sum_{m=1}^{n} \sum_{w=0}^{m} \binom{m}{w} 2^w \gamma \log_2(2^w \cdot \gamma) . \tag{75}$$

We start by bounding the inner sum in (75). We have

$$\sum_{w=0}^{m} \binom{m}{w} 2^w \gamma \log_2(2^w \cdot \gamma)$$
$$= \gamma \sum_{w=0}^{m} \binom{m}{w} 2^w w + \gamma \log_2 \gamma \sum_{w=0}^{m} \binom{m}{w} 2^w . \tag{76}$$

The first sum on the RHS of (76) is bounded by

$$\begin{aligned}
\sum_{w=0}^{m} \binom{m}{w} 2^w w &= \sum_{w=1}^{m} \binom{m}{w} 2^w w \\
&= \sum_{w=1}^{m} \frac{m!}{w!(m-w)!} 2^w w \\
&= \sum_{w=1}^{m} \frac{m \cdot (m-1)!}{(w-1)!(m-w)!} 2^w \\
&= \sum_{w=1}^{m} m \binom{m-1}{w-1} 2^w \\
&= 2m \sum_{w=1}^{m} \binom{m-1}{w-1} 2^{w-1} 1^{(m-1)-(w-1)} \\
&= 2m \sum_{k=0}^{m-1} \binom{m-1}{k} 2^k 1^{(m-1)-k} \\
&\overset{\text{(a)}}{=} 2m(1+2)^{m-1} \\
&= 2m \cdot 3^{m-1} ,
\end{aligned}$$

where (a) follows by the binomial theorem.

The second sum on the RHS of (76) is bounded by

$$\sum_{w=0}^{m} \binom{m}{w} 2^w = \sum_{w=0}^{m} \binom{m}{w} 2^w 1^{m-w} = 3^m .$$

Plugging the above simplifications into (75) yields an overall complexity of order

$$\begin{aligned}
&\sum_{m=1}^{n} \gamma \cdot 2m \cdot 3^{m-1} + \gamma \log_2 \gamma \cdot 3^m \\
&= \gamma \sum_{m=1}^{n} \left(\frac{2}{3} \cdot m + \log_2 \gamma\right) \cdot 3^m \\
&\overset{\text{(a)}}{=} \gamma \left(\frac{2}{3} \cdot \frac{3}{4}(1 + (2n-1) \cdot 3^n) + \log_2 \gamma \cdot \frac{3^{n+1} - 3}{3 - 1}\right) \\
&\leq \gamma \left(\frac{1}{2}(2n \cdot 3^n) + \log_2 \gamma \cdot 3^{n+1}\right) \\
&= \gamma \left(n \cdot 3^n + \log_2 \gamma \cdot 3^{n+1}\right) \\
&= \gamma \left(n \cdot 2^{n \log_2 3} + \log_2 \gamma \cdot 3 \cdot 2^{n \log_2 3}\right) \\
&= \gamma \left(\log_2 N \cdot N^{\log_2 3} + \log_2 \gamma \cdot 3 \cdot N^{\log_2 3}\right) \\
&= \mathcal{O}(\gamma \cdot N^{\log_2 3} \log N + \gamma \log \gamma \cdot N^{\log_2 3}) ,
\end{aligned}$$

where (a) follows by the methods in [30, Section 2.6]. The above can be further bounded as $\mathcal{O}(\gamma \log \gamma \cdot N^{\log_2 3} \log N)$. Since $1.585 > \log_2 3$, we can also bound this as $\mathcal{O}(\gamma \log \gamma \cdot N^{1.585})$. ■

C. *Proofs for Section V*

proof of Corollary 15: The corollary is obtained by a reduction to Proposition 14 with appropriate parameters. Since the same symbols n', η, ϵ', and δ' are used in both Corollary 15 and Proposition 14, we apply hats to all symbols in Corollary 15 to avoid ambiguity.

Thus, our setting is that we are given $\hat{\epsilon}'$, $\hat{\eta}$, and must find $\hat{n}'(\hat{\epsilon}',\hat{\eta})$ such that if $S_0 \leq \hat{\eta}$, then

$$\Pr\left(S_n \leq \hat{\epsilon}' \text{ for all } n \geq \hat{n}'\right) \geq 1 - \hat{\delta}'(\eta) \; ,$$

where

$$\hat{\delta}'(\hat{\eta}) = 2 \cdot (8\hat{\eta})^{\log_2 \varphi} \; . \tag{77}$$

We now consider Proposition 14 with $\epsilon' = \hat{\epsilon}'$, $\delta' = \hat{\delta}'(\hat{\eta})$, and $\kappa = 2$. Consider first the corresponding η. By (38) and (77), this is

$$\begin{aligned}\eta = \eta(\epsilon',\delta') = \frac{1}{8}\left(\frac{\delta'}{2}\right)^{1/\log_2 \varphi} &= \frac{1}{8}\left(\frac{\hat{\delta}'(\hat{\eta})}{2}\right)^{1/\log_2 \varphi} \\ = \frac{1}{8}\left(\frac{1}{2}\cdot 2(8\cdot\hat{\eta})^{\log_2 \varphi}\right)^{1/\log_2 \varphi} &= \hat{\eta} \; .\end{aligned}$$

Thus, if $S_0 \leq \hat{\eta}$, we have for $n'(\epsilon',\delta',\kappa = 2)$ that (37) holds. Comparing (37) and (39), we deduce that we may take $\hat{n}'(\hat{\epsilon}',\hat{\eta}) \triangleq n'(\hat{\epsilon}',\hat{\delta}'(\hat{\eta}),\kappa = 2)$, where n' is the function promised in Proposition 14. ■

proof of Proposition 16: We prove (41) by following similar steps as in [31]. The proof is given for completeness. Let $\varepsilon_a, \varepsilon_b > 0$ and $n_a < n_b$ be parameters. Define the following events:

$$A: \quad S_n \leq \varepsilon_a \text{ for all } n \geq n_a \; , \tag{78}$$

$$B: \quad \left|\frac{|\{n_a < i < n : T_i = t\}|}{n - n_a} - \frac{1}{2}\right| \leq \varepsilon_b \; , \quad \text{for all } \; n \geq n_b \text{ and all } t \in \{0,1\} \; . \tag{79}$$

We will use the following three observations shortly:

1) By Corollary 15, for given $\varepsilon_a > 0$ and $\eta > 0$ there exists an n_a such that
$$\Pr(A) \geq 1 - \delta'(\eta) \; . \tag{80}$$

2) By the strong law of large numbers, for given ε_b, n_a, and $\delta - \delta'(\eta) > 0$ there exists $n_b > n_a$ such that
$$\Pr(B) \geq 1 - (\delta - \delta'(\eta)) \; . \tag{81}$$

3) If the inequalities (80) and (81) hold, then
$$\begin{aligned}\Pr(A \cap B) &= \Pr(A) + \Pr(B) - \Pr(A \cup B) \\ &= \Pr(A) + \Pr(B) - \big(1 - \Pr(\bar{A} \cap \bar{B})\big) \\ &\overset{\text{(a)}}{=} \Pr(A) + \Pr(B) - 1 + \Pr(\bar{A} \cap \bar{B}) \\ &\overset{\text{(b)}}{\geq} \Pr(A) + \Pr(B) - 1 \\ &\overset{\text{(c)}}{\geq} \Big(1 - \delta'(\eta)\Big) + \Big(1 - (\delta - \delta'(\eta))\Big) - 1 \\ &= 1 - \delta \; ,\end{aligned} \tag{82}$$
where (a) is by De Morgan's laws, (b) is since we throw away a non-negative term, and (c) is by (80) and (81).

Define the shorthand

$$\theta \triangleq -\log_{\varepsilon_a}(\kappa) = \log_{1/\varepsilon_a}(2) \; .$$

Note that for $\varepsilon_a < 1$ we have $\theta > 0$ and $\lim_{\varepsilon_a \to 0} \theta = 0$. Moreover define the shorthands $d_0 = 1$ and $d_1 = 2$.

Following the same steps in [31], we require that ε_a is small enough such that $d_t - \theta > 0$ for $t \in \{0,1\}$, and also require $\varepsilon_a, \varepsilon_b < 1/2$. For ε_a, ε_b satisfying the above and $n_b > n_a$ that are yet to be fixed, we have under $A \cap B$ that for all $n > n_b$,

$$S_n \leq 2^{-2^{\left(\frac{1}{2}-\Delta\right)n}} \; ,$$

where

$$\begin{aligned}\Delta = \sum_{t \in \{0,1\}} \frac{1}{2}\log_2\left(\frac{d_t}{d_t - \theta}\right) - \sum_{t \in \{0,1\}} \pm\varepsilon_b \log_2(d_t - \theta) \\ + \sum_{t \in \{0,1\}} \frac{n_a}{n}\left(\frac{1}{2} \pm \varepsilon_b\right)\log_2(d_t - \theta) \; ,\end{aligned} \tag{83}$$

and

$$\pm \triangleq \begin{cases} + & \text{if } \; d_t - \theta \leq 1, \\ - & \text{otherwise.}\end{cases}$$

Now, for a given $0 < \beta < 1/2$, $\eta > 0$ and $\delta > \delta'(\eta)$, our aim is to show that there exists a choice of parameters $\varepsilon_a, \varepsilon_b > 0$ and $n_a < n_b$ such that (82) holds and $\Delta < 1/2 - \beta$. We choose ε_a small enough such that the first sum in (83) is less than $\frac{1/2-\beta}{3}$, $d_t - \theta > 0$ for $t \in \{0,1\}$, and $\varepsilon_a < 1/2$. We choose ε_b small enough such that the second sum in (83) is less than $\frac{1/2-\beta}{3}$ and that $\varepsilon_b < 1/2$. Note that ε_a has been set and η is given. As justified by Observation 1, we choose n_a large enough such that (80) holds. Then, we choose n_b large enough such that both (81) holds (as justified by Observation 2) and the third sum in (83) is less than $\frac{1/2-\beta}{3}$. The above choices indeed satisfy our aim: by Observation 3, (82) holds since both (80) and (81) hold, and $\Delta < 1/2 - \beta$ since each one of the three sums in (83) is smaller than $\frac{1}{3}(1/2 - \beta)$. Therefore, setting $n_0 = n_b$ ensures (41) holds. ■

D. Additional proofs for Section VI

proof of Proposition 20: By (47) we have

$$\begin{aligned}R_{\mathrm{U}}(\mathcal{G}') = R_{\mathrm{U}}(\mathcal{G}) - \frac{1}{2^{d'}} \cdot I\left(\tilde{Q}^{(j')}_{2^{d'}}\right) \\ + \frac{1}{2^{d'+1}} \cdot I\left(\tilde{Q}^{(2j')}_{2^{d'+1}}\right) + \frac{1}{2^{d'+1}} \cdot I\left(\tilde{Q}^{(2j'+1)}_{2^{d'+1}}\right)\end{aligned}$$

Therefore, the claim will follow by showing that

$$I\left(\tilde{Q}^{(2j')}_{2^{d'+1}}\right) + I\left(\tilde{Q}^{(2j'+1)}_{2^{d'+1}}\right) \leq 2 \cdot I\left(\tilde{Q}^{(j')}_{2^{d'}}\right) \; .$$

For brevity denote

$$\begin{aligned}\Gamma \; &\triangleq \tilde{Q}^{(j')}_{2^{d'}} \; , \\ \Lambda^- &\triangleq \tilde{Q}^{(2j')}_{2^{d'+1}} \; , \\ \Lambda^+ &\triangleq \tilde{Q}^{(2j'+1)}_{2^{d'+1}} \; ,\end{aligned}$$

and

$$u \triangleq u_{2j'} \; , \qquad v \triangleq u_{2j'+1} \; .$$

Hence, our goal is to show that

$$I(\Lambda^-) + I(\Lambda^+) \le 2 \cdot I(\Gamma) \, .$$

For this, we further denote by Γ^- and Γ^+ the minus and plus transforms of Γ, respectively. That is,

$$\Gamma^-(t_a, t_b; u) = \sum_v \Gamma(t_a; u \oplus v) \cdot \Gamma(t_b; v) \, , \tag{84a}$$
$$\Gamma^+(t_a, t_b, u; v) = \Gamma(t_a; u \oplus v) \cdot \Gamma(t_b; v) \, . \tag{84b}$$

By the chain rule,

$$I(\Gamma^-) + I(\Gamma^+) = 2 \cdot I(\Gamma) \, .$$

Since (stochastic) degradation reduces mutual information, we will be done once we prove that Λ^- is degraded with respect to Γ^-, and Λ^+ is degraded with respect to Γ^+. By inspection of (14) versus (84), this is indeed the case. Namely, Γ^- is degraded to Λ^- by deterministically mapping (t_a, t_b) to $\tilde{f}(t_a, t_b)$ while Γ^+ is degraded to Λ^+ by deterministically mapping (t_a, t_b, u) to $g_u(t_a, t_b)$. ■

proof of Proposition 21: By (44) we have

$$\begin{aligned} R_{\mathrm{L}}(\mathcal{G}, \mathcal{E}') = R_{\mathrm{L}}(\mathcal{G}, \mathcal{E}) &- \frac{1}{2^{d'}} \cdot \max\left\{1 - \delta'(\zeta_{2^{d'}}^{(j')}), 0\right\} \\ &+ \frac{1}{2^{d'+1}} \cdot \max\left\{1 - \delta'(\zeta_{2^{d'+1}}^{(2j')}), 0\right\} \\ &+ \frac{1}{2^{d'+1}} \cdot \max\left\{1 - \delta'(\zeta_{2^{d'+1}}^{(2j'+1)}), 0\right\} \, . \end{aligned}$$

Therefore, the claim will follow by showing that

$$\begin{aligned} 2 \cdot \max\left\{1 - \delta'(\zeta_{2^{d'}}^{(j')}), 0\right\} \le \max&\left\{1 - \delta'(\zeta_{2^{d'+1}}^{(2j')}), 0\right\} \\ &+ \max\left\{1 - \delta'(\zeta_{2^{d'+1}}^{(2j'+1)}), 0\right\} \, . \end{aligned}$$

Recall (42) and denote for brevity

$$\begin{aligned} \zeta \; &\triangleq (\zeta_{2^{d'}}^{(j')}) \, , \\ \zeta^- &\triangleq (\zeta_{2^{d'+1}}^{(2j')}) \quad = 2 \cdot \zeta \, , \\ \zeta^+ &\triangleq (\zeta_{2^{d'+1}}^{(2j'+1)}) = \zeta^2 \, . \end{aligned}$$

Hence, our goal is to show that

$$\begin{aligned} 2 \cdot \max\{1 - \delta'(\zeta), 0\} \le \max&\{1 - \delta'(\zeta^-), 0\} \\ &+ \max\{1 - \delta'(\zeta^+), 0\} \, . \end{aligned} \tag{85}$$

We assume that the LHS is positive, otherwise the claim is trivial. Under this assumption, we show that

$$2 \cdot (1 - \delta'(\zeta)) \le \big(1 - \delta'(\zeta^-)\big) + \big(1 - \delta'(\zeta^+)\big) \, ,$$

which implies (85). Using the definition of $\delta'(\cdot)$ in (40) and plugging $\zeta^- = 2 \cdot \zeta$ and $\zeta^+ = \zeta^2$, the above simplifies to

$$x \cdot \big(x - (2 - 2^{\log_2 \varphi})\big) \le 0 \, , \tag{86}$$

where we use the shorthand $x \triangleq \zeta^{\log_2 \varphi}$. Therefore, the inequality holds for $x \in [0, 2 - 2^{\log_2 \varphi}]$, which is $\zeta \in [0, 1/4]$. That is, (85) holds if $\zeta \in [0, 1/4]$. Since ζ is non-negative, it remains to show that $\zeta \le 1/4$. Indeed, by our assumption that the LHS in (85) is positive and by (40),

$$\zeta < \frac{1}{8} \cdot \left(\frac{1}{2}\right)^{1/\log_2 \varphi} \approx 0.046 \, .$$

■

proof of Proposition 22: Recall (44) and consider the calculation of $R_{\mathrm{L}}(\mathcal{G}, \mathcal{E})$ versus $R_{\mathrm{L}}(\mathcal{G}', \mathcal{E})$. To distinguish and compare between the terms of these two sums, we use the notation $\hat{\zeta}$ for $\mathcal{G}'$. That is,

$$R_{\mathrm{L}}(\mathcal{G}, \mathcal{E}) = \sum_{(d,j) \in \mathcal{E}} \frac{1}{2^d} \cdot \max\left\{1 - \delta'(\zeta_{2^d}^{(j)}), 0\right\} \, ,$$
$$R_{\mathrm{L}}(\mathcal{G}', \mathcal{E}) = \sum_{(d,j) \in \mathcal{E}} \frac{1}{2^d} \cdot \max\left\{1 - \delta'(\hat{\zeta}_{2^d}^{(j)}), 0\right\} \, .$$

By Definition 3 of $\mathcal{G}'(d', j')$, only nodes $(d, j) \in \mathcal{E}$ which are descendants of (d', j') contribute differently to the sum. The proof will follow by showing that for each such term, the contribution does not decrease when changing $\mathcal{G}$ to $\mathcal{G}'$. Thus, our goal is to show that for all $(d, j) \in \mathcal{E}$ which is a descendant of (d', j') it holds that

$$\delta'(\hat{\zeta}_{2^d}^{(j)}) \le \delta'(\zeta_{2^d}^{(j)}) \, .$$

By the monotonicity of $\delta'(\cdot)$, defined in (40), it is sufficient to show that

$$\hat{\zeta}_{2^d}^{(j)} \le \zeta_{2^d}^{(j)} \, . \tag{87}$$

Recall how $\zeta_{2^d}^{(j)}$ is calculated: we use (42), where the base of the recursion is node (d', j') whose corresponding $\zeta_{2^{d'}}^{(j')}$ equals $Z^\star(\tilde{Q}_{2^{d'}}^{(j')})$. In contrast, for $\hat{\zeta}_{2^d}^{(j)}$, we use the same recursive relations in (42), but the base case is one level deeper: either $\hat{\zeta}_{2^{d'+1}}^{(2j')} = Z^\star(\tilde{Q}_{2^{d'+1}}^{(2j')})$ or $\hat{\zeta}_{2^{d'+1}}^{(2j'+1)} = Z^\star(\tilde{Q}_{2^{d'+1}}^{(2j'+1)})$, depending on the value of j. By inspection of (42) versus (26), we have that $\hat{\zeta}_{2^{d'+1}}^{(2j')} \le \zeta_{2^{d'+1}}^{(2j')}$ and $\hat{\zeta}_{2^{d'+1}}^{(2j'+1)} \le \zeta_{2^{d'+1}}^{(2j'+1)}$. By the monotonicity of the two operations in (42), doubling and squaring, we deduce that this inequality persists throughout the path to (d, j), and hence (87) indeed holds. ■

proof of Corollary 23: To prove (52), recall the definition of $\mathcal{G}^*$. That is, for some finite integer T, there exists a sequence

$$\mathcal{G} = \mathcal{G}_0, \mathcal{G}_1, \ldots, \mathcal{G}_T = \mathcal{G}^* \, , \tag{88}$$

where for each $0 \le t < T$ we have $\mathcal{G}_{t+1} = \mathcal{G}_t'(d, j)$, for some $(d, j) \in \mathcal{G}_t$. Thus, by (49), for each $0 \le t < T$ we have $R_{\mathrm{U}}(\mathcal{G}_{t+1}) \le R_{\mathrm{U}}(\mathcal{G}_t)$. Hence, (52) follows.

To prove (53) we show that

$$R_{\mathrm{L}}(\mathcal{G}, \mathcal{E}) \le R_{\mathrm{L}}(\mathcal{G}, \mathcal{E}^*) \le R_{\mathrm{L}}(\mathcal{G}^*, \mathcal{E}^*) \, . \tag{89}$$

Consider the first inequality. For this, note that as before, we have for some finite integer S the sequence

$$\mathcal{E} = \mathcal{E}_0, \mathcal{E}_1, \ldots, \mathcal{E}_S = \mathcal{E}^* \, , \tag{90}$$

where for each $0 \le s < S$ we have $\mathcal{E}_{s+1} = \mathcal{E}_s'(d, j)$, for some $(d, j) \in \mathcal{E}_s$. Thus, by (50), for each $0 \le s < S$ we have

$R_{\mathrm{L}}(\mathcal{G}, \mathcal{E}_{s+1}) \geq R_{\mathrm{L}}(\mathcal{G}, \mathcal{E}_s)$. Hence, the first inequality in (89) follows.

Consider now the second inequality in (89). Recall that $(\mathcal{G}^*, \mathcal{E}^*)$ is a valid pair. Next, note that if $(\mathcal{G}_{t+1}, \mathcal{E}^*)$ is a valid pair, then so is $(\mathcal{G}_t, \mathcal{E}^*)$. Hence, since $\mathcal{G}^* = \mathcal{G}_T$, all pairs $(\mathcal{G}_t, \mathcal{E}^*)$ are valid for $0 \leq t \leq T$. We wish to apply (51) repeatedly to show that $R_{\mathrm{L}}(\mathcal{G}_{t+1}, \mathcal{E}^*) \geq R_{\mathrm{L}}(\mathcal{G}_t, \mathcal{E}^*)$, from which the second inequality in (89) follows. Recalling the conditions in Proposition 22, we must show that $\mathcal{G}_{t+1}$ is not constructed from $\mathcal{G}_t$ by a node $(d', j') \in \mathcal{E}^*$. Indeed, if this were the case, then $(\mathcal{G}_{t+1}, \mathcal{E}^*)$ would not be valid, contradicting what we have already established. ■

proof of Proposition 19: Note that $I(\tilde{Q}_1^{(0)}) \leq I(W) \triangleq C$, since $\tilde{Q}_1^{(0)}$ is obtained by stochastically degrading W using the labeling function $\lambda(\cdot)$. Next, we define the set $\mathcal{G}_{(0,0)} \triangleq \{(0,0)\}$. By inspection of (47), $R_{\mathrm{U}}(\mathcal{G}_{(0,0)}) = I(\tilde{Q}_1^{(0)})$. Hence, $R_{\mathrm{U}}(\mathcal{G}_{(0,0)}) \leq C$. The claim follows by noting that each valid $\mathcal{G}$ satisfies $\mathcal{G} \geq \mathcal{G}_{(0,0)}$. Therefore by applying (52) we have $R_{\mathrm{U}}(\mathcal{G}) \leq R_{\mathrm{U}}(\mathcal{G}_{(0,0)}) \leq C$. ■

References

[1] E. Arıkan, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.

[2] J. Honda and H. Yamamoto, “Polar coding without alphabet extension for asymmetric channels,” *IEEE Trans. Inform. Theory*, vol. 59, no. 12, pp. 7829–7838, December 2012.

[3] E. Şaşoğlu, E. Telatar, and E. Arıkan, “Polarization for arbitrary discrete memoryless channels,” pp. 144–148, October 2009.

[4] E. Şaşoğlu, “Polar codes for discrete alphabets,” in *Proc. IEEE Int’l Symp. Inform. Theory (ISIT’2012)*, Cambridge, Massachusetts, 2012, pp. 2137–2141.

[5] E. Şaşoğlu and I. Tal, “Polar coding for processes with memory,” *IEEE Trans. Inform. Theory*, vol. 65, no. 4, pp. 1994–2003, April 2019.

[6] B. Shuval and I. Tal, “Fast polarization for processes with memory,” *IEEE Trans. Inform. Theory*, vol. 65, no. 4, pp. 2004–2020, April 2019.

[7] R. Wang, J. Honda, H. Yamamoto, R. Liu, and Y. Hou, “Construction of polar codes for channels with memory,” in *Proc. IEEE Inform. Theory Workshop (ITW’2015)*, Jeju Island, Korea, 2015, pp. 187–191.

[8] Y. Wang, M. Qin, K. R. Narayanan, A. Jiang, and Z. Bandic, “Joint source-channel decoding of polar codes for language-based sources,” in *Proc. IEEE Global Telecommun. Conf. (Globecom’2016)*, Washington, DC, 2016.

[9] I. Tal, H. D. Pfister, A. Fazeli, and A. Vardy, “Polar codes for the deletion channel: weak and strong polarization,” *IEEE Trans. Inform. Theory*, vol. 68, no. 4, pp. 2239–2265, April 2022.

[10] D. Arava and I. Tal, “Stronger polarization for the deletion channel,” in *2023 IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 1711–1716.

[11] H. D. Pfister and I. Tal, “Polar codes for channels with insertions, deletions, and substitutions,” in *Proc. IEEE Int’l Symp. Inform. Theory (ISIT’2021)*, Melbourne, Victoria, Australia, 2021, pp. 2554–2559.

[12] E. Arıkan, “Source polarization,” in *Proc. IEEE Int’l Symp. Inform. Theory (ISIT’2010)*, Austin, Texas, 2010, pp. 899–903.

[13] S. B. Korada and R. Urbanke, “Polar codes for Slepian-Wolf, Wyner-Ziv, and Gelfand-Pinsker,” in *Proc. IEEE Inform. Theory Workshop (ITW’2010)*, Cairo, Egypt, 2010.

[14] E. Hof and S. Shamai, “Secrecy-achieving polar-coding for binary-input memoryless symmetric wire-tap channels,” `arXiv:1005.2759v2`, 2010.

[15] H. Mahdavifar and A. Vardy, “Achieving the secrecy capacity of wiretap channels using polar codes,” *IEEE Trans. Inform. Theory*, vol. 57, pp. 6428–6443, 2011.

[16] B. Shuval and I. Tal, “Strong polarization for shortened and punctured polar codes,” in *Proc. IEEE Int’l Symp. Inform. Theory (ISIT’2024)*, Athens, Greece, 2024.

[17] C. Leroux, I. Tal, A. Vardy, and W. J. Gross, “Hardware architectures for successive cancellation decoding of polar codes,” in *Proc. IEEE Int’l Conf. Acoust. Speech Signal Process. (ICASSP’2011)*, Prague, Czech Republic, 2011, pp. 1665–1668.

[18] M. Fossorier, M. Mihaljevic, and H. Imai, “Reduced complexity iterative decoding of low-density parity check codes based on belief propagation,” *IEEE Transactions on Communications*, vol. 47, no. 5, pp. 673–680, 1999.

[19] I. Tal and A. Vardy, “List decoding of polar codes,” *IEEE Trans. Inform. Theory*, vol. 61, no. 5, pp. 2213–2226, May 2015.

[20] K. Niu and K. Chen, “Stack decoding of polar codes,” *Electronics letters*, vol. 48, no. 12, pp. 695–697, 2012.

[21] C. Leroux, A. J. Raymond, G. Sarkis, and W. J. Gross, “A semi-parallel successive-cancellation decoder for polar codes,” *IEEE Transactions on Signal Processing*, vol. 61, no. 2, pp. 289–299, 2013.

[22] N. Miki, S. Suyama, and S. Nagata, “Performance of polar codes under successive cancellation decoding employing approximation algorithm,” in *2019 13th International Conference on Signal Processing and Communication Systems (ICSPCS)*, 2019, pp. 1–6.

[23] A. Balatsoukas-Stimming, M. B. Parizi, and A. Burg, “LLR-based successive cancellation list decoding of polar codes,” *IEEE Transactions on Signal Processing*, vol. 63, no. 19, pp. 5165–5179, 2015.

[24] I. Tal and A. Vardy, “How to construct polar codes,” *IEEE Trans. Inform. Theory*, vol. 59, no. 10, pp. 6562–6582, October 2013.

[25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[26] B. Shuval and I. Tal, “Universal polarization for processes with memory,” *Accepted for publication in IEEE Trans. Inform. Theory*, 2025, available online at https://ieeexplore.ieee.org/document/10836796.

[27] R. G. Gallager, *Information Theory and Reliable Communications*. New York: John Wiley, 1968.

[28] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[29] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 4th ed. Cambridge, Massachusetts: The MIT Press, 2022.

[30] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*, 2nd ed. Reading, Massachusetts: Addison-Wesley, 1994.

[31] I. Tal, “A simple proof of fast polarization,” *IEEE Trans. Inform. Theory*, vol. 63, no. 12, pp. 7617–7619, December 2017.