Stronger Polarization for the Deletion Channel

Dar Arava, Ido Tal

Department of Electrical and Computer Engineering,
Technion, Haifa 32000, Israel.
{aravadar@campus, idotal@ee}.technion.ac.il

Abstract—In this paper we show a polar coding scheme for the deletion channel with a probability of error that decays roughly like $2^{-\sqrt{\Lambda}}$, where Λ is the length of the codeword. That is, the same decay rate as that of seminal polar codes for memoryless channels. This is stronger than prior art in which the square root is replaced by a cube root. Our coding scheme is similar yet distinct from prior art. The main differences are: 1) Guard-bands are placed in almost all polarization levels; 2) Trellis decoding is applied to the whole received word, and not to segments of it. As before, the scheme is capacity-achieving. The price we pay for this improvement is a higher decoding complexity, which is nonetheless still polynomial, $O(\Lambda^4)$.

I. INTRODUCTION

Deletion errors, along with insertion errors, arise in communication channels with symbol-timing mismatch [1]. These synchronization errors are also common in polymer-based storage solutions [2].

The simplest theoretical model for these errors is the deletion channel with a constant deletion probability. The channel output is a sub-string of the symbols in the input. Deletions occur according to an i.i.d. process that deletes each input symbol with probability δ.

Polar codes [3] for a deletion channel with a fixed deletion probability were first presented in [4]. See also [5]–[8], which use polar codes for weaker settings. In [4], the authors show that for a fixed regular hidden-Markov input process and a fixed parameter $\nu \in (0, \frac{1}{3})$, their coding scheme approaches the mutual information rate between the input process and the channel output. The encoding and decoding complexities are $O(\Lambda \log \Lambda)$ and $O(\Lambda^{1+3\nu})$, respectively, where Λ is the codeword length. Furthermore, for any $0 < \nu' < \nu$ and large enough Λ, the probability of a decoding block error is at most $2^{-\Lambda^{\nu'}}$. The authors also show that there exists a sequence of regular hidden-Markov input processes for which the mutual information rate approaches the deletion channel capacity. This result also follows as a special case of the work of Li and Tan [9], which proved the above for finite-order Markov processes.

We extend [4], and show that with simple alternations to the encoding and decoding schemes, the error probability decreases as $2^{-\Lambda^{\beta'}}$ where $\beta' \in (0, \frac{1}{2})$. Recall that this is

stronger than the result in [4], in which the error exponent is $\nu' \in (0, \frac{1}{3})$.

The structure of our paper is as follows. Section II presents the main results of our work, still without all the definitions in place. Section III completes these definitions and sets up key notation. It also introduces key concepts we inherit from [4], and compares and contrasts our results to those in [4]. The first three subsections of Section IV detail both our encoding and decoding schemes. Thus, the practitioner who only wishes to implement our scheme may decide to only read up to this point. The section concludes with a complexity analysis and simulation results. Section V contains our first key lemma, Lemma 5. This lemma bounds the evolution of the Bhattacharyya parameter when applying '−' and '+' polarization operations. It contains an additive term which is not present in simpler channels. Section VI contains our second key lemma, Lemma 7, which shows that despite this additive penalty, strong polarization still occurs. Section VII uses these two key lemmas in order to prove our main result.

II. MAIN THEOREM

Our main result builds upon the guard-band function g introduced in [4]. We will define g shortly. For now, we note that $g(\mathbf{x}, n_0, \xi)$ recursively transforms $\mathbf{x}$, a word of length 2^n, into a slightly longer word, where the length is controlled by the parameter ξ, and $n - n_0$ is the recursion depth. We say that $g(\mathbf{x}, n_0, \xi)$ is the result of adding guard-bands to $\mathbf{x}$.

Throughout the paper, we assume a deletion channel with a *fixed* deletion probability. We also assume a fixed regular hidden Markov input distribution (see [4, Subsection II-D] for the formal definition). Denote by $\mathcal{I}$ the information rate between an input distributed according to this distribution and the corresponding output of the deletion channel. Denote by Z and K the Bhattacharyya parameter and the total-variation, respectively (see, for example, [10, Section III]).

Here is our "stronger polarization" theorem. It is stronger than the "weaker polarization" theorem proved in [4]. In [4], the parameters $0 < \beta' < \beta < \frac{1}{2}$ below are replaced with $0 < \nu' < \nu < \frac{1}{3}$, which implies a weaker decay to the error probability bound.

Theorem 1 (Stronger polarization): Let $\delta \in (0, 1)$ be a fixed deletion probability of the deletion channel. Fix $\epsilon \in (0, 1)$, $\xi \in (0, \frac{1}{6})$, and $0 < \beta' < \beta < \frac{1}{2}$. There exist $n_0^{\text{th}}(\epsilon, \delta, \xi)$ and $n^{\text{th}}(\epsilon, \beta, \beta', n_0)$ such that the following holds. Take $n_0 \geq n_0^{\text{th}}(\epsilon, \delta, \xi)$ and $n \geq n^{\text{th}}(\epsilon, \beta, \beta', n_0)$. Let $\mathbf{X}$ be of length $N = 2^n$. The vector $\mathbf{X}$ is partitioned into blocks of length

2^{n_0}, and each block is independently distributed according to the fixed regular hidden Markov input distribution. Let $\mathbf{U}$ be the polar transform of $\mathbf{X}$. Denote by $\mathbf{Y}$ the result of transmitting $g(\mathbf{X}, n_0, \xi)$ through the deletion channel. The fraction of indices i for which

$$Z(U_i|U_1^{i-1}, \mathbf{Y}) < 2^{-N^\beta} < \frac{1}{2N} \cdot 2^{-\Lambda^{\beta'}} \qquad (1)$$

$$K(U_i|U_1^{i-1}) < 2^{-N^\beta} < \frac{1}{2N} \cdot 2^{-\Lambda^{\beta'}} \qquad (2)$$

is at least $\mathcal{I} - \epsilon$, where Λ is the length of $g(\mathbf{X}, n_0, \xi)$. Furthermore,

$$\frac{N}{\Lambda} > 1 - \epsilon \;.$$

By using the Honda-Yamamoto scheme [11], [12], we get the following corollary.

Corollary 2: The above implies a coding scheme with rate $\mathcal{I} - 2\epsilon$ and probability of error at most $2^{-\Lambda^{\beta'}}$, where Λ is the length of the transmitted codeword.

III. PRELIMINARIES

In this section we set up some notation and summarize key concepts from [4].

A. Three related channels

We now introduce three related channels: the deletion channel, the trimming channel, and their composition, the trimmed deletion channel.

Deletion Channel (DC) The deletion channel is the channel we are to code over. As its name implies, it takes a binary vector and deletes each bit with probability δ. Thus, the output of the channel is typically shorter than its input. We will often denote a random vector that is an input to such a channel by $\mathbf{G}$ and denote the corresponding output by $\mathbf{Y}$.

The following two channels were introduced in [4], and are concepts we will need for our results as well.

Trimming Channel (TC) The trimming channel takes a binary vector and removes from it all leading and trailing zeros. Note that the trimming channel is deterministic. We will often denote the input to this channel by either $\mathbf{Y}$ or $\mathbf{Z}$. We denote the trimming operation by appending a '$*$' as a superscript. Thus, the outputs corresponding to $\mathbf{Y}$ and $\mathbf{Z}$ will be $\mathbf{Y}^*$ and $\mathbf{Z}^*$, respectively.

Trimmed Deletion Channel (TDC) The trimmed deletion channel is the composition of the above two channels. Thus, if the input to the channel is $\mathbf{G}$, then we first pass $\mathbf{G}$ through the deletion channel and obtain $\mathbf{Y}$, and then pass $\mathbf{Y}$ through the trimming channel, which yields $\mathbf{Y}^*$.
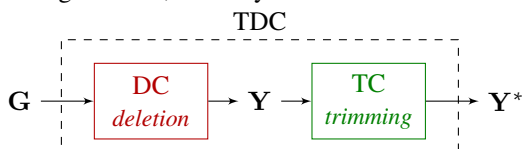
B. Weak polarization for the TDC

We next state a weak polarization theorem for the trimmed deletion channel. The theorem follows easily by combining the weak polarization theorem [4, Theorem 20] with [13, Lemma 1], which bounds the Bhattacharyya parameter Z and total variation K by monotonic functions of the conditional entropy. As we will see, we will use this theorem to reach the stronger polarization rate stated in our main theorem.

Theorem 3 (Weak polarization for the trimmed deletion channel): Fix $\eta \in (0, 1)$ and let $N_0 = 2^{n_0}$. For a given fixed regular hidden Markov input distribution, let $\mathbf{X}$ be a random vector of length N_0 distributed according to the input distribution. Let $\mathbf{Y}^*$ be the result of passing $\mathbf{X}$ through a trimmed deletion channel with deletion probability δ. Denote $\mathbf{U}$ as the polar transform of $\mathbf{X}$. For all $\epsilon' > 0$ there exists an $n_0^{\text{th}}(\epsilon', \eta, \delta)$ s.t. if $n_0 \geq n_0^{\text{th}}$ then the fraction of indices i for which

$$Z(U_i|U_1^{i-1}, \mathbf{Y}^*) < \eta \quad \text{and} \quad K(U_i|U_1^{i-1}) < \eta$$

is larger than $\mathcal{I} - \epsilon'$.

Note that in the above, and throughout the paper, the information rate $\mathcal{I}$ is

$$\mathcal{I} \triangleq \lim_{N_0 \to \infty} \frac{I(\mathbf{X}; \mathbf{Y})}{N_0} \;,$$

where $\mathbf{X}$ is of length $|\mathbf{X}| = N_0$ and is drawn according to the fixed regular hidden Markov input distribution and $\mathbf{Y}$ is the corresponding output of the deletion channel. In [4, Section V] it was shown that $\mathcal{I}$ is indeed well defined.

We end this section by noting that Theorem 3 does not employ guard-bands. However, in order to prove stronger polarization, we will employ guard-bands (as was also done in [4]).

C. Blocks and guard-bands

Recall that in the main theorem, $\mathbf{X}$ was partitioned into independent blocks of length $N_0 = 2^{n_0}$. There are $N_1 = \frac{N}{N_0}$ such blocks, and we denote them by $\mathbf{X}(1), \mathbf{X}(2), \ldots, \mathbf{X}(N_1)$. That is, $\mathbf{X}$ is the concatenation of the above N_1 blocks,

$$\mathbf{X} = \mathbf{X}(1) \odot \mathbf{X}(2) \odot \cdots \odot \mathbf{X}(N_1).$$

We denote the first and second halves of $\mathbf{X}$ by $\mathbf{X}_{\text{I}}$ and $\mathbf{X}_{\text{II}}$. Denoting the length of a vector by $|\cdot|$, we have $|\mathbf{X}_{\text{I}}| = |\mathbf{X}_{\text{II}}| = \frac{N}{2}$ and

$$\mathbf{X} = \mathbf{X}_{\text{I}} \odot \mathbf{X}_{\text{II}}.$$

Note that $\mathbf{X}_{\text{I}}$ and $\mathbf{X}_{\text{II}}$ are independent, a convention that will also hold in other places in which we use the "I" and "II" subscripts.

Recall that the function g mentioned previously transforms a vector $\mathbf{X}$ of length 2^n into a slightly longer vector with "guard-bands". We now define g recursively, and note that it adds the guard-bands between blocks. For a vector $\mathbf{X}$ of length $\leq 2^{n_0}$, $g(\mathbf{X}, n_0, \xi)$ is simply the identity function. For a vector $\mathbf{X}$ of length greater than 2^{n_0},

$$g(\mathbf{X}, n_0, \xi) \triangleq g(\mathbf{X}) \triangleq \underbrace{g(\mathbf{X}_{\text{I}})}_{\triangleq \mathbf{G}_{\text{I}}} \odot \overbrace{\underbrace{000\ldots00}_{\triangleq \mathbf{G}_\Delta}}^{\ell_n} \odot \underbrace{g(\mathbf{X}_{\text{II}})}_{\triangleq \mathbf{G}_{\text{II}}} \;. \qquad (3)$$

That is, we add

$$\ell_n = \lfloor 2^{(1-\xi)(n-1)} \rfloor \qquad (4)$$

'0' symbols between the first and second halves of $\mathbf{X}$, and apply g recursively on each half. Note that $\xi > 0$ is a "small" constant that we will constrain later. To summarize: $\mathbf{X}$ is a concatenation of 2^{n-n_0} independent blocks, each of length $N_0 = 2^{n_0}$. The function $g(\mathbf{X}, n_0, \xi)$ adds a guard-band of '0' symbols between each two blocks, and the length of these guard-bands varies. Here is an illustration, for the case in which $n = n_0 + 2$:
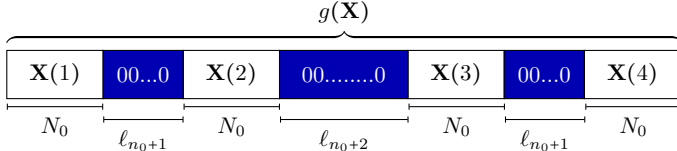
We are not concerned by the added GB bits, since for a large enough n_0, the effect they have on the rate is negligible. The following lemma shows this.

Lemma 4 (GBs have a negligible effect on the rate): For all $\epsilon > 0$ and $\xi > 0$ there exists an $n_0^{\text{th}}(\epsilon, \xi)$ such that if $n_0 \geq n_0^{\text{th}}$ then

$$\frac{|\mathbf{X}|}{|g(\mathbf{X}, n_0, \xi)|} \triangleq \frac{N}{\Lambda} > 1 - \epsilon .$$

Proof:
We set $n_0^{\text{th}} \triangleq -\frac{1}{\xi} \log_2(\epsilon \cdot (1 - 2^{-\xi}))$ and bound Λ as follows:

$$\Lambda \triangleq |g(\mathbf{X})| \overset{(a)}{<} |\mathbf{X}| \cdot \left(1 + \frac{2^{-(\xi n_0 + 1)}}{1 - 2^{-\xi}} \right)$$

$$< |\mathbf{X}| \cdot \left(1 + \frac{2^{-\xi n_0}}{1 - 2^{-\xi}} \right) \qquad (4a)$$

$$\overset{(b)}{\leq} N \cdot (1 + \epsilon) , \qquad (4b)$$

where (a) is by [4, Lemma 22] and (b) follows by recalling that $n_0 \geq n_0^{\text{th}}$. Finally:

$$\frac{N}{\Lambda} \overset{(5)}{>} \frac{1}{1 + \epsilon} = 1 - \frac{\epsilon}{1 + \epsilon} > 1 - \epsilon .$$

∎

We remind the reader that $\mathbf{G} = \mathbf{G}_{\text{I}} \odot \mathbf{G}_\Delta \odot \mathbf{G}_{\text{II}}$ is passed through the DC. We denote the output of this channel by $\mathbf{Y}$, and denote the parts corresponding to $\mathbf{G}_{\text{I}}$, $\mathbf{G}_\Delta$, and $\mathbf{G}_{\text{II}}$ by $\mathbf{Y}_{\text{I}}$, $\mathbf{Y}_\Delta$, and $\mathbf{Y}_{\text{II}}$, respectively. We further denote the application of the TC on $\mathbf{Y}$ by $\mathbf{Z} \triangleq \mathbf{Y}^*$, and denote the parts corresponding to $\mathbf{Y}_{\text{I}}$, $\mathbf{Y}_\Delta$, and $\mathbf{Y}_{\text{II}}$ by $\mathbf{Z}_{\text{I}}$, $\mathbf{Z}_\Delta$, and $\mathbf{Z}_{\text{II}}$, respectively. See Fig. 1, which is essentially [4, Figure 5]. Note that, in general, $\mathbf{Z}_{\text{I}}$ is formed by trimming off only the left side of $\mathbf{Y}_{\text{I}}$. Hence, typically, $\mathbf{Z}_{\text{I}} \neq (\mathbf{Y}_{\text{I}})^*$ and $\mathbf{Z}_{\text{II}} \neq (\mathbf{Y}_{\text{II}})^*$. Also, note that in the typical case, $\mathbf{Z}_\Delta = \mathbf{Y}_\Delta$.

D. An overview of this work compared to [4]

Throughout this paper we inherit concepts from [4], both in our coding scheme and in our analysis. All the concepts we use from [4] are defined in this work as well, to ensure this paper is self-contained. To further highlight what is new, this section outlines the main ideas in [4] and the differences in this work. Every topic mentioned in this section will be explained in more detail later on.
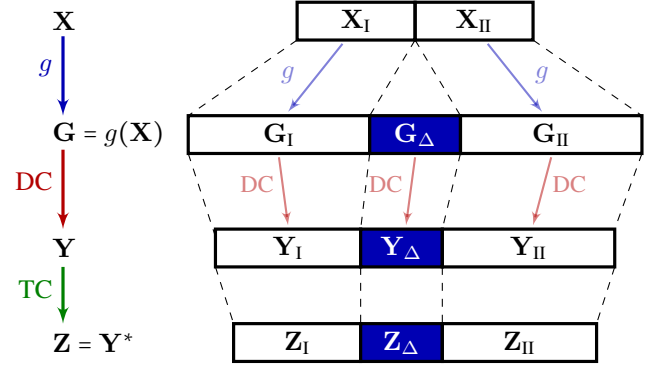

Fig. 1. The random variables $\mathbf{X}$, $\mathbf{G}$, $\mathbf{Y}$, and $\mathbf{Z}$.

The encoder in [4] adds GBs to separate the polar coded word into blocks. We do the same, while slightly changing constrains on parameters that control how many GBs are added and what their lengths are (ξ and n_0). We also slightly change the rule on how to select the information indices in the polar coding scheme. However, in essence, we use the same encoder.

The decoding steps in [4] are as follows. First, the decoder uses the GBs to process the received word into trimmed blocks. In a good scenario, this processing yields the TDC outputs of the blocks of $\mathbf{X}$. Next, the decoder builds a trellis for each trimmed block. The trellisses incorporate the joint probability of the deletion channel input and output. Lastly, the trellisses are processed according to the successive cancellation polar decoder.

Our decoder does not perform the first step of partitioning and trimming the output. Instead, we build one big trellis using the whole received word. As the trellisses in [4], this trellis incorporates the joint probability of the deletion channel input and output. As in [4], we then process the trellis according to the successive cancellation algorithm. Unlike in [4], our trellis includes sections that correspond to GBs. Thus, we expand the definition from [4] on how to build the trellis and how to process it such that it deals with the GB sections as well.

The processing of $\mathbf{Y}$ into trimmed blocks in [4] is subject to errors. For example, if the middle index of the received word does not fall within the middle GB, the decoder in [4] will likely fail. Thus, our coding scheme which does not contain such a preliminary step has a higher probability of decoding successfully. This comes at a cost in decoding complexity, $\mathcal{O}(\Lambda^4)$ in our case compared to at most $\mathcal{O}(\Lambda^2)$ in [4].

Regarding the analysis of the error probability, [4] shows that weak polarization occurs for the TDC after n_0 polar transforms (without adding GBs). Then, they show strong polarization for a "block-TDC" output. That is, essentially, for input $g(\mathbf{X})$ the "block-TDC" output is the TDC outputs of each block in $\mathbf{X}$. We remind that, in a good scenario, the result of the partition and trimming operation in the first step of the decoder of [4] is exactly the "block-TDC" output. Lastly, they bound the probability of failing in partitioning and trimming $\mathbf{Y}$ into the "block-TDC" output (i.e. the probability of not being in the good scenario). This error term is what ultimately constrained their error exponent to at most $\frac{1}{3}$.

We remind that our decoder does not perform partition and trimming to the received vector $\mathbf{Y}$. However, when analyzing the performance of our coding scheme we will use the TDC. Since the TDC is a degraded version of the deletion channel, we may use the bounds on its error probability to bound the error probability of the DC as well. We will show a recursive relation on how the TDC polarizes. This recursive relation will be pivotal in our proof. We will see that the recursive relation is reached while bounding the probability of successfully partitioning the TDC output of $g(\mathbf{X})$ to the TDC outputs of $g(\mathbf{X}_{\mathrm{I}})$ and $g(\mathbf{X}_{\mathrm{II}})$. This is not to be confused with the requirement in [4] of successfully partitioning the received vector to the TDC outputs of the blocks $\mathbf{X}(1), \ldots, \mathbf{X}(N_1)$, which is less probable. When bounding the probability of the former, we will use similar arguments as in the bound of the latter in [4]. To conclude — in [4] it was natural to understand why the authors analyze a channel that trims the output (since the decoder applies trimming), whereas in our setting this might seem like a detour to the reader. However, we do not know how to reach our destination (an analysis of the error probability) without taking this detour.

IV. Coding scheme for the Deletion Channel

A. Encoder

The encoding steps are as follows. The vector $\mathbf{u} = u_1^N$ is produced successively, starting from u_1 and ending in u_N. If the current index i satisfies (1) and (2), then u_i is set to an information bit. Otherwise, u_i is a "frozen bit". Its value is set according to the distribution $\mathbb{P}(U_i = u_i | U_1^{i-1} = u_1^{i-1})$, where u_1^{i-1} are the realizations that were set in the previous indices. Note that this distribution is derived from the regular hidden Markov input distribution we set for the independent blocks. That is, if all bits were frozen, we would get that the blocks $\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(N_1)$ of $\mathbf{x} = \mathcal{A}(\mathbf{u})$ are independently drawn from the regular hidden Markov process. For the calculation of $\mathbb{P}(U_i = u_i | U_1^{i-1} = u_1^{i-1})$ we may use the Markov generalization of Honda-Yamamoto successive cancellation encoding for input distributions with memory [14], [15]. This is done by recursively tracking the distribution of the hidden Markov input and does not require the deletion channel trellis. We assume the random draws are common to both the encoder and the decoder. This is typically implemented using a pseudo-random number generator, common to both sides: if the pseudo-random number $0 \leq r_i \leq 1$ drawn for this stage is such that $\mathbb{P}(U_i = 0 | U_1^{i-1} = u_1^{i-1}) \leq r_i$, we set $u_i = 0$. Otherwise, we set $u_i = 1$. We note that in practice, one may wish to simplify the above by picking a frozen u_i so that it maximizes $\mathbb{P}(U_i = u_i | U_1^{i-1} = u_1^{i-1})$ [16], [17]. We next transform $\mathbf{u}$ to $\mathbf{x} = \mathcal{A}(\mathbf{u})$, using the standard Arıkan transform presented in the seminal paper [3]. For the final step, we add guard-bands of 0's to $\mathbf{x}$, creating $\mathbf{g} = g(\mathbf{x}, n_0, \xi)$. We transmit $\mathbf{g}$ through the deletion channel. The transform adding guard-bands is as in (3).

Our encoder is the same as that in [4], with the following differences. In our encoder, the selection of information indices is according to (1) and (2). In [4], these conditions are

---

Algorithm 1 Encoder

1: **function** ENCODE(information bits, n_0, ξ)
2:     **for** $i = 0$ **to** $N - 1$ **do**
3:         **if** i satisfies (1) and (2) **then**
4:             $u_i \leftarrow$ next information bit
5:         **else**   ▷ u_i is dynamically frozen
6:             $u_i \leftarrow$ draw from input distribution, given u_1^{i-1}
7:     $\mathbf{x} \leftarrow \mathcal{A}(\mathbf{u})$ ▷ Arıkan transform
8:     $\mathbf{g} \leftarrow g(\mathbf{x}, n_0, \xi)$ ▷ add GBs by (3)
9:     **return** $\mathbf{g}$

---

replaced with (essentially, see Equations (101) and (102) in [4]):

$$Z(U_i | U_1^{i-1}, (\mathbf{Y}(1))^*, \ldots, (\mathbf{Y}(N_1))^*) < 2^{-N^\nu}$$
$$\text{and } K(U_i | U_1^{i-1}) < 2^{-N^\nu} ,$$

where $\nu \in \left(0, \frac{1}{3}\right)$. That is, the deletion channel output $\mathbf{Y}$ in (1) is replaced with the TDC outputs of the blocks of $\mathbf{X}$, and the upper bounds in (1) and (2) are replaced with a weaker bound. We also differ in the selection of n_0, i.e. the step from which we start adding guard-bands. Still, the encoding complexity remains $O(\Lambda \log \Lambda)$ for a codeword length of $|\mathbf{g}| = \Lambda$ [4, Sub-claim 9]. We note that the encoding complexity also increases polynomially with the number of states. This is due to the computation of $\mathbb{P}(U_i = u_i | U_1^{i-1} = u_1^{i-1})$ when setting the frozen bit values [14, Theorem 2]. In total, the encoding complexity is $O(|\mathcal{S}|^3 \Lambda \log \Lambda)$, where $|\mathcal{S}|$ is the number of states in the hidden Markov process. Thus, for a fixed input distribution we get $O(\Lambda \log \Lambda)$.

B. High level decoder description

Our decoder is a generalization of the one described in [4, Subsection IV]. That is, a base trellis is constructed, and then '−' and '+' operations are applied to it. One major difference is that in our case, the base trellis corresponds to all of the received word (Panel (a) in Figure 3). This is in contrast to [4], in which N/N_0 base trellises are constructed — one for each block (Figure 4). Since we operate on a larger trellis, our complexity is $O(\Lambda^4)$, as opposed to at most $O(\Lambda^2)$ in [4]. As explained in Theorem 1, this added complexity is compensated for by a reduced probability of error. That is, we reach the same asymptotic bound as seminal polar codes [18].

As in [4], we will use a trellis $\mathcal{T}$ to represent the joint probability of the deletion channel input and output. We perform '−' and '+' operations on $\mathcal{T}$, which merge two-edge paths in $\mathcal{T}$ and result in trellises with a reduced number of columns: $\mathcal{T}^{[0]}$ and $\mathcal{T}^{[1]}$, respectively. The mechanics of our decoder will differ from those of [4, Section IV] in one main point: our trellis contains sections corresponding to guard-bands, and such sections will be treated differently than what was described in [4, Section IV] when applying '−' and '+' operations.

The decoder recursively performs '−' and '+' transforms on $\mathcal{T}$ as follows. First, we perform n '−' transforms, creating $\mathcal{T}^{[000\ldots00]}$. For simplicity, assume for now that the hidden
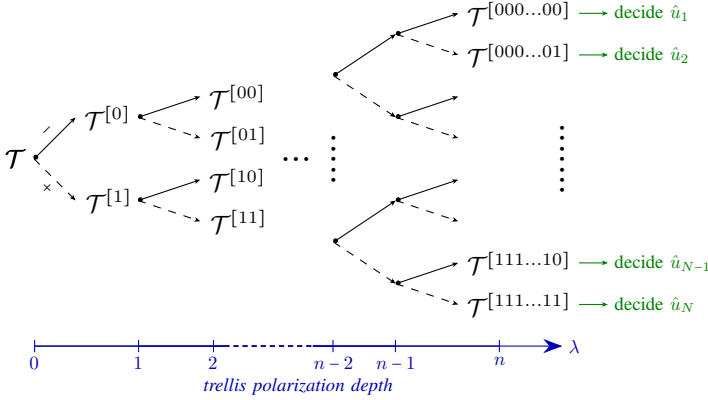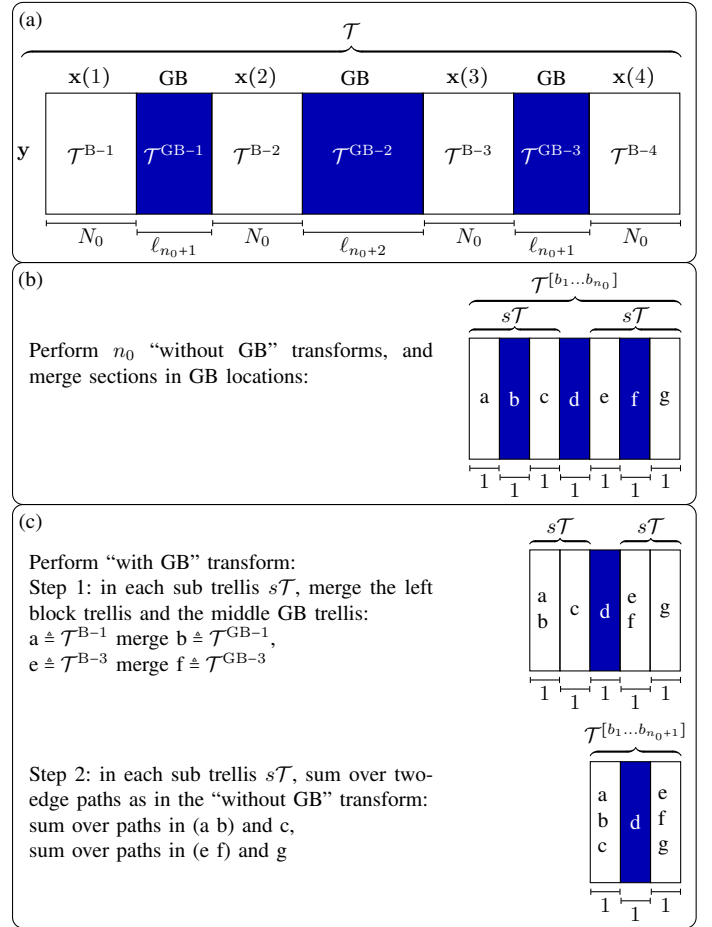
Fig. 2. Recursive trellis transforms.

Fig. 3. Trellis evolution in the decoder. Panel (a) is the initial trellis $\mathcal{T}$. White rectangles are the block trellises which correspond to the blocks of $\mathbf{X}$. In block trellises, both edges with label '1' and edges with label '0' exist. The blue rectangles mark the GB trellises, which correspond to the GBs (have only '0' labels). The number of sections in each block trellis or GB trellis in $\mathcal{T}$ is the length of the corresponding block or GB. Panel (b) shows the trellis after collapsing each of the GB trellises into a single section and after the first n_0 polar transforms. Panel (c) shows how the next polar transform is applied, and is divided into two steps: merging a block trellis with a GB trellis and then merging the resulting trellis with a block trellis (either a '−' or a '+' merge).

Markov input distribution contains only one state (that is, the input distribution is memoryless). In the resulting trellis, we consider the pair of single-edge paths from the upper-left vertex to the lower-right vertex, which represent the two possible values for $\hat{u}_1$. The decision on $\hat{u}_1$ (if it is not frozen) is by the most probable value, i.e. the edge with the largest probability. Using $\hat{u}_1$, we next create:

$$\mathcal{T}^{\overrightarrow{[000...01]}^{n}} = \left(\mathcal{T}^{\overrightarrow{[000...0]}^{n-1}}\right)^{[1]}.$$

We use $\mathcal{T}^{[000...01]}$ to decide on $\hat{u}_2$. We repeat this procedure such that with trellis $\mathcal{T}^{[b_1 b_2 ... b_n]}$ we decide on the value of $\hat{u}_{i(b_1,...,b_n)}$ (if it is not frozen), where $i(b_1,...,b_n)$ is the index with binary representation $b_1,...,b_n$, see (7). If i is a frozen index, we mimic the encoder operation to get $\hat{u}_i$ based on the previous $\hat{u}_1^{i-1}$. See Figure 2 for an illustration of the decoding process. The next three subsections fill in the details missing in the above description.

C. Trellis operations

We first describe how to build the initial trellis $\mathcal{T}$. We will then describe the transforms performed on the trellis. For an extended description on how to build $\mathcal{T}$ and how to perform '−' and '+' operations on it, see [4, Sections III and IV]. We will present the main idea and provide a visual example, while emphasizing the differences in our case.

The trellis $\mathcal{T}$ is built according to the received binary vector $\mathbf{y}$. It also factors in the deletion probability δ, the stationary probabilities of the input process states, and the transition probabilities between them.

1) Vertices: The trellis consists of a grid of vertices with $|\mathbf{y}| + 1$ rows and $|\mathbf{g}| + 1$ columns (where $\mathbf{g} \triangleq g(\mathbf{x})$ is the transmitted codeword we are to decode). We will denote by i a vertex row index, by j a vertex column index and by $\jmath$ (a dot-less j) the corresponding index of the bit $g_\jmath$ in the vector $\mathbf{x}$. That is, if $g_\jmath$ is a bit that originates from the blocks of $\mathbf{x}$ (i.e. is not a GB bit) then $g_\jmath = x_\jmath$. For some of the columns there will also be $|\mathcal{S}|$ "layers" of vertices, which correspond to the possible states of the input distribution. Specifically, for a column $j - 1$, where $j \in \{1,...,|\mathbf{g}|\}$, if $g_\jmath$ originates from

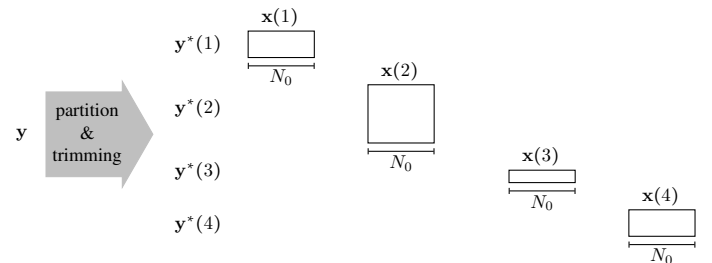Fig. 4. An illustration of the prior-art decoder [4]. The initial step partitions the channel output $\mathbf{y}$ into trimmed outputs of each block. In this example $n = n_0 + 2$, hence there are four blocks. The next step builds a block trellis for each block. These trellises are then processed according to the successive cancellation decoder, by recursively performing '−' and '+' transforms to them.

$\mathbf{x}$ then we must consider the possible input states of S_{j-1} in the trellis. For this column we denote by $v_{i,j-1,s}$ the vertex in row i column $j-1$ and layer (state) $s \in \mathcal{S}$. The vertex $v_{i,j-1,s}$ corresponds to the event in which the output corresponding to g_1^{j-1} is y_1^i and the state of the input process before transmitting x_j was $S_{j-1} = s$. Otherwise, if g_j is a GB bit, we denote by $v_{i,j-1}$ the vertex in row i and column $j-1$ (no layers in this column). The vertex $v_{i,j-1}$ corresponds to the event in which the output corresponding to g_1^{j-1} is y_1^i. The last column in $\mathcal{T}$ will also not have multiple layers, $v_{i,|\mathbf{g}|}$. As we will shortly see, the only vertex of interest in this column is the vertex $v_{|\mathbf{y}|,|\mathbf{g}|}$, which corresponds to the event in which the output corresponding to $\mathbf{g}$ is $\mathbf{y}$, which is always the case. The total set of vertices is:

$$
\begin{aligned}
&\left\{ v_{i,j-1,s} : \begin{array}{l} i \in \{0, \ldots, |\mathbf{y}|\}, \\ j \in \{1, \ldots, |\mathbf{g}|\} \text{ is an index in a block} \\ s \in \mathcal{S} \end{array} \right\} \\
&\quad \bigcup \left\{ v_{i,j-1} : \begin{array}{l} i \in \{0, \ldots, |\mathbf{y}|\}, \\ j \in \{1, \ldots, |\mathbf{g}|\} \text{ is an index in a GB} \end{array} \right\} \\
&\qquad \bigcup \left\{ v_{i,|\mathbf{g}|} : i \in \{0, \ldots, |\mathbf{y}|\} \right\} .
\end{aligned}
$$

We set weights for the vertices in the first and last columns of the trellis as in [4]. That is, for $v_{0,0,s}$, where $s \in \mathcal{S}$, we set $\mathrm{weight}(v_{0,0,s}) = \pi(s)$, the stationary probability of state s in the input process. All other vertices in the first column have $\mathrm{weight}(v_{i,0,s}) = 0$. By this we effectively force all paths to start at a vertex $v_{0,0,s}$, and incorporate the probability of starting the path at the state s. For the last column we set $\mathrm{weight}(v_{|\mathbf{y}|,|\mathbf{g}|}) = 1$ and $\mathrm{weight}(v_{i,|\mathbf{g}|}) = 0$ for $i \in \{0, 1, \ldots, |\mathbf{y}| - 1\}$. By this we effectively force all paths to end at the vertex $v_{|\mathbf{y}|,|\mathbf{g}|}$. That is, at the end of a path, $|\mathbf{g}|$ symbols have been transmitted, and of these, $|\mathbf{y}|$ have been received.

2) Edges: We briefly explain how the edges are set in the trellis $\mathcal{T}$. Each edge e connects a vertex in some column $j-1$ to a vertex in column j, and corresponds to the input bit g_j ($j \in \{1, \ldots, |\mathbf{g}|\}$). Hence, the edge label is g_j. We refer to the edges from column $j-1$ to column j as the edges in trellis section j. Notice that if e is an outgoing edge of some vertex $v_{i,j-1,s}$ and is an incoming edge of some other vertex $v_{i',j,s'}$, it corresponds to the transition of states in the input process from $S_{j-1} = s$ to $S_j = s'$.

- A *diagonal* edge connects a vertex in row $i-1$ to a vertex in row i, and corresponds to output bit y_i ($i \in \{1, \ldots, |\mathbf{y}|\}$). This edge corresponds to the event where the bit y_i is the result of g_j passing through the deletion channel and not being deleted. As in [4], the edge label is the value y_i.
- A *horizontal* edge connects a vertex in row $i-1$ to a vertex in row $i-1$. This edge corresponds to the event where g_j was deleted. If g_j originates from $\mathbf{x}$, there are two parallel horizontal edges for the two possible cases, one with the label '0' for $g_j = 0$ and one with the label '1' for $g_j = 1$. Conversely, if g_j is a GB bit, there is a single horizontal edge with label '0'.

Notice that if g_j is a GB bit, all the edges in section j (from column $j-1$ to column j) are labeled '0' (no '1' edges), since $g_j = 0$.

The edge weights in the initial trellis $\mathcal{T}$ are the conditional probabilities of the events they correspond to. For example, the edge

$$
v_{i-1,j-1,s} \xrightarrow{e} v_{i,j,s'}
$$

connects the vertex in row $i-1$, column $j-1$, and layer s to the vertex in row i, column j, and layer s'. This edge corresponds to the "no deletion" event: $g_j = x_j$ is the input to the channel, there was no deletion hence $y_i = g_j = x_j$, and the input process states before and after this happened were $S_{j-1} = s$ and $S_j = s'$. The edge weight is the probability of this event, conditioned on $S_{j-1} = s$:

$$
\mathrm{weight}(e) = (1 - \delta) \cdot \mathbb{P}(X_j = y_i, S_j = s' | S_{j-1} = s) .
$$

Note that the term $\mathbb{P}(X_j = y_i, S_j = s' | S_{j-1} = s)$ is determined by the input process. See Table I for the weights and labels of all the edges in $\mathcal{T}$.

3) Block and GB trellises: Up to this point, in this section we have defined how to build the initial trellis $\mathcal{T}$. We now set up the definitions of a 'block' trellis and a 'guard-band' (GB) trellis which we will use soon to define the polar transforms performed on $\mathcal{T}$. The initial trellis $\mathcal{T}$ that we have described is a concatenation of 'block' trellises and 'guard-band' trellises. A block trellis consists of consecutive sections with edges starting in vertices of the form $v_{i,j-1,s}$. Consequently, section j is part of a 'block' trellis iff j is an index originating from a block in $\mathbf{x}$.

Informally, a GB trellis is a block trellis without the edges labeled '1'. Formally, it consists of consecutive sections with edges starting in vertices of the form $v_{i,j-1}$. Consequently, section j is part of a 'guard-band' trellis iff j is an index originating from a guard-band added by the function g.

The first N_0 sections in $\mathcal{T}$ are the block trellis $\mathcal{T}^{\mathrm{B}-1}$. The following ℓ_{n_0+1} sections are the GB trellis $\mathcal{T}^{\mathrm{GB}-1}$. The next N_0 sections are $\mathcal{T}^{\mathrm{B}-2}$, followed by ℓ_{n_0+2} sections of $\mathcal{T}^{\mathrm{GB}-2}$ and so on (where the lengths of the guard-bands are according to the recursive function g defined in (3)). That is,

$$
\mathcal{T} \triangleq \mathcal{T}^{\mathrm{B}-1} \odot \mathcal{T}^{\mathrm{GB}-1} \odot \mathcal{T}^{\mathrm{B}-2} \odot \mathcal{T}^{\mathrm{GB}-2} \ldots \odot \mathcal{T}^{\mathrm{GB}-(N_1-1)} \odot \mathcal{T}^{\mathrm{B}-N_1} .
$$

See an example of $\mathcal{T}$ in Panel (a) in Figure 3 where $n = n_0 + 2$. See an example of a block trellis in Figure 5, and an example of a GB trellis in Figure 6. In the latter two figures, the code length is $|\mathbf{g}| = 11$ bits. Specifically, $N_0 = 2^{n_0} = 4$ and $n = n_0 + 1 = 3$. That is, only one GB is added to $\mathbf{x}$ (see also Panel (a) in Figure 7) and it is of length $\ell_3 = 3$ bits. In this example (that is, in Figures 5–7) three bits were deleted in the channel, and the received vector is $\mathbf{y} = 01100010$ ($|\mathbf{y}| = 8$). Lastly we note that in this example the input distribution was memoryless, $|\mathcal{S}| = 1$, hence only a single layer is depicted in Figure 5.

4) Trellis polar transforms: We now briefly describe the polar transforms performed on $\mathcal{T}$. Our trellis consists of both block trellises and GB trellises, whereas in [4] there were only block trellises in a "trimmed" version (contrast Figure 3 with

TABLE I
EDGE WEIGHTS AND LABELS IN TRELLIS $\mathcal{T}$.

	horizontal edge (deletion) $i' = i$	diagonal edge (non-deletion) $i' = i + 1$	
$v_{i,j-1,s} \xrightarrow{e} v_{i',j,s'}$ (g_j is a bit from $\mathbf{x}$)	$x \triangleq \text{label}(e) \in \{0,1\}$ (two parallel edges), $\text{weight}(e) = \delta \cdot \mathbb{P}(X_J = x, S_J = s' \mid S_{J-1} = s)$	$\text{label}(e) = y_{i'},$ $\text{weight}(e) = (1-\delta) \cdot \mathbb{P}(X_J = y_{i'}, S_J = s' \mid S_{J-1} = s)$	
$v_{i,j-1} \xrightarrow{e} v_{i',j}$ (g_j is a bit from a GB)	$\text{label}(e) = 0,$ $\text{weight}(e) = \delta$	if $y_{i'} = 0$: $\text{label}(e) = y_{i'} = 0,$ $\text{weight}(e) = 1 - \delta$	else: no edge
$v_{i,j-1} \xrightarrow{e} v_{i',j,s'}$ (g_j is a last bit in a GB)	$\text{label}(e) = 0,$ $\text{weight}(e) = \delta \cdot \pi(s')$	if $y_{i'} = 0$: $\text{label}(e) = y_{i'} = 0,$ $\text{weight}(e) = (1-\delta) \cdot \pi(s')$	else: no edge
$v_{i,j-1,s} \xrightarrow{e} v_{i',j}$ (g_j is a last bit in a block of $\mathbf{x}$)	$x \triangleq \text{label}(e) \in \{0,1\}$ (two parallel edges), $\text{weight}(e) = \delta \cdot \mathbb{P}(X_J = x \mid S_{J-1} = s)$	$\text{label}(e) = y_{i'},$ $\text{weight}(e) = (1-\delta) \cdot \mathbb{P}(X_J = y_{i'} \mid S_{J-1} = s)$	

The first row regards edges in the sections of the trellis that correspond to the blocks of $\mathbf{x}$. The weights and labels of the edges in these sections are as defined in [4]. The second row regards edges in GB sections of the trellis. The third and forth rows regard the edges in the "stitches" between blocks and GBs. The third row describes edges in the last section of each GB. We denoted by $\pi(s)$ the probability that the regular hidden-Markov process starts at state $s \in \mathcal{S}$. We incorporated $\pi(s)$ in the edge weights this way since each block is independently drawn (a new initial state for each block). The fourth row describes the edges in the last section of each block. Here we disregard the state S_j, since it is of no consequence. That is, the GB bits are always '0'; they are *not* drawn from the hidden Markov input distribution. Thus, the weights in the fourth row are equal to the marginalization of the corresponding weights in the first row over all s'.

Figure 4). Hence, we need to define how to process these GB trellises along the polarization transforms.

A key operation which we will define shortly is the "merge" operation, which collapses two sections into one. It will serve two purposes: collapsing a GB trellis into a single section (getting from the left part of Figure 6 to the right part) and merging this section into the block section to its left (getting from Panel (a) in Figure 7 to Panel (b)).

A preliminary step of our decoding algorithm is to collapse each GB trellis into a single section. For each guard-band trellis this is performed as follows, see Figure 6. Recall that merging two sections together results in one section. We repeatedly merge the rightmost section in the GB trellis with the GB section to its left, until only one section remains. That is, first merge the rightmost section with its neighbor to the left, then merge the section created with the next neighbor to the left and so on, up to the leftmost section.

We now define the merge operation. A merge operation is always performed on a pair of consecutive sections, where the right section is from a GB trellis (has only '0' edges). It combines all the possible two-edge paths connecting two vertices,

$$\alpha \xrightarrow{e_1} \beta \xrightarrow{e_2} \gamma, \tag{5}$$

where all e_1 edges have the same label and all e_2 edges are labeled '0', into one single edge

$$\alpha \xrightarrow{\tilde{e}} \gamma.$$

We remind that edges always connect a vertex in some column to a vertex in the following column, hence α is a vertex in the leftmost column in the sections merged, β is a vertex in the middle column, and γ is in the rightmost column. The label and weight of the resulting edge are

$$\text{label}(\tilde{e}) = \text{label}(e_1)$$
$$\text{and weight}(\tilde{e}) = \sum_{e_1,e_2} \text{weight}(e_1) \cdot \text{weight}(e_2).$$

We note that since the right section merged is always from a GB trellis with only '0' labels, setting the label of $\tilde{e}$ to $\text{label}(e_1)$ is the same as setting it to: $\text{label}(e_1) \text{ xor } \text{label}(e_2)$. Thus, the new edges created by the merge operation are created just as in the minus transform defined in [4, Definition 5].

The first n_0 '−' and '+' transforms on $\mathcal{T}$ are as defined in [4, Definitions 5,6], and are performed on the block trellises. We refer to this as the "without GB" phase. After this phase, each block trellis has collapsed into one section, see Figure 5. In the following $n_1 \triangleq n - n_0$ transforms, which we refer to as "with GB" transforms, we perform the following operation on each 'sub-trellis'. Each sub-trellis is denoted by $s\mathcal{T}$ and consists of three consecutive sections. Specifically, we define

$$s\mathcal{T} = \mathcal{T}^{\text{B}-j} \odot \mathcal{T}^{\text{GB}-j} \odot \mathcal{T}^{\text{B}-(j+1)}, \quad j \in \{1, ..., 2^{n-\lambda}\} \text{ is odd}, \tag{6}$$

and $\lambda \geq n_0$ is the polarization depth we are currently at (and wish to enlarge by 1). We remind that after the first n_0 transforms and after merging the GB paths, $\mathcal{T}^{\text{B}-(j+1)}, \mathcal{T}^{\text{GB}-j}$, and $\mathcal{T}^{\text{B}-(j+1)}$ each consist of one single section. That is, each $s\mathcal{T}$ includes a GB section between two non-guard-band sections, see Panel (a) in Figure 7. Thus, we first merge the left section, $\mathcal{T}^{\text{B}-j}$, and the GB section, $\mathcal{T}^{\text{GB}-j}$. This results in two-sections with no GB sections (Panel (b) in Figure 7). We may now perform the '−' or '+' transform as in the "without GB" phase. See Figure 3 for an illustration.

In the resulting trellis, we think of the operation described above on each $s\mathcal{T}$ as producing a block trellis (even though $s\mathcal{T}$ consisted of a middle section that is the result of merging a GB trellis). Since we have not changed the even index GB trellises, we have produced a trellis with a structure as in (6), with $\lambda+1$ in place of λ. Hence, we can continue the recursion.

D. Decoding complexity

In this section we show that the decoding complexity is $O(|\mathcal{S}|^3 \Lambda^4)$. This will follow by first recalling that in [4,
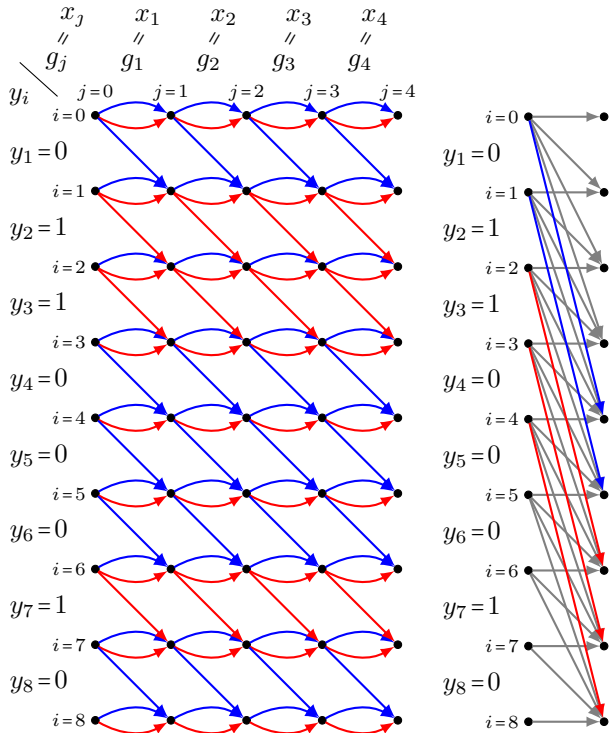
Fig. 5. Left: The first block trellis $\mathcal{T}^{\mathrm{B}-1}$. Right: The same block trellis after $n_0 = 2$ minus polar transforms. Red edges are labeled '1' and blue edges are labeled '0'. To improve legibility, a gray edge represents two parallel edges, red and blue (the two edges do not have the same weight necessarily). In this example $n_0 = 2$, thus there are four sections in each block trellis.

Fig. 6. Left: The GB trellis $\mathcal{T}^{\mathrm{GB}-1}$. Right: The same GB trellis after merging all sections. Blue edges are labeled '0' (in GB trellises all edges are labeled '0'). In this example $n = 3$, $n_0 = 2$ and $\ell_3 = 3$, hence $\mathcal{T}^{\mathrm{GB}-1}$ is the only GB trellis in $\mathcal{T}$ and has 3 sections.

Subsection IV-D], the complexity of decoding a trellis with Λ sections and no GBs is $O(|\mathcal{S}|^3\Lambda^4)$, and then showing that the processing of the GBs does not change the order of the decoding complexity. In [4] this complexity can be further reduced to roughly $O(|\mathcal{S}|^3\Lambda^2)$. On the upside, we do not require a successful partitioning of $\mathbf{y}$ to the TDC outputs of each block $\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(N_1)$, whereas in [4] it is crucial for correct decoding. Thus, and as proven herein, we reach a better probability of correct decoding.

Regarding the details, we will first show that the complexity of collapsing each of the GB trellises into one section is $\mathcal{O}(\Lambda^3)$. Then, we will analyze the complexity of performing the polar transforms on the trellis (both "with GBs" and "without GBs") and show that it is $\mathcal{O}(|\mathcal{S}|^3\Lambda^4)$.

During the process of collapsing the largest GB trellis in $\mathcal{T}$, which consists of ℓ_n sections, we also get the results of collapsing each of the smaller GB trellises. In detail, recall that in a GB trellis with ℓ sections, the first $\ell - 1$ sections (from the left) are identical and created according to the second row in Table I. The right-most section in each GB trellis is created according to the third row in Table I. Observe that in the second and third rows of Table I the section index j does not play a part. Hence, any two suffixes of equal length corresponding to different GB trellises are the same. When collapsing the largest GB trellis, we merge the right-most section with the one to its left, then with the next section to its left and so on, until ℓ_n sections were merged. Hence, we can pause the merging iterations after
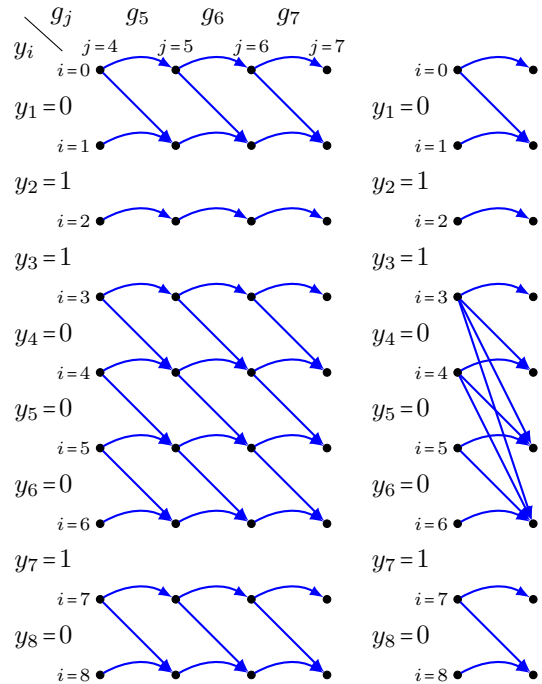
$\ell_{n_0+1}, \ell_{n_0+2}, \ldots, \ell_{n-1}$ sections were merged and extract the rightmost section as the result of collapsing GB trellises which consist of $\ell_{n_0+1}, \ell_{n_0+2}, \ldots, \ell_{n-1}$ sections, respectively.

We bound the number of calculations when collapsing a GB trellis with ℓ_n sections by

$$\sum_{m=1}^{\ell_n-1} (\Lambda + 1) \cdot 2 \cdot (m+1) \in \mathcal{O}(\Lambda^3) \,,$$

where m is the iteration of the merge operations. In each merge we sum over all possible two-edge paths. The paths may start at $|\mathbf{y}| + 1$ possible starting vertices (in each row of the trellis) which is bounded by $\Lambda + 1$. The number of outgoing edges of the starting vertex is at most 2 (diagonal and horizontal edges). The number of outgoing edges in the middle vertex of a path is at most $m + 1$ (since these are edges of a section created after performing $m - 1$ merges).

We now bound the complexity of the remaining decoding steps, after the GB trellises are collapsed into single sections. Let λ denote the polarization depth as before, i.e. the number of '+' or '−' transforms performed on $\mathcal{T}$ (see Figure 2). We will now show that for a fixed regular hidden Markov input distribution with $|\mathcal{S}|$ states, the complexity of our decoder is bounded by

$$\overbrace{\sum_{\lambda=0}^{n_0-1} \underbrace{2^{\lambda+1}}_{(a)} \cdot \underbrace{4(\Lambda+1)|\mathcal{S}|^3(2^\lambda+1)^2}_{(b1)} \cdot \underbrace{2^{n-\lambda-1}}_{(c1)}}^{\text{"without GB" phase}} +$$
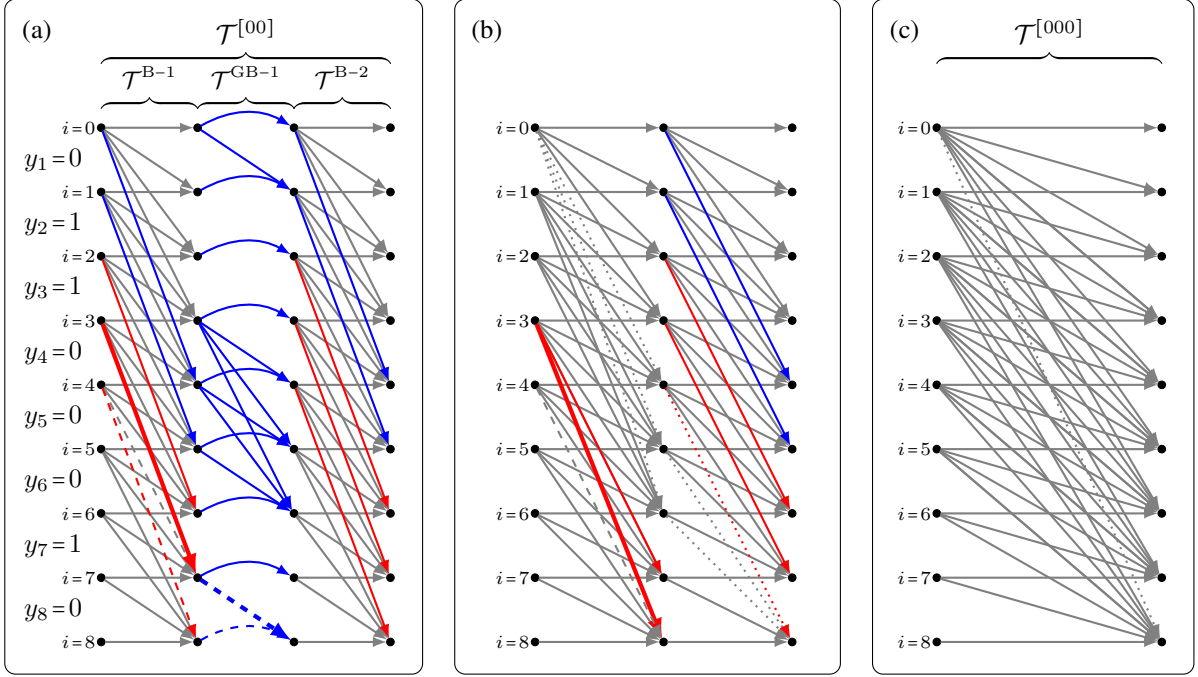
Fig. 7. Panel (a) shows the trellis $\mathcal{T}^{[00]}$. That is, $\mathcal{T}$ after $n_0 = 2$ minus transforms and after merging all the GB paths in the GB trellis, $\mathcal{T}^{\mathrm{GB}-1}$. The two block trellises $\mathcal{T}^{\mathrm{B}-1}$ and $\mathcal{T}^{\mathrm{B}-2}$ are identical since they are identical in the initial trellis $\mathcal{T}$, and we performed only minus transforms. This is not the case for other transforms (for example in the other trellises $\mathcal{T}^{[01]}, \mathcal{T}^{[10]}$ and $\mathcal{T}^{[11]}$ the block trellises are not generally identical). The next polar transform, performed on $\mathcal{T}^{[00]}$, will be a "with GB" transform. That is, first merge the paths of the left block trellis $\mathcal{T}^{\mathrm{B}-1}$ with the GB trellis $\mathcal{T}^{\mathrm{GB}-1}$. The result of this step is shown in Panel (b). Then perform a regular polar transform ("without GB" transform) on the result. After this step we get $\mathcal{T}^{[000]}$ which is shown in Panel (c). In this example $n = 3$, $n_0 = 2$ and $\ell_3 = 3$, hence we get this result for the received vector $\mathbf{y}$. We note that $n_1 \triangleq n - n_0 = 1$ in this example, thus there is only one sub-trellis $s\mathcal{T}$ in $\mathcal{T}^{[00]}$, which is $\mathcal{T}^{[00]}$ itself. As the polarization steps progress, the trellis becomes quite busy and hard to follow. To clarify the tangled scheme, we highlighted some edges as examples. The bold red edge in Panel (b) was created from the bold edges in Panel (a). As shown, only one path exists between row $i = 3$ and $i = 8$ in the block trellis and GB trellis we merge. This edge takes the label of the block trellis edge in the path (which is labeled '1'). The dashed gray arrow in Panel (b), which represents two parallel edges with different labels and likely different weights, was created from the dashed edges in Panel (a). As shown in Panel (a), there exists a path from row $i = 4$ to $i = 8$ which gives the label '0' (the gray arrow from $i = 4$ to $i = 7$ also represents an edge with label '0'). There also exist two paths that give the label '1' (one passes through $i = 7$ along the same gray arrow and one goes straight from $i = 4$ to $i = 8$). The dotted arrow in Panel (c) was created from the dotted paths in Panel (b). This arrow represents two parallel edges: blue and red. The decision on $\hat{u}_1$, if it is not frozen, is according to the higher weighted edge of the two (a higher weight indicates a higher likelihood).

$$\sum_{\lambda=n_0}^{n-1} \underbrace{2^{\lambda+1}}_{(a)} \cdot \underbrace{8(\Lambda+1)|\mathcal{S}|^3(C \cdot 2^{\lambda+1}+1)^2}_{(b2)} \cdot \underbrace{2^{n-\lambda-1}}_{(c2)}$$

$$\in O(|\mathcal{S}|^3 \Lambda^4) .$$

(a) is the number of times we perform a polar transform on trellises of depth λ, see an illustration in Figure 2.

(b1) bounds the number of calculations performed on each pair of sections in a trellis of depth $\lambda < n_0$.

(c1) is the number of pairs of consecutive sections (first odd then even indexed) in a trellis of depth $\lambda < n_0$.

(b2) bounds the number of calculations on each sub-trellis of a given trellis of depth $\lambda \geq n_0$. The constant C will soon be defined.

(c2) is the number of sub-trellises in a trellis of depth λ, see (6).

Before proving (b1) and (b2), we remind the reader that we sum over all possible two-edge paths in pairs of consecutive sections. The proof will follow from bounding the number of two-edge paths in each pair of sections in a trellis of polarization depth λ.

We first show (b1). In (b1) we have $\lambda < n_0$, i.e. we only

work on block trellis sections after λ "without GB" transforms. Each section in a block trellis of depth $\lambda \leq n_0$ was created from 2^λ sections in the base trellis $\mathcal{T}$. Thus, in depth λ, we have $2^\lambda + 1$ possible 'types' of edges, where the edge type refers to how many rows the edge crosses. For example, in the initial trellis $\lambda = 0$ there are two types: horizontal edges and diagonal edges that cross one row. In a block trellis after $\lambda = 2$ "without GB" transforms we have five types: horizontal edges and diagonal edges which cross one, two, three or four rows (see Figure 5).

Each edge can be labeled either '0' or '1', and edges of a given type can enter at most $|\mathcal{S}|$ vertices. Thus, we get at most $2 \cdot (2^\lambda + 1) \cdot |\mathcal{S}|$ outgoing edges for each vertex. Hence, there are at most $(2 \cdot (2^\lambda + 1)|\mathcal{S}|)^2$ two-edge paths starting from each vertex.

There are $|\mathbf{y}| + 1$ starting vertex rows, which is upper bounded by $\Lambda + 1$, and each row is at most $|\mathcal{S}|$ layers deep. In total, the number of calculations performed when collapsing two sections into one for $\lambda < n_0$ is bounded by (b1), that is

$$4(\Lambda + 1)|\mathcal{S}|^3(2^\lambda + 1)^2 .$$

For (b2), we first bound the number of edge types in each sub-trellis. As before, the number of possible edge types in

a section of depth λ is equal to the number of sections in $\mathcal{T}$ that produced it plus 1. For depth λ:

- The block sections were each created from Λ_λ sections in the base trellis $\mathcal{T}$, where Λ_λ is the length of $g(\mathbf{x})$ and $\mathbf{x}$ is of length 2^λ. Thus, we have $\Lambda_\lambda + 1$ types in each block section.
- The middle GB section in each sub-trellis $s\mathcal{T}$ was created from $\ell_{\lambda+1}$ sections in the base trellis (which we collapsed into a single section as described previously), see (3) and (6). Thus, we have at most $\ell_{\lambda+1} + 1$ edge types in the GB section of each sub-trellis. For example, see Panel (a) in Figure 7 which shows a sub-trellis of depth $\lambda = 2$. In this figure the middle GB trellis of $s\mathcal{T}$ was created from $\ell_3 = 3$ GB sections which collapsed into one, and there are four edge types (horizontal edges and diagonal edges that cross 1-3 rows). Since $\ell_{\lambda+1} \overset{(4)}{<} 2^\lambda \le \Lambda_\lambda$, we can bound the possible types of edges in the GB section by $\Lambda_\lambda + 1$ (which is the bound of types in the block sections).

We have shown that the number of edge types in each sub-trellis of depth λ is at most $\Lambda_\lambda + 1$. By the same logic used for (b1), in each sub-trellis there are now at most

$$(2 \cdot (\Lambda_\lambda + 1)|\mathcal{S}|)^2$$

two-edge paths from each starting vertex. Thus, using the same arguments as the (b1) case, the number of steps for merging the GB trellises into the block trellises ("step 1" in Panel (c), Figure 3, and transitioning from Panel (a) to Panel (b) in Figure 7) is at most

$$4(\Lambda + 1)|\mathcal{S}|^3(\Lambda_\lambda + 1)^2 .$$

We bound the number of edge types in a sub-trellis after the merge by $\Lambda_{\lambda+1} + 1$, since the resulting block section was created from $\Lambda_\lambda + \ell_{\lambda+1} < \Lambda_\lambda + \ell_{\lambda+1} + \Lambda_\lambda = \Lambda_{\lambda+1}$ sections of $\mathcal{T}$. Now recall that we must also apply a '$-$' or '$+$' transform to the resulting trellis ("step 2" in Panel (c), Figure 3, and transitioning from Panel (b) to Panel (c) in Figure 7). By the same logic as before, this requires at most

$$4(\Lambda + 1)|\mathcal{S}|^3(\Lambda_{\lambda+1} + 1)^2$$

steps. Thus, since $\Lambda_\lambda \le \Lambda_{\lambda+1}$, we may bound the total number of steps by

$$8(\Lambda + 1)|\mathcal{S}|^3(\Lambda_{\lambda+1} + 1)^2 .$$

Lastly, we loosely bound Λ_λ for all $n_0 \ge 1$ and $\xi > 0$ by $\Lambda_\lambda \overset{(4a)}{<} C \cdot 2^\lambda$, where $C \triangleq \frac{1}{1-2^{-\xi}}$ is a constant, which gives (b2).

E. Simulation results

Figure 8 shows simulation results comparing the performance of two decoders: the one of prior art [4] and the one described and analyzed in this paper. Our decoder outperforms the previous one in all the scenarios tested. This is expected, since we decode based on the trellis of the entire output as opposed to processing the output into blocks, and then using trellises for each block.

It became apparent after establishing numerical results that parameters in short code lengths need to be optimized. For the asymptotic case, when $N \to \infty$, we will detail in the proof of Theorem 1 how to select n_0 and ξ, the parameters controlling the number of GBs and their lengths, and how to select a large enough n to reach the desired code rate and block error probability. However, when working with shorter lengths, we have no simple analytical way of choosing optimal values for these parameters.

Consider the choice of n_0 for some finite n. This paper shows that in an asymptotic setting ($n \to \infty$), the parameter n_0 should be "large enough", for two reasons. First, if n_0 is small the rate is negatively affected: Lemma 4 promises that adding guard-bands has a negligible effect on the rate, if n_0 is large enough. Secondly, a small probability of the complement of the 'GBM' event is key to proving Lemma 5 below, and this probability goes to 0 as n_0 increases (Lemma 6). On the other hand, for a fixed n, we cannot take n_0 too large, since Lemma 7 below only guarantees a strong polarization phase when n is sufficiently far from n_0. That is, the inequality on Z_n in (15) is only guaranteed to hold for $n \ge n_0$. The above trade-off is demonstrated in Figure 9. For example, in the left panel ($N = 128$) transitioning from $n_0 = 3$ to $n_0 = 4$ demonstrates improvement, whereas continuing to $n_0 = 5$ shows a decline.

V. First key lemma — the Bhattacharyya evolution for the trimmed-deletion channel

The proof of our main theorem hinges on two key lemmas. The first lemma is specific to our setting, the trimmed-deletion channel, and is stated and proved in this section. The second lemma is more general, and will be stated and proved in Section VI. These two lemmas will be used in our proof of Theorem 1, in Section VII.

In the seminal paper [3, Proposition 5], it was shown that in a memoryless setting, a '$+$' transform squares the Bhattacharyya parameter, while a '$-$' transform at most doubles it. This was the key property used to prove strong polarization in [18]. Our first key lemma states a similar — yet weaker — claim for our setting. The second key lemma shows that such a claim is still strong enough to prove strong polarization with the same exponent as in the memoryless case, $\beta \in \left(0, \frac{1}{2}\right)$.

We first set up some additional notation. We denote the Arıkan transform of the vector $\mathbf{X}$ by $\mathbf{U} = \mathcal{A}(\mathbf{X})$. Recall that the two halves of $\mathbf{X}$ are $\mathbf{X}_\mathrm{I}$ and $\mathbf{X}_\mathrm{II}$. Their Arıkan transforms are denoted $\mathbf{V} \triangleq \mathcal{A}(\mathbf{X}_\mathrm{I})$ and $\mathbf{V}' \triangleq \mathcal{A}(\mathbf{X}_\mathrm{II})$, and we have

$$U_{2j-1} = V_j + V'_j , \quad U_{2j} = V'_j ,$$

where addition is modulo 2. As in the seminal paper, the binary vector corresponding to $i-1$ is denoted $b_1, \ldots, b_n$. That is, for $1 \le i \le N = 2^n$,

$$i = i(b_1, \ldots, b_n) = 1 + \sum_{k=1}^{n} b_k 2^{n-k}. \tag{7}$$

The following lemma is cardinal to proving the stronger polarization stated in Theorem 1. The proof will be given in the following subsections. Recall that δ is the deletion rate, and that the guard-band length is given in (4), and is a function of ξ.
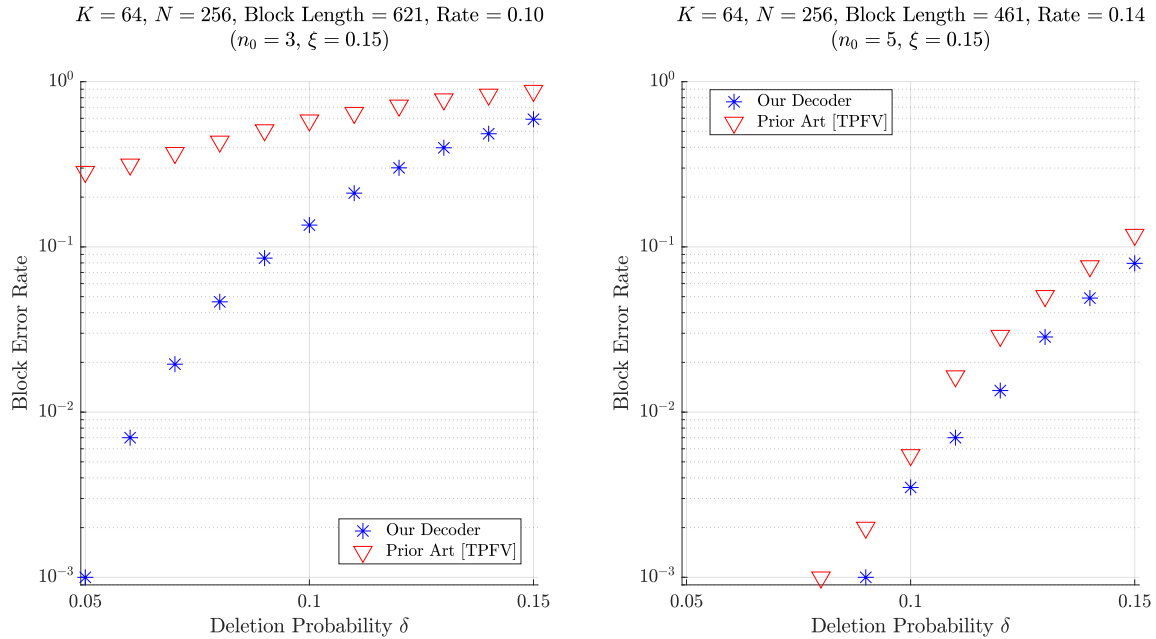
Fig. 8. The block error rate of the decoder in this work compared to the prior-art decoder [4]. The input distribution was fixed to a memoryless uniform distribution, Bernoulli($\frac{1}{2}$). The block-error rate was averaged over 2000 simulations. In the comparison above between our decoder and the one of [4], the same encoder was used and the channel deletions were the same. That is, the decoders in this figure were fed with the same channel output $\mathbf{y}$ and the only difference was the decoding algorithm applied to it. The information indices were selected as the $K = 64$ indices with the lowest error probability in a genie decoder. The genie decoder error probability for each index was estimated via Monte Carlo using 2000 test blocks with deletion probability $\delta = 0.1$.

Lemma 5 (Recursive bounds for the Bhattacharyya parameter of the TDC): Fix a regular and non-degenerate hidden-Markov input distribution. Let $\mathbf{X} = \mathcal{A}(\mathbf{U})$ be of length $N = 2^n$, comprised of i.i.d. blocks of length $N_0 = 2^{n_0}$, each distributed according to the input distribution. Let $\xi \in (0, \frac{1}{6})$ be a fixed parameter and let $\mathbf{Y}^* = \mathbf{Z}_{\mathrm{I}} \odot \mathbf{Z}_{\Delta} \odot \mathbf{Z}_{\mathrm{II}}$ be the result of transmitting $g(\mathbf{X}, n_0, \xi)$ through the TDC. Then, there exists $n_0^{\mathrm{th}}(\xi, \delta)$ s.t. for $n_0 \geq n_0^{\mathrm{th}}$ and all $n > n_0$ the following holds. Let $1 \leq i \leq N$ and $j = \lfloor (i+1)/2 \rfloor$. Then,

$$Z(U_i | U_1^{i-1}, \mathbf{Y}^*) \leq Z(U_i | U_1^{i-1}, \mathbf{Z}_{\mathrm{I}}^*, \mathbf{Z}_{\mathrm{II}}^*) + 2^{-N^{\frac{2}{3}}} \quad (8a)$$

$$\leq \begin{cases} 2 \cdot Z(V_j | V_1^{j-1}, \mathbf{Z}_{\mathrm{I}}^*) + 2^{-N^{\frac{2}{3}}} & \text{if } b_n = 0 \ (\text{`--'}) \\ Z(V_j | V_1^{j-1}, \mathbf{Z}_{\mathrm{I}}^*)^2 + 2^{-N^{\frac{2}{3}}} & \text{if } b_n = 1 \ (\text{`+'}). \end{cases} \quad (8b)$$

We draw the reader's attention to several important points. First, note that in (8a), there is an additive penalty of $2^{-N^{\frac{2}{3}}}$, associated with conditioning on $\mathbf{Z}_{\mathrm{I}}^*, \mathbf{Z}_{\mathrm{II}}^*$ as opposed to conditioning on $\mathbf{Y}^*$. That is, there is a price to be paid for conditioning on the pair of TDC outputs corresponding to $g(\mathbf{X}_{\mathrm{I}})$ and $g(\mathbf{X}_{\mathrm{II}})$, as opposed to conditioning on the TDC output corresponding to $g(\mathbf{X}_{\mathrm{I}} \odot \mathbf{X}_{\mathrm{II}})$. Informally, this is because in the former we have been given the correct partitioning of the output into two halves (that are then further processed by the TC). The inequality (8b) shows us why such a penalty is worth paying: since $\mathbf{Z}_{\mathrm{I}}^*$ and $\mathbf{Z}_{\mathrm{II}}^*$ are independent and identically distributed, we may now use the standard arguments in [3] to reach a recursive relation. This relation is indeed weaker than the one of the seminal paper for memoryless channels [3, Proposition 5] in which the additive penalty of $2^{-N^{\frac{2}{3}}}$ does not

exist. To conclude, the lemma allows us to track the evolution of the Bhattacharyya parameter after each polarization step.

The proof of Lemma 5 will be broken into three conceptual parts. In the first part, we define the "Guard-Band in Middle" event, termed GBM. That is, the event that the middle symbol of the TDC output originated from the middle guard-band. In the second part, we show that under GBM, $\mathbf{Y}^*$ can be processed to yield $\mathbf{Z}_{\mathrm{I}}^*$ and $\mathbf{Z}_{\mathrm{II}}^*$. This essentially gives us the first term on the RHS of (8a). In the third part, we show that the GBM event is very likely. That is, the additive penalty of $2^{-N^{\frac{2}{3}}}$ in (8a) comes from bounding the probability that GBM does not occur.

We note that in [19] we have shown a weaker bound for the Bhattacharyya parameter of this setting. A multiplicative penalty of $\frac{3N}{2}$ existed in the upper bound in (8). That is,

$$Z(U_i | U_1^{i-1}, \mathbf{Y}^*) \leq \frac{3N}{2} \cdot Z(U_i | U_1^{i-1}, \mathbf{Z}_{\mathrm{I}}^*, \mathbf{Z}_{\mathrm{II}}^*) + 2^{-N^{\frac{2}{3}}}.$$

We will now be able to avoid this harsh multiplicative penalty by employing the effect of channel degradation [20, page 207], [21, Lemma 3], which will be demonstrated in the upcoming proof.

A. The GBM event

In this subsection we define the "Guard-band in Middle" (GBM) event, related notation, and consequences. Recall from Section III and Figure 1 that $\mathbf{Y}^* = \mathbf{Z} = \mathbf{Z}_{\mathrm{I}} \odot \mathbf{Z}_{\Delta} \odot \mathbf{Z}_{\mathrm{II}}$ is the

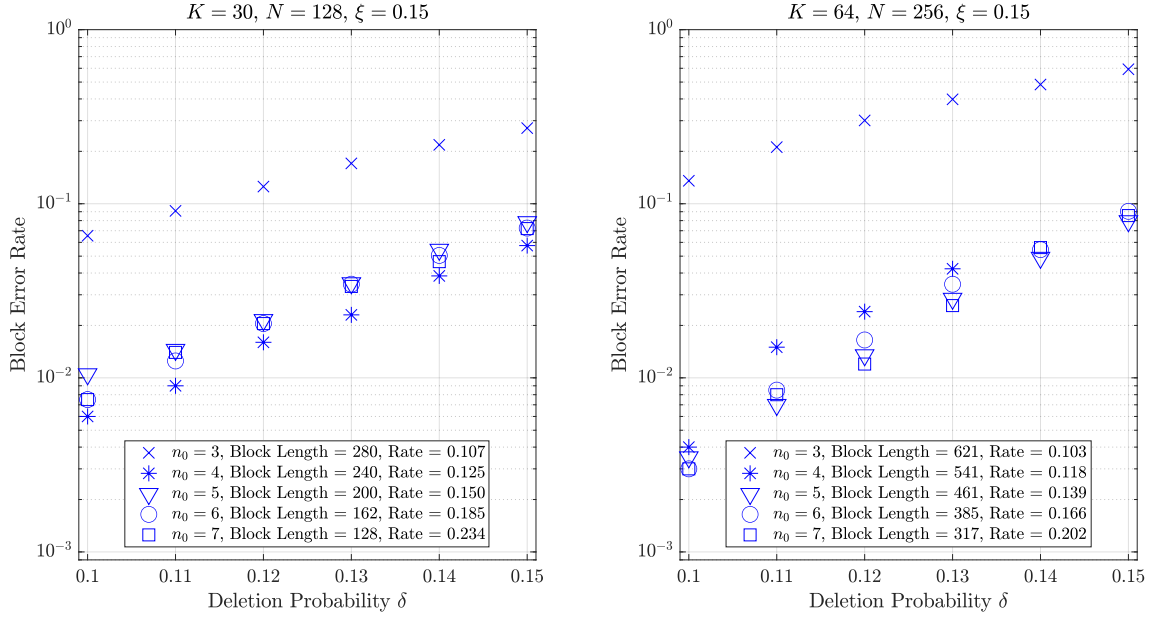$K = 30, N = 128, \xi = 0.15$

$K = 64, N = 256, \xi = 0.15$

Fig. 9. The block error rate of the decoder in this work over various selections of n and n_0. As in Figure 8, the input distribution was fixed to a memoryless uniform distribution, Bernoulli($\frac{1}{2}$). The block-error rate was averaged over 2000 simulations. The information indices were selected as the K indices with the lowest error probability in a genie decoder. The genie decoder error probability for each index was estimated via Monte Carlo using 2000 test blocks with deletion probability $\delta = 0.1$.

result of passing $\mathbf{G}_{\text{I}} \odot \mathbf{G}_{\Delta} \odot \mathbf{G}_{\text{II}}$ through the TDC. The GBM event occurs if $\mathbf{Z}$ is not empty and its middle index

$$i_{\text{mid}} \triangleq \left\lfloor \frac{|\mathbf{Z}| + 1}{2} \right\rfloor, \qquad (9)$$

falls within $\mathbf{Z}_{\Delta}$. That is, GBM occurs iff $Z_{i_{\text{mid}}}$ originates from the middle guard-band $\mathbf{G}_{\Delta}$, see Figure 10. The complementary event is denoted ¬GBM, see Figure 11.

We denote the left and right halves of $\mathbf{Z} = \mathbf{Y}^*$ as $\mathbf{Z}_{\text{L}} = (Z_1, \ldots, Z_{i_{\text{mid}}})$ and $\mathbf{Z}_{\text{R}} = (Z_{i_{\text{mid}}+1}, \ldots, Z_{|\mathbf{Z}|})$, see Figure 10. The main utility of the GBM event is this (again, see Figure 10): since $\mathbf{Z}_{\Delta}$ contains only '0' symbols, under GBM

$$(\mathbf{Z}_{\text{L}})^* = (\mathbf{Z}_{\text{I}})^* , \qquad (10a)$$

$$(\mathbf{Z}_{\text{R}})^* = (\mathbf{Z}_{\text{II}})^* . \qquad (10b)$$

That is, under GBM, the simple operation of trimming the two halves of $\mathbf{Y}^*$ is assured to give us $\mathbf{Z}_{\text{I}}^* \triangleq (\mathbf{Z}_{\text{I}})^*$ and $\mathbf{Z}_{\text{II}}^* \triangleq (\mathbf{Z}_{\text{II}})^*$. This simple observation will be used in the next subsection in order to state a recursive relation.

B. Bounding the Bhattacharyya parameter using GBM

In this subsection, we derive an upper bound on the Bhattacharyya parameter $Z(U_i|U_1^{i-1}, \mathbf{Y}^*)$. In the previous subsection we defined $\mathbf{Z}_{\text{L}}^*$ and $\mathbf{Z}_{\text{R}}^*$ as the left and right halves of $\mathbf{Z} \triangleq \mathbf{Y}^*$ after trimming their trailing and leading zeros, respectively. That is, $\mathbf{Z}_{\text{L}}^*, \mathbf{Z}_{\text{R}}^*$ are the result of some processing of $\mathbf{Y}^*$. This processing is illustrated in Fig 10 and Fig 11. We now consider a channel which is a degradation of the TDC channel. The input of the channel is the same as the case of the TDC, but the outputs of the channel are $\mathbf{Z}_{\text{L}}^*, \mathbf{Z}_{\text{R}}^*$. Clearly
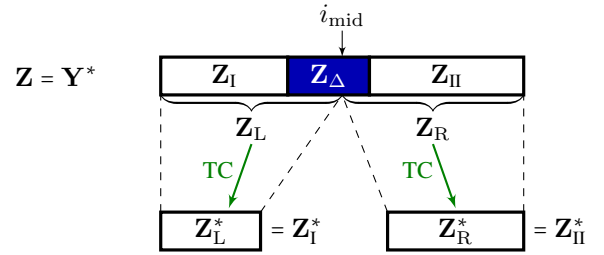
Fig. 10. An illustration of the GBM event.
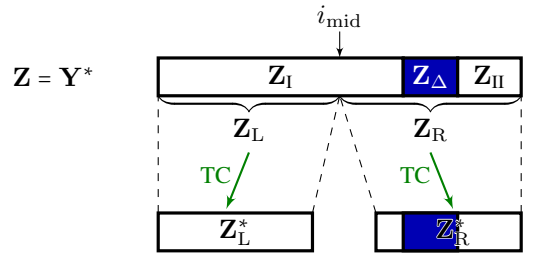
Fig. 11. An illustration of the complementary event, ¬GBM.

this is a degraded channel of the TDC, since $\mathbf{Y}^*$, the output of the TDC is further processed into $\mathbf{Z}_{\text{L}}^*, \mathbf{Z}_{\text{R}}^*$. The Bhattacharyya parameter of the degraded channel is larger or equal to the Bhattacharyya of the TDC [20, page 207] [21, Lemma 3], hence:

$$Z(U_i|U_1^{i-1}, \mathbf{Y}^*) \le Z(U_i|U_1^{i-1}, \mathbf{Z}_{\text{L}}^*, \mathbf{Z}_{\text{R}}^*) . \qquad (11)$$

In order to save space we use the following shorthand

in the upcoming probability expressions: u_1^{i-1} is short for $U_1^{i-1} = u_1^{i-1}$, $\mathbf{z}_L^*$ and $\mathbf{z}_R^*$ are short for $\mathbf{Z}_L^* = \mathbf{z}_L^*$ and $\mathbf{Z}_R^* = \mathbf{z}_R^*$, respectively, and 0 and 1 are short for $U_i = 0$ and $U_i = 1$, respectively. Furthermore, in our shorthand notation, a sum without subscripts is always over $u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*$. To illustrate, we use both the long and short notation in the following expression for our quantity of interest.

$$Z(U_i | U_1^{i-1}, \mathbf{Z}_L^*, \mathbf{Z}_R^*)$$
$$\triangleq \sum_{u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*} \sqrt{\begin{array}{l} \mathbb{P}(U_i = 0, U_1^{i-1} = u_1^{i-1}, \mathbf{Z}_L^* = \mathbf{z}_L^*, \mathbf{Z}_R^* = \mathbf{z}_R^*) \\ \times \mathbb{P}(U_i = 1, U_1^{i-1} = u_1^{i-1}, \mathbf{Z}_L^* = \mathbf{z}_L^*, \mathbf{Z}_R^* = \mathbf{z}_R^*) \end{array}}$$
$$\triangleq \sum \sqrt{\mathbb{P}(0, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*) \cdot \mathbb{P}(1, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*)}$$

By the law of total probability over $\{\text{GBM}, \neg\text{GBM}\}$, the RHS above equals (for $n \geq n_0 + 1$, assuring a guard-band was added):

$$= \sum \sqrt{\begin{array}{l} \left(\mathbb{P}(0, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \text{GBM}) + \mathbb{P}(0, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \neg\text{GBM})\right) \\ \times \left(\mathbb{P}(1, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \text{GBM}) + \mathbb{P}(1, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \neg\text{GBM})\right) \end{array}}$$
$$\leq \sum \sqrt{\mathbb{P}(0, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \text{GBM}) \cdot \mathbb{P}(1, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \text{GBM})} \quad +$$
$$\sum \sqrt{\begin{array}{l} \mathbb{P}(0, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \text{GBM}) \cdot \mathbb{P}(1, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \neg\text{GBM}) \\ + \mathbb{P}(0, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \neg\text{GBM}) \cdot \mathbb{P}(1, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \text{GBM}) \\ + \mathbb{P}(0, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \neg\text{GBM}) \cdot \mathbb{P}(1, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \neg\text{GBM}) \end{array}} \quad (12)$$

We will now bound both sums in (12). For the first sum we have:

$$\sum_{u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*} \sqrt{\begin{array}{l} \mathbb{P}(0, u_1^{i-1}, \mathbf{Z}_L^* = \mathbf{z}_L^*, \mathbf{Z}_R^* = \mathbf{z}_R^*, \text{GBM}) \\ \times \mathbb{P}(1, u_1^{i-1}, \mathbf{Z}_L^* = \mathbf{z}_L^*, \mathbf{Z}_R^* = \mathbf{z}_R^*, \text{GBM}) \end{array}}$$
$$\overset{(10)}{=} \sum_{u_1^{i-1}, \mathbf{z}_I^*, \mathbf{z}_{II}^*} \sqrt{\begin{array}{l} \mathbb{P}(0, u_1^{i-1}, \mathbf{Z}_I^* = \mathbf{z}_I^*, \mathbf{Z}_{II}^* = \mathbf{z}_{II}^*, \text{GBM}) \\ \times \mathbb{P}(1, u_1^{i-1}, \mathbf{Z}_I^* = \mathbf{z}_I^*, \mathbf{Z}_{II}^* = \mathbf{z}_{II}^*, \text{GBM}) \end{array}}$$
$$\leq \sum_{u_1^{i-1}, \mathbf{z}_I^*, \mathbf{z}_{II}^*} \sqrt{\begin{array}{l} \mathbb{P}(0, u_1^{i-1}, \mathbf{Z}_I^* = \mathbf{z}_I^*, \mathbf{Z}_{II}^* = \mathbf{z}_{II}^*) \\ \times \mathbb{P}(1, u_1^{i-1}, \mathbf{Z}_I^* = \mathbf{z}_I^*, \mathbf{Z}_{II}^* = \mathbf{z}_{II}^*) \end{array}}$$
$$= Z(U_i | U_1^{i-1}, \mathbf{Z}_I^*, \mathbf{Z}_{II}^*)$$

For the second sum in (12) we have:

$$\sum \sqrt{\begin{array}{l} \mathbb{P}(0, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \text{GBM}) \cdot \mathbb{P}(1, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \neg\text{GBM}) \\ + \mathbb{P}(0, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \neg\text{GBM}) \cdot \mathbb{P}(1, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \text{GBM}) \\ + \mathbb{P}(0, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \neg\text{GBM}) \cdot \mathbb{P}(1, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \neg\text{GBM}) \end{array}}$$
$$= \sum \sqrt{\begin{array}{l} \mathbb{P}(0, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \text{GBM}) \cdot \mathbb{P}(1, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \neg\text{GBM}) \\ + \mathbb{P}(0, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \neg\text{GBM}) \cdot \mathbb{P}(1, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*) \end{array}}$$
$$\leq \sum \sqrt{\begin{array}{l} \mathbb{P}(u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*) \cdot \mathbb{P}(1, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \neg\text{GBM}) \\ + \mathbb{P}(0, u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \neg\text{GBM}) \cdot \mathbb{P}(u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*) \end{array}}$$
$$= \sum \sqrt{\mathbb{P}(u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*) \cdot \mathbb{P}(u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*, \neg\text{GBM})}$$
$$= \sum \mathbb{P}(u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*) \sqrt{\mathbb{P}(\neg\text{GBM} | u_1^{i-1}, \mathbf{z}_L^*, \mathbf{z}_R^*)}$$
$$\leq \sqrt{\mathbb{P}(\neg\text{GBM})} \,,$$

where last inequality follows by the Jensen inequality, applied to the concave function $\sqrt{\cdot}$.

Combining the bounds for the two sums in (12) yields

$$Z(U_i | U_1^{i-1}, \mathbf{Y}^*) \leq Z(U_i | U_1^{i-1}, \mathbf{Z}_I^*, \mathbf{Z}_{II}^*) + \sqrt{\mathbb{P}(\neg\text{GBM})} \,.$$

To complete the proof of (8a), it remains to show that the term $\sqrt{\mathbb{P}(\neg\text{GBM})}$ is smaller than $2^{-N^{\frac{2}{3}}}$, for large enough n and n_0. This will be shown in the upcoming lemma. Lastly, since $\mathbf{Z}_I^*$ and $\mathbf{Z}_{II}^*$ are i.i.d., the second inequality in our lemma, (8b), is a direct consequence of [3, Proposition 5]. Thus, the following lemma completes our proof.

Lemma 6 (Upper bounding $\sqrt{\mathbb{P}(\neg\text{GBM})}$): Let $\mathbf{X}$ be of length $N = 2^n$ and drawn as described in Lemma 5. Let $\mathbf{Y}^*$ be the TDC output for input $g(\mathbf{X}, n_0, \xi)$. Let GBM be the event defined in Subsection V-A. For a fixed deletion rate $\delta \in (0, 1)$ and a guard-band length parameter $0 < \xi < \frac{1}{6}$ used in (4), there exists an $n_0^{\text{th}}(\xi, \delta)$, which is a function of the input distribution as well, such that:

$$\sqrt{\mathbb{P}(\neg\text{GBM})} \leq 2^{-N^{\frac{2}{3}}} \tag{13}$$

for all $n_0 \geq n_0^{\text{th}}(\xi, \delta)$ and $n > n_0$.

Proving this proves Lemma 5: for it, we take n_0^{th} to be the one in Lemma 6.

The proof of Lemma 6 follows from a strengthening of [4, Lemma 23]. That is, we show that there exists a threshold n_0^{th} and a constant $\theta > 0$ such that

$$\mathbb{P}(\neg\text{GBM}) < 2^{-\theta \cdot 2^{(1-2\xi)n}} \,, \tag{14}$$

for all $n_0 \geq n_0^{\text{th}}$ and $n > n_0$. Thus, since we've required that the constant ξ satisfy $\xi \in (0, \frac{1}{6})$, standard manipulations yield (13), for large enough n.

In [4, Lemma 23], the RHS of (14) is weaker: n is replaced by n_0. We give an outline of the differences between our proof of (14) and the proof of the weaker claim in [4, Lemma 23]. For the full proof of Lemma 6 see Appendix A. The main difference lies in bounding the probability that too much of $\mathbf{Z}_I$ is lost due to trimming. The weaker result follows by showing that the probability of a certain prefix of the leftmost block being completely lost due to trimming and deletion is upper bounded by a term that decays exponentially with N_0, the length of the block. In our proof, we show that for *any* prefix of $\mathbf{G}$, the number of block symbols is always greater than the number of guard-band symbols. Thus, the probability of such a prefix being lost due to deletion and trimming decays exponentially with its length. The stronger bound then follows by taking the prefix length to be proportional to N, as opposed to N_0.

VI. SECOND KEY LEMMA — STRONG POLARIZATION DESPITE A SMALL ADDITIVE PENALTY

In the previous section, we've stated Lemma 5, which gave upper bounds on the evolution of the Bhattacharyya parameter. Due to the added penalty in these bounds, we cannot use prior art in order to claim a polarization rate of roughly $2^{-\sqrt{N}}$. Indeed, in this section we state the second key lemma in the paper, Lemma 7, which implies such a rate for processes that evolve as in Lemma 5. Lemma 7 is stated quite generally, in the hope that it will be useful for other settings.

Lemma 7 (Strong polarization despite an additive penalty): Let $B_1, B_2, \ldots$ be i.i.d. uniformly distributed Bernoulli random variables. Fix $\beta \in (0, \frac{1}{2})$, $\epsilon' > 0$, $\kappa \geq 1$, and $\gamma > \frac{1}{2}$. There exist a positive constant $\eta = \eta(\epsilon', \kappa) > 0$ and a threshold $n_0^{\text{th}} =$

$n_0{}^{\text{th}}(\epsilon', \kappa, \gamma)$ such that the following holds. Fix $n_0 \geq n_0{}^{\text{th}}$ and let $Z_{n_0}, Z_{n_0+1}, Z_{n_0+2}, \ldots$ be a random process that satisfies:

$$Z_{n+1} \leq \begin{cases} \kappa \cdot Z_n + 2^{-N^\gamma} & \text{if } B_{n+1} = 0 \text{ ('–')} \\ \kappa \cdot Z_n^2 + 2^{-N^\gamma} & \text{if } B_{n+1} = 1 \text{ ('+')} \end{cases}, \quad n \geq n_0 \quad (15)$$

and

$$Z_{n_0} \leq \eta . \quad (16)$$

There exists a threshold $n^{\text{th}} = n^{\text{th}}(\epsilon', \beta, \kappa, \gamma, n_0) \geq n_0$ starting from which strong polarization occurs with probability at least $1 - \epsilon'$:

$$\mathbb{P}\left(Z_n < 2^{-N^\beta}, \quad \forall n \geq n^{\text{th}} \right) \geq 1 - \epsilon' . \quad (17)$$

We notice that Z_n provides a rather "weak" polarization step in (15). In memoryless channels, $Z_{n+1}^- \leq 2 \cdot Z_n$, and $Z_{n+1}^+ \leq Z_n^2$ [18, z.2], which is a "cleaner" and stronger polarization step than our case. In the memoryless cases, strong polarization can be proven as in [22]. In our case, we also have an additive term, 2^{-N^γ}. Still, because the additive term decays to 0 fast enough ($\gamma > 1/2$), strong polarization occurs, with the same exponent as the memoryless channels.

We now prove Lemma 7.

Proof: Let $n_0 \geq n_0{}^{\text{th}}$ and let $Z_{n_0}, Z_{n_0+1}, Z_{n_0+2}, \ldots$ satisfy (15) and (16), for $n_0{}^{\text{th}}$ and η we have yet to commit to. We first define an auxiliary process, starting at index n_0 with initial value η:

$$\bar{Z}_{n+1} = 2 \cdot \kappa \cdot \begin{cases} \bar{Z}_n & \text{if } B_{n+1} = 0 \text{ ('–')} \\ \bar{Z}_n^2 & \text{if } B_{n+1} = 1 \text{ ('+')} \end{cases}, \quad n \geq n_0 \quad (18a)$$

$$\bar{Z}_{n_0} = \eta \quad (18b)$$

We now fix $\gamma_a = (\gamma + 1/2)/2$ such that,

$$\frac{1}{2} < \gamma_a < \gamma . \quad (19)$$

Let $n^{\text{th}} \geq n_0$ be a parameter that will be committed to later on as well. We define the following events for the processes $Z_n, \bar{Z}_n$:

$$\Sigma_a \triangleq \left\{ \bar{Z}_n \geq 2^{-N^{\gamma_a}}, \quad \forall n \geq n_0 \right\}, \quad (20a)$$

$$\Sigma_b \triangleq \left\{ \bar{Z}_n < 2^{-N^\beta}, \quad \forall n \geq n^{\text{th}} \right\}, \quad (20b)$$

$$\Sigma_c \triangleq \left\{ Z_n \leq \bar{Z}_n, \quad \forall n \geq n_0 \right\} . \quad (20c)$$

The first two events bound the new process $\bar{Z}_n$, and the third event discusses a relation between $\bar{Z}_n$ and the original process Z_n. For the events above we list the following claims:

Claim 6.1: For all $\epsilon_a > 0$ there exist $\eta_a(\kappa) > 0$ and $n_{0_a}{}^{\text{th}}(\gamma_a, \epsilon_a, \kappa, \eta)$ s.t. if $0 < \eta \leq \eta_a$ and $n_0 \geq n_{0_a}{}^{\text{th}}$, then:

$$\mathbb{P}(\Sigma_a) > 1 - \epsilon_a . \quad (21)$$

Claim 6.2: For all $\epsilon_b > 0$ and $n_0 > 0$ there exist $\eta_b(\epsilon_b, \kappa) > 0$ and $n_b{}^{\text{th}}(\beta, \epsilon_b, \kappa, \eta, n_0) \geq n_0$ s.t. if $0 < \eta \leq \eta_b$ and $n^{\text{th}} \geq n_b{}^{\text{th}}$ then:

$$\mathbb{P}(\Sigma_b) > 1 - \epsilon_b . \quad (22)$$

Claim 6.3: There exists an $n_{0_c}{}^{\text{th}}(\gamma, \gamma_a)$ s.t. if $n_0 \geq n_{0_c}{}^{\text{th}}$, then event Σ_a implies Σ_c, i.e.:

$$\mathbb{P}(\Sigma_c | \Sigma_a) = 1 . \quad (23)$$

The proofs of Claims 6.1–6.3 are given in the upcoming subsections. The combination of the claims above proves our lemma. For this, we set:

$$\epsilon_a = \epsilon_b = \frac{\epsilon'}{2} \quad (24)$$

and set:

$$\eta(\epsilon', \kappa) \triangleq \min\left\{ \eta_a, \eta_b, \frac{1}{2} \right\}, \quad (25a)$$

$$n_0{}^{\text{th}}(\epsilon', \kappa, \gamma) \triangleq \max\{n_{0_a}{}^{\text{th}}, n_{0_c}{}^{\text{th}}\}, \quad (25b)$$

$$n^{\text{th}}(\epsilon', \beta, \kappa, \gamma, n_0) \triangleq \max\{n_b{}^{\text{th}}, n_0\} \geq n_0 . \quad (25c)$$

Using the three claims above, we can bound the probabilities of events Σ_b and Σ_c. For Σ_b, we have:

$$\mathbb{P}(\Sigma_b) \overset{(a)}{>} 1 - \epsilon_b , \quad (26)$$

where in (a) we applied (22) from Claim 6.2, since the conditions are satisfied by our selection of η and n^{th} in (25a) and (25c), respectively. We next note that:

$$\begin{aligned} \mathbb{P}(\Sigma_c) &\geq \mathbb{P}(\Sigma_c | \Sigma_a) \cdot \mathbb{P}(\Sigma_a) \\ &\overset{(a)}{>} 1 \cdot (1 - \epsilon_a) = 1 - \epsilon_a . \end{aligned} \quad (27)$$

In (a) we applied (21) from Claim 6.1, and (23) from Claim 6.3, since their conditions are satisfied by our selection of η and n_0 in (25a) and (25b), respectively.

By inspection, the intersection of Σ_b and Σ_c implies the event in (17), by (25c). Thus,

$$\begin{aligned} &\mathbb{P}\left(Z_n < 2^{-N^\beta}, \quad \forall n \geq n^{\text{th}} \right) \\ &\geq \mathbb{P}\left(\left\{ \bar{Z}_n < 2^{-N^\beta}, \quad \forall n \geq n^{\text{th}} \right\} \bigcap \left\{ Z_n \leq \bar{Z}_n, \quad \forall n \geq n_0 \right\} \right) \\ &= \mathbb{P}(\Sigma_b \cap \Sigma_c) \\ &= 1 - \mathbb{P}(\neg \Sigma_b \cup \neg \Sigma_c) . \end{aligned}$$

In the last equality we denoted events $\neg \Sigma_b, \neg \Sigma_c$ as the complementary events of Σ_b, Σ_c respectively. We now upper bound $\mathbb{P}(\neg \Sigma_b \cup \neg \Sigma_c)$:

$$\begin{aligned} \mathbb{P}(\neg \Sigma_b \cup \neg \Sigma_c) &\leq \mathbb{P}(\neg \Sigma_b) + \mathbb{P}(\neg \Sigma_c) \\ &\overset{(26),(27)}{<} \epsilon_a + \epsilon_b \\ &\overset{(24)}{=} \epsilon' . \end{aligned}$$

Thus:

$$\mathbb{P}\left(Z_n < 2^{-N^\beta}, \quad \forall n \geq n^{\text{th}} \right) \geq \mathbb{P}(\Sigma_b \cap \Sigma_c) \geq 1 - \epsilon' . $$

■

A. Proof of Claim 6.1

Proof: We first set:

$$\eta_a \triangleq \frac{1}{2 \cdot \kappa} \in \left(0, \frac{1}{2} \right], \quad (28)$$

where the inclusion follows from requiring that $\kappa \geq 1$ in Lemma 7.

We note that the process $\bar{Z}_n$ is positive for all $n \geq n_0$, since it starts from a positive value of $\bar{Z}_{n_0} = \eta \in (0, \eta_a]$ and in each step is either multiplied by $2 \cdot \kappa > 0$ or squared and multiplied

by $2 \cdot \kappa > 0$ (both operations preserve positivity). By (18a) and since $\kappa \geq 1$:

$$\bar{Z}_n = 2 \cdot \kappa \cdot \bar{Z}_{n-1}^{2^{B_n}} \geq \bar{Z}_{n-1}^{2^{B_n}} . \tag{29}$$

By recursively applying (29) we get:

$$\begin{aligned}
\bar{Z}_n &\geq \bar{Z}_{n-1}^{2^{B_n}} \\
&\geq \left(\bar{Z}_{n-2}^{2^{B_{n-1}}} \right)^{2^{B_n}} = \bar{Z}_{n-2}^{2^{B_{n-1}+B_n}} \\
&\geq \bar{Z}_{n-3}^{2^{B_{n-2}+B_{n-1}+B_n}} \\
&\geq \ldots \geq \bar{Z}_{n_0}^{2^{\sum_{i=n_0}^{n-1} B_{i+1}}} .
\end{aligned} \tag{30}$$

We also set $\rho \triangleq \gamma_a - \frac{1}{2} \overset{(19)}{>} 0$, and define:

$$\Delta_a \triangleq \left\lceil \frac{1}{2\rho^2} \ln\left(\frac{1}{\epsilon_a \cdot (1 - e^{-2\rho^2})} \right) \right\rceil > 0, \tag{31}$$

$$n_{0_a}^{\text{th}} \triangleq \left\lceil \frac{1}{\gamma_a} \log_2(-\log_2 \eta) + \frac{1}{\gamma_a} \Delta_a \right\rceil, \tag{32}$$

which are well defined since $\gamma_a \overset{(19)}{>} \frac{1}{2}$, $\rho = \gamma_a - \frac{1}{2} > 0$, $0 < \eta \leq \eta_a \overset{(28)}{\leq} \frac{1}{2}$ and $\epsilon_a > 0$.

We first bound the probability of $\bar{Z}_n$ not satisfying the lower bound in our claim (recall (20a)):

$$\begin{aligned}
\mathbb{P}\left(\bar{Z}_n < 2^{-N^{\gamma_a}} \right) &\overset{(30)}{\leq} \mathbb{P}\left(\bar{Z}_{n_0}^{2^{\sum_{i=n_0}^{n-1} B_{i+1}}} < 2^{-N^{\gamma_a}} \right) \\
&\overset{(a)}{=} \mathbb{P}\left(\sum_{i=n_0}^{n-1} B_{i+1} > n \cdot \gamma_a - \log_2(-\log_2 \eta) \right) \\
&\overset{(b)}{\leq} \mathbb{P}\left(\sum_{i=n_0}^{n-1} B_{i+1} > (n - n_0) \cdot \gamma_a \right) \\
&\overset{(c)}{=} \mathbb{P}\left(\sum_{i=n_0}^{n-1} B_{i+1} > (n - n_0) \cdot \left(\frac{1}{2} + \rho \right) \right) \\
&\leq \mathbb{P}\left(\sum_{i=n_0}^{n-1} B_{i+1} \geq (n - n_0) \cdot \left(\frac{1}{2} + \rho \right) \right) \\
&\overset{(d)}{\leq} e^{-2(n-n_0)\rho^2}
\end{aligned} \tag{33}$$

In (a) we applied $\log_2(-\log_2(\cdot))$ to both sides of the inequality, while recalling that $\bar{Z}_{n_0} \overset{(18b)}{=} \eta \overset{(28)}{\in} \left(0, \frac{1}{2} \right]$. In (b), we enlarged the probability by decreasing the lower bound, since $n_0 \geq n_{0_a}^{\text{th}} \overset{(32)}{>} \frac{1}{\gamma_a} \log_2(-\log_2 \eta)$. Inequality (c) holds by our selection of $\rho = \gamma_a - \frac{1}{2}$. Lastly, in (d) we applied the Hoeffding bound [23, page 78].

Next, we set:

$$n_a^{\text{th}} \triangleq n_0 + \Delta_a, \tag{34}$$

and address the probability of our lower bound holding for all $n \geq n_a^{\text{th}}$:

$$\begin{aligned}
\mathbb{P}\big(\bar{Z}_n &\geq 2^{-N^{\gamma_a}}, \quad \forall n \geq n_a^{\text{th}} \big) \\
&= 1 - \mathbb{P}\big(\exists n \geq n_a^{\text{th}}, \quad \bar{Z}_n < 2^{-N^{\gamma_a}} \big) \\
&= 1 - \mathbb{P}\left(\bigcup_{n=n_a^{\text{th}}}^{\infty} \bar{Z}_n < 2^{-N^{\gamma_a}} \right)
\end{aligned}$$

$$\begin{aligned}
&\overset{(a)}{\geq} 1 - \sum_{n=n_a^{\text{th}}}^{\infty} \mathbb{P}\big(\bar{Z}_n < 2^{-N^{\gamma_a}} \big) \\
&\overset{(33)}{\geq} 1 - \sum_{n=n_a^{\text{th}}}^{\infty} e^{-2(n-n_0)\rho^2} \\
&\overset{(b)}{=} 1 - \frac{1}{1 - e^{-2\rho^2}} e^{-2(n_a^{\text{th}}-n_0)\rho^2} \\
&\overset{(34)}{=} 1 - \frac{1}{1 - e^{-2\rho^2}} e^{-2\Delta_a \rho^2} \\
&\overset{(31)}{\geq} 1 - \epsilon_a
\end{aligned} \tag{35}$$

where (a) follows from the union bound, and for (b) and we remind that $\rho > 0$.

For the initial period of $n \in \left[n_0, n_a^{\text{th}} \right)$, we lower bound $\bar{Z}_n$ by the extreme event of drawing only $B_{i+1} = 1$ ('+') for $i = n_0, \ldots, n-1$. This event indeed minimizes $\bar{Z}_n$ for a low enough starting point $\eta \leq \eta_a$, as shown in Lemma 10 in Appendix B. That is, for $n \in \left[n_0, n_a^{\text{th}} \right)$:

$$\begin{aligned}
\bar{Z}_n &\geq \bar{Z}_n \text{ for } B_{i+1} = 1 \text{ ('+')}, \ \forall i = n_0, \ldots, n-1 \\
&\overset{(30)}{\geq} \bar{Z}_{n_0}^{2^{\sum_{i=n_0}^{n-1} B_{i+1}}} \text{ for } B_{i+1} = 1 \text{ ('+')}, \ \forall i = n_0, \ldots, n-1 \\
&= \bar{Z}_{n_0}^{2^{(n-n_0)}} \\
&\overset{(18b)}{=} \eta^{2^{(n-n_0)}} \\
&\overset{(a)}{>} \eta^{2^{(n_a^{\text{th}}-n_0)}} \\
&\overset{(34)}{=} \eta^{2^{\Delta_a}} \\
&\overset{(32)}{\geq} 2^{-2^{\gamma_a n_0}} \\
&\overset{(b)}{\geq} 2^{-2^{\gamma_a n}} = 2^{-N^{\gamma_a}} .
\end{aligned} \tag{36}$$

For (a) and (b) we remind that we are considering the case of $n \in \left[n_0, n_a^{\text{th}} \right)$.

In (36) we have shown that if $n_0 \geq n_{0_a}^{\text{th}}$, then $\bar{Z}_n \geq 2^{-N^{\gamma_a}}$ for all $n \in \left[n_0, n_a^{\text{th}} \right)$. In (35) we have shown the same lower bound holds for all $n \geq n_a^{\text{th}}$ with probability of at least $1 - \epsilon_a$. By combining the two results we complete the proof of our claim. ∎

B. Proof of Claim 6.2

Claim 6.2 is a specialization of [24, Proposition 49]. Namely, [24, Equation 166] as well as the constraint on Z_0 in [24, Proposition 49] have inequalities which in (18) are specialized to equalities. The fact that our indexing starts at n_0 as opposed to 0 in [24, Proposition 49] is of no consequence.

The above paragraph proves our claim, and the reader may elect to leave it at that. However, since the thresholds η and n_0 are not given explicitly in [24, Proposition 49], we choose to do so here, for completeness. That is, we give explicit expressions for η_b (our analog of the η in [24]) and n_b^{th} (our analog of the n_0 in [24]).

To set η_b and n_b^{th} explicitly, we carefully inspect the proof of [24, Proposition 49] and the strong polarization proof in [22] on which it relies. For the diligent reader wishing to correspond our steps to the mentioned sources, we note that the process $\bar{Z}_n$ here is written without the "bar" in [24, Proposition 49] and in [22]. In addition, σ_a and σ_b which will be used here are denoted ϵ_a and ϵ_b, respectively, in [24, Proposition 49] and

in [22]. This replacement was performed to avoid confusion with ϵ_a and ϵ_b from Claim 6.1 and Claim 6.2, respectively. We also note that [22] discusses a general process of the form:

$$Z_{n+1} \leq K \cdot Z_n^{d_{T_n}},$$

where $K \geq 1$, T_n is uniformly distributed over $\{1, 2, \ldots, \ell\}$ and $d_1, d_2, \ldots, d_\ell \geq 1$. In our case, and in [24, Proposition 49] as well, $\ell = 2$ and:

$$d_{T_n} = \begin{cases} 1 & \text{if } T_n = 1 \,, \\ 2 & \text{if } T_n = 2 \,. \end{cases}$$

This simplifies some of the expressions we cite from [22]. Specifically, we note that the parameter $E \triangleq \frac{1}{\ell} \sum_{t=1}^{\ell} \log_\ell d_t > \beta$, defined in [22], equals $\frac{1}{2}$ in our case. Lastly, in our case $K = 2 \cdot \kappa$, not to be confused with the substitution of K with κ, used in [24, Proposition 49]. We note that the multiplicative factor of 2 added to the process $\bar{Z}_n$ (see (18)) essentially originates from "getting rid" of the additive penalty in the process Z_n (see (15)).

We now summarize in four steps how to set η_b and n_b^{th}, in order to prove our claim. By following these steps we satisfy two conditions. The first is that for all $n \geq n_b^{\text{th}}$, the process $\bar{Z}_n$ is upper bounded by $2^{-2^{(\frac{1}{2}-\Delta)n}}$, with probability at least $1 - \epsilon_b$. The expression for Δ is calculated in [22, Equation 16]. We rewrite it here for our case, i.e. for $\ell = 2$, $d_1 = 1$ and $d_2 = 2$:

$$\Delta = \frac{1}{2} \left(\log_2 \left(\frac{1}{1-\theta} \right) + \log_2 \left(\frac{2}{2-\theta} \right) \right) \tag{37}$$
$$+ \sigma_b \cdot \left(-\log_2(1-\theta) + \log_2(2-\theta) \right) \tag{38}$$
$$+ \frac{m_a}{n} \left(\left(\frac{1}{2} + \sigma_b \right) \log_2(1-\theta) + \left(\frac{1}{2} - \sigma_b \right) \log_2(2-\theta) \right) \tag{39}$$

where $\theta \triangleq -\log_{\sigma_a} 2\kappa$ is defined by a small parameter $\sigma_a > 0$ which we will select shortly. The parameters $\sigma_b > 0$ and m_a will be set soon as well.

Just as in [22], the second condition we will satisfy is that each of the three terms in the expression of Δ, i.e. (37), (38) and (39), is upper bounded by $\frac{\frac{1}{2}-\beta}{3}$. Hence, $\frac{1}{2} - \Delta > \beta$. By combining these two conditions we get the desired upper bound on $\bar{Z}_n$.

Throughout these four steps, all the introduced parameters (σ_a, σ_b and m_a) are functions of the parameters in our claim: $\kappa, \epsilon_b, \eta, n_0$ and β.

1) We first set a small enough $\sigma_a > 0$ such that, for $\theta \triangleq -\log_{\sigma_a} 2\kappa \geq 0$, we have $\theta < 1$ and the term in (37) is $< \frac{\frac{1}{2}-\beta}{3}$. For example, set:

$$\theta \triangleq \frac{2}{3} \left(1 - 2^{-\frac{1-2\beta}{3}} \right) > 0$$

and $\sigma_a = (2\kappa)^{-\frac{1}{\theta}} > 0$.

2) We set a small enough η_b and a large enough m_a such that if $\bar{Z}_{n_0} = \eta \leq \eta_b$ then

$$\mathbb{P}\left(\bar{Z}_n < \sigma_a, \quad \forall n \geq m_a \right) > 1 - \frac{\epsilon_b}{2}. \tag{40}$$

The above, up to renaming of parameters, is [24, Equation 171]. A possible selection for these parameters, suggested at the very end of the proof of [24, Proposition 49], is:

$$\eta_b(\epsilon_b, \kappa) \triangleq \frac{1}{2(2\kappa)^2} \left(\frac{\epsilon_b}{4} \right)^{1/r},$$

where $r = r(\kappa)$ is the largest positive solution of $(2\kappa)^r + (4\kappa)^{-r} = 2$, and

$$m_a = n_0 + \left\lceil \max \left\{ \begin{array}{l} \frac{4}{\ln 2}(\ln \eta - \ln \sigma_a), \\ \frac{2}{s^2} \ln \left(\frac{4}{\epsilon_b(1-e^{-s^2/2})} \right) \end{array} \right\} \right\rceil,$$

where $s = s(\kappa) \triangleq \frac{\ln 2}{2 \ln(2(2\kappa)^2)}$.

3) We set a small enough $\sigma_b > 0$ such that the term in (38) is $< \frac{\frac{1}{2}-\beta}{3}$. For example:

$$\sigma_b = \frac{1}{2} \cdot \frac{\frac{\frac{1}{2}-\beta}{3}}{\log_2\left(\frac{2-\theta}{1-\theta} \right)}.$$

We remind that θ was set in the first step and satisfies $\theta < 1$.

4) We set a large enough n_b^{th} such that for all $n \geq n_b^{\text{th}}$ the term in (39) is $< \frac{\frac{1}{2}-\beta}{3}$ and

$$\mathbb{P}\left(\begin{array}{c} \left| \frac{|\{m_a \leq i < n: \, T_i = t\}|}{n - m_a} - \frac{1}{2} \right| \leq \sigma_b \\ \text{for all } n \geq n_b^{\text{th}} \text{ and } t \in \{1, 2\} \end{array} \right) > 1 - \frac{\epsilon_b}{2}.$$

That is, take n_b^{th} to be the maximum of

$$1 + \left\lceil \frac{m_a\left(\left(\frac{1}{2} + \sigma_b \right) \log_2(1-\theta) + \left(\frac{1}{2} - \sigma_b \right) \log_2(2-\theta) \right)}{\frac{\frac{1}{2}-\beta}{3}} \right\rceil$$

and

$$m_a + \left\lceil \frac{1}{2\sigma_b^2} \ln \left(\frac{4}{\epsilon_b(1-e^{-2\sigma_b^2})} \right) \right\rceil, \tag{41}$$

where (41) follows from Hoeffding's inequality [23, Theorem 4.12] and the union bound (see (33) and (35) in which we applied similar arguments).

C. Proof of Claim 6.3

Proof: Under event Σ_a (20a): $\bar{Z}_n \geq 2^{-N^{\gamma_a}}$ for all $n \geq n_0$. Since $\gamma_a \overset{(19)}{<} \gamma$, for $n \geq n_0 \geq n_{0_c}^{\text{th}}(\gamma, \gamma_a) \triangleq \frac{1}{\gamma - \gamma_a}$ we get

$$\gamma n \geq \gamma_a n + 1 \,.$$

Thus, if Σ_a occurs then $\bar{Z}_n$ satisfies:

$$\begin{array}{llll} \kappa \cdot \bar{Z}_n & \geq \bar{Z}_n \geq & 2^{-2^{\gamma_a n}} & \geq 2^{-2^{\gamma n}} \,, \\ \kappa \cdot \bar{Z}_n^2 & \geq \bar{Z}_n^2 \geq & 2^{-2^{\gamma_a n+1}} & \geq 2^{-2^{\gamma n}} \,. \end{array} \tag{42}$$

From the above we get that under Σ_a, if $\bar{Z}_n \geq Z_n$, then $\bar{Z}_{n+1} \geq Z_{n+1}$ as well (regardless of B_{n+1}). That is:

$$\bar{Z}_{n+1} \overset{(18a)}{=} \begin{cases} \kappa \cdot \bar{Z}_n + \kappa \cdot \bar{Z}_n & \text{if } B_{n+1} = 0 \text{ ('−')} \\ \kappa \cdot \bar{Z}_n^2 + \kappa \cdot \bar{Z}_n^2 & \text{if } B_{n+1} = 1 \text{ ('+')} \end{cases}, \quad n \geq n_0$$

$$\overset{(42)}{\geq} \begin{cases} \kappa \cdot \bar{Z}_n + 2^{-2^{\gamma n}} & \text{if } B_{n+1} = 0 \text{ ('−')} \\ \kappa \cdot \bar{Z}_n^2 + 2^{-2^{\gamma n}} & \text{if } B_{n+1} = 1 \text{ ('+')} \end{cases}, \quad n \geq n_0$$

$$\overset{(a)}{\geq} \begin{cases} \kappa \cdot Z_n + 2^{-2^{\gamma n}} & \text{if } B_{n+1} = 0 \text{ ('−')} \\ \kappa \cdot Z_n^2 + 2^{-2^{\gamma n}} & \text{if } B_{n+1} = 1 \text{ ('+')} \end{cases}, \quad n \geq n_0$$

$$\overset{(15)}{\geq} Z_{n+1} \,,$$

where (a) holds under the hypothesis that $\bar{Z}_n \geq Z_n$.

We remind that: $\bar{Z}_{n_0} \overset{(18b)}{=} \eta \overset{(16)}{\geq} Z_{n_0}$. Thus, we may show by induction that under Σ_a, $\bar{Z}_n \geq Z_n$ for all $n \geq n_0$. We have shown that under Σ_a:

$$Z_n \leq \bar{Z}_n, \quad \forall n \geq n_0 .$$

∎

VII. PROOF OF OUR MAIN THEOREM

Proof: The proof of Theorem 1 follows by combining Theorem 3, Lemma 4, Lemma 5, Lemma 7, and Lemma 11. Let $\delta \in (0,1)$ be a fixed deletion probability of the deletion channel. Fix $\epsilon \in (0,1)$, $\xi \in (0, \frac{1}{6})$, and $0 < \beta' < \beta < \frac{1}{2}$. We take ϵ' in both Theorem 3 and Lemma 7 to be $\frac{\epsilon}{3}$, and fix κ and γ from Lemma 7 to 2 and $\frac{2}{3}$, respectively. We set η in Theorem 3 to be $\eta(\epsilon', \kappa)$ from Lemma 7. Next, we take $n_0^{\text{th}}(\epsilon, \delta, \xi)$ to be the maximum of

$n_0^{\text{th}}(\epsilon', \eta, \delta)$ from Theorem 3,

$\qquad n_0^{\text{th}}(\epsilon, \xi)$ from Lemma 4,

$\qquad n_0^{\text{th}}(\xi, \delta)$ from Lemma 5,

$\qquad \qquad$ and $n_0^{\text{th}}(\epsilon', \kappa, \gamma)$ from Lemma 7.

By this selection we are promised for $n_0 \geq n_0^{\text{th}}$ that

$$\frac{N}{\Lambda} = \frac{|\mathbf{X}|}{|\mathbf{G}|} > 1 - \epsilon ,$$

by Lemma 4. For a fixed $n_0 \geq n_0^{\text{th}}$ we take $n^{\text{th}}(\epsilon, \beta, \beta', n_0)$ to be the maximum of $n^{\text{th}}(\epsilon', \beta, \kappa, \gamma, n_0)$ from Lemma 7 and $n^{\text{th}}(\beta, \beta', \epsilon)$ from Lemma 11.

We now prove that if we take $n \geq n^{\text{th}}$, then

$$Z(U_i|U_1^{i-1}, \mathbf{Y}^*) \leq 2^{-N^\beta} \text{ and } K(U_i|U_1^{i-1}) \leq 2^{-N^\beta} \quad (43)$$

for at least $\mathcal{I} - \epsilon$ of the indices. For this, we define the processes $Z_{n_0}, Z_{n_0+1}, Z_{n_0+2} \ldots$ and $K_{n_0}, K_{n_0+1}, K_{n_0+2} \ldots$ which start from the $n_0 \geq n_0^{\text{th}}$ we fixed above. We couple these processes to the Bernoulli($\frac{1}{2}$) i.i.d process $B_1, B_2, \ldots$ by defining:

$$Z_n = Z(U_i|U_1^{i-1}, \mathbf{Y}^*) \text{ and } K_n = K(U_i|U_1^{i-1})$$

for $i = i(B_1, B_2, \ldots, B_n)$, see (7). That is, the drawn $B_1, \ldots, B_n$ are the index bits $b_1, \ldots, b_n$ of i.

Let η and n_0 be as fixed above. We denote by $\mathcal{W}$ the set of "weak polarization" indices. That is, for $|\mathbf{U}| = |\mathbf{X}| = N_0$, $\mathcal{W}$ is the set of indices $1 \leq j \leq N_0$ which satisfy:

$$Z(U_j|U_1^{j-1}, \mathbf{Y}^*) \leq \eta \text{ and } K(U_j|U_1^{j-1}) \leq \eta .$$

In terms of our random processes we get

$$i(B_1, \ldots, B_{n_0}) \in \mathcal{W} \iff Z_{n_0} \leq \eta \text{ and } K_{n_0} \leq \eta . \quad (44)$$

We now show that for all $n \geq n^{\text{th}}$ our processes are upper bounded by 2^{-N^β} with probability $\geq \mathcal{I} - \epsilon$.

$$\mathbb{P}(Z_n < 2^{-N^\beta}, K_n < 2^{-N^\beta}, \forall n \geq n^{\text{th}})$$
$$\geq \mathbb{P}(Z_n < 2^{-N^\beta}, K_n < 2^{-N^\beta}, \forall n \geq n^{\text{th}}, Z_{n_0} \leq \eta, K_{n_0} \leq \eta)$$
$$= \sum_{j \in \mathcal{W}} \mathbb{P}(Z_n < 2^{-N^\beta}, K_n < 2^{-N^\beta}, \forall n \geq n^{\text{th}}, i(B_1, \ldots, B_{n_0}) = j)$$
$$= \sum_{j \in \mathcal{W}} \mathbb{P}(Z_n < 2^{-N^\beta}, K_n < 2^{-N^\beta}, \forall n \geq n^{\text{th}} | i(B_1, \ldots, B_{n_0}) = j)$$
$$\times \mathbb{P}(i(B_1, \ldots, B_{n_0}) = j)$$
$$\overset{(a)}{\geq} \sum_{j \in \mathcal{W}} \Big[\mathbb{P}(Z_n < 2^{-N^\beta}, \forall n \geq n^{\text{th}} | i(B_1, \ldots, B_{n_0}) = j)$$
$$+ \mathbb{P}(K_n < 2^{-N^\beta}, \forall n \geq n^{\text{th}} | i(B_1, \ldots, B_{n_0}) = j) - 1 \Big]$$
$$\times \mathbb{P}(i(B_1, \ldots, B_{n_0}) = j)$$
$$\overset{(b)}{\geq} \sum_{j \in \mathcal{W}} ((1 - \epsilon') + (1 - \epsilon') - 1) \times \mathbb{P}(i(B_1, \ldots, B_{n_0}) = j)$$
$$= (1 - 2\epsilon') \cdot \sum_{j \in \mathcal{W}} \mathbb{P}(i(B_1, \ldots, B_{n_0}) = j)$$
$$= (1 - 2\epsilon') \cdot \frac{|\mathcal{W}|}{N_0}$$
$$\overset{(c)}{\geq} (1 - 2\epsilon') \cdot (\mathcal{I} - \epsilon')$$
$$\geq \mathcal{I} - 3\epsilon'$$
$$= 1 - \epsilon$$

Inequality (a) follows by noting that $P(A \cap B) \geq P(A) + P(B) - 1$, since $P(A) + P(B) - P(A \cap B) = P(A \cup B) \leq 1$. Inequality (c) is by Theorem 3. Inequality (b) follows by applying Lemma 7 to each[1] of the processes $Z_{n_0}, Z_{n_0+1}, Z_{n_0+2} \ldots$ and $K_{n_0}, K_{n_0+1}, K_{n_0+2} \ldots$. Indeed, first note that our choices of n_0, n, and η satisfy the conditions on them in Lemma 7. We now show that each of the two processes satisfy the two conditions, (15) and (16), in Lemma 7. To see that each process satisfies (16) recall that we are under the event $i(B_1, \ldots, B_{n_0}) = j$ where $j \in \mathcal{W}$, and thus (16) is satisfied by (44). For the Bhattacharyya process $Z_{n_0}, Z_{n_0+1}, \ldots$ condition (15) holds by Lemma 5 ($\kappa = 2$ and $\gamma = \frac{2}{3}$) and our choice of n_0. For the total variation process $K_{n_0}, K_{n_0+1}, \ldots$ condition (15) holds by [13, Proposition 4]. That is, since the blocks of $\mathbf{X}$ are i.i.d, for $n > n_0$ we have [13, Proposition 4]

$$K(U_i|U_1^{i-1}) \leq \begin{cases} K(V_j|V_1^{j-1})^2 & \text{if } b_n = 0 \text{ ('−')}, \\ 2 \cdot K(V_j|V_1^{j-1}) & \text{if } b_n = 1 \text{ ('+')}, \end{cases} \quad (45)$$

where $\mathbf{U}, \mathbf{V}, i, j$ and b_n are as in Lemma 5. To summarize, we have proved that (43) holds for a fraction of at least $\mathcal{I} - \epsilon$ of the indices.

We now show that for $n \geq n^{\text{th}}$,

$$Z(U_i|U_1^{i-1}, \mathbf{Y}) \overset{(a)}{\leq} Z(U_i|U_1^{i-1}, \mathbf{Y}^*) \overset{(b)}{\leq} 2^{-N^\beta} \overset{(c)}{\leq} \frac{1}{2N} 2^{-\Lambda^{\beta'}}$$

$$\text{and} \quad K(U_i|U_1^{i-1}) \overset{(b)}{\leq} 2^{-N^\beta} \overset{(c)}{\leq} \frac{1}{2N} 2^{-\Lambda^{\beta'}} ,$$

---

[1] We note that the use of Lemma 7 for upper bounding the process K is a bit of an overkill. The evolution of K in each polarization step is better than the evolution of the Bhattacharyya parameter $Z(U_I|U_1^{i-1}, \mathbf{Y}^*)$. That is, there is no small additive penalty of $2^{-N^{2/3}}$, contrast (8) with its analog for K in (45). Thus, instead of using Lemma 7 we could have used [24, Proposition 49].

for at least $\mathcal{I} - \epsilon$ of the indices $1 \le i \le N$. Inequality (a) results from the TDC being a degradation of the deletion channel. That is, $\mathbf{X} - \mathbf{Y} - \mathbf{Y}^*$ form a Markov chain in that order. Inequalities (b), that is (43), indeed hold for the above fraction of indices, as we have just proved. Lastly, (c) is the result of combining Lemma 4 and Lemma 11, which may be applied by our selection of n_0 and n, respectively. ∎

APPENDIX A
HIGH PROBABILITY OF GBM

We now prove Lemma 6. This lemma states that the GBM event, i.e. the event where the middle bit in the TDC output is a bit from the middle GB, occurs with high probability.

Proof: As mentioned previously in Subsection V-B, parts of the proof resemble the steps taken in the proof of [4, Lemma 23]. We define the following length differences due to channel deletion (recall illustration of $\mathbf{G}_\mathrm{I}, \mathbf{Y}_\mathrm{I}, \mathbf{Z}_\mathrm{I}$, etc. in Figure 1):

$$
\begin{aligned}
\alpha &= |\mathbf{G}_\mathrm{I}| - |\mathbf{Y}_\mathrm{I}| \\
\beta &= |\mathbf{G}_\Delta| - |\mathbf{Y}_\Delta| \\
\gamma &= |\mathbf{G}_\mathrm{II}| - |\mathbf{Y}_\mathrm{II}|,
\end{aligned}
\tag{46a}
$$

and the following length differences due to trimming:

$$
\begin{aligned}
\alpha' &= |\mathbf{Y}_\mathrm{I}| - |\mathbf{Z}_\mathrm{I}| \\
\beta' &= |\mathbf{Y}_\Delta| - |\mathbf{Z}_\Delta| \\
\gamma' &= |\mathbf{Y}_\mathrm{II}| - |\mathbf{Z}_\mathrm{II}|.
\end{aligned}
\tag{46b}
$$

We will be interested in the event: $A \cap A' \cap B \cap B' \cap C \cap C'$, where A, B, C will be events constraining the number of deletions in $\mathbf{G}_\mathrm{I}, \mathbf{G}_\Delta, \mathbf{G}_\mathrm{II}$, respectively, and A', B', C' are events constraining the number of bits trimmed in $\mathbf{Y}_\mathrm{I}, \mathbf{Y}_\Delta, \mathbf{Y}_\mathrm{II}$, respectively. These events will be defined explicitly in a moment, but first, we note that the main property of these events is:

$$
A \cap A' \cap B' \cap B \cap C \cap C' \Rightarrow \mathrm{GBM}.
$$

That is, under all of the events A, A', B, B', C, C' we are under the GBM event. The justification of this property will soon be given in (50).

We define:

$$
\begin{aligned}
A &= \{\delta|\mathbf{G}_\mathrm{I}| - \hat{\ell} < \alpha < \delta|\mathbf{G}_\mathrm{I}| + \hat{\ell}\} \tag{47a} \\
A' &= \{0 \le \alpha' < \hat{\ell}\} \tag{47b} \\
B &= \{\beta < \delta|\mathbf{G}_\Delta| + \hat{\ell}\} \tag{47c} \\
B' &= \{\beta' = 0\} \tag{47d} \\
C &= \{\delta|\mathbf{G}_\mathrm{II}| - \hat{\ell} < \gamma < \delta|\mathbf{G}_\mathrm{II}| + \hat{\ell}\} \tag{47e} \\
C' &= \{0 \le \gamma' < \hat{\ell}\}. \tag{47f}
\end{aligned}
$$

Intuitively, a small $\hat{\ell}$ implies a tight tolerance on the events defined in (47) — and a small enough tolerance will imply GBM. On the other hand, we must not take $\hat{\ell}$ too small, since we want each of the probabilities of these events to be close to 1. As we will see, both aims are achieved by taking

$$
\hat{\ell} = \frac{1-\delta}{4} \ell_n. \tag{48}
$$

For this selection, and since

$$
\ell_n \overset{(4)}{<} 2^{(n-1)} = |\mathbf{X}_\mathrm{I}| \le |\mathbf{G}_\mathrm{I}|,
$$

we notice that under the event A, the number of bits deleted from $\mathbf{G}_\mathrm{I}$ is less than $\frac{1+\delta}{2}|\mathbf{G}_\mathrm{I}|$. Also, under A' we will trim less than $\frac{1-\delta}{4}|\mathbf{G}_\mathrm{I}|$ bits from $\mathbf{Y}_\mathrm{I}$. Thus, in total, under $A \cap A'$, the total number of bits deleted or trimmed from $\mathbf{G}_\mathrm{I}$ is less than

$$
\left(\left(\frac{1+\delta}{2}\right) + \left(\frac{1-\delta}{4}\right)\right)|\mathbf{G}_\mathrm{I}| = \frac{3+\delta}{4}|\mathbf{G}_\mathrm{I}| \le |\mathbf{G}_\mathrm{I}|.
$$

That is, under $A \cap A'$, the trimming of $\mathbf{Y}_\mathrm{I}$ has stopped prior to the received GB bits in $\mathbf{Y}_\Delta$. By symmetry, under $A \cap A' \cap C \cap C'$, we do not trim $\mathbf{Y}_\Delta$ from either side. Hence,

$$
A \cap A' \cap C \cap C' \Rightarrow B'. \tag{49}
$$

We use this now to show the following:

$$
\begin{aligned}
&\{A \cap A' \cap B \cap C \cap C'\} \\
&\overset{(49)}{\Rightarrow} \{A \cap A' \cap B \cap B' \cap C \cap C'\} \\
&\overset{(a)}{\Rightarrow} \left\{
\begin{array}{c}
\alpha + \alpha' < \gamma + \gamma' + \ell_n - \beta - \beta' \\
\text{and } \gamma + \gamma' < \alpha + \alpha' + \ell_n - \beta - \beta'
\end{array}
\right\} \\
&\overset{(46)}{\Rightarrow} \{|\mathbf{Z}_\mathrm{II}| < |\mathbf{Z}_\Delta| + |\mathbf{Z}_\mathrm{I}| \quad \text{and} \quad |\mathbf{Z}_\mathrm{I}| < |\mathbf{Z}_\Delta| + |\mathbf{Z}_\mathrm{II}|\} \\
&\overset{(b)}{\Rightarrow} \mathrm{GBM}. \tag{50}
\end{aligned}
$$

(a) holds since under the event $A \cap A' \cap B \cap B' \cap C \cap C'$:

$$
\begin{aligned}
\gamma + \gamma' + \ell_n - \beta - \beta' &\overset{(47d)}{=} \gamma + \gamma' + \ell_n - \beta \\
&\overset{(47c),(47e),(47f)}{>} \delta|\mathbf{G}_\mathrm{II}| - \hat{\ell} + 0 + \ell_n - \delta|\mathbf{G}_\Delta| - \hat{\ell} \\
&\overset{(3)}{=} \delta|\mathbf{G}_\mathrm{II}| - \hat{\ell} + \ell_n - \delta\ell_n - \hat{\ell} \\
&\overset{(48)}{=} \delta|\mathbf{G}_\mathrm{II}| - \hat{\ell} + 4\hat{\ell} - \hat{\ell} \\
&= \delta|\mathbf{G}_\mathrm{II}| + 2\hat{\ell} \\
&\overset{(47a),(47b)}{>} \alpha + \alpha',
\end{aligned}
$$

and $\gamma + \gamma' < \alpha + \alpha' + \ell_n - \beta - \beta'$ by the same steps. For (b), the middle index i_mid does not fall in $\mathbf{Z}_\mathrm{I}$, by

$$
\begin{aligned}
i_\mathrm{mid} &\overset{(9)}{=} \left\lfloor \frac{|\mathbf{Z}| + 1}{2} \right\rfloor \\
&= \left\lfloor \frac{|\mathbf{Z}_\mathrm{I}| + |\mathbf{Z}_\Delta| + |\mathbf{Z}_\mathrm{II}| + 1}{2} \right\rfloor \\
&\ge \left\lfloor \frac{|\mathbf{Z}_\mathrm{I}| + |\mathbf{Z}_\mathrm{I}| + 1 + 1}{2} \right\rfloor \\
&= |\mathbf{Z}_\mathrm{I}| + 1,
\end{aligned}
$$

and the middle index i_mid does not fall in $\mathbf{Z}_\mathrm{II}$, by

$$
\begin{aligned}
i_\mathrm{mid} &\overset{(9)}{=} \left\lfloor \frac{|\mathbf{Z}| + 1}{2} \right\rfloor \\
&= \left\lfloor \frac{|\mathbf{Z}_\mathrm{I}| + |\mathbf{Z}_\Delta| + |\mathbf{Z}_\mathrm{II}| + 1}{2} \right\rfloor \\
&\le \frac{|\mathbf{Z}_\mathrm{I}| + |\mathbf{Z}_\Delta| + |\mathbf{Z}_\mathrm{II}| + 1}{2} \\
&< \frac{|\mathbf{Z}_\mathrm{I}| + |\mathbf{Z}_\Delta| + |\mathbf{Z}_\mathrm{I}| + |\mathbf{Z}_\Delta| + 1}{2} \\
&< |\mathbf{Z}_\mathrm{I}| + |\mathbf{Z}_\Delta| + 1.
\end{aligned}
$$

Thus, i_mid must fall within the middle guard-band $\mathbf{Z}_\Delta$. We note that the outputs middle guard-band $\mathbf{Z}_\Delta$ is not an empty

string under the events above. If it was an empty string, i.e. $|\mathbf{Z}_\Delta| = 0$, we get from (50): $|\mathbf{Z}_I| < |\mathbf{Z}_{II}|$ and $|\mathbf{Z}_I| > |\mathbf{Z}_{II}|$ which is not possible.

From (50) we get: $\mathbb{P}(\text{GBM}) \geq \mathbb{P}(A \cap A' \cap B \cap C \cap C')$. We are interested in the complementary event, which satisfies:

$$\mathbb{P}(\neg\text{GBM}) \leq \mathbb{P}(\neg\{A \cap A' \cap B \cap C \cap C'\})$$
$$\overset{(a)}{\leq} \mathbb{P}(\neg A) + \mathbb{P}(\neg A') + \mathbb{P}(\neg B) + \mathbb{P}(\neg C) + \mathbb{P}(\neg C')$$
$$\overset{(b)}{=} 2 \cdot \mathbb{P}(\neg A) + 2 \cdot \mathbb{P}(\neg A') + \mathbb{P}(\neg B)$$

(a) is by the union bound and (b) results from the symmetry between events A, A' and C, C' respectively, by (47).

We now set:

$$n_0^{\text{th}}(\xi, \delta) \triangleq \max \left\{ \begin{array}{l} \frac{1}{\xi} \log_2\left(1 + \frac{1}{1-2^{-\xi}}\right), \ \frac{1}{1-\xi} \log_2\left(\frac{64\tau}{1-\delta}\right) + 1, \\ \frac{1}{\xi} \log_2\left(\frac{(1-\delta)^2}{128 \cdot D}\right), \ \frac{1}{1-2\xi-\frac{2}{3}} \log_2\left(\frac{128 \cdot 2}{(1-\delta)^2}\right), \\ \frac{1}{1-2\xi} \log_2\left(\frac{128 \log_2(7)}{(1-\delta)^2 (\log_2(e)-1)}\right) \end{array} \right\},$$
(51)

and proceed to upper bound the probabilities of each of the events $\neg A, \neg A'$, and $\neg B$, for $n > n_0 \geq n_0^{\text{th}}$. We note that $\tau > 0$ is a constant dependent on the input distribution that will be defined later on. We will also set the constant $D > 0$ later on, which is dependent on the input distribution and on the deletion rate δ. In Lemma 8 we will bound $\mathbb{P}(\neg A')$. $\mathbb{P}(\neg A)$ and $\mathbb{P}(\neg B)$ may be bounded using Hoeffding [25, Theorem 4.12], as in [4, equations (89),(90)]. That is,

$$\mathbb{P}(\neg A) \leq 2 \cdot e^{-\frac{2\hat{\ell}^2}{|\mathbf{G}_I|}}$$
$$\overset{(a)}{\leq} 2 \cdot e^{-\frac{2\hat{\ell}^2}{2^n}}$$
$$\overset{(48)}{\leq} 2 \cdot e^{-\frac{2((1-\delta)\ell_n/4)^2}{2^n}}$$
$$= 2 \cdot e^{-\frac{(1-\delta)^2}{8}\ell_n^2 2^{-n}}$$
$$\overset{(4)}{<} 2 \cdot e^{-\frac{(1-\delta)^2}{8}2^{2((1-\xi)(n-1)-1)}2^{-n}}$$
$$= 2 \cdot e^{-\frac{(1-\delta)^2}{32}2^{2(1-\xi)(n-1)-n}}$$
$$\leq 2 \cdot e^{-\frac{(1-\delta)^2}{128}2^{(1-2\xi)n}},$$
(52)

where in (a) we bounded $|\mathbf{G}_I| \leq 2 \cdot 2^{n-1}$, which results from (4a) and $n_0 \geq n_0^{\text{th}} \overset{(51)}{\geq} \frac{1}{\xi} \log_2\left(1 + \frac{1}{1-2^{-\xi}}\right) > \frac{1}{\xi} \log_2\left(\frac{1}{1-2^{-\xi}}\right)$. Similarly for the event $\neg B$ we have (using the penultimate displayed equation in the proof of [25, Theorem 4.12])

$$\mathbb{P}(\neg B) \leq e^{-\frac{2\hat{\ell}^2}{|\mathbf{G}_\Delta|}}$$
$$\overset{(48)}{=} e^{-\frac{2((1-\delta)\ell_n/4)^2}{\ell_n}}$$
$$= e^{-2\frac{(1-\delta)^2}{16}\ell_n}$$
$$\overset{(4)}{=} e^{-2\frac{(1-\delta)^2}{16}\lfloor 2^{(1-\xi)(n-1)} \rfloor}$$
$$\overset{(a)}{\leq} e^{-2\frac{(1-\delta)^2}{16}2^{(1-\xi)(n-1)-1}}$$
$$= e^{-\frac{(1-\delta)^2}{16}2^{(1-\xi)(n-1)}}$$
$$\leq e^{-\frac{(1-\delta)^2}{32}2^{(1-\xi)n}},$$
(53)

where in (a) we used $\lfloor 2^x \rfloor > 2^{x-1}$ which holds for any $x \geq 0$ (for $x \in [0, 1)$ we have $1 > 2^{x-1}$ and for $x \geq 1$ we have $\lfloor 2^x \rfloor > 2^x - 1 \geq 2^{x-1}$).

In total, we reach the following upper bound for $\mathbb{P}(\neg\text{GBM})$:

$$\mathbb{P}(\neg\text{GBM}) \leq 2 \cdot \mathbb{P}(\neg A) + 2 \cdot \mathbb{P}(\neg A') + \mathbb{P}(\neg B)$$
$$\overset{(52),(53), \text{ Lemma 8}}{\leq} 2 \cdot 2 \cdot e^{-\frac{(1-\delta)^2}{128}2^{(1-2\xi)n}}$$
$$+ 2 \cdot e^{-D \cdot 2^{(1-\xi)n}}$$
$$+ e^{-\frac{(1-\delta)^2}{32}2^{(1-\xi)n}},$$

where $D > 0$ is the same constant from (51). The value of D will be given explicitly in the proof of Lemma 8. We note that when bounding $\mathbb{P}(A')$ we will use the fact that $n > n_0 \geq n_0^{\text{th}} \overset{(51)}{\geq} \max\left\{\frac{1}{\xi} \log_2\left(1 + \frac{1}{1-2^{-\xi}}\right), \ \frac{1}{1-\xi} \log_2\left(\frac{64\tau}{1-\delta}\right) + 1\right\}$, and the qualities of the input distribution we fixed. The definition of τ will also be given in the proof of Lemma 8.

Finally, for $0 < \xi < \frac{1}{6}$ and $n > n_0 \geq n_0^{\text{th}}$ we have

$$\mathbb{P}(\neg\text{GBM}) \overset{(a)}{\leq} 7e^{-\frac{(1-\delta)^2}{128}2^{(1-2\xi)n}} \overset{(b)}{\leq} 2^{-\frac{(1-\delta)^2}{128}2^{(1-2\xi)n}} \overset{(c)}{\leq} 2^{-2 \cdot N^{\frac{2}{3}}}.$$

Specifically, inequality (a) holds for $n_0^{\text{th}} \overset{(51)}{\geq} \frac{1}{\xi} \log_2\left(\frac{(1-\delta)^2}{128 \cdot D}\right)$, (b) holds for $n_0^{\text{th}} \overset{(51)}{\geq} \frac{1}{1-2\xi} \log_2\left(\frac{128}{(1-\delta)^2} \cdot \frac{\log_2(7)}{\log_2(e)-1}\right)$, and (c) holds for $n_0^{\text{th}} \overset{(51)}{\geq} \frac{1}{1-2\xi-\frac{2}{3}} \log_2\left(\frac{128 \cdot 2}{(1-\delta)^2}\right)$. ∎

A. Bounding the probability of event $\neg A'$

The following lemma was used in the proof of Lemma 6. In this lemma we develop a bound on $\mathbb{P}(\neg A')$, the probability that "too many" bits were trimmed in $\mathbf{Y}_I$. The bound we reach decays with n (in contrast to the weaker bound in [4, equation (94)] which decays with n_0).

Lemma 8 (Upper bounding $\mathbb{P}(\neg A')$): Let A' be as in (47b). There exists $n_0^{\text{th}}(\xi)$, which is also dependent on the input distribution, such that for $n > n_0 \geq n_0^{\text{th}}(\xi)$

$$\mathbb{P}(\neg A') \leq e^{-D \cdot 2^{(1-\xi)n}},$$

where $D > 0$ is a constant dependent on the input distribution and on the deletion rate δ.

Proof: We consider the event A'', defined as follows. Under the event A'', some index $j < \hat{\ell}$ in $\mathbf{G}_I$ is a '1' bit and was not deleted by the channel (where $\hat{\ell}$ was set in (48)). Clearly $A'' \Rightarrow A'$, hence,

$$\mathbb{P}(\neg A') \leq \mathbb{P}(\neg A'').$$

$\neg A''$ is the complementary event where no index $j < \hat{\ell}$ in $\mathbf{G}_I$ is a '1' bit that was not deleted.

We denote by $\#_{\mathbf{X}}^j$ the number of bits to the left of index j in $g(\mathbf{X}, n_0, \xi)$ that originate from $\mathbf{X}$, and denote by $\#_{\text{GB}}^j$ the number of GB bits to the left of index j. For $n_0 \geq n_0^{\text{th}} \geq \frac{1}{\xi} \log_2\left(1 + \frac{1}{1-2^{-\xi}}\right)$

$$\#_{\mathbf{X}}^j \geq \#_{\text{GB}}^j, \quad \forall j \in \{1, 2, \dots, \Lambda\}. \tag{54}$$

That is, there are more bits from $\mathbf{X}$ than GB bits, for any prefix of $\mathbf{G}$. The proof of (54) is given in Lemma 9. By (54), there are at least $\frac{j}{2}$ bits from $\mathbf{X}$ prior to index j in $\mathbf{G}$, i.e.

$$\#_{\mathbf{X}}^j \geq \frac{j}{2}. \tag{55}$$

For the case of $\mathbf{X}$ distributed according to a regular Markov input distribution with states $\mathcal{S}$, which we assumed is not degenerate, there exists an integer $\tau > 0$ and a probability $0 < p_0 < 1$ s.t. for any state $s \in \mathcal{S}$

$$\mathbb{P}((X_1, X_2, \ldots, X_\tau) = (0, 0, \ldots, 0)|S_0 = s) < p_0 . \quad (56)$$

That is, the probability of a '1' bit in a series of τ bits in $\mathbf{X}$ is greater than $1 - p_0$. For each τ bits in $\mathbf{X}$, the probability of at least one of them being a '1' bit that was not deleted in the channel is greater than

$$(1 - p_0)(1 - \delta) .$$

In our setting, $\mathbf{X}$ consists of blocks of length N_0, each drawn independently from the regular hidden Markov input distribution. We consider two cases: $\#_{\mathbf{X}}^{\lfloor \hat{\ell} \rfloor} \leq N_0$ and $\#_{\mathbf{X}}^{\lfloor \hat{\ell} \rfloor} > N_0$. For the first case, all the bits up to index $\lfloor \hat{\ell} \rfloor$ that originate from $\mathbf{X}$ were drawn according to the underlying regular Markov input distribution (that is, originate from the same block in $\mathbf{X}$). There are $\left\lfloor \#_{\mathbf{X}}^{\lfloor \hat{\ell} \rfloor}/\tau \right\rfloor$ segments of $\mathbf{X}$ bits of length τ up to index $\lfloor \hat{\ell} \rfloor$.

For the case where $\#_{\mathbf{X}}^{\lfloor \hat{\ell} \rfloor} > N_0$, we count the number of segments of τ bits up to index $\lfloor \hat{\ell} \rfloor$ in $\mathbf{G}$ that originate from $\mathbf{X}$ and require that each segment was taken from the same block in $\mathbf{X}$. There are at least

$$\left\lfloor \#_{\mathbf{X}}^{\lfloor \hat{\ell} \rfloor}/N_0 \right\rfloor \cdot \lfloor N_0/\tau \rfloor$$

such segments of length τ. We lower bound by

$$\left\lfloor \#_{\mathbf{X}}^{\lfloor \hat{\ell} \rfloor}/N_0 \right\rfloor \cdot \lfloor N_0/\tau \rfloor \overset{(a)}{\geq} \frac{\#_{\mathbf{X}}^{\lfloor \hat{\ell} \rfloor}}{2N_0} \cdot \frac{N_0}{2\tau} = \frac{\#_{\mathbf{X}}^{\lfloor \hat{\ell} \rfloor}}{4\tau} \geq \left\lfloor \frac{\#_{\mathbf{X}}^{\lfloor \hat{\ell} \rfloor}}{4\tau} \right\rfloor .$$

In (a) we applied $\lfloor x \rfloor \geq \frac{x}{2}$ which holds for $x \geq 1$, which is satisfied for $\#_{\mathbf{X}}^{\lfloor \hat{\ell} \rfloor} > N_0$ and $n_0 \geq n_0^{\text{th}} \geq \log_2 \tau$.

For both cases we get that there are at least $\left\lfloor \frac{\#_{\mathbf{X}}^{\lfloor \hat{\ell} \rfloor}}{4\tau} \right\rfloor$ segments of $\mathbf{X}$ bits of length τ up to index $\lfloor \hat{\ell} \rfloor$. Thus, by the Markov property:

$$\mathbb{P}(\neg A'') \leq (1 - (1 - p_0)(1 - \delta))^{\left\lfloor \frac{\#_{\mathbf{X}}^{\lfloor \hat{\ell} \rfloor}}{4\tau} \right\rfloor}$$

$$\overset{(55)}{\leq} (p_0(1 - \delta) + \delta)^{\left\lfloor \frac{\lfloor \hat{\ell} \rfloor}{8\tau} \right\rfloor}$$

$$\overset{(a)}{=} (p_0(1 - \delta) + \delta)^{\left\lfloor \frac{\hat{\ell}}{8\tau} \right\rfloor} ,$$

where (a) follows from [26, equation (3.11)]. We continue to upper bound the RHS from above:

$$\overset{(48)}{=} (p_0(1 - \delta) + \delta)^{\left\lfloor \frac{(1-\delta)\ell_n}{32\tau} \right\rfloor}$$

$$\overset{(4)}{=} (p_0(1 - \delta) + \delta)^{\left\lfloor \frac{1-\delta}{32\tau} \lfloor 2^{(1-\xi)(n-1)} \rfloor \right\rfloor}$$

$$\overset{(a)}{<} (p_0(1 - \delta) + \delta)^{\left\lfloor \frac{1-\delta}{64\tau} 2^{(1-\xi)(n-1)} \right\rfloor}$$

$$\overset{(b)}{\leq} (p_0(1 - \delta) + \delta)^{\frac{1-\delta}{128\tau} 2^{(1-\xi)(n-1)}}$$

$$= (p_0(1 - \delta) + \delta)^{\frac{1-\delta}{128\tau \cdot 2^{(1-\xi)}} 2^{(1-\xi)n}}$$

$$\leq (p_0(1 - \delta) + \delta)^{\frac{1-\delta}{128 \cdot 2\tau} 2^{(1-\xi)n}}$$

$$= e^{-\ln\left(\frac{1}{p_0(1-\delta)+\delta}\right) \frac{1-\delta}{256\tau} 2^{(1-\xi)n}}$$

where (a) is by $\lfloor 2^x \rfloor > 2^{x-1}$ which holds for any $x \geq 0$. For (b) we apply $\lfloor x \rfloor \geq \frac{x}{2}$ which holds for $x \geq 1$, thus inequality (b) holds for $n > n_0 \geq n_0^{\text{th}} \geq \frac{1}{1-\xi} \log_2\left(\frac{64\tau}{1-\delta}\right) + 1$.

We denote $D \triangleq \frac{1-\delta}{256\tau} \ln\left(\frac{1}{p_0(1-\delta)+\delta}\right)$, where τ and p_0 satisfy (56). We note that $D > 0$, since $0 < p_0(1 - \delta) + \delta < 1$. Finally,

$$\mathbb{P}(\neg A') \leq \mathbb{P}(\neg A'') \leq e^{-D \cdot 2^{(1-\xi)n}} ,$$

for

$$n > n_0 \geq n_0^{\text{th}}$$
$$\triangleq \max\left\{\frac{1}{\xi} \log_2\left(1 + \frac{1}{1 - 2^{-\xi}}\right), \frac{1}{1-\xi} \log_2\left(\frac{64\tau}{1-\delta}\right) + 1\right\} .$$

$\blacksquare$

B. Guard-band presence in $g(\mathbf{X})$

To show (54), we state and prove the following lemma.

Lemma 9 (Majority of $\mathbf{X}$ bits in any prefix of $\mathbf{G}$): If $n_0 \geq \frac{1}{\xi} \log_2\left(1 + \frac{1}{1-2^{-\xi}}\right)$, then for $n > n_0$ and for any given index j in $g(\mathbf{X}, n_0, \xi)$,

$$\#_{\mathbf{X}}^j \geq \#_{\text{GB}}^j ,$$

where $\#_{\mathbf{X}}^j$ is the number of $\mathbf{X}$ bits in the prefix up to j in $g(\mathbf{X}, n_0, \xi)$, and $\#_{\text{GB}}^j$ is the number of GB bits up to j. That is, in any prefix of $\mathbf{G}$ the number of bits from $\mathbf{X}$ is greater or equal to the number of GB bits.

Proof: We divide our proof to three claims.

Claim A.1: Assume to the contrary that there exists an index j_0 for which our lemma does not hold, i.e. $\#_{\mathbf{X}}^{j_0} < \#_{\text{GB}}^{j_0}$. Then, there must exist some index j_1 which is located at the **right edge of some guard-band** that also does not satisfy the lemma, i.e. $\#_{\mathbf{X}}^{j_1} < \#_{\text{GB}}^{j_1}$.

Proof: If j_0 is an index of a GB bit, we may continue to the right edge of the GB containing j_0, making the rightmost index of this GB the desired j_1. This j_1 satisfies:

$$\#_{\mathbf{X}}^{j_1} = \#_{\mathbf{X}}^{j_0} < \#_{\text{GB}}^{j_0} \leq \#_{\text{GB}}^{j_1} .$$

If j_0 is an index of an $\mathbf{X}$ bit, we may continue to the left edge of the block of $\mathbf{X}$ containing j_0, making the rightmost index of the GB to the left of this block the desired j_1. This j_1 satisfies:

$$\#_{\mathbf{X}}^{j_1} < \#_{\mathbf{X}}^{j_0} < \#_{\text{GB}}^{j_0} = \#_{\text{GB}}^{j_1} .$$

$\blacksquare$

Claim A.2: We define index j_{mid} as the rightmost index of the **middle** GB of $g(\mathbf{X})$. We remind that $g(\mathbf{X})$ is created from $N_1 = 2^{n_1}$ blocks of $\mathbf{X}$, each block of length $N_0 = 2^{n_0}$. If $\#_{\mathbf{X}}^{j_{\text{mid}}} \geq \#_{\text{GB}}^{j_{\text{mid}}}$ for all n_1, then,

$$\#_{\mathbf{X}}^j \geq \#_{\text{GB}}^j$$

for any index j in $g(\mathbf{X})$.

Proof: Recall that $g(\mathbf{X})$ was generated according to a given n_0 and n_1. We denote by

$$(n_1, n_0)_{\text{series}}$$

the prefix of $g(\mathbf{X})$ of length j_{mid}. For a general n_1, the full $g(\mathbf{X})$ will be the concatenation:

$$(n_1, n_0)_{\text{series}} \odot (n_1 - 1, n_0)_{\text{series}} \odot \ldots \odot (2, n_0)_{\text{series}} \odot (1, n_0)_{\text{series}} \odot \mathbf{X}(N_1) .$$

See Figure 12, for $g(\mathbf{X}) = (2, n_0)_{\text{series}} \odot (1, n_0)_{\text{series}} \odot \mathbf{X}(4)$.
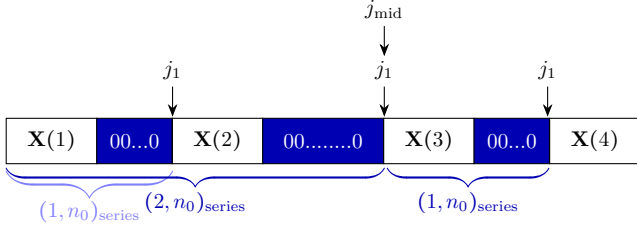

Fig. 12. The $(n_1, n_0)_{\text{series}}$ portions of $g(\mathbf{X})$. In this example $n = n_0 + 2$ and the vector $g(\mathbf{X})$ is a concatenation of $(2, n_0)_{\text{series}}$, $(1, n_0)_{\text{series}}$, and $\mathbf{X}(4)$.

We notice the following property: for each index j_1 located at the right edge of some guard-band in $g(\mathbf{X})$, the series of bits to the left of j_1 are concatenations of the building-blocks

$$\{(i, n_0)_{\text{series}}\}_{i \in \mathcal{J}} \ ,$$

where $\mathcal{J}$ is some subset of $\{1, 2, \ldots, n_1\}$. Therefore, if $\#_{\mathbf{X}}^{j_{\text{mid}}} \geq \#_{\text{GB}}^{j_{\text{mid}}}$ is satisfied for any n_1, then each building-block $(n_1, n_0)_{\text{series}}$ contains at least as many bits from $\mathbf{X}$ as guard-band bits. Thus,

$$\#_{\mathbf{X}}^{j_1} \geq \#_{\text{GB}}^{j_1} \ ,$$

for any rightmost index j_1 of a GB in $g(\mathbf{X})$. By Claim A.1, this leads to: $\#_{\mathbf{X}}^{j} \geq \#_{\text{GB}}^{j}$ $\quad \forall j \in \{1, \ldots, \Lambda\}$. ∎

Claim A.3: For any n_1, $\#_{\mathbf{X}}^{j_{\text{mid}}} \geq \#_{\text{GB}}^{j_{\text{mid}}}$.

Proof: In the series of bits up to j_{mid}, there are half of the bits of $\mathbf{X}$. That is,

$$\#_{\mathbf{X}}^{j_{\text{mid}}} = \frac{1}{2} |\mathbf{X}| = 2^{n-1} \ . \tag{57}$$

Also, up to j_{mid}, there are half of the GB bits of $g(\mathbf{X})$, plus the additional bits from the middle GB,

$$\#_{\text{GB}}^{j_{\text{mid}}} = \frac{1}{2} (\Lambda - N) + \frac{1}{2} \ell_n \ . \tag{58}$$

The total number of GB bits satisfies

$$\begin{aligned} |g(\mathbf{X})| - |\mathbf{X}| \quad &= \quad \Lambda - N \\ &\overset{(4a)}{\leq} \quad 2^n \cdot \left(\frac{2^{-\xi n_0}}{1 - 2^{-\xi}} \right) . \end{aligned} \tag{59}$$

The length of the middle GB satisfies:

$$\ell_n \overset{(4)}{\leq} 2^{(1-\xi)(n-1)} \leq 2^{n-1} \cdot 2^{-\xi n_0} \ , \tag{60}$$

where the last inequality is by $n \geq n_0 + 1$. Thus,

$$\begin{aligned} \#_{\mathbf{X}}^{j_{\text{mid}}} \quad &\overset{(57)}{=} \quad 2^{n-1} \\ &\overset{(a)}{\geq} \quad \frac{1}{2} \cdot 2^n \cdot \left(\frac{2^{-\xi n_0}}{1 - 2^{-\xi}} \right) + 2^{n-1} \cdot 2^{-\xi n_0} \\ &\overset{(59),(60)}{\geq} \quad \frac{1}{2} (\Lambda - N) + \ell_n \\ &\overset{(58)}{\geq} \quad \#_{\text{GB}}^{j_{\text{mid}}} \ , \end{aligned} \tag{61}$$

where (a) is satisfied for $n_0 \geq \frac{1}{\xi} \log_2 \left(1 + \frac{1}{1 - 2^{-\xi}} \right)$ and any $\xi > 0$. ∎

By combining the results from Claims A.2 and A.3, we have proven Lemma. 9. ∎

APPENDIX B
PLUS TRANSFORMS DECREASE THE PROCESS $\bar{Z}_n$

Recall that Claim 6.1 was stated and proved as part of the proof of Lemma 7. That proof relied on Lemma 10, which we prove here. We remind the reader that $B_1, B_2, \ldots$ is a random process with i.i.d. entries, each with distribution Bernoulli$(\frac{1}{2})$. That is,

$$B_n = \begin{cases} 0 \ \text{w.p.} \ \frac{1}{2} \ , \\ 1 \ \text{w.p.} \ \frac{1}{2} \ . \end{cases}$$

We also remind that the process B_n is coupled with the process $\bar{Z}_n$ in (18). In Claim 6.1 we lower bound $\bar{Z}_n$. A pivotal part in lower bounding $\bar{Z}_n$ is showing that the event of drawing only 1's in the process B_n minimizes the process $\bar{Z}_n$. This will be proven herein.

For this, we denote by $\Sigma_1(n)$ the event of drawing only 1's up to n, starting from n_0. The parameter n_0 is the starting index of the process $\bar{Z}_n$ (18). That is,

$$\Sigma_1(n) = \{B_{i+1} = 1, \quad \forall n_0 \leq i \leq n - 1\} \ . \tag{62a}$$

The following lemma shows that under this event, $\bar{Z}_n$ is minimized.

Lemma 10: Let $\bar{Z}_{n_0}, \bar{Z}_{n_0+1}, \ldots$ be the process defined in (18). We denote by $\bar{Z}_n^{\Sigma_1}$ the value of $\bar{Z}_n$ under event $\Sigma_1(n)$ in (62a). Note that $\bar{Z}_n^{\Sigma_1}$ is a constant. Let $\eta_a = \frac{1}{2 \cdot \kappa}$ as in (28). For $\eta \leq \eta_a$ and for any $n \geq n_0$:

$$\bar{Z}_n \geq \bar{Z}_n^{\Sigma_1} \quad \text{with probability 1} \ .$$

Proof: We denote the value of $\bar{Z}_{n+1}$ when $B_{n+1} = 1$ ('+') as $\bar{Z}_{n+1}^+$, and the value of $\bar{Z}_{n+1}$ when $B_{n+1} = 0$ ('−') as $\bar{Z}_{n+1}^-$.
Sub-Claim B.1:

$$\text{For } \eta \leq \eta_a : \quad \bar{Z}_n \leq \eta \Rightarrow \bar{Z}_{n+1}^+ \leq \eta \ . \tag{63}$$

Proof:

$$\begin{aligned} \bar{Z}_{n+1}^+ \quad &\overset{(18a)}{=} \quad 2 \cdot \kappa \cdot \bar{Z}_n^2 \\ &\leq \quad 2 \cdot \kappa \cdot \eta^2 \\ &\leq \quad 2 \cdot \kappa \cdot \eta_a \cdot \eta \\ &\overset{(28)}{=} \quad \eta \ . \end{aligned}$$

Sub-Claim B.2: For $\eta \leq \eta_a$:

$$\bar{Z}_n^{\Sigma_1} \leq \eta, \quad \forall n \geq n_0 \ .$$

Sub-claim B.2 may be easily proven by induction, where the induction basis is given by our starting point at $\bar{Z}_{n_0} \overset{(18b)}{=} \eta$. The induction step is given by (63).

Sub-Claim B.3: Under $\Sigma_1(n)$, for $\eta \leq \eta_a$:

$$\bar{Z}_{n+1}^+ \leq \bar{Z}_n < \bar{Z}_{n+1}^- \ . \tag{64}$$

The '−' case in Sub-claim B.3 is true since $\kappa \geq 1$ and since the process $\bar{Z}_n$ is positive for all n:

$$\bar{Z}_n < 2 \cdot \kappa \cdot \bar{Z}_n \overset{(18a)}{=} \bar{Z}_{n+1}^- \ .$$

For the '+' case, we must show that for $\eta \leq \eta_a$ and under $\Sigma_1(n)$,

$$\bar{Z}_{n+1}^+ \overset{(18a)}{=} 2 \cdot \kappa \bar{Z}_n^2 \leq \bar{Z}_n \ .$$

Hence, it suffices to prove that under $\Sigma_1(n)$:

$$2 \cdot \kappa \cdot \bar{Z}_n \le 1 \ .$$

We use Sub-claim B.2 to recall that $\bar{Z}_n^{\Sigma_1} \le \eta$. Recalling again that $\eta \overset{(28)}{\le} \frac{1}{2\cdot\kappa}$, we deduce the above.

Now to the proof of our lemma. We will prove by contradiction. Assume that $\bar{Z}_n$ is minimized for some series of draws $\mathcal{B} = \{B_{i+1}\}_{i=n_0}^{n-1}$ which is not all 1's. For this series, for some i: $B_{i+1} = 0$ ('−'). We denote the first index in $\mathcal{B}$ where 0 was drawn as i_0, i.e,

$$B_{i_0+1} = 0 \ (\text{'−'}) \ \text{ and } \ B_{i+1} = 1 \ (\text{'+'}) \ \ \forall n_0 \le i < i_0 \ .$$

Notice that up to i_0 we have performed only '+' operations. Thus, by Sub-claim B.2, $\bar{Z}_{i_0} \le \eta$. We now replace $B_{i_0+1} = 0$ with $B_{i_0+1} = 1$. By (64), $\bar{Z}_{i_0+1}$ is now smaller. We perform all other transforms dictated from $\mathcal{B}$ without change. The key thing to note is that both operations in (18a) are monotonically increasing in $\bar{Z}_n$. Thus, by a simple induction argument, $\bar{Z}_n$ will also be decreased by our change at step i_0. We arrive at a contradiction to $\bar{Z}_n$ being minimized by the original series $\mathcal{B}$. Hence, the series minimizing $\bar{Z}_n$ is $B_{i+1} = 1$ for all $n_0 \le i \le n - 1$ (the draw of event $\Sigma_1(n)$). ∎

APPENDIX C
AUXILIARY THRESHOLD CALCULATION FOR THEOREM 1

The following lemma, combined with Lemma 4, proves the rightmost inequalities in (1) and (2) for a large enough n. This lemma is used in Section VII, i.e. in the proof of our main theorem, Theorem 1.

Lemma 11: Fix $\beta > \beta' > 0$ and $\epsilon \in (0,1)$. Let $N \triangleq 2^n$. There exists a threshold $n^{\text{th}}(\beta, \beta', \epsilon)$ such that if $n \ge n^{\text{th}}$ then

$$2^{-N^\beta} < \frac{1}{2N} \cdot 2^{-(\frac{N}{1-\epsilon})^{\beta'}} \ .$$

Proof: We set

$$n^{\text{th}}(\beta, \beta', \epsilon) \triangleq$$
$$\max\left\{ \frac{1}{(\ln(2)\cdot\beta')^2} \ , \ 1 + \frac{1}{\beta - \beta'} \log_2\left(1 + \frac{1}{(1-\epsilon)^{\beta'}}\right) \right\} . \quad (65)$$

For $n \ge n^{\text{th}}$:

$$2^{-N^\beta} \quad < \frac{1}{2N} \cdot 2^{-(\frac{N}{1-\epsilon})^{\beta'}}$$

$$\overset{-\log_2(\cdot)}{\Leftrightarrow} \quad 2^{\beta n} \quad > \frac{1}{(1-\epsilon)^{\beta'}} \cdot 2^{\beta' n} + 1 + n$$

$$\overset{(a)}{\Leftarrow} \quad 2^{\beta n} \quad > \frac{1}{(1-\epsilon)^{\beta'}} \cdot 2^{\beta' n} + 1 + (\ln(2)\beta' n)^2$$

$$\overset{(b)}{\Leftarrow} \quad 2^{\beta n} \quad > \frac{1}{(1-\epsilon)^{\beta'}} \cdot 2^{\beta' n} + e^{\ln(2)\beta' n}$$

$$\Leftrightarrow \quad 2^{\beta n} \quad > 2^{\beta' n} \cdot \left(1 + \frac{1}{(1-\epsilon)^{\beta'}}\right)$$

$$\Leftrightarrow \quad 2^{(\beta-\beta') n} > \left(1 + \frac{1}{(1-\epsilon)^{\beta'}}\right)$$

$$\Leftarrow \quad n \quad \ge n^{\text{th}} \overset{(65)}{>} \frac{1}{\beta - \beta'} \log_2\left(1 + \frac{1}{(1-\epsilon)^{\beta'}}\right) .$$

Transition (a) is satisfied for $n \ge n^{\text{th}} \overset{(65)}{\ge} \frac{1}{(\ln(2)\cdot\beta')^2}$. For transition (b), recall that by Taylor's theorem applied to $f(x) = e^x$, for $x \ge 0$ we have:

$$e^x = 1 + x + \frac{x^2}{2} + E_2(x), \quad (66)$$

where $E_2(x) \triangleq \frac{f^{(3)}(c)}{3!} x^3 = \frac{e^c}{3!} x^3$ for $c \in [0, x]$ [27, p. 283]. Since all the terms on the RHS of (66) are nonnegative, we have:

$$e^x \ge 1 + x^2, \quad \forall x \ge 0 \ . \quad (67)$$

That is, for $x = \ln(2)\beta' n \ge 0$ inequality (67) yields transition (b). ∎

REFERENCES

[1] H. Mercier, V. K. Bhargava, and V. Tarokh, "A survey of error-correcting codes for channels with symbol synchronization errors," *IEEE Communications Surveys Tutorials*, vol. 12, no. 1, pp. 87–96, 2010.

[2] R. Heckel, G. Mikutis, and R. N. Grass, "A characterization of the dna data storage channel," *Scientific Reports*, vol. 9, 2019.

[3] 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.

[4] 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.

[5] E. K. Thomas, V. Y. F. Tan, A. Vardy, and M. Motani, "Polar coding for the binary erasure channel with deletions," *IEEE Communications Letters*, vol. 21, pp. 710–713, 2017.

[6] K. Tian, A. Fazeli, and A. Vardy, "Polar coding for deletion channels: Theory and implementation," in *Proc. IEEE Int'l Symp. Inform. Theory (ISIT'2017)*, Aachen, Germany, 2017, pp. 1869–1873.

[7] K. Tian, A. Fazeli, A. Vardy, and R. Liu, "Polar codes for channels with deletions," in *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2017, pp. 572–579.

[8] K. Tian, A. Fazeli, and A. Vardy, "Polar coding for channels with deletions," *IEEE Trans. Inform. Theory*, vol. 67, no. 11, pp. 7081–7095, November 2021.

[9] Y. Li and V. Y. F. Tan, "On the capacity of channels with deletions and states," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2663–2679, 2021.

[10] B. Shuval and I. Tal, "Fast polarization for processes with memory," *IEEE Transactions on Information Theory*, vol. 65, no. 4, pp. 2004–2020, 2019.

[11] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 7829–7838, 2013.

[12] S. B. Korada and R. L. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1751–1768, 2010.

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

[14] 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.

[15] R. Wang, R. Liu, and Y. Hou, "Joint successive cancellation decoding of polar codes over intersymbol interference channels," *CoRR*, vol. abs/1404.3001, 2014. [Online]. Available: http://arxiv.org/abs/1404.3001

[16] M. Mondelli, S. H. Hassani, and R. Urbanke, "How to achieve the capacity of asymmetric channels," *IEEE Trans. Inform. Theory*, vol. 64, no. 5, pp. 3371–3393, May 2018.

[17] A. Chou and M. R. Bloch, "Using deterministic decisions for low-entropy bits in the encoding and decoding of polar codes," in *Proc. Annu. Allerton Conf. Commun. Control Computing (Allerton'2015)*, New York, NY, USA, 2015.

[18] E. Arıkan and E. Telatar, "On the rate of channel polarization," in *Proc. IEEE Int'l Symp. Inform. Theory (ISIT'2009)*, Seoul, South Korea, 2009, pp. 1493–1495.

[19] D. Arava and I. Tal, "Stronger polarization for the deletion channel," in *Proc. IEEE Int'l Symp. Inform. Theory (ISIT'2023)*, Taipai, Taiwan, 2023.

[20] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge, UK: Cambridge University Press, 2008.

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

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

[23] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomizition and Probabilistic Techniques in Algorithms and Data Analysis*, 2nd ed. Cambridge, UK: Cambridge University Press, 2017.

[24] B. Shuval and I. Tal, "Universal polarization for processes with memory," 2018. [Online]. Available: https://arxiv.org/abs/1811.05727v4

[25] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge, UK: Cambridge University Press, 2005.

[26] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, 2nd ed. Reading, Massachusetts: Addison-Wesley, 1994.

[27] T. M. Apostol, *Calculus*, 2nd ed. Wiley, 1967, vol. 1.