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

A. The deletion channel

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.

B. Polar codes for the deletion channel

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 $2^{-\Lambda^{\nu'}}$. For completeness, the authors 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 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 for a more elaborate decoding scheme, the error probability decreases as $2^{-\Lambda^{\beta'}}$ where $\beta' \in (0, \frac{1}{2})$ instead of the previous decay coefficient $\nu' \in (0, \frac{1}{3})$.

Supporting grants: 2018218 (BSF), CIF-2212437 (NSF), and DIP (DFG).

II. Main theorem

Our main result builds upon the 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. As we will see, the proof uses a previously proven "weaker polarization" theorem as a bootstrap.

Theorem 1 (Stronger Polarization): Fix $\epsilon > 0$, $\xi \in (0, \frac{1}{6})$, and parameters $0 < \beta' < \beta < \frac{1}{2}$. There exist $n^{\text{th}}(\beta', \beta, \epsilon, \xi)$ and $n_0^{\text{th}}(\beta', \beta, \epsilon, \xi)$ such that the following holds. Take $n \geq n^{\text{th}}$ and $n_0 \geq n_0^{\text{th}}$. 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 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:

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

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

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.

Here is the "weaker polarization" theorem which we will build on. This theorem follows from the proof of [4, Theorem 1], and by recalling that the Bhattacharyya parameter is upper bounded by twice the square-root of the probability of error [13, by combining (4a) and (4c)].

Theorem 3 (Prior-art Polarization): Fix $\epsilon' > 0$ and $0 < \nu < \frac{1}{3}$. There exists an $n^{\text{pa-th}}(\nu, \epsilon', \xi)$ such that the following holds. Take $n \geq n^{\text{pa-th}}$ and $n_0 = \lfloor \nu n \rfloor$. Let $\mathbf{U}$, $\mathbf{X}$, $\mathbf{Y}$, N, and Λ be as in Theorem 1. The fraction of indices i for which:

$$Z(U_i | U_1^{i-1}, \mathbf{Y}) < 2^{-N^\nu} \quad (3)$$

$$K(U_i | U_1^{i-1}) < 2^{-N^\nu} \quad (4)$$

is at least $\mathcal{I} - \epsilon'$. Furthermore, $\frac{N}{\Lambda} > 1 - \epsilon'$.

To recap: our stronger result promises a probability of error that decays roughly like the square root of the codeword length, as is the case for the seminal polar codes defined for BMS channels [3]. In contrast, prior art only promises a probability of error that decays roughly like the cube root of the codeword length.

III. Notation

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

A. Three related channels

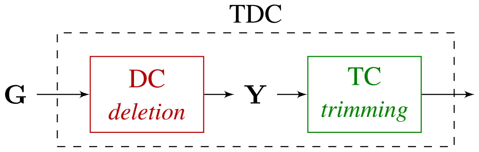
We now introduce three related channels: the deletion channel; the trimming channel; and their composition, the trimmed deletion channel.

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 $\mathbf{Y}^*$.



We end this subsection by noting that Theorem 3 holds for the TDC as well. This follows by carefully reading the proof of [4, Theorem 1], and noting that in [4, Subclaim 2] the initial step involves trimming $\mathbf{Y}$ into $\mathbf{Y}^*$.

Remark 4 (Prior-art Polarization, for the TDC): Theorem 3 continues to hold if we replace $\mathbf{Y}$ with $\mathbf{Y}^*$ in (3).

B. Blocks and guard-bands

Recall that in the previous theorems, $\mathbf{X}$ 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 $\mathbf{X}_{\mathrm{I}}$ and $\mathbf{X}_{\mathrm{II}}$. Denoting the length of a vector by $|\cdot|$, we have $|\mathbf{X}_{\mathrm{I}}| = |\mathbf{X}_{\mathrm{II}}| = \frac{N}{2}$ and

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

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

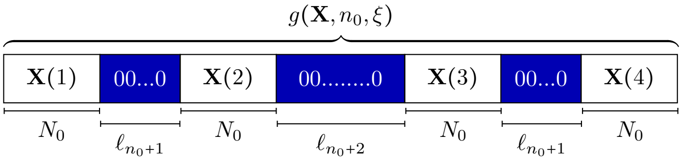
Recall that the function g mentioned previously transforms a vector $\mathbf{X}$ of length 2^n into a slightly longer vector with "guard-bands". We now define g recursively, and note that it adds the guard-bands between blocks. For a vector $\mathbf{X}$ of length $\leq 2^{n_0}$, $g(\mathbf{X}, n_0, \xi)$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

$$g(\mathbf{X}, n_0, \xi) \triangleq g(\mathbf{X}) \triangleq \underbrace{g(\mathbf{X}_{\mathrm{I}})}_{\triangleq \mathbf{G}_{\mathrm{I}}} \odot \underbrace{\overbrace{000\ldots00}^{\ell_n}}_{\triangleq \mathbf{G}_{\Delta}} \odot \underbrace{g(\mathbf{X}_{\mathrm{II}})}_{\triangleq \mathbf{G}_{\mathrm{II}}} \quad (5)$$

That is, we add

$$\ell_n = \left\lfloor 2^{(1-\xi)(n-1)} \right\rfloor \quad (6)$$

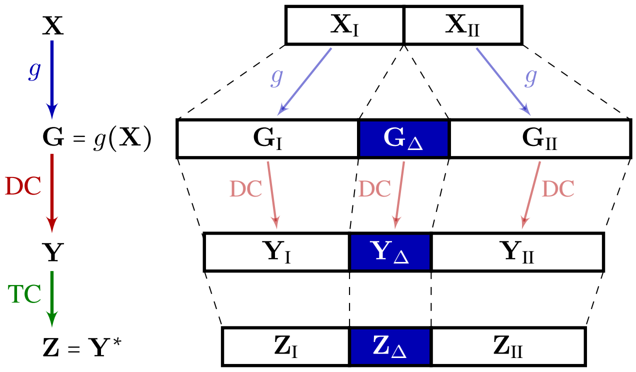
"0" symbols between the first and second halves of $\mathbf{X}$, and apply g recursively on each half. Note that $\xi > 0$ is a "small" constant that we will define later. To summarize: $\mathbf{X}$ is a concatenation of 2^{n-n_0} independent blocks, each of length $N_0 = 2^{n_0}$. The function $g(\mathbf{X}, n_0, \xi)$ 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 remind the reader that $\mathbf{G} = \mathbf{G}_{\mathrm{I}} \odot \mathbf{G}_{\Delta} \odot \mathbf{G}_{\mathrm{II}}$ is passed through the DC. We denote the output of this channel by $\mathbf{Y}$, and denote the parts corresponding to $\mathbf{G}_{\mathrm{I}}$, $\mathbf{G}_{\Delta}$, and $\mathbf{G}_{\mathrm{II}}$ by $\mathbf{Y}_{\mathrm{I}}$, $\mathbf{Y}_{\Delta}$, and $\mathbf{Y}_{\mathrm{II}}$, respectively. We further denote the application of the TC on $\mathbf{Y}$ by $\mathbf{Z} \triangleq \mathbf{Y}^*$, and denote the parts corresponding to $\mathbf{Y}_{\mathrm{I}}$, $\mathbf{Y}_{\Delta}$, and $\mathbf{Y}_{\mathrm{II}}$ by $\mathbf{Z}_{\mathrm{I}}$, $\mathbf{Z}_{\Delta}$, and $\mathbf{Z}_{\mathrm{II}}$, respectively. See Fig. 1, which is essentially [4, Figure 5]. Note that, in general, $\mathbf{Z}_{\mathrm{I}}$ is formed by trimming off only the left side of $\mathbf{Y}_{\mathrm{I}}$. Hence, typically, $\mathbf{Z}_{\mathrm{I}} \neq (\mathbf{Y}_{\mathrm{I}})^*$ and $\mathbf{Z}_{\mathrm{II}} \neq (\mathbf{Y}_{\mathrm{II}})^*$. Also, we note that in the typical case, $\mathbf{Z}_{\Delta} = \mathbf{Y}_{\Delta}$.

IV. Two key lemmas

In this section we state the two lemmas that are key to our main result. As we will see, the first lemma is specific to our setting, while the second is more general.



$\mathbf{X}$
g
$\mathbf{G} = g(\mathbf{X})$
DC
$\mathbf{Y}$
TC
$\mathbf{Z} = \mathbf{Y}^*$
$\mathbf{X}_{\mathrm{I}}$
$\mathbf{X}_{\mathrm{II}}$
g
g
$\mathbf{G}_{\mathrm{I}}$
$\mathbf{G}_{\triangle}$
$\mathbf{G}_{\mathrm{II}}$
DC
DC
DC
$\mathbf{Y}_{\mathrm{I}}$
$\mathbf{Y}_{\triangle}$
$\mathbf{Y}_{\mathrm{II}}$
$\mathbf{Z}_{\mathrm{I}}$
$\mathbf{Z}_{\triangle}$
$\mathbf{Z}_{\mathrm{II}}$

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

A. Single-step bounds for the TDC

In the seminal paper [3, Proposition 5], it was shown that a '+' transform squares the Bhattacharyya parameter, while a '−' transform at most doubles it. This was the key property used to prove strong polarization in [14]. We will soon state a similar claim for our setting. Our claim is significantly weaker than the one derived for a memoryless channel in [3] and also from the one derived for a Markovian setting in [15, Section VI], but still strong enough to imply strong polarization.

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 $b_1, \dots, b_n$. That is, for $1 \le i \le N = 2^n$,

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

The following lemma is cardinal to proving the stronger polarization stated in Theorem 1. Recall that δ is the deletion rate, and that the guard-band length is given in (6), and is a function of ξ. The proof will be given in Section V.

Lemma 5 (Bhattacharyya single-step bounds for 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 $\mathbf{Y}^* = \mathbf{Z}_{\mathrm{I}} \odot \mathbf{Z}_{\triangle} \odot \mathbf{Z}_{\mathrm{II}}$ be the result of transmitting $g(\mathbf{X}, n_0, \xi)$ through the TDC. There exist $m_0^{\mathrm{th}}(\xi)$ and $m^{\mathrm{th}}(\xi, \delta)$ s.t. for $n_0 \ge m_0^{\mathrm{th}}$ and all $n \ge \max\{m^{\mathrm{th}}, n_0 + 1\}$ the following holds. Let $1 \le i \le N$ and $j = \lfloor (i+1)/2 \rfloor$. Then,

$$Z(U_i | U_1^{i-1}, \mathbf{Y}^*) \le \frac{3}{2} N \cdot Z(U_i | U_1^{i-1}, \mathbf{Z}_{\mathrm{I}}^*, \mathbf{Z}_{\mathrm{II}}^*) + 2^{-N^{\frac{2}{3}}} \quad \text{(7a)}$$

$$\le \begin{cases} \frac{3}{2} N \cdot 2 \cdot Z(V_j | V_1^{j-1}, \mathbf{Z}_{\mathrm{I}}^*) + 2^{-N^{\frac{2}{3}}} & \text{if } b_n = 0 \text{ ('−')} \\ \frac{3}{2} N \cdot Z(V_j | V_1^{j-1}, \mathbf{Z}_{\mathrm{I}}^*)^2 + 2^{-N^{\frac{2}{3}}} & \text{if } b_n = 1 \text{ ('+')} \ . \end{cases} \quad \text{(7b)}$$

We draw the reader's attention to several important points. First, note that in (7a), there is both an additive penalty of $2^{-N^{\frac{2}{3}}}$ as well as a multiplicative penalty of $\frac{3N}{2}$, 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 in (7b) shows us why such a penalty is worth paying: since $\mathbf{Z}_{\mathrm{I}}^*$ and $\mathbf{Z}_{\mathrm{II}}^*$ are independent, we may now use the standard arguments in [3] to reach a recursive relation. To conclude, the lemma allows us to track the evolution of the Bhattacharyya parameter after each polarization step.

B. The walking-to-running lemma

In the previous subsection, we've stated Lemma 5, which gave upper bounds on the evolution of the Bhattacharyya parameter. Due to the added penalties in these bounds, we cannot use prior art in order to claim a polarization rate of roughly $2^{-\sqrt{N}}$. Indeed, in this subsection we state the second key lemma in the paper, Lemma 6, which implies such a rate for the process in Lemma 5. Lemma 6 is stated quite generally, in the hope that it will be useful to other settings. We have termed it the "walking-to-running" lemma, since we show that if we have "walking-speed" polarization (for example, $\approx 2^{-\sqrt[3]{N}}$) at some stage of the process, this implies "running-speed" polarization ($\approx 2^{-\sqrt{N}}$) during later stages. In our setting, the "walking-speed" is guaranteed by [4, Theorem 1].

Lemma 6 (walking-to-running): Let $B_1, B_2, \dots$ be i.i.d. uniformly distributed Bernoulli random variables. Fix constants $\kappa \ge 1, d \ge 0, \gamma > \frac{1}{2}$ and $m^{\mathrm{th}} > 0$. Let $Z_0, Z_1, Z_2, \dots$ be a random process s.t. for all $n \ge m^{\mathrm{th}}$,

$$Z_{n+1} \le \begin{cases} \kappa N^d \cdot Z_n + 2^{-N^\gamma} & \text{if } B_{n+1} = 0 \text{ ('−')} \\ \kappa N^d \cdot Z_n^2 + 2^{-N^\gamma} & \text{if } B_{n+1} = 1 \text{ ('+')} \ . \end{cases} \quad (8)$$

Fix $\beta \in (0, \frac{1}{2})$, the "running speed" parameter, and $\nu > 0$, the "walking speed" parameter. For all $\epsilon' > 0$ there exists a threshold $n_{\mathrm{w}}^{\mathrm{th}} = n_{\mathrm{w}}^{\mathrm{th}}(\epsilon', \beta, \nu, \kappa, d, \gamma, m^{\mathrm{th}}) \ge m^{\mathrm{th}}$ such that if for some $n_{\mathrm{w}} \ge n_{\mathrm{w}}^{\mathrm{th}}$ we are assured "walking speed":

$$Z_{n_{\mathrm{w}}} \le 2^{-(2^{n_{\mathrm{w}}})^\nu} \ , \quad (9)$$

then there exists $n_{\mathrm{r}}^{\mathrm{th}} = n_{\mathrm{r}}^{\mathrm{th}}(\epsilon', \beta, \nu, \kappa, d, n_{\mathrm{w}}) > n_{\mathrm{w}}$ such that above this threshold, with high probability, we are indefinitely at "running speed":

$$\mathbb{P}\left(Z_n < 2^{-N^\beta}, \quad \forall n \ge n_{\mathrm{r}}^{\mathrm{th}}\right) \ge 1 - \epsilon' \ . \quad (10)$$

V. Proof of Lemma 5

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 after trimming the output, it holds that the middle symbol originated from the outermost guard-band. In the second part, we show that under GBM, we have a recursive relation for Z similar to the memoryless case, up to an extra multiplicative factor of $\frac{3N}{2}$, see (7b). In the third part, we show that the GBM event is

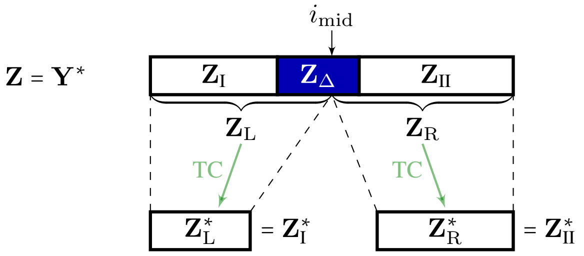

i_{mid}
$\mathbf{Z} = \mathbf{Y}^*$
$\mathbf{Z}_{\mathrm{I}}$
$\mathbf{Z}_{\triangle}$
$\mathbf{Z}_{\mathrm{II}}$
$\mathbf{Z}_{\mathrm{L}}$
$\mathbf{Z}_{\mathrm{R}}$
TC
TC
$\mathbf{Z}_{\mathrm{L}}^*$
$= \mathbf{Z}_{\mathrm{I}}^*$
$\mathbf{Z}_{\mathrm{R}}^*$
$= \mathbf{Z}_{\mathrm{II}}^*$

Fig. 2. The GBM event.

very likely. That is, the additive penalty of $2^{-N^{\frac{2}{3}}}$ in (7) comes from bounding the probability that GBM does not occur.

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 $\mathbf{Y}^* = \mathbf{Z} = \mathbf{Z}_{\mathrm{I}} \odot \mathbf{Z}_{\triangle} \odot \mathbf{Z}_{\mathrm{II}}$ is the result of passing $\mathbf{G}_{\mathrm{I}} \odot \mathbf{G}_{\triangle} \odot \mathbf{G}_{\mathrm{II}}$ through the TDC. The GBM event occurs if $\mathbf{Z}$ is not empty and its middle index $i_{\mathrm{mid}} \triangleq \left\lfloor \frac{|\mathbf{Z}|+1}{2} \right\rfloor$, falls within $\mathbf{Z}_{\triangle}$. That is, GBM occurs if $Z_{i_{\mathrm{mid}}}$ originates from the outermost guard-band $\mathbf{G}_{\triangle}$. The complementary event is denoted ¬GBM.

We denote the left and right halves of $\mathbf{Z} = \mathbf{Y}^*$ as $\mathbf{Z}_{\mathrm{L}} = (Z_1, \ldots, Z_{i_{\mathrm{mid}}})$ and $\mathbf{Z}_{\mathrm{R}} = (Z_{i_{\mathrm{mid}}+1}, \ldots, Z_{|\mathbf{Z}|})$, see Figure 2. The main utility of the GBM event is this (again, see Figure 2): since $\mathbf{Z}_{\triangle}$ contains only '0' symbols, under GBM

$$(\mathbf{Z}_{\mathrm{L}})^* = (\mathbf{Z}_{\mathrm{I}})^* , \tag{11a}$$
$$(\mathbf{Z}_{\mathrm{R}})^* = (\mathbf{Z}_{\mathrm{II}})^* . \tag{11b}$$

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

$$L_0 \triangleq |\mathbf{Y}^*| - |\mathbf{Z}_{\mathrm{I}}^*| - |\mathbf{Z}_{\mathrm{II}}^*| .$$

Since $\mathbf{Y}^*$ does not contain leading nor trailing '0' symbols, L_0 equals the sum of the following: the number of trailing '0' symbols in $\mathbf{Z}_{\mathrm{I}}$, the length of $\mathbf{Z}_{\triangle}$, and the number of leading '0' symbols in $\mathbf{Z}_{\mathrm{II}}$. Hence,

$$\mathbf{Y}^* = \mathbf{Z}_{\mathrm{I}}^* \odot \overbrace{000\ldots00}^{L_0} \odot \mathbf{Z}_{\mathrm{II}}^* . \tag{12}$$

Thus, by (12) and (11):

$$\mathrm{GBM} \Rightarrow \mathbf{Y}^* = (\mathbf{Z}_{\mathrm{L}})^* \odot \overbrace{000\ldots00}^{L_0} \odot (\mathbf{Z}_{\mathrm{R}})^* . \tag{13}$$

B. *Bounding the Bhattacharyya parameter using* GBM

In this subsection, we derive an upper bound on the Bhattacharyya parameter corresponding to an index i. In order to save space we use the following shorthand in the upcoming probability expressions: u_1^{i-1} is short for $U_1^{i-1} = u_1^{i-1}$, $\mathbf{z}$ is short for $\mathbf{Y}^* = \mathbf{z}$, and 0 and 1 are short for $U_i = 0$ and $U_i = 1$, respectively. To illustrate, we use both the long and short notation in the following expression for the Bhattacharyya parameter corresponding to index i.

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

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

$$\begin{aligned} &= \sum_{u_1^{i-1}, \mathbf{z}} \sqrt{\begin{array}{l} \left(\mathbb{P}(0, u_1^{i-1}, \mathbf{z}, \mathrm{GBM}) + \mathbb{P}(0, u_1^{i-1}, \mathbf{z}, \neg\mathrm{GBM})\right) \\ \times \left(\mathbb{P}(1, u_1^{i-1}, \mathbf{z}, \mathrm{GBM}) + \mathbb{P}(1, u_1^{i-1}, \mathbf{z}, \neg\mathrm{GBM})\right) \end{array}} \\ &\leq \sum_{u_1^{i-1}, \mathbf{z}} \sqrt{\mathbb{P}(0, u_1^{i-1}, \mathbf{z}, \mathrm{GBM}) \cdot \mathbb{P}(1, u_1^{i-1}, \mathbf{z}, \mathrm{GBM})} \qquad \text{(14a)} \\ &\quad + \sum_{u_1^{i-1}, \mathbf{z}} \sqrt{\begin{array}{l} \mathbb{P}(0, u_1^{i-1}, \mathbf{z}, \mathrm{GBM}) \cdot \mathbb{P}(1, u_1^{i-1}, \mathbf{z}, \neg\mathrm{GBM}) \\ + \mathbb{P}(0, u_1^{i-1}, \mathbf{z}, \neg\mathrm{GBM}) \cdot \mathbb{P}(1, u_1^{i-1}, \mathbf{z}, \mathrm{GBM}) \\ + \mathbb{P}(0, u_1^{i-1}, \mathbf{z}, \neg\mathrm{GBM}) \cdot \mathbb{P}(1, u_1^{i-1}, \mathbf{z}, \neg\mathrm{GBM}) \end{array}} \qquad \text{(14b)} \end{aligned}$$

We will bound both the sum in (14a) and the sum in (14b). For the sum in (14a), we have:

$$\begin{aligned} &\sum_{u_1^{i-1}, \mathbf{z}} \sqrt{\mathbb{P}(0, u_1^{i-1}, \mathbf{Y}^* = \mathbf{z}, \mathrm{GBM}) \cdot \mathbb{P}(1, u_1^{i-1}, \mathbf{Y}^* = \mathbf{z}, \mathrm{GBM})} \\ &\overset{(13)}{=} \sum_{u_1^{i-1}, \mathbf{z}} \sqrt{\begin{array}{l} \mathbb{P}\left(0, u_1^{i-1}, \mathbf{Z}_{\mathrm{L}}^* {=} \mathbf{z}_{\mathrm{L}}^*, \mathbf{Z}_{\mathrm{R}}^* {=} \mathbf{z}_{\mathrm{R}}^*, L_0 {=} |\mathbf{z}| {-} |\mathbf{z}_{\mathrm{L}}^*| {-} |\mathbf{z}_{\mathrm{R}}^*|, \mathrm{GBM}\right) \\ \times \mathbb{P}\left(1, u_1^{i-1}, \mathbf{Z}_{\mathrm{L}}^* {=} \mathbf{z}_{\mathrm{L}}^*, \mathbf{Z}_{\mathrm{R}}^* {=} \mathbf{z}_{\mathrm{R}}^*, L_0 {=} |\mathbf{z}| {-} |\mathbf{z}_{\mathrm{L}}^*| {-} |\mathbf{z}_{\mathrm{R}}^*|, \mathrm{GBM}\right) \end{array}} \\ &\overset{(a)}{=} \sum_{\ell=1}^{\frac{3}{2}N} \sum_{u_1^{i-1}, \mathbf{z}_{\mathrm{L}}^*, \mathbf{z}_{\mathrm{R}}^*} \sqrt{\begin{array}{l} \mathbb{P}(0, u_1^{i-1}, \mathbf{Z}_{\mathrm{L}}^* {=} \mathbf{z}_{\mathrm{L}}^*, \mathbf{Z}_{\mathrm{R}}^* {=} \mathbf{z}_{\mathrm{R}}^*, L_0 {=} \ell, \mathrm{GBM}) \\ \times \mathbb{P}(1, u_1^{i-1}, \mathbf{Z}_{\mathrm{L}}^* {=} \mathbf{z}_{\mathrm{L}}^*, \mathbf{Z}_{\mathrm{R}}^* {=} \mathbf{z}_{\mathrm{R}}^*, L_0 {=} \ell, \mathrm{GBM}) \end{array}} \\ &\leq \sum_{\ell=1}^{\frac{3}{2}N} \sum_{u_1^{i-1}, \mathbf{z}_{\mathrm{L}}^*, \mathbf{z}_{\mathrm{R}}^*} \sqrt{\begin{array}{l} \mathbb{P}(0, u_1^{i-1}, \mathbf{Z}_{\mathrm{L}}^* = \mathbf{z}_{\mathrm{L}}^*, \mathbf{Z}_{\mathrm{R}}^* = \mathbf{z}_{\mathrm{R}}^*, \mathrm{GBM}) \\ \times \mathbb{P}(1, u_1^{i-1}, \mathbf{Z}_{\mathrm{L}}^* = \mathbf{z}_{\mathrm{L}}^*, \mathbf{Z}_{\mathrm{R}}^* = \mathbf{z}_{\mathrm{R}}^*, \mathrm{GBM}) \end{array}} \\ &\overset{(11)}{=} \frac{3}{2}N \cdot \sum_{u_1^{i-1}, \mathbf{z}_{\mathrm{I}}^*, \mathbf{z}_{\mathrm{II}}^*} \sqrt{\begin{array}{l} \mathbb{P}(0, u_1^{i-1}, \mathbf{Z}_{\mathrm{I}}^* {=} \mathbf{z}_{\mathrm{I}}^*, \mathbf{Z}_{\mathrm{II}}^* {=} \mathbf{z}_{\mathrm{II}}^*, \mathrm{GBM}) \\ \times \mathbb{P}(1, u_1^{i-1}, \mathbf{Z}_{\mathrm{I}}^* {=} \mathbf{z}_{\mathrm{I}}^*, \mathbf{Z}_{\mathrm{II}}^* {=} \mathbf{z}_{\mathrm{II}}^*, \mathrm{GBM}) \end{array}} \\ &\leq \frac{3}{2}N \cdot \sum_{u_1^{i-1}, \mathbf{z}_{\mathrm{I}}^*, \mathbf{z}_{\mathrm{II}}^*} \sqrt{\begin{array}{l} \mathbb{P}(0, u_1^{i-1}, \mathbf{Z}_{\mathrm{I}}^* = \mathbf{z}_{\mathrm{I}}^*, \mathbf{Z}_{\mathrm{II}}^* = \mathbf{z}_{\mathrm{II}}^*) \\ \times \mathbb{P}(1, u_1^{i-1}, \mathbf{Z}_{\mathrm{I}}^* = \mathbf{z}_{\mathrm{I}}^*, \mathbf{Z}_{\mathrm{II}}^* = \mathbf{z}_{\mathrm{II}}^*) \end{array}} \\ &= \frac{3}{2}N \cdot Z(U_i | U_1^{i-1}, \mathbf{Z}_{\mathrm{I}}^*, \mathbf{Z}_{\mathrm{II}}^*) \end{aligned}$$

In (a), the length of L_0 under GBM is at least 1 (the middle bit of $\mathbf{Y}^*$ is a GB bit, under GBM), and is at most $\frac{3}{2}N$, since:

$$\begin{aligned} L_0 \leq |\mathbf{G}| &\overset{(i)}{\leq} |\mathbf{X}| \cdot \left(1 + \frac{2^{-(\xi n_0 + 1)}}{1 - 2^{-\xi}}\right) \\ &= N \cdot \left(1 + \frac{2^{-(\xi n_0 + 1)}}{1 - 2^{-\xi}}\right) \overset{(ii)}{\leq} \frac{3}{2}N . \end{aligned} \tag{15}$$

(i) follows from (5) and (6), and by summing all GB lengths as in [4, Lemma 22]. For (ii), recall that $\xi > 0$ is a constant and $n_0 \geq m_0^{\mathrm{th}}(\xi)$. Thus, we take m_0^{th} large enough such that (ii) holds.

For the sum in (14b) we have:

$$\begin{aligned}
&\sum_{u_1^{i-1},\mathbf{z}} \sqrt{\begin{array}{l}\mathbb{P}(0,u_1^{i-1},\mathbf{z},\mathrm{GBM})\cdot\mathbb{P}(1,u_1^{i-1},\mathbf{z},\neg\mathrm{GBM})\\ \quad+\mathbb{P}(0,u_1^{i-1},\mathbf{z},\neg\mathrm{GBM})\cdot\mathbb{P}(1,u_1^{i-1},\mathbf{z},\mathrm{GBM})\\ \quad+\mathbb{P}(0,u_1^{i-1},\mathbf{z},\neg\mathrm{GBM})\cdot\mathbb{P}(1,u_1^{i-1},\mathbf{z},\neg\mathrm{GBM})\end{array}}\\
&\quad=\sum_{u_1^{i-1},\mathbf{z}} \sqrt{\begin{array}{l}\mathbb{P}(0,u_1^{i-1},\mathbf{z},\mathrm{GBM})\cdot\mathbb{P}(1,u_1^{i-1},\mathbf{z},\neg\mathrm{GBM})\\ \quad+\mathbb{P}(0,u_1^{i-1},\mathbf{z},\neg\mathrm{GBM})\cdot\mathbb{P}(1,u_1^{i-1},\mathbf{z})\end{array}}\\
&\quad\le\sum_{u_1^{i-1},\mathbf{z}} \sqrt{\begin{array}{l}\mathbb{P}(u_1^{i-1},\mathbf{z})\cdot\mathbb{P}(1,u_1^{i-1},\mathbf{z},\neg\mathrm{GBM})\\ \quad+\mathbb{P}(0,u_1^{i-1},\mathbf{z},\neg\mathrm{GBM})\cdot\mathbb{P}(u_1^{i-1},\mathbf{z})\end{array}}\\
&\quad=\sum_{u_1^{i-1},\mathbf{z}} \sqrt{\mathbb{P}(u_1^{i-1},\mathbf{z})\cdot\mathbb{P}(u_1^{i-1},\mathbf{z},\neg\mathrm{GBM})}\\
&\quad=\sum_{u_1^{i-1},\mathbf{z}} \mathbb{P}(u_1^{i-1},\mathbf{z})\sqrt{\mathbb{P}(\neg\mathrm{GBM}|u_1^{i-1},\mathbf{z})}\\
&\quad\le\sqrt{\mathbb{P}(\neg\mathrm{GBM})}
\end{aligned}$$

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

Combining the bounds for the two sums in (14) yields

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

To complete the proof of (7a), it remains to show that the term $\sqrt{\mathbb{P}(\neg\mathrm{GBM})}$ is smaller than $2^{-N^{\frac{2}{3}}}$, for large enough n and n_0. This will be shown in the next subsection. Lastly, since $\mathbf{Z}_{\mathrm{I}}^*$ and $\mathbf{Z}_{\mathrm{II}}^*$ are i.i.d., the second inequality in our lemma, (7b), is a direct consequence of [3, Proposition 5].

C. The GBM *event occurs with high probability*

In this section we show that there exist $m_0^{\mathrm{th}}(\xi)$ and $m^{\mathrm{th}}(\xi,\delta)$ such that for $n_0\ge m_0^{\mathrm{th}}$ and $n\ge\max\{m^{\mathrm{th}},n_0+1\}$ we have

$$\sqrt{\mathbb{P}(\neg\mathrm{GBM})}\le 2^{-N^{\frac{2}{3}}}\,. \tag{16}$$

Proving this proves Lemma 5: for it, we take $m^{\mathrm{th}}(\xi,\delta)$ equal to the one developed in this subsection and take m_0^{th} to be the maximum of the m_0^{th} appearing in the previous subsection and this subsection.

The proof follows from a strengthening of [4, Lemma 23]. That is, we show that there exist thresholds m_0^{th}, m^{th} and a constant $\theta>0$ such that

$$\mathbb{P}(\neg\mathrm{GBM})<2^{-\theta\cdot 2^{(1-2\xi)n}}\,, \tag{17}$$

for all $n_0\ge m_0^{\mathrm{th}}$ and $n\ge\max\{m^{\mathrm{th}},n_0{+}1\}$. Thus, if we require that the constant ξ satisfy $\xi\in(0,\frac{1}{6})$, standard manipulations yield (16), for large enough n.

In [4, Lemma 23], the RHS of (17) is weaker: n is replaced by n_0. For lack of space, we only give an outline of the differences between our proof of (17) and the proof of the weaker claim in [4, Lemma 23]. The main difference lies in bounding the probability that too much of $\mathbf{Z}_{\mathrm{I}}$ is lost due to trimming. That is, event A' in [4, Lemma 23, Subsection VII.C]. 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. Proof outline for our main theorem

The proof of Theorem 1 follows by combining Theorem 3, Remark 4, Lemma 5, and Lemma 6. In essence, relabel the n in Theorem 3 as $\tilde{n}$ and take $\epsilon'=\frac{\epsilon}{3}$. Next, fix $\nu\in(0,\frac{1}{3})$ and take $\tilde{n}$ large enough so that (3) and (4) hold, with $\mathbf{Y}^*$ in place of $\mathbf{Y}$, for a fraction of at least $\mathcal{I}-\epsilon'=\mathcal{I}-\frac{\epsilon}{3}$ indices. Recall that this dictates $n_0=\lfloor\nu\tilde{n}\rfloor$ by Theorem 3. Furthermore, take $\tilde{n}$ large enough such that $n_0>m_0^{\mathrm{th}}$ and $\tilde{n}>m^{\mathrm{th}}$, where the right-hand sides are given in Lemma 5. We also take $\tilde{n}$ large enough such that $\frac{N}{\Lambda}=\frac{|\mathbf{X}|}{|\mathbf{G}|}>1-\epsilon$, which is possible by (15). We now show that if we take $\tilde{n}\ge n_{\mathrm{w}}^{\mathrm{th}}$, then

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

for at least $(\mathcal{I}-\frac{\epsilon}{3})-\frac{\epsilon}{3}=\mathcal{I}-\frac{2\epsilon}{3}$ of the indices as $n\to\infty$. The 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. Inequality (b) indeed holds for the above fraction of indices by Lemma 6. That is, the $B_1,\ldots,B_n$ in Lemma 6 correspond to the index bits $b_1,\ldots,b_n$ of i, the Z_n process is set to $Z(U_i|U_1^{i-1},\mathbf{Y}^*)$ and recall that $\tilde{n}\ge n_{\mathrm{w}}^{\mathrm{th}}$. Notice (8) is satisfied by Lemma 5 ($\kappa=3,d=1$ and $\gamma=\frac{2}{3}$) and condition (9) is satisfied from (3) for a fraction of at least $\mathcal{I}-\frac{\epsilon}{3}$ indices. We have proven (1) from Theorem 1, i.e. the strong polarization of the Bhattacharyya parameter. The proof of (2) follows along the same lines as that of (4), and will be given in the full version. In total, both (1) and (2) are satisfied for at least $\mathcal{I}-\epsilon$ of the indices, as $n\to\infty$.

VII. Coding scheme and complexity

Our encoder is the same as that in [4], where we only 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 $|\mathbf{G}|=\Lambda$.

Our decoder is essentially 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. This is in contrast to [4], in which N/N_0 base trellises are constructed — one for each block. 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 [14].

Note that in our analysis, we've analyzed the probability of the middle index falling within the outermost guard-band (the GBM event). This was important in order to prove Theorem 1. However, as opposed to [4], no corresponding operation of partitioning the output is carried out by the decoder.

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] 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.

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

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

Appendix A
Decoder

As in [4], we will use a trellis $\mathcal{T}$ to represent the joint probability of the deletion channel input and output. For simplicity of exposition and lack of space, we describe the memoryless input case here. We perform '−' and '+' operations on $\mathcal{T}$, which merge two-edge paths in $\mathcal{T}$ and result in trellises with half the number of sections: $\mathcal{T}^{[0]}$ and $\mathcal{T}^{[1]}$, respectively. A pair/triplet of sections will be referred to as a sub-trellis $s\mathcal{T}$. After a '−' or '+' transform, each $s\mathcal{T}$ is merged into one section. Our decoder will differ from that of [4, Section IV] in one main point: we incorporate the probabilities of the GB bits into our trellis $\mathcal{T}$. That is, $\mathcal{T}$ is one big trellis encompassing all of $\mathbf{Y}$.

The decoder recursively performs '−' and '+' transformations on $\mathcal{T}$ as follows. First, we perform n '−' transforms, creating $\mathcal{T}^{[000\ldots00]}$. We consider the two single-edge paths from the left upper vertex to the right bottom vertex, which represent the two possible values for $\hat{U}_0$. The decision on $\hat{U}_0$ (if it is not frozen) is by the most probable value, i.e. the edge with the largest probability. Using $\hat{U}_0$, we next create: $\mathcal{T}^{[\overbrace{000\ldots01}^{n}]} = \left(\mathcal{T}^{[\overbrace{000\ldots0}^{n-1}]}\right)^{[1]}$. We use $\mathcal{T}^{[000\ldots01]}$ to decide on $\hat{U}_1$. We repeat this procedure such that with trellis $\mathcal{T}^{[b_0b_1\ldots b_n]}$ we decide on the value of $\hat{U}_{i(b_1,\ldots,b_n)}$ (if it is not frozen). See Figure 3 for an illustration of the decoding process.

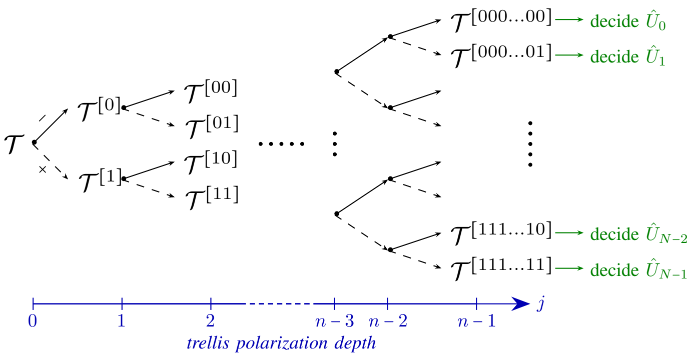

$\mathcal{T}$
$\mathcal{T}^{[0]}$
$\mathcal{T}^{[1]}$
$\mathcal{T}^{[00]}$
$\mathcal{T}^{[01]}$
$\mathcal{T}^{[10]}$
$\mathcal{T}^{[11]}$
$\mathcal{T}^{[000\ldots00]}$ decide $\hat{U}_0$
$\mathcal{T}^{[000\ldots01]}$ decide $\hat{U}_1$
$\mathcal{T}^{[111\ldots10]}$ decide $\hat{U}_{N-2}$
$\mathcal{T}^{[111\ldots11]}$ decide $\hat{U}_{N-1}$
0
1
2
$n-3$
$n-2$
$n-1$
j
trellis polarization depth

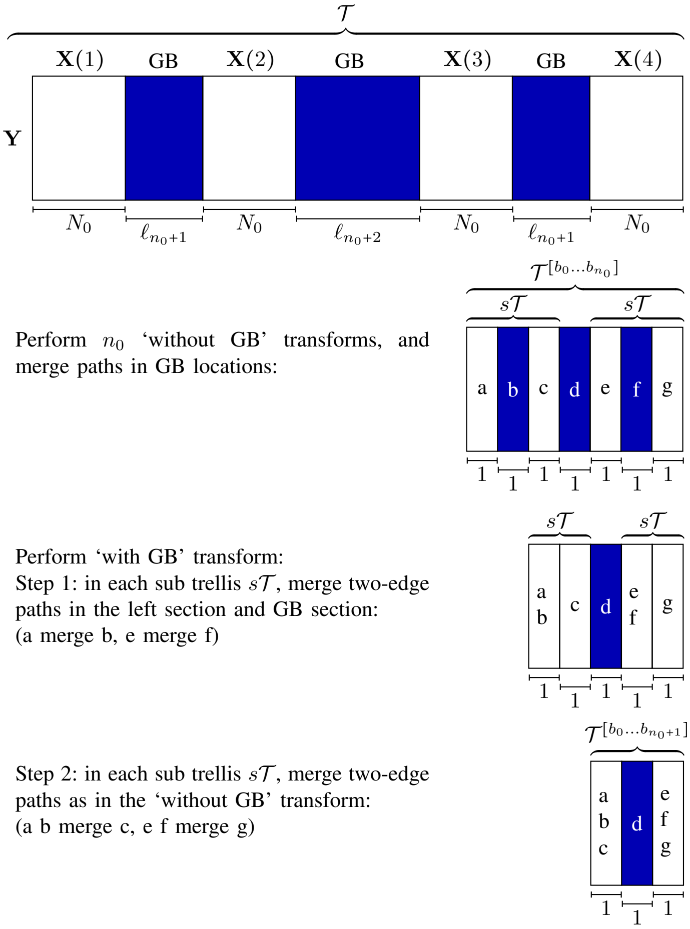
Fig. 3. Recursive trellis transforms.

For an input distribution with $|\mathcal{S}|$ states, the complexity of our decoder is bounded by

$$\sum_{j=0}^{n-1}\underbrace{2^{j+1}}_{\text{(a)}}\cdot\underbrace{8N|\mathcal{S}|^3(2^j+1)^2}_{\text{(b)}}\cdot\underbrace{2^{n-j}}_{\text{(c)}}\in O(N^4)$$

where j is the trellis polarization depth (i.e. the number of '+' or '−' transforms performed on $\mathcal{T}$). (a) is the number of times we return to the trellises of depth j. (b) bounds the number of calculations on each sub-trellis of a given trellis of depth j. (c) is the number of sub-trellises in a trellis of depth j. Note that we think of $|\mathcal{S}|$ as a constant.

In the first n_0 in polarization steps on $\mathcal{T}$, the '−' and '+' transformations are as defined in [4, Definitions 5,6]. We refer to this as the 'without GB' phase. Next, we merge all paths in the GB locations in the trellis, such that each GB is merged into one section. In the following n_1 transforms, referred to as the 'with GB' transforms, each $s\mathcal{T}$ includes a GB section between two non-guard-band sections. Thus, we first merge the two-edge paths in the left section and the GB section. This results in a two-section $s\mathcal{T}$, as in the 'without GB' case. We may now perform the '−' or '+' transformation as before. See Fig. 4 for an illustration.



Perform n_0 'without GB' transforms, and merge paths in GB locations:

Perform 'with GB' transform:
Step 1: in each sub trellis $s\mathcal{T}$, merge two-edge paths in the left section and GB section:
(a merge b, e merge f)

Step 2: in each sub trellis $s\mathcal{T}$, merge two-edge paths as in the 'without GB' transform:
(a b merge c, e f merge g)

Fig. 4. Trellis evolution in the decoder.

Appendix B
High probability of GBM – proof sketch

We state and prove the following Lemma, as a supplementary reading for Subsection V-C. This lemma states that the GBM event, i.e. the event where the middle bit in the TDC output is a GB bit, occurs with high probability.

Lemma 7 (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 (6), there exists an $m^{\text{th}}(\xi,\delta)$, which is a function of the input distribution as well, such that:

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

for all $n \geq m^{\text{th}}(\xi,\delta)$ and $n_0 \geq m_0^{\text{th}}(\xi) = \log_{2^{-\xi}}\left(\frac{1-2^{-\xi}}{2}\right)$.

Proof: As mentioned previously, the proof resembles the steps taken in the proof of [4, Lemma 23]. We define the following length differences due to channel deletion:

$$\begin{aligned} \alpha &= |\mathbf{G}_{\mathrm{I}}| - |\mathbf{Y}_{\mathrm{I}}| \\ \beta &= |\mathbf{G}_{\Delta}| - |\mathbf{Y}_{\Delta}| \\ \gamma &= |\mathbf{G}_{\mathrm{II}}| - |\mathbf{Y}_{\mathrm{II}}| \,, \end{aligned} \tag{18a}$$

and the following length differences due to trimming:

$$\begin{aligned} \alpha' &= |\mathbf{Y}_{\mathrm{I}}| - |\mathbf{Z}_{\mathrm{I}}| \\ \beta' &= |\mathbf{Y}_{\Delta}| - |\mathbf{Z}_{\Delta}| \\ \gamma' &= |\mathbf{Y}_{\mathrm{II}}| - |\mathbf{Z}_{\mathrm{II}}| \,. \end{aligned} \tag{18b}$$

We observe the event: $A \cap A' \cap B \cap B' \cap C \cap C'$, where A, B, C are events constraining the number of deletions in $\mathbf{G}_{\mathrm{I}}, \mathbf{G}_{\Delta}, \mathbf{G}_{\mathrm{II}}$, and A', B', C' are events constraining the number of bits trimmed in $\mathbf{Y}_{\mathrm{I}}, \mathbf{Y}_{\Delta}, \mathbf{Y}_{\mathrm{II}}$. These events will be defined explicitly in a moment, but first, 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 (21).

We define:

$$A = \{\delta|\mathbf{G}_{\mathrm{I}}| - \hat{\ell} < \alpha < \delta|\mathbf{G}_{\mathrm{I}}| + \hat{\ell}\} \tag{19a}$$

$$A' = \{0 \le \alpha' < \hat{\ell}\} \tag{19b}$$

$$B = \{\beta < \delta|\mathbf{G}_{\Delta}| + \hat{\ell}\} \tag{19c}$$

$$B' = \{\beta' = 0\} \tag{19d}$$

$$C = \{\delta|\mathbf{G}_{\mathrm{II}}| - \hat{\ell} < \gamma < \delta|\mathbf{G}_{\mathrm{II}}| + \hat{\ell}\} \tag{19e}$$

$$C' = \{0 \le \gamma' < \hat{\ell}\} \tag{19f}$$

where $\hat{\ell}$ is some length which is chosen such that not all of the GB will be removed under event B. Specifically we select:

$$\hat{\ell} = \frac{1-\delta}{4}\ell_n \tag{20}$$

For this selection, we notice that under the event $A \cap C$, the deletions in $\mathbf{G}_{\mathrm{I}}, \mathbf{G}_{\mathrm{II}}$ are less than $\frac{1+\delta}{2}|\mathbf{G}_{\mathrm{I}}|$ of the bits. Also, under $A' \cap C'$ we will trim less than $\frac{1-\delta}{2}|\mathbf{G}_{\mathrm{I}}|$ bits from $\mathbf{Z}_{\mathrm{I}}, \mathbf{Z}_{\mathrm{II}}$. Thus, under $A \cap A' \cap C \cap C'$, we stop the trimming of $\mathbf{Y}$ prior to the received GB bits in $\mathbf{Y}_{\Delta}$. We get: $A \cap A' \cap C \cap C' \Rightarrow B'$. Next, we notice:

$$\begin{aligned} &\{A \cap A' \cap B \cap C \cap C'\} \\ &\Leftrightarrow \{A \cap A' \cap B \cap B' \cap C \cap C'\} \\ &\overset{(a)}{\subset} \left\{ \begin{array}{l} \alpha + \alpha' < \gamma + \gamma' + \ell_n - \beta \\ \quad \text{and } \gamma + \gamma' < \alpha + \alpha' + \ell_n - \beta \end{array} \right\} \\ &\overset{(18)}{\Leftrightarrow} \{|\mathbf{Z}_{\mathrm{I}}| < |\mathbf{Z}_{\Delta}| + |\mathbf{Z}_{\mathrm{II}}| \quad \text{and} \quad |\mathbf{Z}_{\mathrm{II}}| < |\mathbf{Z}_{\Delta}| + |\mathbf{Z}_{\mathrm{I}}|\} \\ &\Leftrightarrow \mathrm{GBM} \end{aligned} \tag{21}$$

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

$$\begin{aligned} \gamma + \gamma' + \ell_n - \beta &\overset{(19c),(19e),(19f)}{>} \delta|\mathbf{G}_{\mathrm{II}}| - \hat{\ell} + 0 + \ell_n - \delta\ell_n - \hat{\ell} \\ &\overset{(20)}{=} \delta|\mathbf{G}_{\mathrm{II}}| - \hat{\ell} + 4\hat{\ell} - \hat{\ell} \\ &= \delta|\mathbf{G}_{\mathrm{II}}| + 2\hat{\ell} \\ &\overset{(19a),(19b)}{>} \alpha + \alpha'. \end{aligned}$$

and $\gamma + \gamma' < \alpha + \alpha' + \ell_n - \beta$ by the same steps.

From (21) we get: $\mathbb{P}(\mathrm{GBM}) \ge \mathbb{P}(A \cap A' \cap B \cap C \cap C')$. We are interested in the complementary event, which will satisfy:

$$\begin{aligned} \mathbb{P}(\neg\mathrm{GBM}) &\le \mathbb{P}(\neg\{A \cap A' \cap B \cap C \cap C'\}) \\ &\overset{(a)}{\le} \mathbb{P}(\neg A) + \mathbb{P}(\neg A') + \mathbb{P}(\neg B) + \mathbb{P}(\neg C) + \mathbb{P}(\neg C') \\ &\overset{(b)}{=} 2\mathbb{P}(\neg A) + 2\mathbb{P}(\neg A') + \mathbb{P}(\neg B) \end{aligned}$$

(a) is by the union bound and (b) results from the symmetry between events A, A' and C, C' respectively, by (19).

$\mathbb{P}(\neg A)$ and $\mathbb{P}(\neg B)$ may be bounded using Hoeffding [16, Theorem 4.12], as in [4, equations (89),(90)]. In Lemma 8 we bound $\mathbb{P}(\neg A^{'})$. In total, we reach the following upper bound for $\mathbb{P}(\neg\mathrm{GBM})$:

$$\begin{aligned} \mathbb{P}(\neg\mathrm{GBM}) &\le 2\mathbb{P}(\neg A) + 2\mathbb{P}(\neg A') + \mathbb{P}(\neg B) \\ &\le 2 \cdot 2e^{-\frac{(1-\delta)^2}{128}2^{(1-2\xi)n}} \\ &\quad + 2 \cdot e^{-D \cdot 2^{(1-\xi)n}} \\ &\quad + 2e^{-\frac{(1-\delta)^2}{32}2^{(1-\xi)n}} \,, \end{aligned}$$

where $D > 0$ is a constant dependent on the input distribution and on the deletion rate δ. The value of D is given explicitly in the proof of Lemma 8. We note that when bounding $\mathbb{P}(A')$ we used the fact that $n_0 \ge m_0^{\mathrm{th}}(\xi) \ge \log_{2^{-\xi}}\left(\frac{1-2^{-\xi}}{2}\right)$, and the qualities of the input distribution we fixed.

Finally, for $0 < \xi < \frac{1}{6}$ and for a large enough n:

$$\mathbb{P}(\neg\mathrm{GBM}) \le 8e^{-\frac{(1-\delta)^2}{128}2^{(1-2\xi)n}} \le 2^{-2 \cdot N^{\frac{2}{3}}}$$

specifically, this holds for:

$$n \ge m^{\mathrm{th}}(\xi, \delta) \triangleq \max \left\{ \begin{array}{l} \frac{1}{\xi}\log_2\left(\frac{(1-\delta)^2}{128 \cdot D}\right), \\ \frac{1}{1-2\xi}\log_2\left(\frac{128 \cdot \log_2(5)}{(1-\delta)^2(\log_2(e)-1)}\right), \\ \frac{1}{1-2\xi-\frac{2}{3}}\log_2\left(\frac{128 \cdot 2}{(1-\delta)^2}\right) \end{array} \right\}.$$

■

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

The following lemma is used for the proof of Lemma 7. In this lemma we develop a bound on $\mathbb{P}(\neg A^{'})$, the probability that 'too many' bits were trimmed in $\mathbf{Y}_{\mathrm{I}}$. 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 (19b), and let $m_0^{\mathrm{th}}(\xi) \ge \log_{2^{-\xi}}\left(\frac{1-2^{-\xi}}{2}\right)$. Then, for $n_0 \ge m_0^{\mathrm{th}}(\xi)$:

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

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 $\mathbf{G}_{\mathrm{I}}$ is a '1' and was not deleted (where $\hat{\ell}$ was set in (20)). Clearly: $A'' \Rightarrow A'$, hence,

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

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

We denote $\#^{j}_{\mathbf{X}}$ as the number of bits to the left of index j in $g(\mathbf{X}, n_0, \xi)$ that originate from $\mathbf{X}$, and denote $\#^{j}_{\mathrm{GB}}$ as the number of GB bits to the left of index j. For $n_0 \geq m_0^{\mathrm{th}}(\xi)$:

$$\#^{j}_{\mathbf{X}} \geq \#^{j}_{\mathrm{GB}}, \quad \forall j \in \{1, 2, \ldots, \Lambda\} . \tag{22}$$

That is, there are more bits from $\mathbf{X}$ than GB bits, for any prefix of $\mathbf{G}$. The proof of (22) is given in Lemma 9. The proof follows from the recursive manner in which the GBs are added and by $|\mathbf{G}| \overset{(15)}{\leq} |\mathbf{X}|\left(1 + \frac{2^{-(\xi n_0+1)}}{1-2^{-\xi}}\right)$ holding in each recursive step.

By (22), there are at least $\frac{j}{2}$ bits from $\mathbf{X}$ prior to index j in $\mathbf{G}$, i.e.:

$$\#^{j}_{\mathbf{X}} \geq \frac{j}{2} . \tag{23}$$

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 $s \in \mathcal{S}$:

$$\mathbb{P}((X_1, X_2, \ldots, X_\tau) = (0, 0, \ldots, 0) | S_0 = s) < p_0 . \tag{24}$$

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:

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

There are $\left\lfloor \#^{\hat{\ell}}_{\mathbf{X}} / \tau \right\rfloor$ series of $\mathbf{X}$ bits (of length τ) up to index $\hat{\ell}$. Thus, by the Markov property:

$$\begin{aligned}\mathbb{P}(\neg A'') &\leq (1 - (1 - p_0)(1 - \delta))^{\left\lfloor \#^{\hat{\ell}}_{\mathbf{X}} / \tau \right\rfloor} \\ &\overset{(23)}{\leq} (p_0(1 - \delta) + \delta)^{\left\lfloor \hat{\ell} / 2\tau \right\rfloor}\end{aligned}$$

We continue to upper bound the RHS from above:

$$\begin{aligned}&\overset{(20)}{=} (p_0(1 - \delta) + \delta)^{\left\lfloor \frac{(1-\delta)\ell_n}{8\tau} \right\rfloor} \\ &\leq (p_0(1 - \delta) + \delta)^{\frac{(1-\delta)\ell_n}{16\tau}} \\ &\overset{(6)}{\leq} (p_0(1 - \delta) + \delta)^{\frac{1-\delta}{16\tau} 2^{(1-\xi)(n-1)-1}} \\ &= (p_0(1 - \delta) + \delta)^{\frac{1-\delta}{32\tau \cdot 2^{(1-\xi)}} 2^{(1-\xi)n}} \\ &\leq (p_0(1 - \delta) + \delta)^{\frac{1-\delta}{32 \cdot 2\tau} 2^{(1-\xi)n}} \\ &= e^{-\frac{1}{2\tau} \ln\left(\frac{1}{p_0(1-\delta)+\delta}\right) \frac{1-\delta}{32} 2^{(1-\xi)n}}\end{aligned}$$

We mark: $D \triangleq \frac{1}{2\tau} \ln\left(\frac{1}{p_0(1-\delta)+\delta}\right) \frac{1-\delta}{32}$, where τ, p_0 satisfy (24). We note that $D > 0$, since $0 < p_0(1 - \delta) + \delta < 1$.

Finally,

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

■

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

To show (22), we state and prove the following lemma.

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

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

where $\#^{j}_{\mathbf{X}}$ is the number of $\mathbf{X}$ bits in the prefix up to j in $g(\mathbf{X}, n_0, \xi)$, and $\#^{j}_{\mathrm{GB}}$ is the number of GB bits up to j.

Proof: We divide our proof to three claims.

Claim B.1: We assume there exists an index j_0 for which our lemma does not hold, i.e. $\#^{j_0}_{\mathbf{X}} < \#^{j_0}_{\mathrm{GB}}$. 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. $\#^{j_1}_{\mathbf{X}} < \#^{j_1}_{\mathrm{GB}}$.

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:

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

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:

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

■

Claim B.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 data, each block of length $N_0 = 2^{n_0}$.

If, $\#^{j_{\mathrm{mid}}}_{\mathbf{X}} \geq \#^{j_{\mathrm{mid}}}_{\mathrm{GB}}$ for all n_1, then,

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

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

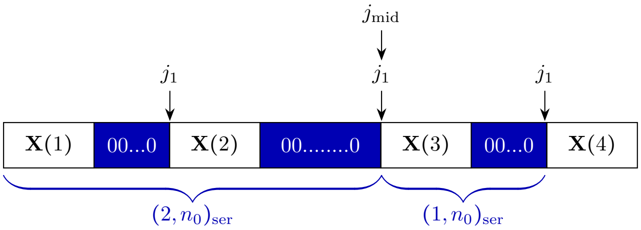
Proof: We name the series of bits leading to j_{mid}, where $g(\mathbf{X})$ was generated according to a given n_0 and n_1, as:

$$(n_1, n_0)_{\mathrm{ser}}$$

For a general n_1, the full $g(\mathbf{X})$ will be the concatenation:

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

See example below, for $g(\mathbf{X}) = (2, n_0)_{\mathrm{ser}} \odot (1, n_0)_{\mathrm{ser}} \odot \mathbf{X}(4)$.



We notice the following quality. 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:

$$\{(i, n_0)_{\mathrm{ser}}\}_{i \in \mathcal{J}}$$

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 $(n_1, n_0)_{\text{ser}}$ consists of more (or equal) bits of $\mathbf{X}$ than GB bits, leading to:

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

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

Claim B.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 $\mathbf{X}$:

$$\#_{\mathbf{X}}^{j_{\text{mid}}} = \frac{1}{2}|\mathbf{X}| = 2^{n-1} \tag{25}$$

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

$$\#_{\text{GB}}^{j_{\text{mid}}} = \frac{1}{2}(\Lambda - N) + \frac{1}{2}\ell_n \tag{26}$$

The total number of GB bits satisfies:

$$\begin{array}{rcl} |g(\mathbf{X})| - |\mathbf{X}| & = & \Lambda - N \\ & \overset{(15)}{\leq} & 2^n \cdot \left(\frac{2^{-(\xi n_0 + 1)}}{1 - 2^{-\xi}}\right) \\ & \leq & 2^n \cdot \left(\frac{2^{-\xi n_0}}{1 - 2^{-\xi}}\right) \end{array} \tag{27}$$

where the last inequality holds since $2^{-\xi} \in (0,1)$ for $\xi > 0$.

The length of the middle GB satisfies:

$$\ell_n \overset{(6)}{\leq} 2^{(1-\xi)(n-1)} \leq 2^{n-1} \cdot 2^{-\xi n_0} \ , \tag{28}$$

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

$$\begin{array}{rcl} \#_{\mathbf{X}}^{j_{\text{mid}}} & \overset{(25)}{=} & 2^{n-1} \\ & \overset{(a)}{\geq} & \frac{1}{2} \cdot 2^n \cdot \left(\frac{2^{-\xi n_0}}{1 - 2^{-\xi}}\right) + 2^{n-1} \cdot 2^{-\xi n_0} \\ & \overset{(27),(28)}{\geq} & \frac{1}{2}(\Lambda - N) + \ell_n \\ & \overset{(26)}{\geq} & \#_{\text{GB}}^{j_{\text{mid}}} \end{array} \tag{29}$$

where (a) is satisfied for $n_0 \geq \log_{2^{-\xi}}(\frac{1-2^{-\xi}}{2})$ and any $\xi > 0$. ■

By combining the results from Claims B.2 and B.3, we have proven the Lemma. ■

Appendix C
Proof sketch of the walking-to-running lemma

We now prove Lemma 6.

Proof: We will assume WLOG that $\nu \in \left(0, \frac{1}{3}\right)$. For example, set $\nu := \min\left\{\nu, \frac{1}{4}\right\}$, and note that if (9) holds for the “old" value of ν, then it surely holds for the “new" value as well.

Let us first define the process $\bar{Z}_n$ as:

$$\bar{Z}_{n+1} = 2 \cdot \kappa N^d \begin{cases} \bar{Z}_n & \text{if } B_{n+1} = 0 \text{ (`}-\text{')} \\ \bar{Z}_n^2 & \text{if } B_{n+1} = 1 \text{ (`}+\text{')} \end{cases}, \quad n \geq n_{\text{w}} \tag{30a}$$

$$\bar{Z}_{n_{\text{w}}} = 2^{-(2^{n_{\text{w}}})^{\nu}} \tag{30b}$$

This process is defined from some starting point $n_{\text{w}} \geq n_{\text{w}}^{\text{th}}$, where n_{w}^{th} is a parameter that will be fixed later on. Note that the process $\bar{Z}_n$ is “simpler" than Z_n: the inequalities in (8) and (9) have been replaced by equalities, and the additive term $2^{-N^{\gamma}}$ in (8) has been removed from (30a). The price we pay for this simplification is a multiplicative factor of 2.

Let $n_{\text{r}}^{\text{th}} > n_{\text{w}}$ be a parameter that will be fixed later on as well. We now fix γ_a and ν_b such that,

$$\frac{1}{2} < \gamma_a < \gamma \tag{31a}$$

$$0 < \nu_b < \nu < \frac{1}{3} \tag{31b}$$

We define the following events for the processes $Z_n, \bar{Z}_n$:

$$\Sigma_a \triangleq \{\bar{Z}_n \geq 2^{-N^{\gamma_a}}, \qquad \forall n \geq n_{\text{w}}\} \tag{32a}$$

$$\Sigma_b \triangleq \{\bar{Z}_n \leq 2^{-N^{\nu_b}}, \qquad \forall n \geq n_{\text{w}}\} \tag{32b}$$

$$\Sigma_c \triangleq \{\bar{Z}_n < \frac{1}{2N} 2^{-N^{\beta}}, \quad \forall n \geq n_{\text{r}}^{\text{th}}\} \tag{32c}$$

$$\Sigma_d \triangleq \{Z_n \leq \bar{Z}_n, \qquad \forall n \geq n_{\text{w}}\} \tag{32d}$$

The first three events discuss bounds concerning the new process $\bar{Z}_n$, and the forth discusses a relation between $\bar{Z}_n$ and the original process Z_n. For the events above we list the following claims:

Claim C.1: For all $\epsilon_a > 0$ there exists an $n_{a_{\text{I}}}^{\text{th}}(\epsilon_a, \nu, \kappa, d, \gamma_a)$ s.t. if $n_{\text{w}} \geq n_{a_{\text{I}}}^{\text{th}}$, then:

$$\mathbb{P}(\Sigma_a) > 1 - \epsilon_a \tag{33}$$

Claim C.2: For all $\epsilon_b > 0$ there exists an $n_{b_{\text{I}}}^{\text{th}}(\epsilon_b, \nu, \kappa, d, \nu_b)$ s.t. if $n_{\text{w}} \geq n_{b_{\text{I}}}^{\text{th}}$, then:

$$\mathbb{P}(\Sigma_b) > 1 - \epsilon_b \tag{34}$$

Claim C.3: For all $\epsilon_c > 0$ there exist $n_{c_{\text{I}}}^{\text{th}}(\beta, \kappa, d, \nu_b)$ and $n_{c_{\text{II}}}^{\text{th}}(\beta, \epsilon_c, n_{\text{w}}, \kappa, d, \nu_b)$ s.t. if $n_{\text{w}} \geq n_{c_{\text{I}}}^{\text{th}}$, $n_{\text{r}}^{\text{th}} > n_{c_{\text{II}}}^{\text{th}}$ and $n_{\text{r}}^{\text{th}} > n_{\text{w}}$, then:

$$\mathbb{P}(\Sigma_c) > 1 - \epsilon_c - \mathbb{P}(\neg\Sigma_b) \tag{35}$$

Claim C.4: There exists an $n_{d_{\text{I}}}^{\text{th}}(\gamma, \gamma_a)$ s.t. if $n_{\text{w}} \geq n_{d_{\text{I}}}^{\text{th}}$ and if $\bar{Z}_{n_{\text{w}}} \geq Z_{n_{\text{w}}}$, then event Σ_a implies Σ_d, i.e.:

$$\bar{Z}_{n_{\text{w}}} \geq Z_{n_{\text{w}}} \Rightarrow \mathbb{P}(\Sigma_d | \Sigma_a) = 1 \tag{36}$$

The proof for Claims C.1–C.3 is briefly discussed in the following subsection. The proof of Claim C.4 is given in Subsection C-B. We set:

$$\epsilon_a = \epsilon_b = \epsilon_c = \frac{\epsilon'}{3} \tag{37}$$

We also set the starting point:

$$n_{\text{w}} \geq n_{\text{w}}^{\text{th}} \triangleq \max\left\{n_{a_{\text{I}}}^{\text{th}}, n_{b_{\text{I}}}^{\text{th}}, n_{c_{\text{I}}}^{\text{th}}, n_{d_{\text{I}}}^{\text{th}}, m^{\text{th}}\right\} \tag{38}$$

and set:

$$n_{\text{r}}^{\text{th}} \triangleq \max\{n_{c_{\text{II}}}^{\text{th}}, n_{\text{w}} + 1\} \tag{39}$$

Notice that if $Z_{n_{\text{w}}}$ satisfies (9), then by (30b), $\bar{Z}_{n_{\text{w}}} \geq Z_{n_{\text{w}}}$.

Using the four claims above, we can bound the probabilities of events Σ_c, Σ_d. For Σ_c, we have:

$$\begin{array}{rl} \mathbb{P}(\Sigma_c) & \overset{(a)}{>} 1 - \epsilon_c - \mathbb{P}(\neg\Sigma_b) \\ & \overset{(b)}{>} 1 - \epsilon_c - \epsilon_b \end{array} \tag{40}$$

Where in (a) we applied (35) from Claim C.3, and in (b) we applied (34) from Claim C.2, since their conditions are satisfied by our selection of n_{w} and $n_{\mathrm{r}}^{\mathrm{th}}$ in (38) and (39).

We next note that:

$$\begin{aligned} \mathbb{P}(\Sigma_d) &\geq \mathbb{P}(\Sigma_d|\Sigma_a)\cdot\mathbb{P}(\Sigma_d) \\ &\overset{(a)}{>} 1\cdot(1-\epsilon_a)=1-\epsilon_a \end{aligned} \tag{41}$$

In (a) we applied (33) from Claim C.1, and (36) from Claim C.4, since their conditions are satisfied by our selection of n_{w} in (38), and by (9) and (30b).

By inspection, the intersection of Σ_c and Σ_d implies the event in (10). Thus,

$$\begin{aligned} &\mathbb{P}\Big(Z_n < \frac{1}{2N}2^{-N^{\beta}}, \quad \forall n \geq n_{\mathrm{r}}^{\mathrm{th}}\Big) \\ &\geq \mathbb{P}\Big(\Big\{\bar{Z}_n < \frac{1}{2N}2^{-N^{\beta}}, \quad \forall n \geq n_{\mathrm{r}}^{\mathrm{th}}\Big\}\bigcap\big\{Z_n \leq \bar{Z}_n, \quad \forall n \geq n_{\mathrm{w}}\big\}\Big) \\ &= \mathbb{P}(\Sigma_c \cap \Sigma_d) \\ &= 1-\mathbb{P}(\neg\Sigma_c \cup \neg\Sigma_d) \end{aligned}$$

In the last equality we denoted events $\neg\Sigma_c, \neg\Sigma_d$ as the complementary events of Σ_c, Σ_d respectively.

We now upper bound $\mathbb{P}(\neg\Sigma_c \cup \neg\Sigma_d)$:

$$\begin{aligned} \mathbb{P}(\neg\Sigma_c \cup \neg\Sigma_d) &\leq \mathbb{P}(\neg\Sigma_c)+\mathbb{P}(\neg\Sigma_d) \\ &\overset{(40),(41)}{<} \epsilon_a+\epsilon_b+\epsilon_c \\ &\overset{(37)}{=} \epsilon' \end{aligned}$$

Thus:

$$\mathbb{P}\Big(Z_n < \frac{1}{2N}2^{-N^{\beta}}, \quad \forall n \geq n_{\mathrm{r}}^{\mathrm{th}}\Big) \geq \mathbb{P}(\Sigma_c \cap \Sigma_d) \geq 1-\epsilon'$$

■

A. High-level discussion on the proof of claims C.1–C.3

The proofs of Claims C.1 and C.2 are similar and will be given in the full version. For now, we give an outline of the main steps. The first step is setting some threshold Δ_{th} for which Hoeffding [16, Theorem 4.12] assures us that the fraction of $\{B_{i+1}\}_{i=n_{\mathrm{w}}}^{n}$ which are 0 ('−') and the fraction of $\{B_{i+1}\}_{i=n_{\mathrm{w}}}^{n}$ which are 1 ('+') are both close to half, for all $n \geq n_{\mathrm{w}}+\Delta_{\mathrm{th}}$. When this occurs, we can derive the "soft" bounds in (32a) and (32b) for all $n \geq n_{\mathrm{w}}+\Delta_{\mathrm{th}}$. For the initial period of $n_{\mathrm{w}} \leq n \leq n_{\mathrm{w}}+\Delta_{\mathrm{th}}$, we use the initial condition (30b) and take a large enough n_{w} to set a "low enough" starting point. The low starting point promises we will not cross the "soft" bounds during the initial period, even for the most problematic cases (which we can prove are when only '+' or only '−' are drawn).

The proof of Claim C.3 will also be given in the full version. The main step in the proof is using event Σ_b in order to replace Z_n in (30a) with the bound from (32b). This leads to a bound on $\bar{Z}_{n+1}$, which, given B_{n+1}, is a deterministic function of n. Since $\bar{Z}_n$ is bounded by $2^{-2^{\nu_b n}}$, the multiplicative factor of $2\cdot\kappa N^d$ is neglectable for a large enough n. That is, the function of n bounding $\bar{Z}_{n+1}$ is monotonically decreasing for a large enough n. Next, using Hoeffding [16, Theorem 4.14] once more, we complete the proof for Claim C.3.

B. Proof of Claim C.4

Proof: We define processes $Z_n^{'}, Z_n^{''}$ which will assists us in proving the claim. First, $Z_n^{'}$ is defined to be:

$$\begin{aligned} Z_{n+1}^{'} &= \begin{cases} \kappa N^d\cdot Z_n^{'}+2^{-2^{\gamma n}} & \text{if } B_{n+1}=0\ (\text{'}-\text{'}) \\ \kappa N^d\cdot {Z_n^{'}}^{2}+2^{-2^{\gamma n}} & \text{if } B_{n+1}=1\ (\text{'+'}) \end{cases}, \quad n \geq n_{\mathrm{w}} \\ Z_{n_{\mathrm{w}}}^{'} &= Z_{n_{\mathrm{w}}} \end{aligned}$$

i.e. Z_n from (8), with the weak inequality replaced by equality. By the monotinicity of the terms in (8) and the above, we easily prove by induction that: $Z_n \leq Z_n^{'}, \quad \forall n \geq n_{\mathrm{w}}$.

Next, $Z_n^{''}$ is set to be:

$$\begin{aligned} Z_{n+1}^{''} &= \begin{cases} \kappa N^d\cdot Z_n^{''}+2^{-2^{\gamma n}} & \text{if } B_{n+1}=0\ (\text{'}-\text{'}) \\ \kappa N^d\cdot {Z_n^{''}}^{2}+2^{-2^{\gamma n}} & \text{if } B_{n+1}=1\ (\text{'+'}) \end{cases}, \quad n \geq n_{\mathrm{w}} \\ Z_{n_{\mathrm{w}}}^{''} &= 2^{-(2^{n_{\mathrm{w}}})^{\nu}} \end{aligned} \tag{42}$$

For any given draw of $B_{n_{\mathrm{w}}}, \ldots, B_n$, the processes $Z_n^{'}, Z_n^{''}$ go through the same transformations, and the only difference is that the starting point of $Z_n^{''}$ is at a higher value, by (9). Again, by monotinicity we prove by induction that $Z_n^{'} \leq Z_n^{''}, \quad \forall n \geq n_{\mathrm{w}}$. Next, we prove that under event Σ_a from (32a), $\bar{Z}_n$ of (30) dominates $Z_n^{''}$. That is,

$$\Sigma_a \Rightarrow Z_n^{''} \leq \bar{Z}_n, \quad \forall n \geq n_{\mathrm{w}}\ .$$

Under Σ_a: $\bar{Z}_n \geq 2^{-N^{\gamma_a}}, \quad \forall n \geq n_{\mathrm{w}}$. Since $\gamma_a \overset{(31a)}{<} \gamma$, for $n \geq n_{d_{\mathrm{I}}}^{\mathrm{th}}(\gamma,\gamma_a) \triangleq \frac{1}{\gamma-\gamma_a}$ we get $\gamma n \geq \gamma_a n+1$. Meaning $\bar{Z}_n$ satisfies:

$$\begin{aligned} \kappa N^d\cdot\bar{Z}_n &\geq \bar{Z}_n &&\geq 2^{-2^{\gamma_a n}} &&\geq 2^{-2^{\gamma n}} \\ \kappa N^d\cdot\bar{Z}_n^2 &\geq \bar{Z}_n^2 &&\geq 2^{-2^{\gamma_a n+1}} &&\geq 2^{-2^{\gamma n}} \end{aligned} \tag{43}$$

If we assume $\bar{Z}_n \geq Z_n^{''}$, then $\bar{Z}_{n+1} \geq Z_{n+1}^{''}$(regardless of B_{n+1}), since:

$$\begin{aligned} \bar{Z}_{n+1} &= \begin{cases} \kappa N^d\cdot\bar{Z}_n+\kappa N^d\cdot\bar{Z}_n & \text{if } B_{n+1}=0\ (\text{'}-\text{'}) \\ \kappa N^d\cdot\bar{Z}_n^2+\kappa N^d\cdot\bar{Z}_n^2 & \text{if } B_{n+1}=1\ (\text{'+'}) \end{cases}, \quad n \geq n_{\mathrm{w}} \\ &\overset{(43)}{\geq} \begin{cases} \kappa N^d\cdot\bar{Z}_n+2^{-2^{\gamma n}} & \text{if } B_{n+1}=0\ (\text{'}-\text{'}) \\ \kappa N^d\cdot\bar{Z}_n^2+2^{-2^{\gamma n}} & \text{if } B_{n+1}=1\ (\text{'+'}) \end{cases}, \quad n \geq n_{\mathrm{w}} \\ &\overset{(a)}{\geq} \begin{cases} \kappa N^d\cdot Z_n^{''}+2^{-2^{\gamma n}} & \text{if } B_{n+1}=0\ (\text{'}-\text{'}) \\ \kappa N^d\cdot {Z_n^{''}}^{2}+2^{-2^{\gamma n}} & \text{if } B_{n+1}=1\ (\text{'+'}) \end{cases}, \quad n \geq n_{\mathrm{w}} \\ &= Z_{n+1}^{''} \end{aligned}$$

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

We remind that: $\bar{Z}_{n_{\mathrm{w}}} \overset{(30b)}{=} 2^{-(2^{n_{\mathrm{w}}})^{\nu}} \overset{(42)}{=} Z_{n_{\mathrm{w}}}^{''}$. That is, $Z_n^{''}, \bar{Z}_n$ begin at the same value. Thus, we may show by induction that under Σ_a, $\bar{Z}_n \geq Z_n^{''}, \quad \forall n \geq n_{\mathrm{w}}$. We have shown:

$$Z_n \leq Z_n^{'} \leq Z_n^{''} \leq \bar{Z}_n, \quad \forall n \geq n_{\mathrm{w}}\ .$$

■