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 $\Lambda$ 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, $O(\Lambda^3)$.

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 $\nu$. For this improvement is a higher decoding complexity, which is nonetheless still polynomial, $O(\Lambda^3)$.

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 $\nu \in (0, \frac{1}{2})$, 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 $\Lambda$ is the codeword length. Furthermore, for any $0 < \nu' < \nu$ and large enough $\Lambda$, 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 $\nu' \in (0, \frac{1}{2})$.

II. MAIN THEOREM

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

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 $I$ 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 $\epsilon > 0$, $\xi \in (0, \frac{1}{2})$, and parameters $0 < \beta' < \beta < \frac{1}{2}$. There exist $n^1(\beta', \beta, \epsilon, \xi)$ and $n^2(\beta', \beta, \epsilon, \xi)$ such that the following holds. Take $n \geq n^1$ and $n_0 \geq n^2$. Let $X$ be of length $N = 2^n$. The vector $X$ is partitioned into blocks of length $2^{n_0}$, and each block is independently distributed according to the hidden Markov input distribution. Let $U$ be the polar transform of $X$. Denote by $Y$ the result of transmitting $g(X, n_0, \xi)$ through the deletion channel. The fraction of indices $i$ for which:

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

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

is at least $I - \epsilon$, where $\Lambda$ is the length of $g(X, n_0, \xi)$. 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 $I - 2\epsilon$ and probability of error at most $2^{-\Lambda^{\beta'}}$, where $\Lambda$ 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 $0 < \nu < \frac{1}{2}$ 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 $U, X, Y, N,$ and $\Lambda$ be as in Theorem 1. The fraction of indices $i$ for which:

$$Z(U_i U_{i-1}^1, Y) < 2^{-N \nu}$$  \hspace{1cm} (3)

$$K(U_i U_{i-1}^1) < 2^{-N \nu}$$  \hspace{1cm} (4)

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

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 that decays roughly like the square 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 $\delta$. 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 $G$ and denote the corresponding output by $Y$.

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 $Y$ or $Z$. We denote the trimming operation by appending a ‘$\star$’ as a superscript. Thus, the outputs corresponding to $Y$ and $Z$ will be $Y^\star$ and $Z^\star$, 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 $G$, then we first pass $G$ through the deletion channel and obtain $Y$, and then pass $Y$ through the trimming channel, which yields $Y^\star$.

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 $Y$ into $Y^\star$.

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

B. Blocks and guard-bands

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

$$X = X(1) \odot X(2) \odot \cdots \odot X(N_1)$$

We denote the first and second halves of $X$ by $X_I$ and $X_II$. Denoting the length of a vector by $|\cdot|$, we have $|X_I| = |X_II| = \frac{N_0}{2}$ and

$$X = X_I \odot X_II$$

Note that $X_I$ and $X_II$ 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 $X$ 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 $X$ of length $\leq 2^{n_0}$, $g(X, n_0, \xi)$ is simply the identity function. For a vector $X$ of length greater than $2^{n_0}$,

$$g(X, n_0, \xi) \triangleq g(X) \triangleq g(X_I) \odot 000 \ldots 00 \odot g(X_II)$$  \hspace{1cm} (5)

That is, we add

$$\ell_n = \left[ 2^{(1-\xi)(n-1)} \right]$$  \hspace{1cm} (6)

“0” symbols between the first and second halves of $X$, and apply $g$ recursively on each half. Note that $\xi > 0$ is a “small” constant that we will define later. To summarize: $X$ is a concatenation of $2^{n-n_0}$ independent blocks, each of length $N_0 = 2^{n_0}$. The function $g(X, n_0, \xi)$ 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 $n = n_0 + 2$:

$$g(X, n_0, \xi)$$

$$\begin{array}{c|c|c|c|c|c|c}
\hline
& X(1) & 00_0 & X(2) & 00_0 & X(3) & 00_0 \\
\hline
N_0 & n_{m_1} & N_0 & n_{m_2} & N_0 & n_{m_3} & N_0 \\
\hline
\ell_{n_0+1} & \ell_{n_0+2} & \ell_{n_0+3} & \ell_{n_0+4} & \ell_{n_0+5} & \ell_{n_0+6} \\
\hline
\end{array}$$

We remind the reader that $G = G_I \odot G_\Delta \odot G_II$ is passed through the DC. We denote the output of this channel by $Y$, and denote the parts corresponding to $G_I$, $G_\Delta$, and $G_II$ by $Y_I$, $Y_\Delta$, and $Y_II$, respectively. We further denote the application of the TC on $Y$ by $Z \neq Y^\star$, and denote the parts corresponding to $Y_I$, $Y_\Delta$, and $Y_II$ by $Z_I$, $Z_\Delta$, and $Z_II$, respectively. See Fig. 1, which is essentially [4, Figure 5]. Note that, in general, $Z_I$ is formed by trimming off only the left side of $Y_I$. Hence, typically, $Z_I \neq (Y_I)^\star$ and $Z_II \neq (Y_II)^\star$. Also, we note that in the typical case, $Z_\Delta = Y_\Delta$.

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.
First, note that in (7a), there is both an additive penalty of the two halves of \( I \), than the one derived for a memoryless channel in [3] and also a similar claim for our setting. Our claim is significantly weaker.' 

where addition is modulo 1. As in the seminal paper, the binary vector corresponding to \( i \) is denoted \( b_i \). As in the seminal paper, the binary

\[
U_{2j-1} = V_j + V_j', \quad U_{2j} = V_j',
\]

where addition is modulo 2. As in the seminal paper, the binary vector corresponding to \( i \) is denoted \( b_i \). That is, for

\[
i = i(b_1, \ldots, b_n).
\]

The following lemma is cardinal to proving the stronger polarization stated in Theorem 1. Recall that \( \delta \) is the deletion rate, and that the guard-band length is given in (6), and is a function of \( \xi \). 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 \( X = A(U) \) 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 \( Y^* = Z_1 \oplus Z_\Delta \oplus Z_{II} \) be the result of transmitting \( g(X, n_0, \xi) \) through the TDC. There exist \( m_{Ih}(\xi) \) and \( m_{IIh}(\xi, d) \) s.t. for \( n_0 \geq m_{Ih} \) and all \( n \geq \max\{m_{Ih}, n_0 + 1\} \) the following holds. Let \( 1 \leq i \leq N \) and \( j = \lceil (i + 1)/2 \rceil \). Then,

\[
Z(U_i|U_{i-1}, Y_i) \leq \frac{3}{2} N \cdot Z(U_i|U_{i-1}, Z_i, Z_{II}) + 2^{-N/2} \quad (7a)
\]

\[
\leq \begin{cases} \frac{3}{2} N \cdot Z(U_i|U_{i-1}, Z_i) + 2^{-\gamma} & \text{if } b_n = 0 (\cdot) \\ \frac{3}{2} N \cdot Z(V_j|V_{j-1}, Z_j^I) + 2^{-N/2} & \text{if } b_n = 1 (+) \end{cases} \quad (7b)
\]

We draw the reader’s attention to several important points. First, note that in (7a), there is both an additive penalty of \( 2^{-N/2} \) as well as a multiplicative penalty of \( \frac{3N}{2} \), associated with conditioning on \( Z_i^I, Z_{II} \) as opposed to conditioning on \( Y^* \). That is, there is a price to be paid for conditioning on the pair of TDC outputs corresponding to \( g(X_i) \) and \( g(X_{II}) \), as opposed to conditioning on the TDC output corresponding to \( g(X_i \oplus X_{II}) \). 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 TDC). The inequality in (7b) shows us why such a penalty is worth paying: since \( Z_i^I \) and \( Z_{II} \) 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, \( 2^{3/2} \)) at some stage of the process, this implies “running-speed” polarization (\( 2^{1/2} \)) during later stages. In our setting, the “walking-speed” is guaranteed by [4, Theorem 1].

**Lemma 6 (walking-to-running):** Let \( B_1, B_2, \ldots \) be i.i.d. uniformly distributed Bernoulli random variables. Fix constants \( \kappa \geq 1, d \geq 0, \gamma \geq 1/2 \), and \( m_{IIh} > 0 \). Let \( Z_0, Z_1, Z_2, \ldots \) be a random process s.t. for all \( n \geq m_{IIh} \),

\[
Z_{n+1} \leq \begin{cases} \kappa N^d \cdot Z_n + 2^{-N/2} & \text{if } B_{n+1} = 0 (\cdot) \\ \kappa N^d \cdot Z_n^I + 2^{-N/2} & \text{if } B_{n+1} = 1 (+) \end{cases} \quad (8)
\]

Fix \( \beta \in (0, 1/2) \), the “running speed” parameter, and \( \nu > 0 \), the “walking speed” parameter. For all \( \epsilon' > 0 \) there exists a threshold \( n_{IIh} = n_{IIh}(\epsilon', \beta, \nu, \kappa, d, \gamma) \) such that for some \( n_w \geq n_{IIh} \) we are assured “walking speed”:

\[
Z_{n_w} \leq 2^{-2\gamma}, \quad (9)
\]

then there exists \( n_{Ih} = n_{Ih}(\epsilon', \beta, \nu, \kappa, d, n_w) \) such that above this threshold, with high probability, we are indefinitely at “running speed”:

\[
P \left( Z_n < 2^{-N/2}, \quad \forall n \geq n_{Ih} \right) \geq 1 - \epsilon'. \quad (10)
\]

**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
very likely. That is, the additive penalty of $2^{-N^{1/2}}$ 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 $Y^* = Z = Z_L \oplus Z_D \oplus Z_R$ is the result of passing $G_1 \oplus G_2 \oplus G_3$ through the TDC. The GBM event occurs if $Z$ is not empty and its middle index $i_{mid} = \lfloor \frac{Z_{mid}}{2} \rfloor$, falls within $Z_D$. That is, GBM occurs if $Z_{mid}$ originates from the outermost guard-band $G_D$. The complementary event is denoted $\neg$GBM.

We denote the left and right halves of $Z = Y^*$ as $Z_L = (Z_1, \ldots, Z_{i_{mid}})$ and $Z_R = (Z_{i_{mid}+1}, \ldots, Z_{2N})$, see Figure 2. The main utility of the GBM event is this (again, see Figure 2): since $Z_D$ contains only ‘0’ symbols, under GBM

\[
\begin{align*}
(Z_L)^* &= (Z_1)^*, \\
(Z_R)^* &= (Z^*_{i_{mid}})^*.
\end{align*}
\]

That is, under GBM, the simple operation of trimming the two halves of $Y^*$ is assured to give us $Z^*_L \triangleq (Z_1)^*$ and $Z^*_R \triangleq (Z^*_{i_{mid}})^*$. This simple observation will be used in the next subsection in order to state a recursive relation. We end this subsection by defining

\[
L_0 \triangleq |Y^*| - |Z^*_L| - |Z^*_R|.
\]

Since $Y^*$ does not contain leading nor trailing ‘0’ symbols, $L_0$ equals the sum of the following: the number of trailing ‘0’ symbols in $Z_L$, the length of $Z_D$, and the number of leading ‘0’ symbols in $Z_R$. Hence,

\[
Y^* = Z^*_L \oplus 000 \ldots 00 \oplus Z^*_R.
\]

Thus, by (12) and (11):

\[
\text{GBM} \Rightarrow Y^* = (Z^*_L)^* \oplus 000 \ldots 00 \oplus (Z^*_R)^*.
\]

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: $u_{i-1}^1$ is short for $U_{i-1}^{1-1}$, $z$ is short for $Y^* = z$, 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 $i$.

\[
Z(U_i|U_i^{1-1}, Y^*) = \sum_{u_i^{1-1}} \sqrt{P(U_i = 0, u_i^{1-1} = u_i^{1-1}, Y^* = z) \times P(U_i = 1, u_i^{1-1} = u_i^{1-1}, Y^* = z)}
\]

\[
= \sum_{u_i^{1-1}} \sqrt{P(0, u_i^{1-1}, z, \neg\text{GBM}) \times P(1, u_i^{1-1}, z, \neg\text{GBM})}
\]

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

\[
= \sum_{u_i^{1-1}, z} \left( P(0, u_i^{1-1}, z, \text{GBM}) + P(0, u_i^{1-1}, z, \neg\text{GBM}) \right) \\
\leq \sum_{u_i^{1-1}, z} \sqrt{P(0, u_i^{1-1}, z, \text{GBM}) \times P(1, u_i^{1-1}, z, \neg\text{GBM})}
\]

\[
+ \sum_{u_i^{1-1}, z} \sqrt{P(0, u_i^{1-1}, z, \neg\text{GBM}) \times P(1, u_i^{1-1}, z, \text{GBM})}
\]

(14a)

We will bound both the sum in (14a) and the sum in (14b).

For the sum in (14a), we have:

\[
\sum_{u_i^{1-1}, z} \sqrt{P(0, u_i^{1-1}, Y^* = z, \text{GBM}) \times P(1, u_i^{1-1}, Y^* = z, \neg\text{GBM})}
\]

\[
\leq \sum_{u_i^{1-1}, z} \sqrt{P(0, u_i^{1-1}, \text{GBM}) \times P(1, u_i^{1-1}, \neg\text{GBM})}
\]

\[
= \sum_{u_i^{1-1}} \sqrt{P(0, u_i^{1-1}, \text{GBM}) \times P(1, u_i^{1-1}, \neg\text{GBM})}
\]

\[
= \sum_{u_i^{1-1}} \sqrt{P(0, u_i^{1-1}, z, \text{GBM}) \times P(1, u_i^{1-1}, z, \neg\text{GBM})}
\]

\[
= \sum_{u_i^{1-1}} \sqrt{P(0, u_i^{1-1}, z, \neg\text{GBM}) \times P(1, u_i^{1-1}, z, \text{GBM})}
\]

\[
= \sum_{u_i^{1-1}} \sqrt{P(0, u_i^{1-1}, \text{GBM}) \times P(1, u_i^{1-1}, \neg\text{GBM})}
\]

(14b)

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

\[
L_0 \leq G(i) \leq \left| X_i \right| \leq 1 + \frac{2^{-\xi(n_0+1)}}{1 - 2^{-\xi}} = N \left( 1 + \frac{2^{-\xi(n_0+1)}}{1 - 2^{-\xi}} \right) \leq \frac{3}{2} N
\]

(15)
For the sum in (14b) we have:

\[
\sum_{u_1^{-1} \in \mathcal{U}} \sqrt{\mathbb{P}(0, u_1^{-1}, z, \text{GBM}) \cdot \mathbb{P}(1, u_1^{-1}, z, -\text{GBM}) + \mathbb{P}(0, u_1^{-1}, z, -\text{GBM}) \cdot \mathbb{P}(1, u_1^{-1}, z, \text{GBM}) + \mathbb{P}(0, u_1^{-1}, z, -\text{GBM}) \cdot \mathbb{P}(1, u_1^{-1}, z, -\text{GBM})}
\]

\[
= \sum_{u_1^{-1} \in \mathcal{U}} \sqrt{\mathbb{P}(u_1^{-1}, z, \text{GBM}) \cdot \mathbb{P}(1, u_1^{-1}, z, -\text{GBM}) + \mathbb{P}(0, u_1^{-1}, z, -\text{GBM}) \cdot \mathbb{P}(u_1^{-1}, z)}
\]

\[
\leq \sum_{u_1^{-1} \in \mathcal{U}} \sqrt{\mathbb{P}(u_1^{-1}, z) \cdot \mathbb{P}(u_1^{-1}, z, -\text{GBM})}
\]

\[
= \sum_{u_1^{-1} \in \mathcal{U}} \sqrt{\mathbb{P}(u_1^{-1}, z) \mathbb{P}(u_1^{-1}, z, -\text{GBM})}
\]

\[
\leq \sqrt{\mathbb{P}(\text{GBM})}.
\]

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

\[
Z(U_i | U_i^{-1}, \mathbf{Y}^\star) \leq \frac{3}{2} N \cdot Z(U_i | U_i^{-1}, \mathbf{Z}_i^\star, \mathbf{Z}_i) + \sqrt{\mathbb{P}(\text{GBM})}.
\]

To complete the proof of (7a), it remains to show that the term \(\sqrt{\mathbb{P}(\text{GBM})}\) is smaller than \(2^{-N\frac{d}{2}}\), for large enough \(n\) and \(n_0\). This will be shown in the next subsection. Lastly, since \(\mathbf{Z}_i^\star\) and \(\mathbf{Z}_i\) 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 \(m_{1h}(\xi)\) and \(m_{th}(\delta)\) such that for \(n_0 \geq m_{1h}^0\) and \(n \geq \max\{m_{th}, n_0 + 1\}\) we have

\[
\sqrt{\mathbb{P}(\text{GBM})} \leq 2^{-N\frac{d}{2}}.
\]

Proving this proves Lemma 5: for it, we take \(m_{1h}(\xi, \delta)\) equal to the one developed in this subsection and take \(m_{1h}^0\) to be the maximum of the \(m_{1h}^0\) 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 \(m_{th}^0\) and \(m_{th}\) and a constant \(\theta > 0\) such that

\[
\mathbb{P}(\text{GBM}) \leq 2^{-\theta d(1-z)n},
\]

for all \(n_0 \geq m_{1h}^0\) and \(n \geq \max\{m_{th}, n_0 + 1\}\). Thus, if we require that the constant \(\xi\) satisfy \(\xi \in (0, \frac{1}{2})\), 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 \(\mathbf{Z}_i\) 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 \(G\), 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 \(\bar{n}\) and take \(\epsilon' = \frac{\epsilon}{4}\). Next, fix \(\nu \in (0, \frac{1}{4})\) and take \(\bar{n}\) large enough so that (3) and (4) hold, with \(\mathbf{Y}^\star\) in place of \(\mathbf{Y}\), for a fraction of at least \(\mathcal{I} - \epsilon'\) of the indices. Recall that this dictates \(n_0 = [\nu \bar{n}]\) by Theorem 3. Furthermore, take \(\bar{n}\) large enough such that \(n_0 > m_{1h}^0\) and \(\bar{n} > m_{th}\), where the right-hand sides are given in Lemma 5. We also take \(\bar{n}\) large enough such that \(\frac{N}{\bar{n}} = \frac{X_i}{\bar{n}} > 1 - \epsilon\), which is possible by (15).

We now show that if we take \(\bar{n} \geq n_{thw}^0\), then

\[
Z(U_i | U_i^{-1}, \mathbf{Y}^\star) \leq Z(U_i | U_i^{-1}, \mathbf{Y}^\star) \leq 2^{-N^d}
\]

for at least \((\mathcal{I} - \epsilon') - \frac{\epsilon}{4} = \mathcal{I} - \frac{7\epsilon}{4}\) of the indices as \(n \to \infty\). The inequality (a) results from the TDC being a degradation of the deletion channel. That is, \(X - Y - \mathbf{Y}^*\) 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 \(Z(U_i | U_i^{-1}, \mathbf{Y}^\star)\) and recall that \(\bar{n} \geq n_{thw}^0\).

Notice (8) is satisfied by Lemma 5 (\(\kappa = 3, d = 1\) and \(\gamma = \frac{3}{4}\)) and condition (9) is satisfied from (3) for a fraction of at least \(\mathcal{I} - \frac{\epsilon}{4}\) 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 \(O(\log L)\) for a codeword length of \(|G| = L\).

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 \(N_0\) base trellises are constructed — one for each block. Since we operate on a larger trellis, our complexity is \(O(L^3)\), as opposed to at most \(O(L^2)\) 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


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 $\mathcal{T}$.

After a ‘−’ or ‘+’ transform, each $\mathcal{T}$ 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 $\mathcal{Y}$.

The decoder recursively performs ‘−’ and ‘+’ transformations on $\mathcal{T}$ as follows. First, we perform $n$ ‘−’ transforms, creating $\mathcal{T}^{[000...0]}$. 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: $\mathcal{T}^{[000...1]} = \left(\mathcal{T}^{[000...0]}\right)^{[1]}$. We use $\mathcal{T}^{[000...01]}$ to decide on $\hat{U}_1$. We repeat this procedure such that with trellis $\mathcal{T}^{[b_0b_1...b_n]}$, we decide on the value of $\hat{U}_n$ (if it is not frozen). See Figure 3 for an illustration of the decoding process.

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

$$\sum_{j=0}^{n-1} 2^{j+1} \cdot 8N|\mathcal{S}|^3(2^j+1)^2 \cdot 2^{n-j} = O(N^4)$$

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.

![Fig. 3. Recursive trellis transforms.](image)

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 $\sqrt{n(\neg\text{GBM})}$): Let $X$ be of length $N = 2^n$ and drawn as described in Lemma 5. Let $Y^*$ be the TDC output for input $g(X, n_0, \xi)$. 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 $0 < \xi < \frac{1}{8}$ used in (6), there exists an $m^{th}(\xi, \delta)$, which is a function of the input distribution as well, such that:

$$\sqrt{\mathbb{P}(\neg\text{GBM})} \leq 2^{-N^{\frac{1}{2}}}$$

for all $n \geq m^{th}(\xi, \delta)$ and $n_0 \geq m_0^{th}(\xi) = \log_2\left(\frac{1-2^{-\xi}}{2}\right)$. 

![Fig. 4. Trellis evolution in the decoder.](image)
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 deletions:

\[
\alpha = |G_i| - |Y_i| \\
\beta = |G_\alpha| - |Y_\alpha| \\
\gamma = |G_{\|}| - |Y_{\|}|,
\]

and the following length differences due to trimming:

\[
\alpha' = |Y_i| - |Z_i| \\
\beta' = |Y_\alpha| - |Z_\alpha| \\
\gamma' = |Y_{\|}| - |Z_{\|}|.
\]

We define:

\[
\gamma + \gamma' < \alpha + \alpha' + \ell_n - \beta \quad \text{by the same steps.}
\]

From (21) we get: \(P(GBM) \geq P(A \cap A' \cap B \cap C \cap C')\).

We are interested in the complementary event, which will satisfy:

\[
P(\neg GBM) \leq P(\neg(A \cap A' \cap B \cap C \cap C'))
\]

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

\[P(A)\text{ and } P(\neg B)\text{ may be bounded using Hoeffding [16, Theorem 4.11]}\text{, as in [4, equations (89),(90)]. In Lemma 8 we bound } P(\neg A)\text{. In total, we reach the following upper bound for } P(\neg GBM):\]

\[
P(\neg GBM) \leq 2P(\neg A) + 2P(\neg A') + P(\neg B)
\]

\[\leq 2 \cdot 2e^{-\frac{(1-\delta)^2 \ell_n}{128}} + 2 \cdot e^{-D \cdot (1-\delta)n}
\]

where \(D > 0\) is a constant dependent on the input distribution and on the deletion rate \(\delta\). The value of \(D\) is given explicitly in the proof of Lemma 8. We note that when bounding \(P(A')\) we used the fact that \(n_0 \geq m_0^{th}(\xi) \geq \log_2 e \cdot \left(\frac{1-\delta}{2}\right)^n\), and the qualities of the input distribution we fixed.

Finally, for \(0 < \xi < \frac{1}{2}\) and for a large enough \(n:\)

\[P(\neg GBM) \leq 8e^{-\frac{(1-\delta)^2 \ell_n}{128}} \leq 2^{-2N}\frac{\hat{D}}{4}
\]

specifically, this holds for:

\[n \geq m_0^{th}(\xi, \delta) \iff \max \left\lfloor \frac{1}{\log_2 \left(\frac{(1-\delta)^2}{128}\right)} \right\rfloor \left(\frac{128 \log_2^2(\xi)}{1-2\xi} + 1\right)^{\frac{1}{4}} \log_2 \left(\frac{128 \log_2^2(\xi)}{1-2\xi}\right)
\]

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 \(P(\neg A')\), the probability that ‘too many’ bits were trimmed in \(Y_i\). 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 \(P(\neg A')\)): Let \(A'\) be as in (19b), and let \(m_0^{th}(\xi) \geq \log_2 e \cdot \left(\frac{1-\delta}{2}\right)^n\). Then, for \(n_0 \geq m_0^{th}(\xi):\)

\[P(\neg A') \leq e^{-D \cdot (1-\delta)n}\]

where \(D > 0\) is a constant dependent on the input distribution and on the deletion rate \(\delta\).

Proof: We consider the event \(A''\), defined as follows.

Under the event \(A''\), some index \(j < \hat{\ell}\) in \(G_i\) is a ‘1’ and was not deleted (where \(\hat{\ell}\) was set in (20)). Clearly: \(A'' \Rightarrow A'\), hence,

\[P(\neg A') \leq P(\neg A'').\]
\(\neg A'\) is the complementary event where no index \(j < \ell\) in \(G_1\) is a ‘1’ that was not deleted.

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

\[
\#_X^j \geq \#_{GB}^j, \quad \forall j \in \{1, 2, \ldots, A\}. \tag{22}
\]

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

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

\[
\#_X^j \geq \frac{j}{2}. \tag{23}
\]

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

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

That is, the probability of a ‘1’ bit in a series of \(\tau\) bits in \(X\) is greater than \(1 - p_0\). For each \(\tau\) bits in \(X\), the probability of at least one of them being a ‘1’ bit that was not deleted in the channel is greater than:

\[
(1 - p_0)(1 - \delta). \nonumber
\]

There are \(\left\lfloor \frac{j}{\tau}\right\rfloor\) series of \(X\) bits (of length \(\tau\)) up to index \(\ell\). Thus, by the Markov property:

\[
P(\neg A') \leq (1 - (1 - p_0)(1 - \delta))\left\lfloor \frac{\#_X^\ell}{\tau}\right\rfloor \tag{23}
\]

We continue to upper bound the RHS from above:

\[
\leq (p_0(1 - \delta) + \delta)\left(\frac{1 - \delta}{\tau}\right)^{\#_X^\ell} \tag{20}
\]

\[
\leq (p_0(1 - \delta) + \delta)\frac{1 - \delta}{\tau}^{\frac{\#_X^\ell}{\tau}} \tag{6}
\]

\[
= (p_0(1 - \delta) + \delta)\left(\frac{1}{2^{\#_X^\ell}}\right)^{\#_X^\ell} \tag{4}
\]

\[
= (p_0(1 - \delta) + \delta)\left(\frac{1}{2^{\#_X^\ell}}\right)^{\frac{\#_X^\ell}{\tau}} \tag{20}
\]

\[
= e^{-\frac{1}{2} \ln \left(\frac{1}{2}\right) \frac{\#_X^\ell}{\tau} + \frac{1}{2^{\#_X^\ell}}}. \tag{4}
\]

We mark: \(D = \frac{1}{2} \ln \left(\frac{1}{2^{\#_X^\ell}}\right)\), where \(\tau, p_0\) satisfy (24). We note that \(D > 0\), since \(0 < p_0(1 - \delta) + \delta < 1\). Finally,

\[
P(\neg A') \leq e^{-D \cdot 2^{\#_X^\ell}}. \tag{23}
\]

**B. Guard-band presence in \(g(X)\)**

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

**Lemma 9:** If \(n_0 \geq \log_2\left(\frac{1 - 2^\ell}{2^n}\right)\), then for any \(n \geq n_0 + 1\) and for any given index \(j\) in \(g(X, n_0, \xi)\),

\[
\#_X^j \geq \#_{GB}^j, \nonumber
\]

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

**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:

\[
\#_{GB}^j < \#_X^j < \#_{GB}^n = \#_{GB}^j. \nonumber
\]

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

\[
\#_X^j < \#_{GB}^j < \#_{GB}^n = \#_X^j, \nonumber
\]

for any index \(j\) in \(g(X)\).

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

\[
(n_1, n_0)_{ser} \nonumber
\]

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

\[
(n_1, n_0)_{ser} \otimes (n_1 - 1, n_0)_{ser} \otimes \ldots \otimes (2, n_0)_{ser} \otimes (1, n_0)_{ser} \otimes X(N_1). \nonumber
\]

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

\[
\begin{array}{c c c}
& X(1) & 00 \cdots 0 \\
\downarrow & j_1 & j_i \\
& X(2) & 00 \cdots 0 \\
\downarrow & j_i & j_i \\
& X(3) & 00 \cdots 0 \\
\downarrow & j_i & j_i \\
& X(4) & \\
\end{array}
\]

We notice the following quality. For each index \(j_i\) located at the right edge of some guard-band in \(g(X)\), the series of bits to the left of \(j_i\) are concatenations of the building-blocks:

\[
\{i, n_0)_{ser}\}_{i \in J}, \nonumber
\]

where \(J\) is some subset of \{1, 2, \ldots, n_1\}. 

Therefore, if $\#_{\text{mid}}^X \geq \#_{\text{mid}}^GB$ is satisfied for any $n_1$, then each building-block $(n_1, n_0)_{\text{ser}}$ consists of more (or equal) bits of $X$ than GB bits, leading to:

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

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

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

**Proof:** In the series of bits up to $j_{\text{mid}}$, there are half of the bits of $X$:

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

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

$$\#_{\text{mid}}^GB = \frac{1}{2} \Lambda \geq 2 \ell_n$$

The total number of GB bits satisfies:

$$|g(X)| - |X| = \Lambda - N \leq 2^{(\frac{2^{-\xi n_0}+1}{1-2^{-\xi}})} \leq 2^{\Lambda - N + \ell_n} \leq 2^{n-1}$$

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

The length of the middle GB satisfies:

$$|X| \leq 2^{(1-\xi)(n-1)} \leq 2n-1 \cdot 2^{-\xi n_0},$$

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

$$\#_{\text{mid}}^X \leq 2n-1 \leq 2^{n-1} \cdot 2^{-\xi n_0}$$

where (a) is satisfied for $n_0 \geq \log_2(\frac{1-2^{-\xi}}{2})$ 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 $\nu := \min\{\nu, \frac{1}{3}\}$, and note that if (9) holds for the “old” value of $\nu$, then it surely holds for the “new” value as well.

Let us first define the process $Z^n$: $n_{\text{w}}$ is a parameter that will be fixed later on. Note that the process $Z^n$ 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 $n_{w}^{th} > n_{w}$ be a parameter that will be fixed later on as well. We now fix $\gamma_0$ and $\nu_0$ such that:

$$\frac{1}{2} < \gamma_0 < \gamma$$

$$0 < \nu_0 < \nu < \frac{1}{3}$$

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

$$\Sigma_a \triangleq \{ \tilde{Z}_n \geq 2^{-N^{\gamma_0}} \quad \forall n \geq n_{w} \}$$

$$\Sigma_b \triangleq \{ \tilde{Z}_n \leq 2^{-N^{\nu_0}} \quad \forall n \geq n_{w} \}$$

$$\Sigma_c \triangleq \{ \tilde{Z}_n < \frac{1}{2N} 2^{-N^{\gamma_0}} \quad \forall n \geq n_{w} \}$$

$$\Sigma_d \triangleq \{ Z_n \leq \tilde{Z}_n, \forall n \geq n_{w} \}$$

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

**Claim C.1:** For all $\epsilon \nu > 0$ there exists an $n_{w}^{th}(\epsilon, \nu, \kappa, \nu, \gamma_0)$ s.t. if $n_{w} \geq n_{w}^{th}$ then:

$$P(\Sigma_a) > 1 - \epsilon$$

**Claim C.2:** For all $\epsilon \nu > 0$ there exists an $n_{w}^{th}(\epsilon, \nu, \kappa, \nu, \nu_0)$ s.t. if $n_{w} \geq n_{w}^{th}$, then:

$$P(\Sigma_b) > 1 - \epsilon$$

**Claim C.3:** For all $\epsilon \nu > 0$ there exist $n_{w}^{th}(\epsilon, \nu, \kappa, \nu, \nu_0)$ s.t. if $n_{w} \geq n_{w}^{th}$, then:

$$P(\Sigma_c) > 1 - \epsilon - P(\Sigma_0)$$

**Claim C.4:** There exists an $n_{w}^{th}(\gamma_0, \gamma_0)$ s.t. if $n_{w} \geq n_{w}^{th}$ and if $\tilde{Z}_{n_{w}} \geq Z_{n_{w}}$, then event $\Sigma_0$ implies $\Sigma_d$, i.e.:

$$\tilde{Z}_{n_{w}} \geq Z_{n_{w}} \Rightarrow P(\Sigma_0) = 1$$

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:

$$\epsilon_0 = \epsilon_0 = \epsilon_0 = \epsilon' \frac{3}{3}$$

We also set the starting point:

$$n_{w} \geq n_{w}^{th} \geq \max\left\{ n_{w}^{th}, n_{w}^{th}, n_{w}^{th}, n_{w}^{th}, n_{w}^{th} \right\}$$

and set:

$$n_{w}^{th} \geq \max\left\{ n_{w}^{th}, n_{w} + 1 \right\}$$

Notice that if $Z_{n_{w}}$ satisfies (9), then by (30b), $\tilde{Z}_{n_{w}} \geq Z_{n_{w}}$. Using the four claims above, we can bound the probabilities of events $\Sigma_c, \Sigma_d$. For $\Sigma_c$, we have:

$$P(\Sigma_c) \leq 1 - \epsilon_0 - P(\Sigma_0)$$

(37)
Thus:

\[
P\left( Z_n < \frac{1}{2N} 2^{-N^d}, \ \forall n \geq n_1^{th} \right) \\
\geq P \left( \{ Z_n < \frac{1}{2N} 2^{-N^d}, \ \forall n \geq n_1^{th} \} \cap \{ Z_n \leq \hat{Z}_n, \ \forall n \geq n_w \} \right) \\
= P(\Sigma_c \cap \Sigma_d) \\
= 1 - P(\neg \Sigma_c \cup \neg \Sigma_d) \\
\]

In the last equality we denoted events \( \neg \Sigma_c, \neg \Sigma_d \) as the complementary events of \( \Sigma_c, \Sigma_d \) respectively.

We now upper bound \( P(\neg \Sigma_c \cup \neg \Sigma_d) \):

\[
P(\neg \Sigma_c \cup \neg \Sigma_d) \leq P(\neg \Sigma_c) + P(\neg \Sigma_d) \leq \epsilon_a + \epsilon_b + \epsilon_c \\
\leq \epsilon'\]  

Thus:

\[
P\left( Z_n < \frac{1}{2N} 2^{-N^d}, \ \forall n \geq n_1^{th} \right) \geq P(\Sigma_c \cap \Sigma_d) \geq 1 - \epsilon' \\
\]

**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 \( \Delta_{th} \) for which Hoeffding [16, Theorem 4.12] assures us that the fraction of \( \{ B_{1,1} \}_{i=n_w} \) which are \( 0 \) (‘−’)) and the fraction of \( \{ B_{1,1} \}_{i=n_w} \) which are \( 1 \) (‘+’)) are both close to half, for all \( n \geq n_w + \Delta_{th} \). When this occurs, we can derive the “soft” bounds in (32a) and (32b) for all \( n \geq n_w + \Delta_{th} \). For the initial period of \( n_w \leq n \leq n_w + \Delta_{th} \), 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 \( \Sigma_d \) in order to replace \( Z_n \) in (30a) with the bound from (32b). This leads to a bound on \( \hat{Z}_{n+1} \), which, given \( B_{1,1} \), is a deterministic function of \( n \). Since \( Z_n \) is bounded by \( 2^{-2\gamma n} \), the multiplicative factor of \( 2 \cdot \kappa N^d \) is negligible for a large enough \( n \). That is, the function of \( n \) bounding \( \hat{Z}_{n+1} \) 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 assist us in proving the claim. First, \( Z_n \) is defined to be:

\[
Z_{n+1} = \begin{cases} 
\kappa N^d \cdot Z'_{n} + 2^{-2\gamma n} & \text{if } B_{n+1} = 0 (\text{‘−’}) \\
\kappa N^d \cdot Z''_{n} + 2^{-2\gamma n} & \text{if } B_{n+1} = 1 (\text{‘+’}) 
\end{cases} \quad n \geq n_w \\
Z_{n_w} = Z_{n_w}
\]

i.e. \( Z_n \) from (8), with the weak inequality replaced by equality.

By the monotonicity of the terms in (8) and the above, we easily prove by induction that:

\[
Z_n \leq Z'_n, \quad \forall n \geq n_w.
\]

Next, \( Z''_n \) is set to be:

\[
Z''_{n+1} = \begin{cases} 
\kappa N^d \cdot Z''_{n} + 2^{-2\gamma n} & \text{if } B_{n+1} = 0 (\text{‘−’}) \\
\kappa N^d \cdot Z''_{n} + 2^{-2\gamma n} & \text{if } B_{n+1} = 1 (\text{‘+’}) 
\end{cases} \quad n \geq n_w \\
Z_{n_w}'' = 2^{-2\gamma n_w}.
\]

For any given draw of \( B_{n_w}, \ldots, B_{n} \), 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 monotonicity we prove by induction that \( Z_n \leq Z'_n, \quad \forall n \geq n_w \). Next, we prove that under event \( \Sigma_a \) from (32a), \( Z_n \) of (30) dominates \( Z'_n \). That is,

\[
\Sigma_a \implies Z''_n \leq Z'_n, \quad \forall n \geq n_w.
\]

Under \( \Sigma_a \): \( Z'' \geq 2^{-N^d}, \quad \forall n \geq n_w. \) Since \( \gamma < \gamma_a \), for \( n \geq n_1^{th}(\gamma_a) \), we get \( \gamma_n \leq \gamma_{a,n+1} \). Meaning \( Z_n \) satisfies:

\[
\kappa N^d \cdot Z_n \geq Z_n, \geq 2^{-2\gamma n} \geq 2^{-2\gamma n}
\]

If we assume \( \hat{Z}_n \geq Z''_n \), then \( \hat{Z}_{n+1} \geq Z''_{n+1} \) regardless of \( B_{n+1} \), since:

\[
\hat{Z}_{n+1} = \begin{cases} 
\kappa N^d \cdot Z_{n+1} + \kappa N^d \cdot Z_n & \text{if } B_{n+1} = 0 (\text{‘−’}) \\
\kappa N^d \cdot Z_{n+1} + \kappa N^d \cdot Z_n & \text{if } B_{n+1} = 1 (\text{‘+’}) 
\end{cases} \quad n \geq n_w \\
\geq \begin{cases} 
\kappa N^d \cdot Z_{n+1} + 2^{-2\gamma n} & \text{if } B_{n+1} = 0 (\text{‘−’}) \\
\kappa N^d \cdot Z_{n+1} + 2^{-2\gamma n} & \text{if } B_{n+1} = 1 (\text{‘+’}) 
\end{cases} \quad n \geq n_w \\
\geq \hat{Z}_{n+1} = Z''_{n+1}
\]

where (a) holds under the hypothesis that \( \hat{Z}_n \geq Z''_n \).

We remind that: \( Z_{n_w}'' = \frac{2^{-(2\gamma \gamma a)} - 2^{-2\gamma n_w}}{2^{-(2\gamma \gamma a)}} \). That is, \( Z_n, \hat{Z}_n \) begin at the same value. Thus, we may show by induction that under \( \Sigma_a \), \( \hat{Z}_n \geq Z_{n_w} \), \( \forall n \geq n_w \). We have shown:

\[
Z_n \leq Z'_n \leq Z''_n \leq \hat{Z}_n, \quad \forall n \geq n_w.
\]