Relaying One Bit Across a Tandem of Binary-Symmetric Channels

Wasim Huleihel
Tel-Aviv University
wasim8@gmail.edu

Yury Polyanskiy
MIT
yp@mit.edu

Ofer Shayevitz
Tel-Aviv University
ofersha@eng.tau.ac.il

***Abstract*—We consider the problem of transmitting reliably one bit of information across a tandem of binary symmetric channels interconnected by a relay/processor station. In our setting, the relay is instantaneous in the sense that its outputs are allowed to causally depend on previous received noisy bits. For this model, we investigate the optimal exponential decay rate of the average probability of error, when relaying one bit of information using n synchronous channel uses, by devising good relaying schemes.**

I. Introduction

Consider a chain of k binary symmetric channels with crossover probability δ ($\mathsf{BSC}(\delta)$) interconnected by $k-1$ relays/processors, for $k \geq 1$. We wish to propagate reliably one bit of information over this channel. Fig. 1 describes this model for $k = 2$. In the 2015 program on Information Theory (Simons Institute for the Theory of Computing, Berkeley, CA) the second author of this paper asked the following question: Let the *information velocity* for the above k-hop relaying problem be defined as $\mathsf{IV} \triangleq \lim_{k\to\infty} k/n$, where n is the minimal number of channel uses needed to estimate the information bit with probability of error $1/3$.[1] Then, what can be said about this quantity? More generally, as will be explained later on, it is also interesting to study the error-exponent,

$$\mathsf{E}_{\mathsf{opt},k}(\delta) \triangleq \liminf_{n\to\infty} -\frac{1}{n}\log \mathsf{P}_{\mathsf{error},k}, \tag{1}$$

associated with the k-hop relaying problem, where $\mathsf{P}_{\mathsf{error},k}$ is the optimal probability of error. While it is clear that $\mathsf{E}_{\mathsf{opt},1}(\delta) = d_{\mathsf{KL}}(1/2||\delta)$ (see (2) for more details), finding $\mathsf{E}_{\mathsf{opt},k}(\delta)$ for $k \geq 2$ appears challenging.

In fact, even understanding whether the information velocity is non-zero is non-trivial. Indeed, a straightforward repetition approach requires $O(k\log k)$ time: the first relay takes the majority of the first ℓ bits, for some $\ell \in \mathbb{N}$, and repeats its decision for the rest of the transmission duration. Then, the second relay looks at the second batch of ℓ bits (which are the noisy version of the first relays' repetitive decision), takes majority, and repeats, etc. It is clear that each relay will make an error in decoding the message with probability roughly $2^{-\ell\cdot d_{\mathsf{KL}}(1/2||\delta)}$. Hence, the probability that the decoder will make an error is upper bounded by $k\cdot 2^{-\ell\cdot d_{\mathsf{KL}}(1/2||\delta)}$, and this upper bound is finite if $\ell \approx \frac{\log k}{d_{\mathsf{KL}}(1/2||\delta)}$. Therefore, the total transmission time is approximately $\approx \frac{k\log k}{d_{\mathsf{KL}}(1/2||\delta)}$. It is then evident that for this scheme $\mathsf{IV} = 0$. Nonetheless, in the same workshop, L. J. Schulman pointed out that in [1] it has been already shown that $\mathsf{IV} > 0$ using tree codes.

Intuitively, in order for the information velocity to be positive, the error-exponent $\mathsf{E}_{\mathsf{opt},2}(\delta)$, across two-hops should be comparable to $\mathsf{E}_{\mathsf{opt},1}(\delta)$, across a single-hop, in the high-noise regime, so that "information propagation does not slow down". In other words, we conjecture that $\mathsf{E}_{\mathsf{opt},2}(\delta) = \mathsf{E}_{\mathsf{opt},1}(\delta)\cdot[1+o(1)]$, as $\delta \to 1/2$. Although we are unable to resolve this conjecture in this paper, our main contribution is a simple lower-bound (acheivability) on $\mathsf{E}_{\mathsf{opt},2}(\delta)$, holds for any $\delta < 1/2$. In particular, we show that $\mathsf{E}_{\mathsf{opt},2}(\delta) \geq \frac{3}{4}\cdot \mathsf{E}_{\mathsf{opt},1}(\delta)$, as $\delta \to 1/2$. Our acheivability scheme is based on a sequential majority relay which computes the majority of the its last received symbols, and propagates its decisions to the decoder.

II. Problem Setting and Main Results

Consider the model depicted in Fig. 1, shown at the top of the next page. Specifically, a binary message $M \in \{0,1\}$ is mapped by an encoder into a sequence of n symbols X^n. This sequence is then transmitted through a binary symmetric channel (BSC) with crossover probability $\delta \leq 1/2$. The output of this channel, denoted by Y^n, is then observed by a relay/processor, which maps *causally* the sequence Y^n into another binary sequence U^n. To wit, at time $i \in [1:n]$, the relay's output is $U_i = f_i(Y^i)$, where $f_i : \{0,1\}^i \to \{0,1\}$ is a pre-designed Boolean function. The relay output serves as an input to yet another BSC with the same crossover probability .[2] Finally, a decoder observes the sequence Z^n and assigns an estimate $\hat{M}$ to the message M. Our main goal is to understand what is the "best" scheme that the relay should employ so that the decoder could reliably decode the transmitted message. In particular, we are interested in characterizing the optimal achievable error exponent, namely, $\mathsf{E}_{\mathsf{opt}}(\delta) \triangleq \mathsf{E}_{\mathsf{opt},2}(\delta)$, defined in (1), where, once again, $\mathsf{P}^*_{\mathsf{error},2}$ designates the average probability of error associated with the best triplet; encoder, relay, and decoder, namely, those which minimize the average probability of error. By symmetry it is clear that it is suffice to look at the encoder which maps the messages $M = 0$ and $M = 1$ into the all-zero and all-one sequences, respectively. It is also clear that the maximum-likelihood (ML) decoder is optimal. It is a-priori unclear, however, what the relay should convey to achieve $\mathsf{E}_{\mathsf{opt}}(\delta)$.

The work of W. Huleihel and O. Shayevitz was supported by the European Research Council under grant agreement 639573. The work of Y. Polyanskiy was supported by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370.

[1] The choice of $1/3$ is arbitrary.

[2] The generalization of the results in this paper to different crossover probabilities is straightforward.

Fig. 1. The BSC–Relay–BSC channel.

To better understand the communication limits of the considered problem, let us look at the problem of sending one bit of information over the point-to-point $\mathsf{BSC}(\delta)$ channel, with input X^n and output Y^n. In other words, we consider the case where the secondary BSC in Fig. 1 is noiseless, and then it is evident that the relay should forward its observations to the decoder. It is clear that the error exponent achieved in this scenario serves as an upper bound on $\mathsf{E}_{\mathsf{opt}}(\delta)$. Furthermore, since the maximum-likelihood (ML) decoder in this case is the majority rule, $\mathsf{Majority}\left(Y^n\right)$, using standard concentration bounds, the obtained exponent is

$$\mathsf{E}_{\mathsf{converse}} \triangleq d_{\mathsf{KL}}\left(1/2||\delta\right). \tag{2}$$

Therefore, based on the above observation we may conclude that $0 < \mathsf{E}_{\mathsf{opt}}(\delta) \leq d_{\mathsf{KL}}\left(1/2||\delta\right)$. The question is then how the relay can be used to mitigate the noise in the BSC following it? and how far is $\mathsf{E}_{\mathsf{opt}}(\delta)$ from $d_{\mathsf{KL}}\left(1/2||\delta\right)$? Note that (2) can be interpreted also as the exponent achieved in the non-causal setting of the model in Fig. 1, where the relay have access to Y^n in advance.

Before presenting our main result, we mention two sub-optimal simple schemes which serve as a comparison for the proposed scheme. In the first scheme, the relay simply forwards its observations, namely, $U_i = Y_i$, for $1 \leq i \leq n$. Then, similarly to (2), it is clear that

$$\mathsf{E}_{\mathsf{opt}}(\delta) \geq \mathsf{E}_{\mathsf{forward}}(\delta) \triangleq d_{\mathsf{KL}}\left(1/2||\delta \star \delta\right), \tag{3}$$

where $\delta \star \delta \triangleq 2\delta(1-\delta)$. It turns out that the forwarding scheme is far from being optimal. Generally speaking, in the above scheme the relay tries to convey to the decoder the number of 1's he has seen. Indeed, if the relay can do that perfectly then we would achieve (2). It is clear that this information is quite sensitive and redundant. Instead, all we want is to coordinate the decisions of the relay and the decoder about what the message M was. Accordingly, when the relay becomes more certain about its decision being "0" (or "1") it should transmit more 0's (1's) than what he actually received.

To clarify the idea in the previous paragraph, consider the following scheme which produces a larger error exponent for a wide range of the parameter δ. Specifically, let us look at the case where for the first $n/2$ symbols, the relay is silent (or, equivalently, the decoder simply ignores its first $n/2$ received symbols). For the rest $n/2$ symbols, the relay sends repeatedly its majority decision, $\mathsf{Majority}\left(Y^{n/2}\right)$, based on the previously received $n/2$ symbols $Y^{n/2}$. Accordingly, it is an easy exercise to check that

$$\mathsf{E}_{\mathsf{silent}}(\delta) \triangleq 0.5 \cdot d_{\mathsf{KL}}\left(1/2||\delta\right). \tag{4}$$

Therefore, using the above scheme we achieve half the non-causal performance in (2). We emphasize here that no gain can be obtained by other splitting of the silence and the majority periods.

It should be clear that throwing out the first $n/2$ symbols, as in the above scheme, is quite extravagant. While in principle better exponents can be achieved by, for example, "time-sharing" the forwarding and the silent schemes, no gain will be obtained at the vicinity of $\delta = 1/2$. Interestingly, many schemes (and their combinations) give an exponential decay rate of $\frac{1}{2} \cdot d_{\mathsf{KL}}\left(1/2||\delta\right)$ at the vicinity of $\delta = 1/2$. It turns out that beating this result is quite challenging, placing this objective of particular interest.

We next describe our *sequential majority* scheme, which defeats the previous schemes, also at the vicinity of $\delta = 1/2$. At time $i \in [1:n]$, given Y^i, the relay's output is the majority of the sequence Y^i, i.e.,

$$U_i \triangleq \mathsf{Majority}\left(Y^i\right). \tag{5}$$

Basically, this means upon receiving the i'th symbol, the relay's output is simply its best estimate of the transmitted message based on the received information up to time i. It turns out that the underlying mechanism of the sequential majority scheme is as follows: There is a time[3] τ whose distribution is not affected by the transmitted information bit M, such that everything before τ is independent of M, but after τ the relay cleanly forwards M. With this in mind, consider the following composite hypothesis testing problem:

$$\begin{aligned}\mathsf{H}_0 &: X_t = \mathsf{arbitrary},\ t \in [1:\tau]\\ &\quad X_t = 0,\ t \in [\tau+1:n],\ \mathbb{P}(\tau=\ell) \propto [2\sqrt{\delta(1-\delta)}]^{\ell},\end{aligned}$$

versus H_1, defined similarly but with $X_t = 1$, for $t = [\tau+1:n]$. The receiver observes $\bar{Z}^n \triangleq X^n \oplus \mathsf{Bernoulli}(\delta)^{\otimes n}$, and needs to decide on the true hypothesis. The decoder, upon observing the sequence $z^n \in \{0,1\}^n$, decides that $\hat{M}_{\mathsf{LRT}}(z^n) = 1$, if

$$\max_{\tau' \in [n]} \left\{W_{\tau'} + 0.5 \cdot \tau'\right\} > \max_{\tau \in [n]} \left\{n - 0.5 \cdot \tau - W_{\tau}\right\}, \tag{6}$$

otherwise $\hat{M}_{\mathsf{LRT}}(z^n) = 0$, where $W_{\tau} \triangleq \sum_{t=\tau+1}^{n} z_t$. This rule corresponds to an exponential approximation of the likelihood-ratio test (LRT) in testing H_0 vs. H_1 but with $X_t \sim \mathsf{Bernoulli}(1/2)$, for $t \leq \tau$. Due to space limitation, we provide a proof outline only.

Theorem 1. *Let $\mathsf{E}_{\mathsf{SM}}(\delta)$ designate the error exponent associated with the sequential majority scheme described above. Then, for any $\delta < 1/2$,*

$$\mathsf{E}_{\mathsf{SM}}(\delta) \geq \min_{0 \leq \gamma \leq 1} \max_{\beta \geq 0} \left\{\gamma \cdot d_{\mathsf{KL}}\left(1/2||\delta\right) + \beta \cdot T(\gamma) - g(\gamma,\beta)\right\},$$

where $T(\gamma) \triangleq 1 - \frac{\gamma}{2}$, $g(\gamma,\beta) \triangleq \gamma \cdot \log(\delta + \bar{\delta}2^{\beta}) + \bar{\gamma} \cdot \log(\bar{\delta} + \delta 2^{2\beta})$, and $\bar{x} = 1 - x$, for $0 \leq x \leq 1$.

Proof Outline: First, it can be shown that $\mathsf{P}_{\mathsf{error}}^{\mathsf{SM}} \leq \mathsf{P}_{\mathsf{error}}^{\mathsf{HT}}$, where $\mathsf{P}_{\mathsf{error}}^{\mathsf{HT}}$ is the error probability associated with the above

[3]Formally, $\tau \triangleq \inf\left\{t \geq 0 : \mathsf{H}(S_t) = 0,\ \mathsf{H}(S_\ell) = M,\ \ell > t\right\}$, where $S_t \triangleq \sum_{i=1}^{t}(2Y_i - 1)$ is the sum of the first t received bits by the relay shifted to $\{-1,1\}$, and $\mathsf{H}(x) = 1$ if $x > 0$, and zero otherwise.

composite hypothesis testing problem. Accordingly,[4]

$$\begin{aligned}\mathsf{P}_{\mathsf{error}}^{\mathsf{HT}} &= \mathbb{E}_{\tau|H_0}\left[\mathbb{P}\left(\mathsf{error}|\mathsf{H}_0,\tau\right)\right] \\ &\doteq \max_{0\le\tau\le n} 2^{-\tau d_{\mathsf{KL}}(1/2||\delta)}\mathbb{P}\left(\mathsf{error}|\mathsf{H}_0,\tau\right).\end{aligned}$$

Then, it is clear that

$$\mathbb{P}\left(\mathsf{error}|\mathsf{H}_0,\tau\right) \le \max_{\tau'}\mathbb{P}\left(W_\tau + W_{\tau'} \ge n - \frac{\tau+\tau'}{2}\middle|\mathsf{H}_0,\tau\right).$$

Now, use Chernoff's inequality and the fact that $Z_t|\tau \sim \mathsf{Bernoulli}(\delta)$, for $t > \tau$, and $Z_t|\tau \sim \mathsf{Bernoulli}(1-\delta)$, for $t \le \tau$, to overbound the above probability. ■

We next describe our second scheme which is based on a block-wise relaxation of the above sequential relaying scheme. For $1 \le m \le \sqrt{n}-1$, and $m\sqrt{n}+1 \le i \le (m+1)\sqrt{n}$, the relay output at time i is given by[5]

$$U_i \triangleq \mathsf{Majority}\left(Y^{m\sqrt{n}}\right). \tag{7}$$

In words, we split the n length sequence y^n into $\sqrt{n}$ blocks, each block of size $\sqrt{n}$. Then, at the beginning of the m'th block, the relay sends repeatedly, for the next $\sqrt{n}$ symbols, its majority decision based on the last m blocks. It should be emphasized here that the $\sqrt{n}$ choice for the block size, is not special, and in fact any sub-linear choice is equally good in the error exponent sense.

On the other side, it is clear that the optimal decoder, minimizing the average probability of error, is the ML decoder. Specifically, upon observing the sequence $z^n \in \{0,1\}^n$, the ML decoder decision is $\hat{M}_{\mathsf{ML}}(z^n) = 1$ if

$$P_{Z^n|M}\left(z^n|1\right) > P_{Z^n|M}\left(z^n|0\right), \tag{8}$$

otherwise, $\hat{M}_{\mathsf{ML}}(z^n) = 0$. Let us simplify (8). For any $y^n \in \{0,1\}^n$, let $f^n(y^n)$ designate the n-bit output sequence of the relay, namely, $f^n(y^n) = (U_1,\ldots,U_n)$, with U_i defined in (7). We denote by $\mathbb{E}_\ell$ the expectation taken w.r.t. Y^n given $M = \ell$, for $\ell = 0,1$. Finally let $\alpha \triangleq \delta/(1-\delta)$. Straightforward algebra steps reveal that for $\ell = 0,1$,

$$P_{Z^n|M}\left(z^n|\ell\right) = \bar{\delta}^n \cdot \mathbb{E}_\ell\left[\alpha^{\mathsf{d_H}(z^n,f^n(Y^n))}\right], \tag{9}$$

where $\mathsf{d_H}(\cdot,\cdot)$ is the Hamming distance. Therefore, (8) reduces to deciding $\hat{M}_{\mathsf{ML}}(z^n) = 1$, if

$$\mathbb{E}_1\left[\alpha^{\mathsf{d_H}(z^n,f^n(Y^n))}\right] > \mathbb{E}_0\left[\alpha^{\mathsf{d_H}(z^n,f^n(Y^n))}\right], \tag{10}$$

otherwise, $\hat{M}_{\mathsf{ML}}(z^n) = 0$. The difficulty in analyzing the above scheme stems from the fact that the relay operation induces strong memory in the channel, contrary to the previous schemes.

To present our main result, we establish first some notation. For any $0 \le \gamma \le 1$, let

$$\begin{aligned}\Gamma(\gamma,\delta) &\triangleq \gamma \cdot d_{\mathsf{KL}}(1/2||\delta) \\ &- \frac{\log\alpha}{4(1-2\delta)}\frac{\left[\frac{d_{\mathsf{KL}}(1/2||\delta)}{\log\alpha}\gamma + (1-2\delta)(1-\gamma)\right]_+^2}{1-\gamma},\end{aligned} \tag{11}$$

[4]For two positive sequences a_n and b_n, the notation $a_n \doteq b_n$ stands for equality in the exponential scale, namely, $\lim_{n\to\infty}\frac{1}{n}\log\frac{a_n}{b_n} = 0$.

[5]For $1 \le i \le \sqrt{n}$, the relay outputs U_i are defined in an arbitrary fashion, and can be ignored at the decoder. It is evident that this does not affect the achievable exponent.

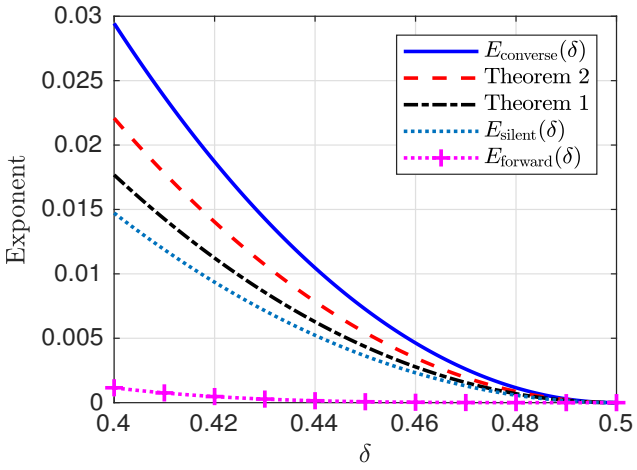

Fig. 2. Comparison of the achievable error exponents as a function of the noise parameter δ, and the trivial converse bound.

where $[a]_+ = \max(0,a)$, for any $a \in \mathbb{R}$. We have the following result, proved in Section III.

Theorem 2. *Let $\mathsf{E_{BW}}(\delta)$ designate the error exponent associated with the block-wise majority scheme described above. Then, for any $\delta < 1/2$,*

$$\mathsf{E_{BW}}(\delta) \ge \min_{0\le\gamma\le1}\Gamma(\gamma,\delta), \tag{12}$$

where $\Gamma(\gamma,\delta)$ is defined in (11).

Fig. 2 presents a comparison of the several exponents discussed in this paper. Specifically, it can be seen that the exponent in Theorem 2 is the best among the considered schemes. Also, it can be shown that at the vicinity of $\delta = 1/2$, the proposed scheme achieves $\mathsf{E_{BW}}(\delta \to 1/2) = \frac{3}{4}\cdot d_{\mathsf{KL}}(1/2||\delta)$, and the minimizer of (12) is at $\gamma = 2/3$, while $\mathsf{E_{SM}}(\delta \to 1/2) \approx 0.6\cdot d_{\mathsf{KL}}(1/2||\delta)$, and the minimizer is at $\gamma \approx 0.51$. Note that the while the error exponent in Theorem 2 is strictly better than the error exponent in Theorem 1, we do not claim that the block-wise majority scheme is in general better than the sequential majority scheme. Indeed, since Theorem 1 provides a lower bound only, it could be the case that the error exponent associated with the sequential majority scheme is larger.

III. Proof of Theorem 2

Due to page limitation some technical details will be omitted. Recall the relay operation in (7), and note that it can be rewritten as follows: For $1 \le m \le \sqrt{n}-1$, and $m\sqrt{n}+1 \le i \le (m+1)\sqrt{n}$,

$$U_i \triangleq \mathsf{H}\left(\frac{1}{m\sqrt{n}}\sum_{j=1}^{m\sqrt{n}} y_j - \frac{1}{2}\right), \tag{13}$$

where $\mathsf{H}(\cdot)$ is the Heaviside step function, namely, $\mathsf{H}(x) = 1$ if $x > 0$, and zero otherwise. We next analyze the probability of error associated with (10). Let $\mathcal{C}_n = \{0,1\}^n$ be the set of all n-length binary sequences. For $0 \le \ell \le 2^{\sqrt{n}}-1$ we let $\mathbf{b}^\ell$ be the *left-to-right* $\sqrt{n}$–binary expansion of ℓ. By "left-to-right"

we mean that the least significant bit in the expansion appears first from the left, for example, $\mathbf{b}^1 = (1, 0, \ldots, 0)$, rather than $(0, 0, \ldots, 1)$ (as is customary for binary expansions). Finally, we define the *profile* of a sequence y^n, as follows

$$\mathcal{A}_\ell \triangleq \Bigg\{ y^n \in \mathcal{C}_n : \frac{1}{j\sqrt{n}} \sum_{i=1}^{j\sqrt{n}} y_i > \frac{1}{2}, \text{ if } \mathbf{b}_j^\ell = 1, \\ \frac{1}{j\sqrt{n}} \sum_{i=1}^{j\sqrt{n}} y_i \leq \frac{1}{2}, \text{ if } \mathbf{b}_j^\ell = 0, \ \forall j \in [\sqrt{n}] \Bigg\}. \quad (14)$$

For example, for $\ell = 2^{\sqrt{n}-1}$ we have $\mathbf{b}^{2^{\sqrt{n}-1}} = (1, \ldots, 1)$, and thus,

$$\mathcal{A}_{2^{\sqrt{n}}-1} = \left\{ y^n \in \mathcal{C}_n : \ \frac{1}{\sqrt{n}} \sum_{i=1}^{\sqrt{n}} y_i > \frac{1}{2}, \ldots, \frac{1}{n} \sum_{i=1}^{n} y_i > \frac{1}{2} \right\}.$$

Therefore, $\{\mathcal{A}_\ell\}$ forms a partition of all the $2^{\sqrt{n}}$ possible majority paths of the received sequence y^n. Recall (10). Then, it is clear that

$$\begin{aligned} \mathsf{P_e} &= \mathbb{P}\left[\hat{M}_{\mathsf{ML}}(Z^n) = 1 \middle| \, M = 0 \right] \\ &= \sum_{\ell=0}^{2^{\sqrt{n}}-1} \mathbb{P}\left[Y^n \in \mathcal{A}_\ell \middle| \, M = 0 \right] \mathbb{P}_{\mathcal{A}_\ell}\left[\hat{M}_{\mathsf{ML}}(Z^n) = 1 \right], \quad (15) \end{aligned}$$

where we denote by $\mathbb{P}_{\mathcal{A}_\ell}(\mathcal{B})$ the probability of $\mathcal{B}$ conditioned on $Y^n \in \mathcal{A}_\ell$, and we have used the fact that $M \multimap (Y^n \in \mathcal{A}_\ell) \multimap \{Z_i\}_i$ forms a Markov chain for any ℓ.

We next analyze the probability that $Y^n \in \mathcal{A}_\ell$ given $M = 0$ for any ℓ. To this end, let $L(\ell)$ be the index of the least significant bit which is 1 in the $\sqrt{n}$ binary expansion of ℓ. For example, the binary expansion of $\ell = 2^{\sqrt{n}} - 1$ is $(111 \ldots 11)$, and thus $L(2^{\sqrt{n}} - 1) = \sqrt{n}$. Similarly, for $\ell = 0$ we have $L(0) = 0$. Accordingly, using a simple concentration bound on the tails of a binomial process we have, for any $\ell \geq 1$,

$$\begin{aligned} \mathbb{P}\left[Y^n \in \mathcal{A}_\ell \middle| \, M = 0 \right] &\leq \mathbb{P}\left[\frac{1}{L(\ell)\sqrt{n}} \sum_{i=1}^{L(\ell)\sqrt{n}} Y_i > \frac{1}{2} \middle| \, M = 0 \right] \\ &\doteq 2^{-L(\ell)\sqrt{n} \cdot d_{\mathsf{KL}}(1/2||\delta)}. \quad (16) \end{aligned}$$

For $\ell = 0$, we trivially have that $\mathbb{P}\left[Y^n \in \mathcal{A}_0 \middle| \, M = 0 \right] \leq 1$ In fact, we claim that the above inequality is tight in the exponential scale. We next show that this is correct for $\ell = 2^{\sqrt{n}} - 1$, namely,

$$\begin{aligned} &\mathbb{P}\left[Y^n \in \mathcal{A}_{2^{\sqrt{n}}-1} \middle| \, M = 0 \right] \\ &= \mathbb{P}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{\sqrt{n}} Y_i > \frac{1}{2}, \ldots, \frac{1}{n} \sum_{i=1}^{n} Y_i > \frac{1}{2} \middle| \, M = 0 \right] \\ &= \mathbb{P}\left[S_{\sqrt{n}} > 0, \ldots, S_n > 0 \middle| \, M = 0 \right], \quad (17) \end{aligned}$$

where the second equality follows by letting $V_i \triangleq 2Y_i - 1$, for any $1 \leq i \leq n$, and $S_i \triangleq \sum_{j=1}^{i} V_i$. Noticing that $(S_i)_i$ forms a simple biased random walk, the above asks for the probability that a random walk stays positive at times $\sqrt{n}, 2\sqrt{n}, \ldots, n$. This can be answered using the reflection principle [2, Chapter 2.7]. Indeed, recall that for any $k > 0$, we have[6]

$$\mathbb{P}\left[S_1 > 0, \ldots, S_{n-1} > 0, S_n = k \right] = \frac{k}{n} \mathbb{P}\left(S_n = k \right), \quad (18)$$

while we are in fact asking for much less restrictive set of conditions. Therefore, on the one hand, we trivially have

$$\mathbb{P}\left[S_{\sqrt{n}} > 0, \ldots, S_n > 0 \middle| \, M = 0 \right] \leq \mathbb{P}\left(S_n > 0 | M = 0 \right), \quad (19)$$

while on the other hand,

$$\begin{aligned} &\mathbb{P}\left[S_{\sqrt{n}} > 0, \ldots, S_n > 0 \middle| \, M = 0 \right] \\ &\geq \mathbb{P}\left[S_1 > 0, S_2 > 0, \ldots, S_n > 0 \middle| \, M = 0 \right] \\ &= \sum_{k>0} \mathbb{P}\left[S_1 > 0, S_2 > 0, \ldots, S_n = k \middle| \, M = 0 \right] \\ &= \sum_{k>0} \frac{k}{n} \mathbb{P}\left(S_n = k | M = 0 \right) \geq \frac{1}{n} \mathbb{P}\left(S_n > 0 | M = 0 \right). \quad (20) \end{aligned}$$

Combining (19) and (20) we obtain

$$\mathbb{P}\left[S_{\sqrt{n}} > 0, \ldots, S_n > 0 \middle| \, M = 0 \right] \doteq \mathbb{P}\left(S_n > 0 | M = 0 \right),$$

as claimed. In other words, the contribution of the intersection gives a polynomial decay factor which do not effect the exponent. Using the above result and (15), we get

$$\begin{aligned} \mathsf{P}_e &\doteq \sum_{\ell=0}^{2^{\sqrt{n}}-1} 2^{-L(\ell)\sqrt{n} \cdot d_{\mathsf{KL}}(1/2||\delta)} \mathbb{P}_{\mathcal{A}_\ell}\left[\hat{M}_{\mathsf{ML}}(Z^n) = 1 \right] \\ &\doteq \max_{0 \leq \ell \leq 2^{\sqrt{n}}-1} 2^{-L(\ell)\sqrt{n} \cdot d_{\mathsf{KL}}(1/2||\delta)} \mathbb{P}_{\mathcal{A}_\ell}\left[\hat{M}_{\mathsf{ML}}(Z^n) = 1 \right]. \quad (21) \end{aligned}$$

We next analyze the probability term at the r.h.s. of (21). To this end, we will use Chernoff's bound. Given $\mathcal{A}_\ell$, the sequence $\{Z_i\}_{i=1}^n$ forms an independent Bernoulli sequence with alternating parameters δ and $1 - \delta$. Let $p_{i,\ell} = \mathbb{P}(Z_i = 1 | \mathcal{A}_\ell)$ and it is clear that $p_{i,\ell} \in \{\delta, 1 - \delta\}$ depending on ℓ. For example, for $\ell = 0$, we have $p_{i,0} = \delta$ for any i, while $\ell = 2^{\sqrt{n}} - 1$ we have $p_{i,0} = 1 - \delta$ for any i. More generally, for a given certain ℓ, in order to understand the corresponding sequence $\{p_{i,\ell}\}_i$, we look at the $\sqrt{n}$ binary expansion of ℓ, and then if at some index j of the binary expansion we have "0" then $p_{i,\ell} = \delta$ for $j\sqrt{n} + 1 \leq i \leq (j+1)\sqrt{n}$, otherwise, if at some index j of the binary expansion we have "1" then $p_{i,\ell} = 1 - \delta$ for $j\sqrt{n} + 1 \leq i \leq (j+1)\sqrt{n}$. Using (10), we have

$$\begin{aligned} \mathbb{P}_{\mathcal{A}_\ell}\left[\hat{M}_{\mathsf{ML}}(Z^n) = 1 \right] &= \mathbb{P}_{\mathcal{A}_\ell}\left[\frac{\mathbb{E}_1\left[\alpha^{\mathsf{d_H}(Z^n, f^n(Y^n))} \right]}{\mathbb{E}_0\left[\alpha^{\mathsf{d_H}(Z^n, f^n(Y^n))} \right]} > 1 \right] \\ &\triangleq \lambda_\ell, \quad (22) \end{aligned}$$

where the inner expectations are evaluated w.r.t. Y^n given $M = m$ while Z^n is treated as a constant, and then as a random variable when calculating the probability $\mathbb{P}_{\mathcal{A}_\ell}$. We next simplify the above expression. To this end, notice that given $Y^n \in \mathcal{A}_\ell$, the sequence $f^n(Y^n)$ is independent of Y^n. In other words, $f^n(Y^n)$ depends on Y^n only through its profile. Accordingly, with some abuse of notation, for a given ℓ, we let

[6]Note that the reflection principle applies here because the distribution of $(S_1, \ldots, S_n)$ given that does not depend on the bias of the walk.

$\mathsf{d}_\mathsf{H}(z^n, \mathbf{c}_\ell) = \mathsf{d}_\mathsf{H}(z^n, f^n(\mathcal{A}_\ell))$ denote the Hamming distance between z^n and the sequence $\mathbf{c}_\ell$ obtained by applying f^n on any sequence in $\mathcal{A}_\ell$. Then,

$$\log \mathbb{E}_0 \left[\alpha^{\mathsf{d}_\mathsf{H}(z^n, f^n(Y^n))}\right] = \log \sum_\ell \alpha^{\mathsf{d}_\mathsf{H}(z^n, \mathbf{c}_\ell)} \mathbb{P}_0(\mathcal{A}_\ell)$$
$$\doteq \max_\ell \left\{\mathsf{d}_\mathsf{H}(z^n, \mathbf{c}_\ell) \log \alpha - L(\ell)\sqrt{n} \cdot d_{\mathsf{KL}}(1/2||\delta)\right\}. \quad (23)$$

Similarly,

$$\log \mathbb{E}_1 \left[\alpha^{\mathsf{d}_\mathsf{H}(z^n, f^n(Y^n))}\right] = \log \sum_\ell \alpha^{\mathsf{d}_\mathsf{H}(z^n, \mathbf{c}_\ell)} \mathbb{P}_1(\mathcal{A}_\ell)$$
$$\doteq \max_\ell \left\{\mathsf{d}_\mathsf{H}(z^n, \mathbf{c}_\ell) \log \alpha - M(\ell)\sqrt{n} \cdot d_{\mathsf{KL}}(1/2||\delta)\right\}, \quad (24)$$

where $M(\ell)$ is the index of the least significant bit which is 0 in the $\sqrt{n}$ binary expansion of ℓ. In the following, instead of analyzing the ML decoder (22), we analyze the approximated-ML decoder which compares (23) and (24). While doing so trivially serves as an upper bound on (22), it can be shown that there is no loss in the attained exponent. Due to page limitation, the proof is omitted. Let

$$T_{\ell',\ell''} \triangleq \sqrt{n} d_{\mathsf{KL}}(1/2||\delta) \frac{M(\ell') - L(\ell'')}{-\log \alpha}. \quad (25)$$

Then, it is clear that $\lambda_\ell \leq \max_{\ell'} \min_{\ell''} \beta_{\ell,\ell',\ell''}$, where

$$\beta_{\ell,\ell',\ell''} \triangleq \mathbb{P}_{\mathcal{A}_\ell} \left[\mathsf{d}_\mathsf{H}(Z^n, \mathbf{c}_{\ell''}) - \mathsf{d}_\mathsf{H}(Z^n, \mathbf{c}_{\ell'}) > T_{\ell',\ell''}\right]. \quad (26)$$

Note that

$$\mu_{\ell,\ell',\ell''} \triangleq \mathbb{E}_{\mathcal{A}_\ell} \left[\mathsf{d}_\mathsf{H}(Z^n, \mathbf{c}_{\ell''}) - \mathsf{d}_\mathsf{H}(Z^n, \mathbf{c}_{\ell'})\right]$$
$$= \sum_{i=1}^n (1 - 2p_{i,\ell}) \left[\mathsf{d}_\mathsf{H}(1, \mathbf{c}_{i,\ell'}) - \mathsf{d}_\mathsf{H}(1, \mathbf{c}_{i,\ell''})\right]. \quad (27)$$

Denote the summand in (27) by $\mu_{i,\ell,\ell',\ell''}$. Then, for any $\beta \geq 0$, we have

$$\beta_{\ell,\ell',\ell''} \leq 2^{-\beta\left(T_{\ell',\ell''} - \mu_{\ell,\ell',\ell''}\right)} \prod_{i=1}^n \mathbb{E}_{\mathcal{A}_\ell} \left[2^{\beta V_i}\right], \quad (28)$$

where $V_i \triangleq \mathsf{d}_\mathsf{H}(Z_i, \mathbf{c}_{i,\ell''}) - \mathsf{d}_\mathsf{H}(Z_i, \mathbf{c}_{i,\ell'}) - \mu_{i,\ell,\ell',\ell''}$. Let $\alpha_i \triangleq \mathsf{d}_\mathsf{H}(1, \mathbf{c}_{i,\ell''}) - \mathsf{d}_\mathsf{H}(1, \mathbf{c}_{i,\ell'})$. It is easy to check that

$$\mathbb{E}_{\mathcal{A}_\ell} \left[2^{\beta V_i}\right] = p_{i,\ell} 2^{2\beta\alpha_i(1-p_{i,\ell})} + (1 - p_{i,\ell}) 2^{-2\beta\alpha_i p_{i,\ell}}. \quad (29)$$

While in principle we can use the above Chernoff's inequality mechanism all the way, it results in more complicated formulas for the error exponent. To this end, we use the following inequality [3] which simplifies our subsequent derivations. Numerical calculations suggest that the differences are insignificant. For all $p \in [0, 1]$ and $t \in \mathbb{R}$,

$$(1 - p)2^{-tp} + p2^{t(1-p)} \leq \exp_2 \left(\frac{1 - 2p}{4 \log \frac{1-p}{p}} t^2\right), \quad (30)$$

where $\exp_2(a) \triangleq 2^a$, for any $a \in \mathbb{R}$. Hence, we get

$$\mathbb{E}_{\mathcal{A}_\ell} \left[2^{\beta V_i}\right] \leq \exp_2 \left(\frac{1 - 2p_{i,\ell}}{\log \frac{1-p_{i,\ell}}{p_{i,\ell}}} \beta^2 \alpha_i^2\right). \quad (31)$$

Therefore,

$$\beta_{\ell,\ell',\ell''} \leq$$
$$\exp_2 \left(-\beta \left(T_{\ell',\ell''} - \mu_{\ell,\ell',\ell''}\right) + \beta^2 \sum_{i=1}^n \frac{1 - 2p_{i,\ell}}{\log \frac{1-p_{i,\ell}}{p_{i,\ell}}} \alpha_i^2\right). \quad (32)$$

Note that since $p_{i,\ell} \in \{\delta, 1 - \delta\}$, whatever is the assignment of the probabilities $\{p_{i,\ell}\}$, we have

$$\sum_{i=1}^n \frac{1 - 2p_{i,\ell}}{\log \frac{1-p_{i,\ell}}{p_{i,\ell}}} \alpha_i^2 = \frac{1 - 2\delta}{\log \frac{1-\delta}{\delta}} \sum_{i=1}^n \alpha_i^2. \quad (33)$$

For simplicity of notation, we let $\nu \triangleq \frac{1-2\delta}{\log \frac{1-\delta}{\delta}}$. Optimizing over β we finally obtain,

$$\beta_{\ell,\ell',\ell''} \leq \exp_2 \left[-\frac{\left[T_{\ell',\ell''} - \mu_{\ell,\ell',\ell''}\right]_+^2}{4\nu \sum_{i=1}^n \alpha_i^2}\right]. \quad (34)$$

Substituting the last result in (21), we obtain

$$\mathsf{P}_e \leq \max_{\ell,\ell'} \min_{\ell''} \exp_2 \left[-\Lambda(\ell, \ell', \ell'')\right] \leq \max_{\ell,\ell'} \exp_2 \left[-\Lambda(\ell, \ell', \ell)\right], \quad (35)$$

where

$$\Lambda(\ell, \ell', \ell'') \triangleq L(\ell)\sqrt{n} \cdot d_{\mathsf{KL}}(1/2||\delta) + \frac{\left[T_{\ell',\ell''} - \mu_{\ell,\ell',\ell''}\right]_+^2}{4\nu \sum_{i=1}^n \alpha_i^2}.$$

The maximization over ℓ and ℓ' can be further simplified. In particular, it can be shown that

$$\min_{\ell,\ell'} \Lambda(\ell, \ell', \ell) = \min_{0 \leq \gamma \leq 1} \left\{n\gamma \cdot d_{\mathsf{KL}}(1/2||\delta) + n \frac{\left[T_\gamma + (1 - 2\delta)(1 - \gamma)\right]_+^2}{4\nu(1 - \gamma)}\right\}, \quad (36)$$

where $T_\gamma = d_{\mathsf{KL}}(1/2||\delta) \frac{\gamma}{\log \alpha}$. Due to length constraints the derivation of (36) is omitted. Thus,

$$\lim_{n \to \infty} -\frac{1}{n} \log \mathsf{P}_e \geq \min_{0 \leq \gamma \leq 1} \left\{\gamma \cdot d_{\mathsf{KL}}(1/2||\delta) + \frac{\log \frac{1-\delta}{\delta}}{4(1 - 2\delta)} \frac{\left[T_\gamma + (1 - 2\delta)(1 - \gamma)\right]_+^2}{1 - \gamma}\right\}, \quad (37)$$

which concludes the proof.

ACKNOWLEDGMENT

The authors would like to thank Or Ordentlich for many fruitful discussions.

REFERENCES

[1] S. Rajagopalan and L. D. Schulman, "A coding theorem for distribute computation," in *Proc. 26th Ann. ACM Symp. Theory Comput.*, 1994, pp. 790–799.

[2] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*. American Mathematical Society, 2009.

[3] M. J. Kearns and L. K. Saul, "Large deviation methods for approximate probabilistic inference," in *Proc. 14th Conference on Uncertainly in Artificial Intelligence*, 1998, pp. 311–319.