

Models and information-theoretic bounds for nanopore sequencing

Wei Mao, Suhas Diggavi, and Sreeram Kannan

Abstract—Nanopore sequencing is an emerging new technology for sequencing DNA, which can read long fragments of DNA ($\sim 50,000$ bases) unlike most current sequencers which can only read hundreds of bases. While nanopore sequencers can acquire long reads, the high error rates ($\approx 30\%$) pose a technical challenge. In a nanopore sequencer, a DNA is migrated through a nanopore and current variations are measured. The DNA sequence is inferred from this observed current pattern using an algorithm called a *base-caller*. In this paper, we propose a mathematical model for the “channel” from the input DNA sequence to the observed current, and calculate bounds on the information extraction capacity of the nanopore sequencer. This model incorporates impairments like inter-symbol interference, deletions, as well as random response. The practical application of such information bounds is two-fold: (1) benchmarking present base-calling algorithms, and (2) offering an optimization objective for designing better nanopore sequencers.

I. INTRODUCTION

DNA sequencing technology has undergone a major revolution with the cost of sequencing plummeting nearly six orders of magnitude. Much of this improvement was made possible by second generation sequencers, utilizing massive parallelization, but these machines can only read short fragments of DNA, typically a few hundred bases long. These short reads are then stitched together with algorithms exploiting the overlap between reads to assemble them into long DNA sequences. This assembly is unreliable because of repeated regions which commonly occur in genomic DNA. These repeated regions play an important role in evolution, development and in the genetics of many diseases.

Nanopore sequencing promises to address this problem, by increasing the read lengths by orders of magnitude (up to 100K bases) [1]. The technology is based on DNA transmigrated through a nanopore, and the ion current variations through the pore are measured [2]. The sequence of DNA bases is inferred from the observed current traces using an algorithm termed as base-caller. Nanopore-sequencing technology is also beginning to be commercialized by Oxford Nanopore Technologies. Nanopore sequencers have enabled new applications such as rapid virology and forensics.

However, despite recent progress, there is an important bottleneck; nanopore sequencers built to date, have high error rate for *de novo* sequencing¹. It is unclear whether the present error-rate (of 30% for single-strand sequencing [3]) is fundamental to the nanopore or due to the limitations of present base-calling algorithms. Thus one goal of the present

direction of work is to understand the information-theoretic limits of the nanopore sequencer, and help benchmark base-calling algorithms. To achieve this goal, standardized signal-level models are needed to analyze such sequencers, which are currently unavailable. Another important benefit of such information theoretic understanding is that it provides a way to optimize the nanopore sequencer (for maximum information capacity).

Motivated by this, our first contribution is in developing a mathematical model for the nanopore sequencer. We use the physics of the nanopore sequencers and experimental data provided by our collaborators [2] to develop a signal-level model for the *nanopore channel*, which captures (non-linear) inter-symbol interference, signal-level synchronization errors, and random response of the nanopore to the same inputs. Our second contribution is capacity upper and lower bounds for the nanopore channel. We develop lower bounds on the information capacity of the nanopore channel using techniques for deletion channels. These lower bounds can be used for the list size estimation for sequencing arbitrary sequences as well as assessing DNA storage capability of nanopore decoding devices. We also develop novel *computable* upper bounds for the decoding capacity of nanopore channel, using a combination of upper-bounding techniques for deletion channels and finite-state channels. We numerically evaluate these bounds for both synthetic data using nanopore models as well as measured responses from nanopore data.

The major technical challenge for the analysis of the nanopore channel lies in the fact that it belongs to the category of channels with synchronization errors. The study of such channels dates back to Gallager [4] and Dobrushin [5], and interest in the problem was revived due to new bounds in [6]. See [7] for a survey of results. However, even the simplest i.i.d. deletion channel capacity is unresolved. In this paper we develop novel lower and upper bounds for a channel model in the this category, which not only are useful to analyze the nanopore sequencers, but are also interesting technical results in their own right. In [8], a nanopore sequencer is modeled at a *hard-decision* level by a simplified insertion-deletion channel, where no run of DNA bases is deleted, to understand how to combine multiple reads. Our channel model, however, is aimed at designing base-callers (decoding DNA from current trace) and therefore operates at fine-grained signal level.

The paper is organized as follows. Section II develops the signal-level model for the nanopore channel. We outline the main technical results of this paper in Section III. We develop the proof of the achievable rates for the nanopore channel in

¹*De novo* sequencing refers to decoding the DNA sequence without the help of any reference DNA.

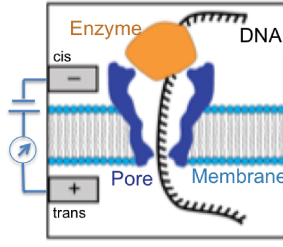


Fig. 1. Nanopore Sequencer

Section IV and the upper bounds in Section V. Numerical evaluation of these bounds, including with real measurements from nanopore sequencers provided by UW nanopore lab [2] are given in Section VI. Details of the proofs and further references are given in [9].

II. MODEL, NOTATION AND PROBLEM FORMULATION

A. Nanopore sequencer

We will use the nanopore sequencer shown in the simplified schematic in Fig. 1, with details in [2]. A salt solution is divided into two wells: *cis* well and *trans* well by a lipid bilayer membrane. A nanopore is inserted into the bilayer, and a voltage applied between the cis and trans wells results in an ionic current. The (single-strand) DNA sequence to be measured is prefixed and affixed with short known DNA pieces (called adaptors) and enters the nanopore. As the DNA migrates through the pore, modulations of the ionic current caused by different nucleotides partially blocking the pore are measured. An enzyme controls the DNA motion through the nanopore, slowing it down so that the current variations can be measured accurately.

An ideal nanopore sequencing system would ensure that DNA migrates through the pore at a constant rate, that only one base of DNA affects the current at a given time and from the observed current, the nucleotide of DNA can be decoded unambiguously with high probability. However, the following phenomena occur due to the physics of the nanopore and the enzyme (see Fig. 2). (i) *Random dwelling time*: Each nucleotide may spend a varying amount of time in the nanopore, resulting in unequal step size/level length. This is because the speed at which the DNA sequence migrates through the enzyme is a stochastic process. (ii) *Inter-symbol interference*: Each observed current value is influenced by multiple bases or “*K*-mer” (*K* bases), because the constriction of the nanopore is thicker than a single nucleotide. In our experiments $K \approx 4$, so we also call them quadromers or “*Q*-mers” for short. As in DNA sequences there are 4 different type of nucleotides, there are $4^4 = 256$ different *Q*-mers all together. (iii) *Backtracking and skipings*: The nucleotides are not necessarily read in order by the nanopore - there is some backtracking and significant mis-stepping in the nanopore induced by the enzyme that draws the DNA through the nanopore. This results in segments that are repeated as well as segments that pass through without registering a current reading, resulting in backstepping as well as skipping (deletion) of segments. (iv) *Q-mer map fading*:

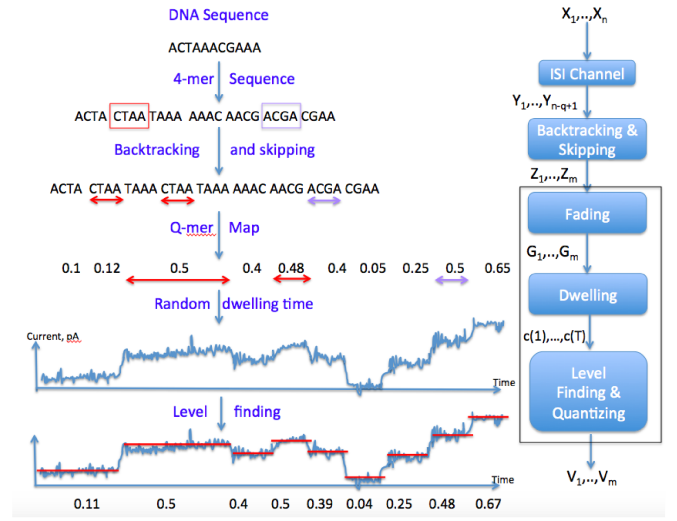


Fig. 2. Model of the Nanopore Sequencer for a toy DNA sequence.

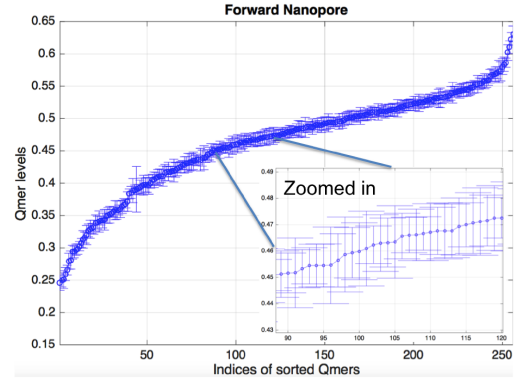


Fig. 3. Mean and standard deviation of 256 *Q*-mer maps obtained from our nanopore. Notice the significant overlap of *Q*-mer maps.

The current level G_i at step i is a function of the *Q*-mer z_i dwelling in the nanopore. However, this function is not deterministic: each time the same dwelling *Q*-mer z produces a current level with some variation, which resembles the *fading* effect in communication channels. In Fig. 3, we plot the mean value of the 256 random *Q*-mer map, together with their variances as error-bar. (v) *Noisy samples*: On top of the *Q*-mer map fading, within the same step/level, each sample of the same level G_i is subject to a random noise. Usually we model the noise as an AWGN process with some variance σ_i^2 , which may vary from level to level.

B. Channel model and Notation

We can model the nanopore experiment as a communication process, with the input being the DNA sequence to be measured and the output being the current samples produced, with the following simplifying assumptions.

- 1) Experiments show that skipping happens much more frequently than backtracking, so we model only skipping as an i.i.d. deletion process.



Fig. 4. Simplified nanopore channel.

- 2) While the nanopore model shown in Fig. 2 produces current waveforms in continuous time, in practice, a level-finding algorithm is usually applied (see [2]) to determine the change-points at which nucleotides migrate in/out of the nanopore, and output the mean current-levels as a discrete-time current-level sequence $V_1, \dots, V_m$. We further assume that the output V_i are discretized. Thus the effects of sample noise, dwelling and fading are modeled using a single discrete memoryless channel.

In this simplified channel (depicted in Fig. 4), the input (corresponding to the bases in the DNA sequence) symbol at time n is denoted by X_n , which has a finite alphabet $\mathcal{X}$. X_n is passed through an ISI channel with memory $K - 1$, to form the K -mer $Y_n = X_{n-K+1}^n = (S_{n-1}, X_n)$, where $S_{n-1} \triangleq X_{n-K+1}^{n-1}$ denotes the ISI channel state at time $n - 1$. The alphabets for the K -mers and ISI states are denoted by $\mathcal{Y} = \mathcal{X}^K$ and $\mathcal{S} = \mathcal{X}^{K-1}$, respectively. The initial state s_0 is known (they are the prefixed adaptors). The K -mers $\{Y_n\}$ are then sent to a deletion channel, where each K -mer is independently deleted with probability p_d . We can represent the deletion process by an i.i.d. Bernoulli(p_d) process $\{D_i\}_{i \geq 1}$, where $D_i = 1$ denotes a deletion at time i . When the input of the deletion channel is a length- n vector Y^n , its output is a K -mer vector of random length, denoted by $Z^{(n)}$. The corresponding alphabet is $\mathcal{Y}^{(n)} \triangleq \emptyset \cup \mathcal{Y} \cup \mathcal{Y}^2 \cup \dots \cup \mathcal{Y}^n$. Similarly when a segment of K -mers Y_m^n is input to the deletion channel, we use $Z_{(m)}^{(n)}$ to denote the output.² Finally, after deletion, a K -mer $Z_m^{(n)}$ is sent to a DMC to produce the channel output V_m (corresponding to the discretized levels), whose alphabet is denoted by $\mathcal{V}$. We also use $V_{(m)}^{(n)}$ to denote the output of DMC corresponding to the random-length K -mer vector $Z_{(m)}^{(n)}$, and denote its alphabet by $\mathcal{V}^{(n-m+1)}$.

III. MAIN RESULTS

In this paper we obtain both capacity lower and upper bounds for the nanopore channel given in Fig. 4. We outline the proofs in this paper with details in [9]. In particular, we derive achievable rates for a cascade of a deletion channel with a DMC, i.e., the channel without the first block in Fig. 4, then apply the result to the nanopore channel. The main result for the lower bound is the following theorem (see Section IV for the definition of $E(\gamma)$ and other notations and definitions):

Theorem 1: For the cascade of a deletion channel with a DMC, the following is an achievable rate for each irreducible Markov transition matrix P on the input alphabet $\mathcal{Y}$:

$$\underline{C}(P) = - \inf_{\gamma > 0} [(1 - \theta)\gamma + \theta E(\gamma)] / \ln 2 - \theta H(\mathcal{Z}|\mathcal{V}). \quad (1)$$

²The sub/superscripts in these notations have parentheses around them to denote the effect of deletion.

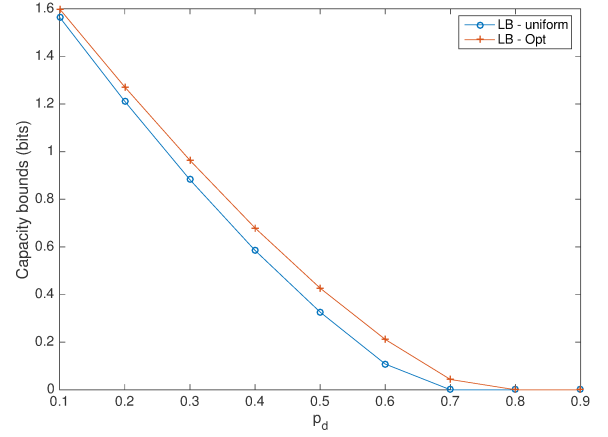


Fig. 5. Achievable rates using Q-mer map data from experiments [2].

Therefore, if $\mathcal{P}$ is a collection of irreducible Markov transition matrices, then $\sup_{P \in \mathcal{P}} \underline{C}(P)$ is also achievable.

For the achievable rates of the nanopore channel, we can apply the results above, with a constraint on the choice of Markov transition matrix P : since the current K -mer takes the form $Y_n = X_{n-K+1}^n$, the next K -mer Y_{n+1} can only be obtained from left-shifting X_{n-K+1}^n by one position and attaching a new nucleotide X_{n+1} to the right. Thus for the nanopore channel the Markov transition matrix can only be supported on such legal transitions.

The interpretation of these rates is as follows. If we use i.i.d. uniform input distribution on the nucleotides to generate the K -mer transition matrix P , then we obtain an achievable rate that reflects the sequencing capability of the nanopore sequencer. If we optimize the transition matrix, then we obtain an achievable rate that can be used to measure the capability of the nanopore sequencer as a reader for a DNA storage system. In Fig. 5, we plot the achievable rates in Theorem 1, both when the input distribution is i.i.d. uniform and when P is optimized. The channel used is constructed from *experimental data*, where the DMC is obtained using the Q-mer map data from experiments in [2] with level discretization. We can see that optimized transition matrices can achieve higher rates than the uniformly generated one.

The main capacity upper bound results are derived in Section V. Our starting point is the following upper bounds in the form of finite block mutual information.

Theorem 2: For each n , $\overline{C}_n$ defined below is a capacity upper bound for the nanopore channel:

$$\overline{C}_n = \frac{1}{n} \left[\max_{s_0} \max_{P_{X^n}} I(X^n; V^{(n)} | s_0) + \log |\mathcal{S}| \right]. \quad (2)$$

Second, since the extra term $\log |\mathcal{S}|$ above considerably affects the effectiveness of the upper bound for small to moderate n , while for large n the computation is not practical due to the exponential complexity of the first term, we seek a relaxation that has much less complexity. For that purpose we introduce periodical synchronization symbols to the channel output (as in [10]), and convert the resulting channel across synchronization

periods to a finite state channel [11] with only ISI memory. In this form the upper bounds can be relaxed (see Theorem 3 in Section V), which upper bounds mutual information by the relative entropy between the “worst-case” block conditional probability and a stationary Markov distribution on the output. The relaxation can then be computed using the Viterbi algorithm with linear complexity, and so we can compute it for very large n , which suppresses the term $\log |\mathcal{S}|$ and yields practically more effective upper bounds.

IV. ACHIEVABLE RATES

Our approach is based on and generalizes the lower bound ideas for (noisy) deletion channels in [6]. The codebook is generated randomly using a stationary ergodic Markov process.³ The decoder outputs an estimated codeword if, roughly, the output length is typical and it is the unique codeword which is jointly typical with the output. With such a coding system we derive an achievable rate by analyzing the average error probability, which utilizes techniques for watched Markov chains in [6], the general AEP [12], and ergodic theory [13].

Consider an i.i.d. deletion channel with input alphabet $\mathcal{Y}$, connected to a DMC with output alphabet $\mathcal{V}$. We assume all alphabets are finite and the deletion probability p_d satisfies $0 \leq p_d < 1$. Let a stationary Markov process $\{Y_i\}_{i \geq 1}$ on $\mathcal{Y}$ with an *irreducible* transition matrix P be input to the deletion channel, and let $\{Z_j\}_{j \geq 1}$ denote the corresponding output of the deletion channel. Let π be the steady state distribution of P . Then $\{Z_j\}$ is a watched Markov chain [6], which is also a stationary Markov process [14, Lemma 6-6]. Define $\theta = 1 - p_d > 0$, then the watched Markov chain $\{Z_j\}$ has a transition matrix

$$\bar{P} = \frac{\theta}{1 - \theta} \sum_{k=1}^{\infty} (1 - \theta)^k P^k = \theta P [I - (1 - \theta)P]^{-1},$$

which also has π as the steady state distribution. Note that as P is irreducible, $\bar{P}$ is primitive, i.e., all of its entries are positive. Hence by [13, Thm 1.19], $\{Z_j\}$ is a stationary ergodic process. The corresponding output process $\{V_j\}_{j \geq 1}$ of the DMC is a hidden Markov process, and the joint process $\{Z_j, V_j\}_{j \geq 1}$ is also stationary and ergodic (see [15]). As in [12] we use $H(\mathcal{V})$ to denote the entropy rate of the stationary ergodic process $\{V_j\}$ and define the conditional entropy rate $H(\mathcal{Z}|\mathcal{V}) = H(\mathcal{Z}, \mathcal{V}) - H(\mathcal{V})$.

Let $a, b \in \mathcal{Y}$. Define $\tilde{P}_b$ to be the submatrix of P obtained from removing the b -th row and b -th column, $\tilde{q}_b$ to be the column vector obtained from removing the b -th entry of the b -th column of P , and $\tilde{p}_{ab}$ to be the row vector obtained from removing the b -th entry of the a -th row of P . Let

$$E_{ab}(\gamma) = \ln \left[P_{ab} + \tilde{p}_{ab} (e^\gamma I - \tilde{P}_b)^{-1} \tilde{q}_b \right]$$

for $\gamma > 0$ and define $E(\gamma) = \sum_{a,b \in \mathcal{Y}} \pi_a \bar{P}_{ab} E_{ab}(\gamma)$. With all these definitions we have the achievable rate theorem (Theorem 1 in Section III), whose proof is presented in [9].

³In the nanopore channel, the ISI channel naturally provides such a Markov structure.

V. CAPACITY UPPER BOUNDS

We use methods in [16] and [11] to derive a series of capacity upper bounds in terms of finite blocks of mutual information (see also [17]), and then seek a relaxation in the formulation of [18] that leads to more effective computation.

Using techniques similar to [16], we can show that the channel capacity is upper bounded by

$$\liminf_{n \rightarrow \infty} C_n, \quad C_n \triangleq \frac{1}{n} \sup_{X^n} I(X^n; V^{(n)}). \quad (3)$$

In general, this upper bound is not computable, as the limiting behavior of C_n is unknown. However, if we can show that for each n , there is a $\bar{C}_n$ such that (i) $C_n \leq \bar{C}_n$ and (ii) $\{n\bar{C}_n\}_{n=1}^{\infty}$ is subadditive, then $\lim_{n \rightarrow \infty} \bar{C}_n$ exists and is equal to $\inf_n \bar{C}_n$. Hence $\bar{C}_n$ is an upper bound for each finite n and is computable.⁴ Using these ideas we derive the upper bound series $\{\bar{C}_n\}$ in Theorem 2 (see Section III), whose proof is presented in [9].

The upper bound (2) suffers from two issues: i) the computational complexity grows exponentially, since the optimization space is $|\mathcal{X}|^n$ -dimensional; ii) for smaller n , the extra term $\frac{1}{n} \log |\mathcal{S}|$ is relatively large and decays very slowly with n , which greatly reduces the effectiveness of the upper bounds. To address these issues, we add periodic synchronization symbols to the output (as in [10]), and use the formulation developed in [18]. Let M denote the period of synchronization and consider block length $N = nM$. Let $\bar{X}_k$ and $\bar{V}_k$ denote $X_{(k-1)M+1}^{kM}$ and $V_{((k-1)M+1)}^{(kM)}$, respectively. Then $I(X^N; V^{(N)} | s_0) \leq I(\bar{X}^n; \bar{V}^n | s_0) = I(\bar{X}^n; \bar{V}^n | \bar{s}_0)$, where $\bar{s}_n \triangleq S_{nM} = X_{nM-K+2}^{nM}$. Thus $\bar{C}_N \leq \frac{1}{M} \bar{C}'_n$, where

$$\bar{C}'_n \triangleq \frac{1}{n} \left[\max_{\bar{s}_0} \max_{P_{\bar{X}^n}} I(\bar{X}^n; \bar{V}^n | \bar{s}_0) + \log |\mathcal{S}| \right]. \quad (4)$$

Note that the channel with input $\bar{X}_n$ and output $\bar{V}_n$ is a finite state channel [11] with state $\bar{S}_n$ and only has inter-symbol interference (ISI) memory. So we can apply methods in [18] to simplify the computation of $\bar{C}'_n$:

$$\bar{C}'_n \leq \tilde{C}'_n \triangleq \frac{1}{n} \left[\max_{\bar{u}^n} D(W(\cdot | \bar{u}^n) \| R(\cdot)) + \log |\mathcal{S}| \right], \quad (5)$$

where $\bar{U}_k = (\bar{S}_{k-1}, \bar{X}_k)$ is the input branch of the finite state channel at time k , W denotes the transition probability $p(\bar{v}^n | \bar{u}^n)$ and R can be any probability distribution on sequences $\bar{v}^n$.

Theorem 3: $\bar{C}_N \leq \frac{1}{M} \tilde{C}'_n$ for all $N = nM$.

This new upper bound admits a *linear complexity algorithm* (the Viterbi algorithm), if we restrict R to be a stationary measure of an L -th order homogeneous Markov chain [18]. Furthermore, a good choice of R can be obtained from the $(L+1)$ -dimensional output distribution of the finite state channel $p(\bar{v}_n \bar{s}_n | \bar{x}_n \bar{s}_{n-1})$ when the input distribution is optimized with respect to the achievable rate.

⁴These upper bounding techniques dates back to Gallager [11] for finite state channels. For discrete memoryless synchronization channels Dobrushin [19] also showed (as explicitly pointed out in [7]) subadditivity of a series of capacity upper bounds, however, these channels are memoryless and have no ISI (or any channels states).

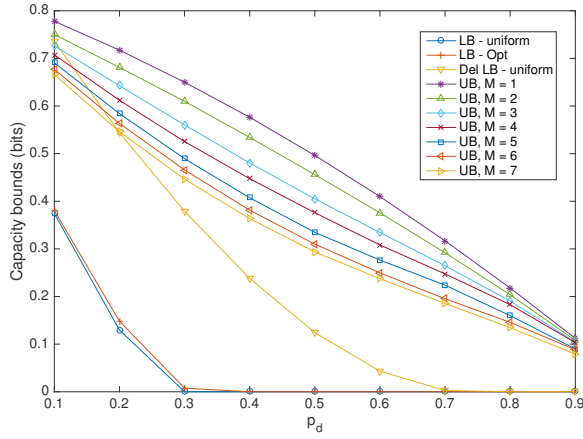


Fig. 6. $K = 2$, $|\mathcal{X}| = 2$, symmetric DMC.

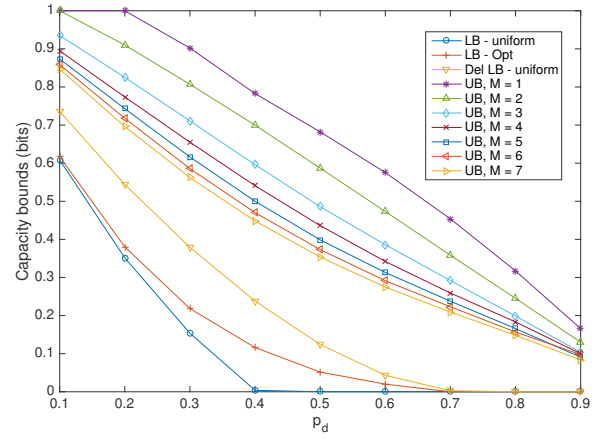


Fig. 7. $K = 2$, $|\mathcal{X}| = 2$, non-symmetric DMC.

VI. NUMERICAL RESULTS

We use two examples to illustrate the computation of the capacity bounds for the nanopore channel. The first example has ISI memory length $K = 2$, input alphabet size $|\mathcal{X}| = 2$, and uses a symmetric DMC: in this case $\mathcal{V} = \mathcal{Y}$, $p(y|y) = c$ for all y and $p(y'|y) = c'$ for all $y' \neq y$, for some constants c and c' . The bounds are plotted in Fig. 6 for different deletion probabilities. The second example has the same ISI memory length and input alphabet size, but the DMC is not symmetric: it is constructed similarly to the example in Section III, where a part of the Q-mer map data in Fig. 3 is extracted to simulate the real world situation. The corresponding bounds are plotted in Fig. 7.

In both examples, we plot the achievable rates in Theorem 1 as the capacity lower bounds, both when the input distribution is i.i.d. uniform and when P is optimized. We also plot the achievable rate of pure deletion channel when the input distribution is i.i.d. uniform, to illustrate the rate loss due to the signal degradation caused the DMC. From Fig. 6 we can see that for the channel with a symmetric DMC, i.i.d. uniformly generated codewords already have a performance very close to the optimal coding scheme in Section IV. From Fig. 7 (and also 5 in Section III), we can see that when the DMC is not symmetric, non-uniformly generated codewords can achieve higher rates than the uniform ones. But when the deletion probability p_d is small, the uniform case is still close to optimal.

We also plot the capacity upper bounds in Theorem 3 with different parameters for these two examples. From the computation results we found that either increasing the synchronization period M or the output Markov order L yields a tighter upper bound, but when M is not too small the bounds become very close to each other for different L . Hence in Fig. 6 and 7 we only plot the upper bounds for different M (with $L = 1$). We note that further computational optimization is needed to calculate the upper bounds for Fig. 5.

REFERENCES

- [1] D. Deamer, M. Akeson, and D. Branton, "Three decades of nanopore sequencing," *Nature biotechnology*, vol. 34, no. 5, pp. 518–524, 2016.
- [2] A. H. Laszlo et al, "Decoding long nanopore sequencing reads of natural dna," *Nature biotechnology*, vol. 32, no. 8, pp. 829–833, 2014.
- [3] I. Sović, M. Šikić, A. Wilm, S. N. Fenlon, S. Chen, and N. Nagarajan, "Fast and sensitive mapping of nanopore sequencing reads with graphmap," *Nature communications*, vol. 7, 2016.
- [4] R. G. Gallager, "Sequential decoding for binary channels with noise and synchronization errors," DTIC Document, Tech. Rep., 1961.
- [5] R. L. Dobrushin, "Shannon's theorems for channels with synchronization errors," *Problemy Peredachi Informatsii*, vol. 3, no. 4, pp. 18–36, 1967.
- [6] S. N. Diggavi and M. Grossglauser, "On information transmission over a finite buffer channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1226–1237, Mar. 2006.
- [7] H. Mercier, V. Tarokh, and F. Labeau, "Bounds on the capacity of discrete memoryless channels corrupted by synchronization and substitution errors," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4306–4330, Jul. 2012.
- [8] A. Magner, J. Duda, W. Szpankowski, and A. Grama, "Fundamental bounds for sequence reconstruction from nanopore sequencers," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 2, no. 1, pp. 92–106, Jun. 2016.
- [9] W. Mao, S. Diggavi, and S. Kannan, "Models and information-theoretic bounds for nanopore sequencing," *available on arXiv.org*, 2017.
- [10] D. Fertonani and T. M. Duman, "Novel bounds on the capacity of binary channels with deletions and substitutions," in *2009 IEEE International Symposium on Information Theory*, Jun. 2009.
- [11] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, 1968.
- [12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley-Interscience, 2006.
- [13] P. Walters, *An Introduction to Ergodic Theory*, ser. Graduate texts in mathematics. New York: Springer-Verlag, 1982, vol. 79.
- [14] J. G. Kemeny, J. L. Snell, and A. W. Knapp, *Denumerable Markov Chains*, 2nd ed. New York: Springer-Verlag, 1976.
- [15] R. M. Gray, *Entropy and Information Theory*. New York: Springer-Verlag, 1990.
- [16] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.
- [17] W. Mao and B. Hassibi, "Capacity bounds for certain channels with states and the energy harvesting channel," in *Proc. of the 2014 Information Theory Workshop*, Hobart, Australia, Nov. 2014.
- [18] P. O. Vontobel and D. M. Arnold, "An upper bound on the capacity of channels with memory and constraint input," in *Proc. of the 2001 Information Theory Workshop*, Cairns, Australia, Sep. 2001.
- [19] R. L. Dobrushin, "Shannon's theorems for channels with synchronization errors shannon's theorems for channels with synchronization errors," *Problems Inform. Transmission*, vol. 3, no. 4, pp. 11–26, 1967.