Guessing noise, not code-words

Guessing noise, not code-words

Ken R. Duffy*, Jiange Li† and Muriel Médard†
*Hamilton Institute, Maynooth University, Ireland.
E-mail: ken.duffy@mu.ie.
†Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, U. S. A.
E-mail: lijiange@mit.edu, medard@mit.edu.

***Abstract*—We introduce a new algorithm for Maximum Likelihood (ML) decoding for channels with memory. The algorithm is based on the principle that the receiver rank orders noise sequences from most likely to least likely. Subtracting noise from the received signal in that order, the first instance that results in an element of the code-book is the ML decoding. In contrast to traditional approaches, this novel scheme has the desirable property that it becomes more efficient as the code-book rate increases. We establish that the algorithm is capacity achieving for randomly selected code-books. When the code-book rate is less than capacity, we identify asymptotic error exponents as the block length becomes large. When the code-book rate is beyond capacity, we identify asymptotic success exponents. We determine properties of the complexity of the scheme in terms of the number of computations the receiver must perform per block symbol. Worked examples are presented for binary memoryless and Markovian noise. These demonstrate that block-lengths that offer a good complexity–rate tradeoff are typically smaller than the reciprocal of the bit error rate.**

***Keywords*—ML Decoding; Noise Guessing; Complexity Analysis; Error and Success Exponents.**

I. Introduction

Consider a channel with inputs, X^n, and outputs, Y^n, consisting of blocks of n symbols from a finite alphabet $\mathbb{A}$ of size $|\mathbb{A}|$. Assume that channel input is altered by random, not necessarily memoryless, noise, N^n, that is independent of the channel input and also takes values in $\mathbb{A}^n$. Assume that the function, $\oplus$, describing the channel's action $Y^n = X^n \oplus N^n$, is invertible, so that knowing the output and input the noise can be recovered: $X^n = Y^n \ominus N^n$. To implement Maximum-Likelihood (ML) decoding, the sender and receiver first share a code-book $\mathscr{C}_n = \{c^{n,1}, \ldots, c^{n,M_n}\}$ consisting of M_n elements of $\mathbb{A}^n$. For a given channel output y^n, denote the conditional probability of the received sequence for each code-word in the code-book by

$$p(y^n|c^{n,i}) = P(y^n = c^{n,i} \oplus N^n) \text{ for } i \in \{1, \ldots, M_n\}. \quad (1)$$

The ML decoding is then an element of the code-book that has the highest conditional likelihood of transmission given what was received,

$$\begin{aligned} c^{n,*} &\in \arg\max \left\{ p(y^n|c^{n,i}) : c^{n,i} \in \mathscr{C}_n \right\} \\ &= \arg\max \left\{ P(N^n = y^n \ominus c^{n,i}) : c^{n,i} \in \mathscr{C}_n \right\}, \end{aligned} \quad (2)$$

where we have used the invertibility of $\oplus$ in the final equation.

To realize this ML decoding, it would appear that the receiver would have to perform the M_n computations described in equation (1) every time a signal is received. Code-book sizes are typically exponential in the block length n, $M_n \sim |\mathbb{A}|^{Rn}$, and, taking logs throughout the article as base $|\mathbb{A}|$, we define the normalized rate of the code-book to be $R = \lim_n 1/n \log(M_n)$. As R increases, the code-book becomes denser in $\mathbb{A}^n$, so the number of computations required in (1) increases exponentially, potentially making this approach infeasible in practice.

In the present paper we propose an entirely distinct algorithm for ML decoding. The principle underlying the approach is for the receiver to rank-order noise sequences from most likely to least likely and then sequentially query whether the sequence that remains when the noise is removed from the received signal is an element of the code-book. For the channel structure described above, the first instance where the answer is in the affirmative corresponds to the ML decoding. More formally, the receiver first creates an ordered list of noise sequences, $G : \mathbb{A}^n \mapsto \{1, \ldots, |\mathbb{A}|^n\}$, from most likely to least likely, breaking ties arbitrarily (throughout lower case letters correspond to realizations of upper-case random variables, apart from for noise where z is used):

$$G(z^{n,i}) \leq G(z^{n,j}) \text{ iff } P(N^n = z^{n,i}) \geq P(N^n = z^{n,j}). \quad (3)$$

For each received signal, the receiver executes the following algorithm:

- Given channel output y^n, initialize $i = 1$ and set z^n to be the most likely noise sequence, i.e. the z^n such that $G(z^n) = i$.
- While $x^n = y^n \ominus z^n \notin \mathscr{C}_n$, increase i by 1 and set z^n to be the next most likely noise sequence, i.e. the z^n such that $G(z^n) = i$.
- The x^n that results from this while loop is the decoded element.

To see that this algorithm corresponds to ML decoding, note that owing to the definition of $c^{n,*}$ in equation (2),

$$P(N^n = y^n \ominus c^{n,*}) \geq P(N^n = y^n \ominus c^{n,i}) \text{ for all } c^{n,i} \in \mathscr{C}_n.$$

Thus the scheme identifies an ML decoding. Our fundamental premise is that this new scheme is practically feasible, even though the more direct approach described in equation (2) is not.

To determine the asymptotic properties of the core scheme for random code-books as the block length becomes large, we recall one theorem from the literature and establish several new ones. As they may appear somewhat mathematically involved, we begin by explaining the intuitive meaning behind them.

Theorem 1 is taken from [1] and provides a Large Deviation Principle (LDP) as the block length, n, becomes large, for the distribution of the logarithm of the number of guesses

needed until the actual noise in the channel is queried, $G(N^n)$. For uniform-at-random code-books, Theorem 2 is new and gives a LDP for the distribution of the logarithm of the number of guesses that would be made until an element of the code-book that was not the channel input is identified. Here we leverage the fact that for uniformly distributed code-books the location of each of these elements in the guessing order is uniform in $\{1,\ldots,|\mathbb{A}|^n\}$. As a result, the distribution of the number of guesses until any non-input element of the code-book is hit upon is distributed essentially the same as the minimum of M_n such uniform random variables. In the asymptotic analysis presented here where $M_n \approx |\mathbb{A}|^{nR}$ and n becomes large, this number is essentially uniformly distributed in $\{1,\ldots,|\mathbb{A}|^{n(1-R)}\}$ so that the receiver will identify a code-word in no more than approximately $|\mathbb{A}|^{n(1-R)}$ guesses. Note, in particular, that as R increases and the code-book becomes more dense and efficient, while the number of computations in traditional approaches to ML decoding increases, the noise guessing approach experiences the reverse phenomenon.

The ML decoder introduced in the present paper is essentially a race between these two guessing processes. If the number of guesses required to identify the true noise is less than the number of guesses to identify any other element of the code-book, then the ML decoder provides the correct answer on termination. Combining the two earlier theorems in two different ways first recovers the Channel Coding Theorem as Proposition 1 via this new guessing argument. Combined in a distinct fashion, Proposition 2 characterizes the asymptotic complexity, in terms of the distribution of the number of guesses to termination, of the scheme.

II. Related work

Large deviation style arguments employed to establish error exponents in both source and channel coding are typically variants of Sanov's Theorem and the method of types. For error exponents in source coding, these methods have been used extensively, originally for asymptotically error-free source coding with IID and Markov sources [2], [3], [4], and then for variable-length and lossy source coding of IID and stationary sources [5]. For channel coding of Discrete Memoryless Channels (DMCs), error exponents were first identified in [6]. More recently, an approach along these lines has been used to study joint source-channel coding [7].

We analyse our proposed approach starting from a completely distinct angle: the recently proved LDP [1] for Massey's guesswork [8]. That LDP is established based on earlier results [9], [10], [11] that identify scaling exponents for moments of guesswork in terms of Rényi entropy rates. The connection between source coding and guesswork was first noted in [12], and has since been established [13]. For channel coding, a connection between guesswork and error exponent analysis was proved by Arikan for sequential decoding of tree codes [9]. A general framework for designing codes that increase the cutoff rate is discussed in [14]. Polar coding, which is capacity achieving, fits into that framework.

III. Analysis

We begin with the assumption we shall make on the noise process. Recall that log is taken base $|\mathbb{A}|$ throughout.

Assumption 1. *Assuming it exists, define the Rényi entropy rate of the noise $\{N^n\}$ process with parameter $\alpha \in (0,1)\cup(1,\infty)$ to be*

$$H_\alpha = \lim_{n\to\infty} \frac{1}{1-\alpha} \log\left(\sum_{z^n\in\mathbb{A}^n} P(N^n = z^n)^\alpha\right),$$

with $H = H_1$ being the Shannon entropy rate of the noise. Denote the min-entropy rate of the noise by $H_{\min} = \lim_{\alpha\to\infty} H_\alpha$. Assume that

$$\Lambda^N(\alpha) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}(G(N^n)^\alpha) = \begin{cases} \alpha H_{\frac{1}{1+\alpha}} & \text{for } \alpha \in (-1,\infty) \\ -H_{\min} & \text{for } \alpha \le -1, \end{cases} \tag{4}$$

and that the derivative of $\Lambda^N(\alpha)$ is continuous on the range $\alpha \in (-1,\infty)$.

Assumption 1 is known to be satisfied for a broad range of sources including i.i.d. [9], Markovian [10], a large class of general, stationary processes [11] and others [15]; the condition for $\alpha \le -1$ is established for all of these processes in [1]. Note that by setting $\alpha = 1$, from equation (4) one has that the average number of guesses required to identify the true noise grows exponentially in block size, n, with Rényi entropy rate at parameter $1/2$, $H_{1/2}$, which is no smaller than the Shannon entropy rate, H, of the noise.

From equation (4), Λ^N can be identified as the scaled cumulant generating function for the process $\{1/n \log G(N^n)\}$ [16] and so Λ^N is necessarily convex. Moreover, that identification suggested that this process may satisfy a Large Deviation Principle (LDP) [12], [17], which was proven in [1].

Theorem 1 (LDP for Guessing the Noise [1])**.** *Under assumption 1, $\{1/n \log G(N^n)\}$ satisfies the Large Deviation Principle with the convex lower-semi continuous rate function, $I^N : [0,1] \to [0,\infty]$,*

$$I^N(x) := \sup_{\alpha\in\mathbb{R}} \{x\alpha - \Lambda^N(\alpha)\}, \tag{5}$$

which is the Legendre-Fenchel transform of Λ^N. In particular: $I^N(0) = H_{\min}$, the min-entropy rate of the noise; $I^N(x)$ is linear on $[0,\gamma]$, where $\gamma := \lim_{\alpha\downarrow -1} d/d\alpha\, \Lambda^N(\alpha)$, and then strictly convex thereafter while finite; and $I^N(x) = 0$ if and only if $x = H$, the Shannon entropy rate of the noise.

The second theorem provides a LDP for the number of guesses on the noise that will be made until identifying an element of the code-book that is not the transmitted code-word. The key realization is that if elements of the code-book have been uniformly selected, the location of the non-transmitted code-book elements in the ordered list of noise guesses is also uniform.

Theorem 2 (LDP for Guessing a Non-transmitted Code-word)**.** *Assume that $\lim_{n\to\infty} n^{-1}\log M_n = R$ for some $R > 0$, and that $U^{n,1},\ldots,U^{n,M_n}$ are independent random variables, each uniformly distributed in $\{1,\ldots,|\mathbb{A}|^n\}$. Define $U^n = \min_i U^{n,i}$. Then $\{1/n \log U^n\}$ satisfies the large deviation principle with lower semi-continuous rate function*

$$I^U(x) = \begin{cases} 1-R-x & \text{if } x \in [0,1-R] \\ +\infty & \text{otherwise.} \end{cases} \tag{6}$$

Proof: It is sufficient [16] to prove that for all $x \in [0,1]$

$$\lim_{\varepsilon\downarrow 0}\liminf_{n\to\infty}\frac{1}{n}\log P\left(\frac{1}{n}\log U^n \in (x-\varepsilon, x+\varepsilon)\right)$$
$$= \lim_{\varepsilon\downarrow 0}\limsup_{n\to\infty}\frac{1}{n}\log P\left(\frac{1}{n}\log U^n \in (x-\varepsilon, x+\varepsilon)\right) = -I^U(x).$$

This begins by observing that

$$P(U^n > |\mathbb{A}|^{xn}) = \prod_{i=1}^{M_n} P\left(U^{n,i} > |\mathbb{A}|^{xn}\right) = \left(1 - \frac{\lceil |\mathbb{A}|^{xn}\rceil}{|\mathbb{A}|^n}\right)^{M_n}.$$

Thus we have the following limiting equality for the complementary cumulative distribution function

$$\lim_{n\to\infty}\frac{1}{n}\log P\left(\frac{1}{n}\log U^n > x\right) = \begin{cases} 0 & \text{if } x \in [0, 1-R] \\ -\infty & \text{if } x \in (1-R, 1), \end{cases}$$

confirming equality for $x \in (1-R, 1]$. The corresponding equality for the cumulative distribution function can be obtained by first noting that by the Binomial theorem

$$\lim_{n\to\infty}\frac{\left(1 - |\mathbb{A}|^{n(x-1)}\right)^{|\mathbb{A}|^{nR}}}{1 - |\mathbb{A}|^{n(R+x-1)}} = 1 \text{ if } x \in [0, 1-R),$$

while if $x = 1-R$ the limit of the numerator in the above equation is $1/e$. Using this, for $x \in [0, 1-R]$ one obtains

$$\lim_{\varepsilon\downarrow 0}\lim_{n\to\infty}\frac{1}{n}\log P\left(\frac{1}{n}\log U^n \in (x-\varepsilon, x+\varepsilon)\right)$$
$$= \lim_{\varepsilon\downarrow 0}\lim_{n\to\infty}\frac{1}{n}\log\left(\left(1 - \left(1 - \frac{\lceil |\mathbb{A}|^{(x+\varepsilon)n}\rceil}{|\mathbb{A}|^n}\right)^{M_n}\right) - \left(1 - \left(1 - \frac{\lceil |\mathbb{A}|^{(x-\varepsilon)n}\rceil}{|\mathbb{A}|^n}\right)^{M_n}\right)\right)$$
$$= \lim_{\varepsilon\downarrow 0}\lim_{n\to\infty}\frac{1}{n}\log\left(|\mathbb{A}|^{n(\min(R+x+\varepsilon-1, 0))} - |\mathbb{A}|^{n(R+x-\varepsilon-1)}\right)$$
$$= -(1-R-x).$$

■

Theorem 2 says that on the scale of large deviations, for large n the first non-transmitted code-word that will be encountered in the noise guesswork list is approximately uniformly distributed in $\{1, \ldots, |\mathbb{A}|^{n(1-R)}\}$. Moreover, on the scale of large deviations, one will never make more than order $|\mathbb{A}|^{n(1-R)}$ guesses before encountering some element of the code-book.

Combining Theorems 1 and 2 enables us to provide a guessing based proof of Channel Coding Theorem, where capacity is upper bounded by $1-H$. The proposition that follows establishes this is achieved for all noise processes satisfying Assumption 1 through the use of a uniformly-at-random code-book and ML decoding.

Proposition 1 (Channel Coding Theorem)**.** *Under the assumptions of Theorems 1 and 2 with I^U is defined in equation* (6) *and I^N in equation* (5)*, we have the following. If the code-book rate is less than the capacity, $R < 1-H$, then*

$$\lim_{n\to\infty}\frac{1}{n}\log P(U^n \le G(N^n)) = -\inf_{a\in[H,1-R]}\{I^U(a) + I^N(a)\} < 0,$$

so that probability that the ML decoding is not the transmitted code-word decays exponentially in the block length n. If, in addition, x^* exists such that $d/dx\ I^N(x)|_{x=x^*} = 1$, then, recalling $H_{1/2}$ is the Rényi entropy rate of the noise with parameter $1/2$, the error rate simplifies further to

$$\varepsilon(R) = -\lim_{n\to\infty}\frac{1}{n}\log P(U^n \le G(N^n)) = \begin{cases} 1-R-H_{1/2} & \text{if } R \in (0, 1-x^*) \\ I^N(1-R) & \text{if } R \in [1-x^*, 1-H). \end{cases} \quad (7)$$

If, instead, the code-book rate is greater than capacity,

$$s(R) = -\lim_{n\to\infty}\frac{1}{n}\log P(U^n \ge G(N^n)) = I^N(1-R), \quad (8)$$

which we call the success rate, is strictly positive.

Proof: As $\{1/n\log G(N^n)\}$ and $\{1/n\log U^n\}$ are independent processes, $\{(1/n\log G(N^n), 1/n\log U^n)\}$ satisfies the LDP with rate function $I^N(x) + I^U(y)$. The LDP for $\{1/n\log U^n/G(N^n)\}$ then follows from an application of contraction principle, [16][Theorem 4.2.1], with the continuous function $f(x,y) = x-y$, giving $I^{U/N}(x) = \inf_{a,b}\{I^N(a)+I^U(b) : f(a,b) = a-b = x\} = \inf_{a\in[0,1-R]}\{I^U(a)+I^N(a-x)\}$.

Noting the following equality $P(U^n \le G(N^n)) = P(1/n\log(U^n/G(N^n)) \le 0)$, we can use the LDP for $\{1/n\log U^n/G(N^n)\}$ to determine asymptotics for the likelihood that fewer queries are necessary to determine a non-transmitted element of the code-book than the truly transmitted element. From the LDP lower and upper bounds,

$$-\inf_{x<0} I^{U/N}(x) = \liminf_{n\to\infty}\frac{1}{n}\log P\left(\frac{1}{n}\log\frac{U^n}{G(N^n)} \le 0\right)$$
$$= \limsup_{n\to\infty}\frac{1}{n}\log P\left(\frac{1}{n}\log\frac{U^n}{G(N^n)} \le 0\right)$$
$$\le -\inf_{x\le 0} I^{U/N}(x).$$

For the limit to exist, we require that $\inf_{x<0} I^{U/N}(x) = \inf_{x\le 0} I^{U/N}(x)$. Consider $I^{U/N}(0) = \inf_{a\in[0,1-R]}\{I^U(a) + I^N(a)\} = I^U(a^*) + I^N(a^*) < \infty$, where a^* necessarily exists as I^U and I^N are lower-semicontinuous. As we have assumed $H > 0$, $a^* > 0$ and $I^U(a^*) + I^N(a^*)$ is then arbitrarily well approximated by $I^U(a^*) + I^N(a^* - \varepsilon)$ as I^N is continuous where it is finite, so the above limit exits. The following simplification is achieved by changing the order the infima are taken in:

$$\lim_{n\to\infty}\frac{1}{n}\log P(U^n \le G(N^n)) = -\inf_{x\le 0} I^{U/N}(x) = -\inf_{a\in[0,1-R]}\{I^U(a) + \inf_{y\ge a} I^N(y)\}. \quad (9)$$

Starting from $P(U^n \ge G(N^n)) = P(1/n\log(U^n/G(N^n)) \ge 0)$, similar logic, but with an additional simplification due to the

form of I^U found equation (6), leads us to

$$\lim_{n\to\infty}\frac{1}{n}\log P(U^n \geq G(N^n)) = -\inf_{x\in[0,1-R]} I^N(x). \qquad (10)$$

For the within-capacity result, if $R < 1-H$, then $H < 1-R$. Considering the right hand side of equation (9) as both I^U and I^N are decreasing on $[0,H]$ and I^N is either infinite or increasing on $[H,1-R]$, $\inf_{a\in[0,1-R]}\{I^U(a)+\inf_{y\geq a} I^N(y)\} = \inf_{a\in[H,1-R]}\{I^U(a)+I^N(a)\}$. This quantity is strictly positive as I^U is strictly decreasing to zero on $[H,1-R]$ while I^N is strictly increasing from zero on the same range. To get the additional simplification to equation (7), note that as I^N is strictly convex to the right of H, I^U is decreasing at rate 1 and x^* is defined to be the value at which I^N is increasing with rate 1, then $\inf_{a\in[H,1-R]}\{I^U(a)+I^N(a)\}$ is either $I^N(1-R)$ if $x^* > 1-R$ or $I^U(x^*)+I^N(x^*)$. Now $I^N(x^*) = x^* - H_{1/2}$, so that $I^U(x^*)+I^N(x^*) = 1-R-x^*+x^*-H_{1/2}$ and the result follows. On the other hand $\inf_{x\in[0,1-R]} I^N(x) = I^N(H) = 0$, and so the right hand side of equation (10) is zero.

For the beyond-capacity result if, alternatively, $R > 1-H$, then $H > 1-R$ and $\inf_{a\in[0,1-R]}\{I^U(a)+\inf_{y\geq a} I^N(y)\} = I^U(1-R)+I^N(H) = 0$, and so the right hand side of equation (9) is zero. While $\inf_{x\in[0,1-R]} I^N(x) = I^N(1-R) > 0$, so that the right hand side of (10) is strictly greater than zero. ■

For memoryless channels, the error rate in equation (7) coincides with that identified by Gallager's [6][Theorem 2]. Proposition 1 establishes that phenomenon for more general noise processes, and also provides success exponents for when the rate is beyond capacity.

The algorithm terminates after $D^n := \min(G(N^n), U^n)$ guesses; i.e. at either determination of the noise or when a non-transmitted element of the code-book is unintentionally identified, whichever occurs first. Combining Theorems 1 and 2 in a distinct way determines the asymptotic complexity of the new decoding scheme for random code-books.

Proposition 2 (Guessing Complexity of ML Decoding)**.** *Under the assumptions of Theorems 1 and 2, if the code-book rate is below capacity, $R < 1-H$, $\{1/n\log D^n\}$ satisfies a LDP with a lower-semicontinuous rate function*

$$I^{ML}(x) = \begin{cases} I^N(x) & \text{if } x \in [0,1-R] \\ +\infty & \text{if } x > 1-R, \end{cases} \qquad (11)$$

where the input code-word will be recovered in the large deviations limit with unaffected likelihoods, and the impact of the code-book is to curtail guessing of unlikely inputs. The average number of guesses until an ML decoding is found satisfies $\lim_{n\to\infty} 1/n\log E(D^n) = \min\left(H_{1/2}, 1-R\right)$.

Proof: As $\{1/n\log G(N^n)\}$ and $\{1/n\log U^n\}$ are independent processes, $\{(1/n\log G(N^n), 1/n\log U^n)\}$ satisfies the LDP with rate function $I^N(x)+I^U(y)$. The LDP for $\{1/n\log D^n = 1/n\log\min(G(N^n),U^n)\}$ follows from an application of contraction principle, [16][Theorem 4.2.1], with the continuous function $f(x,y) = \min(x,y)$, giving

$$\begin{aligned} I^{\mathrm{ML}}(x) &= \inf_{a,b}\left\{I^N(a)+I^U(b) : f(a,b) = \min(a,b) = x\right\} \\ &= \min\left\{I^N(x), \inf_{y\geq x} I^N(y)+I^U(x)\right\}. \end{aligned} \qquad (12)$$

where the last line follows from the form of I^U in equation (6). The simplification of equation (12) into (11) comes about about due to considerations from the following structure. By Theorem 1, the noise guessing rate function starts at the min-entropy rate, $I^N(0) = H_{\min}$. As the min-entropy rate is always less than or equal to the Shannon rate, $H_{\min} \leq H$, $I^N(H) = 0$ and I^N is convex, I^N cannot lie above line from $(0, H_{\min})$ to $(H,0)$. If $R < 1-H$, then $H < 1-R$ and $I^N(x) \leq I^U(x)$ for all $x \leq H$ from the definition of I^U in equation (6). For $H \leq x \leq 1-R$, I^N is non-decreasing and so $\min(I^N(x), \inf_{y\geq x} I^N(y) + I^U(x)) = I^N(x)$.

To obtain the scaling result for $E(D^n)$, one reverses the transformation from the rate function I^{ML} to its Legendre-Fenchel transform, the scaled cumulant generating function of the process $\{n^{-1}\log D^n\}$ and evaluates it at argument 1. ■

One interpretation of the first part of that proposition is that if the code-book is such that $R < 1-H$, and so within capacity, identification of the correct code-word occurs because it is likely that all elements in the typical set of the noise will be queried before a non-transmitted element of the code-book is identified. Owing to the long tail of guesswork, in the absence of the other elements of the code-book stopping the guessing algorithm, the average number of guesses that would be made would grow with rate $H_{1/2}$ [9]. If, however, one minus the normalized code-book rate R is less than that, the long tail of the scheme is clipped. While this clipping is not enough to make an error likely, it is enough to reduce the average number of queries that will be made before an element of the code-book is identified.

IV. Examples and Discussion

We consider binary noise sequences, $\mathbb{A} = \{0,1\}$, with and without memory. For complexity of ML decoding by noise guessing, we use the average number of guesses per symbol to a decoding identified in Proposition 2, $\approx 2^{n\min(1-R, H_{1/2})}/n$. For comparison, we define the complexity of (2) as 2^{nR}/n, the number of conditional probabilities that must be computed per bit before rank ordering and determining the most likely code-book element. Here we are equating the work in one noise guess with one computation of a conditional probability. As both of these schemes result in the same ML decoding, they share the same error and success probabilities. By Proposition 1, the error probability is approximated by $\approx 2^{-n\varepsilon(R)}$ for $R < 1-H$, and the success probability by $\approx 2^{-ns(R)}$ for $R > 1-H$.

Consider binary noise sequences, $\{N^n\}$ whose elements are chosen via a process a Markov chain with transition matrix

$$\Pi = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}.$$

Assuming that $a, b > 0$, the second eigenvalue of this matrix is $1-a-b$, which captures the burstiness of the Markov chain. The Rényi entropy rate of the noise source can be evaluated explicitly [10], so that the scaled cumulant generating function given by equation (4) can be determined. While there is an explicit expression for Λ^N, the rate function I^N cannot be calculated in closed form, but is readily evaluated numerically. If $1-a = b$, this Markovian noise reduces to the BSC.

The left two panels of Fig. 1 consider BSCs with bit error probabilities of $p = 10^{-2}$ and 10^{-3}, respectively, while the

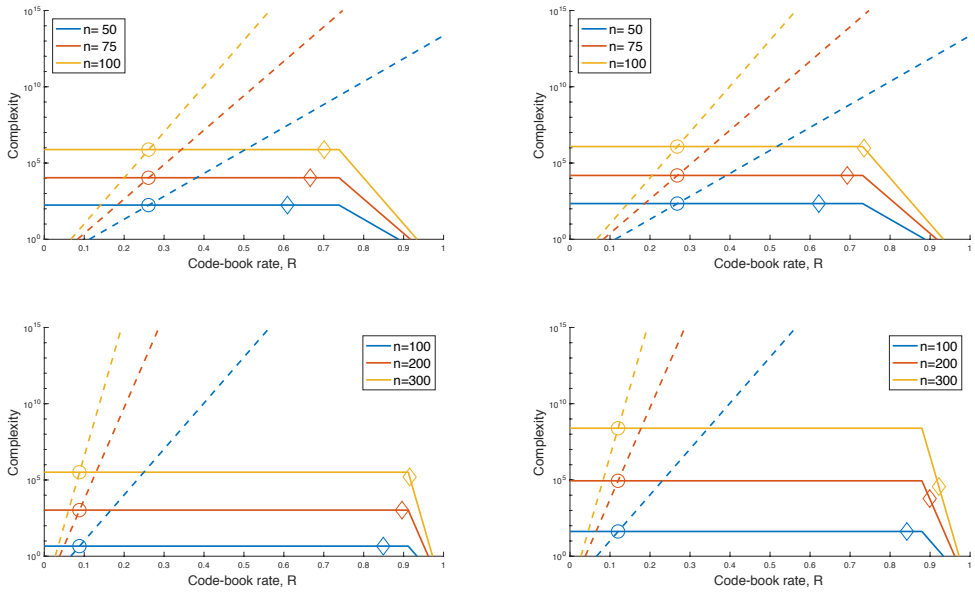

Fig. 1. Complexity of ML decoding by guessing noise (solid lines) vs. computing conditional probabilities (dashed lines). Circles indicate the rate beyond which computing within the code-book has higher complexity than noise guessing. Diamonds indicate the rate below which block error probability is less than 10^{-2}. Upper Left: BSC, $p = 10^{-2}$. Lower Left BSC, $p = 10^{-3}$. Upper Right: Markovian $p = 10^{-2}$ obtained by $a = 10^{-2}/5$ and $b = a(1-p)/p \approx 0.198$. Lower Right: Markovian $p = 10^{-3}$ obtained by setting $a = 10^{-3}/5$ and $b = a(1-p)/p \approx 0.1998$.

right two consider highly bursty Markov channels with the same long run average bit error probabilities. The computational complexity of computing conditional probabilities for the code-book leads it to become infeasible for even modest rates. The complexity of guessing the noise only decreases as rates increase, with the circles indicating the threshold beyond which the complexity of guessing within the code-book exceeds the complexity of guessing from outside it. Several of the block lengths shown would be feasible in practice as the number of computations per bit per second is normally several orders of magnitude greater than the number of bits received over the channel per second. Moreover, guessing is readily parallelizable, offering further efficiencies. Note that here the scheme has been directly applied to Markovian noise with no need for interleaving to reduce correlations.

Rates to the left of the diamonds can be achieved with a block error probability of less than 10^{-2}. These limits are impressive and illustrate the inherent promise of ML decoding, which motivated the introduction of our new approach to unlock that potential. Observe that block lengths are no larger than the reciprocal of the corresponding bit error rate, $1/p$, in these examples. This is an unusual and desirable feature of the scheme that we have consistently observed across scenarios not presented here.

REFERENCES

[1] M. M. Christiansen and K. R. Duffy, "Guesswork, large deviations and Shannon entropy," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 796–802, 2013.

[2] V. Anantharam, "A large deviations approach to error exponents in source coding and hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 36, no. 4, pp. 938–943, Jul 1990.

[3] I. Csiszár and J. Körner, *Information Theory; Coding Theorems for Discrete Memoryless Sources*. New York: Academic, 1981.

[4] L. Davisson, G. Longo, and A. Sgarro, "The error exponent for the noiseless encoding of finite ergodic markov sources," *IEEE Trans. Inf. Theory*, vol. 27, no. 4, pp. 431–438, Jul 1981.

[5] A. Dembo and I. Kontoyiannis, "Source coding, large deviations, and approximate pattern matching," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1590–1615, 2002.

[6] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. 11, no. 1, pp. 3–18, 1965.

[7] R. Yaguchi, V. Y. F. Tan, S. Watanabe, and M. Hayashi, "Large and moderate deviations for joint source-channel coding of systems with markovian memory," *available at https://www.ece.nus.edu.sg/stfpage/vtan/YTWH17.pdf, submitted to IEEE Transactions on Information Theory*, 2017.

[8] J. L. Massey, "Guessing and entropy," *IEEE Int. Symp. Inf Theory*, pp. 204–204, 1994.

[9] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 99–105, 1996.

[10] D. Malone and W. G. Sullivan, "Guesswork and entropy," *IEEE Trans. Inf. Theory*, vol. 50, no. 4, pp. 525–526, 2004.

[11] C.-E. Pfister and W. Sullivan, "Rényi entropy, guesswork moments and large deviations," *IEEE Trans. Inf. Theory*, no. 11, pp. 2794–00, 2004.

[12] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inf. Theory*, vol. 44, pp. 1041–1056, 1998.

[13] M. K. Hanawal and R. Sundaresan, "Guessing and compression subject to distortion," Division of Electrical Sciences, Indian Institute of Science, Bangalore, Tech. Rep., 2010.

[14] E. Arıkan, "On the origin of polar coding," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 2, pp. 209–223, 2016.

[15] M. K. Hanawal and R. Sundaresan, "Guessing revisited: A large deviations approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 70–78, 2011.

[16] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer-Verlag, 1998.

[17] R. Sundaresan, "Guessing based on length functions," in *Proc. 2007 International Symp. on Inf. Th.*, 2007.