Guessing random additive noise decoding with soft detection symbol reliability information - SGRAND

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

***Abstract*—We recently introduced a noise-centric algorithm, Guessing Random Additive Noise Decoding (GRAND), that identifies a Maximum Likelihood (ML) decoding for arbitrary code-books. GRAND has the unusual property that its complexity decreases as code-book rate increases. Here we provide an extension to GRAND, soft-GRAND (SGRAND), that incorporates soft detection symbol reliability information and identifies a ML decoding in that context. In particular, we assume symbols received from the channel are declared to be error free or to have been potentially subject to additive noise. SGRAND inherits desirable properties of GRAND, including being capacity achieving when used with random code-books, and having a complexity that reduces as the code-rate increases.**

***Index Terms*—Channel Coding; Soft detection; Symbol Reliability; ML decoding; Error exponents.**

I. Introduction

Consider a channel with inputs, X^n, and outputs, Y^n, consisting of blocks of n symbols from a finite alphabet $\mathbb{A} = \{0, \ldots, |\mathbb{A}| - 1\}$[1]. Assume that channel input is altered by random 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 channel's output and input the noise can be recovered: $N^n = Y^n \ominus X^n$. In the hard detection setting, we recently introduced a novel noise-centric channel decoding algorithm, GRAND [1], [2]. GRAND, which is described in detail in the next section, has unusual features: it is suitable for use with any code-book; it can be employed without interleaving if noise is bursty; it is capacity achieving when used with random code-books; and its complexity decreases as code-book rates increase.

A significant feature of the GRAND approach is that if information regarding the noise process is garnered at the receiver and passed to the decoder, the algorithm can naturally incorporate it. For example, [1], [2] establish that if the receiver knows noise is bursty and well-modeled by a Markov process, use of that information by GRAND results in more accurate decodings as well as reduced decoding complexity.

In the present paper, we extend the remit of the algorithm's adaptability by establishing its ability to incorporate soft detection symbol reliability information where symbols received from the channel are accurately indicated to be error free or to have possibly been subjected to independent, additive random noise. In practice, this corresponds to a situation where, for example, soft information such as instantaneous Signal to Interference plus Noise Ratio is thresholded so as to provide false negatives with a sufficient small likelihood that they do not dominate the block error probability; this approach is, in effect, a simple code-independent quantization of soft information for decoding [3].

As soft information is known to enhance decoding accuracy, its use has been extensively considered since early on, with Wagner decoding [4] and, later, Chase decoding [5]. It has been extensively investigated in decoding algorithms for linear block [6], [7] and convolutional [8] codes.

Unlike those code-book centric channel decoding algorithms which require algorithmically involved modifications, soft detection symbol reliability information can be incorporated in a straight-forward way into the GRAND approach. In this paper, we determine the gain in capacity, reduction in block error rate, and decrease in decoding complexity that can be obtained by leveraging this soft detection symbol reliability information, leading to soft-GRAND (SGRAND),

II. Guessing Random Additive Noise Decoding

To implement ML decoding, the sender and receiver 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 given the transmitted code-word was $c^{n,i}$ by $p_{Y^n|C^n}(y^n|c^{n,i}) = P(N^n = y^n \ominus c^{n,i})$ for $i \in \{1, \ldots, M_n\}$. A ML decoding, $c^{n,*}$, is then an element of the code-book that has the highest conditional likelihood of transmission given what was received, i.e.

$$c^{n,*} \in \arg\max \left\{ p_{Y^n|C^n}(y^n|c^{n,i}) : c^{n,i} \in \mathscr{C}_n \right\}. \tag{1}$$

The fundamental principle underlying GRAND is that the receiver rank-orders noise sequences from most likely to least likely, breaking ties arbitrarily, and then sequentially queries whether the sequence that remains when the noise is removed from the received signal is an element of the code-book. The first instance where the answer is in the affirmative is the decoded element. To see that GRAND identifies a ML decoding irrespective of how the code-book is constructed, note that owing to the definition of $c^{n,*}$

$$\begin{aligned} p_{Y^n|C^n}(y^n|c^{n,*}) &= P(N^n = y^n \ominus c^{n,*}) \\ &\geq P(N^n = y^n \ominus c^{n,i}) \text{ for all } c^{n,i} \in \mathscr{C}_n \end{aligned}$$

[1]Lower case letters correspond to realizations of upper-case random variables or their normalized limits, apart from for noise where z is used as n denotes the code block-length. Logs are taken base $|\mathbb{A}|$ throughout.

and so by sequentially subtracting noise sequences from the received sequence in decreasing order of likelihood and querying if what remains is in the code-book, the first identified element is a ML decoding.

GRAND can be thought of as a guessing race where the querying process is halted either with success on identifying the true noise, and hence the transmitted code-word, or with an error on identifying a non-transmitted element of the code-book. In the present paper we assume that, in addition to the channel output Y^n, the receiver is provided with a vector of symbol reliability information, S^n taking values in $\{0,1\}^n$ where a 0 truthfully indicates a symbol has not been subject to noise while a 1 indicates it may have been, which is similar in spirit to the well-known Gilbert-Elliott model [9], [10]. The core idea behind SGRAND is for S^n to be used as a mask that separates symbols that require guessing from those that do not.

The adaptation to the symbol reliability information setting results in a ML decoder, SGRAND, that proceeds as follows:

- Given channel output y^n and symbol reliability information s^n, initialize $i=1$, set the non-noise-impacted symbol locations of guessed noise sequence z^n to 0, and set the masked potentially noise-impacted locations z^n to be the most likely noise sequence of length $l^n = \sum_i s_i^n$.
- While $x^n = y^n \ominus z^n \notin \mathscr{C}_n$, increase i by 1 and change the masked symbols of z^n to be the next most likely noise sequence of length l^n.
- The x^n resulting from this while loop is the decoded element.

Based on the same logic as GRAND, this procedure identifies a ML decoding, but it will perform fewer queries and the output element will be more likely to be the transmitted one, given the targeted nature of the querying.

III. Mathematical Analysis

As in [2], for the analysis of SGRAND we exploit the fact that the algorithm is a guessing race. The difference here is that the decoder is only asking questions of the sub-string potentially impacted by noise.

If informed that n symbols have been potentially impacted by noise, the receiver creates a guesswork order [11], $G: \mathbb{A}^n \mapsto \{1,\ldots,|\mathbb{A}|^n\}$, a list of noise sequences ordered from most likely to least likely and breaking ties arbitrarily:

$$G(z^{n,i}) \leq G(z^{n,j}) \text{ iff } P(N^n = z^{n,i}) \geq P(N^n = z^{n,j}), \quad (2)$$

For general i.i.d. noise, exponential families whose elements all possess the same guesswork order have been identified [12].

Assumption 1. *When noise occurs, it is i.i.d., distributed as N_1 where $P(N_1 = i) = p_{N|S}(i|1) = P(N = i|S = 1)$ for $i \in \mathbb{A}$.*

Under assumption 1, if one must guess the entire noise string of length n, Arikan [13] first established how the positive moments of guesswork, $E(G(N^n)^\alpha)$ for $\alpha > 0$, scale in n in terms of Rényi entropies. Building on those and subsequent results that treated negative moments, it was established that the logarithm of guesswork satisfies a Large Deviation Principle (LDP) [14].

Proposition 1 (Guesswork Moments & LDP [13], [15], [14])**.** *Under assumption 1, if $S^n = 1^n$ so that all received symbols are potentially impacted by noise and are distributed as N_1, the scaled Cumulant Generating Function (sCGF) of $\{n^{-1}\log G(N^n)\}$ exists:*

$$\begin{aligned}\Lambda^{N_1}(\alpha) &= \lim_{n\to\infty}\frac{1}{n}\log E(G(N^n)^\alpha | S^n = 1^n) \\ &= \begin{cases} \alpha H_{1/(1+\alpha)} & \text{if } \alpha > -1 \\ -H_{\min} & \text{if } \alpha \leq -1, \end{cases} \end{aligned} \quad (3)$$

where H_α is the Rényi entropy of a single noise element, N_1, with parameter α

$$\begin{aligned} H_\alpha &= \frac{1}{1-\alpha}\log\left(\sum_{i\in\mathbb{A}} p_{N|S}(i|1)^\alpha\right), \\ H = H_1 &= -\sum_{i\in\mathbb{A}} p_{N|S}(i|1)\log p_{N|S}(i|1), \\ \text{and } H_{\min} &= -\max_{i\in\mathbb{A}}\log p_{N|S}(i|1). \end{aligned}$$

Moreover, given $S^n = 1^n$, the process $\{n^{-1}\log G(N^n)\}$ satisfies a LDP with convex rate-function

$$I^{N_1}(x) = \sup_{\alpha\in\mathbb{R}}(x\alpha - \Lambda^{N_1}(\alpha)), \quad (4)$$

where $I^{N_1}(0) = H_{\min}$ and $I^{N_1}(H) = 0$.

Setting $\alpha = 1$ in eq. (3), as Arikan originally did in his investigation of sequential decoding, establishes that the expected guesswork grows exponentially in n with rate $H_{1/2}$, which is no smaller than the Shannon entropy, H. That the zero of the rate-function in eq. (4) occurs at H ensures, however, that the majority of the probability is accumulated by making queries up to and including the Shannon typical set. The apparent discrepancy in these two facts occurs because the guesswork distribution has a long tail that dominates its average [13], [16]. In the soft-detection setting, for a transmitted block of length n it is not necessary to guess a noise-string of length n, only potentially noise-impacted symbols. To that end, we have the following assumption.

Assumption 2. *With $L^n = \sum_{i=1}^n S_i^n$ being the number of potentially noise-impacted symbols in a block of length n, the proportion of them, $\{L^n/n\}$, satisfies a LDP with a strictly convex rate-function $I^L : \mathbb{R} \mapsto [0,\infty]$ such that $I^L(y) = \infty$ if $y \notin [0,1]$ and $I^L(\mu) = 0$, where $\lim_n E(L^n/n) = \mu > 0$. Define the sCGF for $\alpha \in \mathbb{R}$ to be*

$$\Lambda^L(\alpha) = \lim_{n\to\infty}\frac{1}{n}\log E\left(|\mathbb{A}|^{\alpha L^n}\right) = \sup_{x\in[0,1]}\left(\alpha x - I^L(x)\right) \in [-\infty,\infty],$$

which exists due to Varadhan's Lemma, e.g. [17].

Under Assumptions 1 and 2, the channel's capacity with soft detection symbol reliability information is upper bounded by

$$\begin{aligned} C^{\text{Soft}} &\leq \limsup_{n\to\infty}\frac{1}{n}\sup I(X^n;(Y^n,S^n)) \\ &= 1 - \mu h(p_{N|S}(\cdot|1)), \end{aligned} \quad (5)$$

where $h\left(p_{N|S}(\cdot|1)\right) = -\sum_{i\in\mathbb{A}} p_{N|S}(i|1)\log p_{N|S}(i|1)$ is the Shannon entropy of $p_{N|S}(\cdot|1)$; SGRAND will confirm this is achievable. Under Assumptions 1 and 2, for a distinct purpose it has been established that with a random number of characters to be guessed one has the following LDP [18].

Proposition 2 (LDP for guessing subordinated noise [18])**.** *Under assumptions 1 and 2, the joint subordinated guesswork and length process $\{(1/n\log G(N_1^{L^n}), L^n/n)\}$ satisfies a LDP with the jointly convex rate-function $I^{N_1,L}(g,l) = lI^{N_1}(g/l) + I^L(l)$. The subordinated guesswork process $\{1/n\log G(N_1^{L^n})\}$ alone satisfies a LDP with the convex rate function $I^{N_1{}^L}(g) = \inf_{l\in[0,1]}\left(lI^{N_1}(g/l) + I^L(l)\right)$, where $I^{N_1{}^L}(\mu H) = I^{N_1,L}(H,\mu) = 0$. The sCGF for $\{1/n\log G(N_1^{L^n})\}$, the Legendre-Fenchel transform of $I^{N_1{}^L}$, is given by the composition of the sCGF for the length with the sCGF for the guesswork of non-subordinated noise*

$$\begin{aligned}\Lambda^{N_1{}^L}(\alpha) &= \lim_{n\to\infty}\frac{1}{n}\log E\left(G\left(N_1^{L^n}\right)^{\alpha}\right) = \Lambda^L(\Lambda^{N_1}(\alpha)) \\ &= \sup_g(g\alpha - I^{N_1{}^L}(g)) \text{ for } \alpha\in\mathbb{R}. \end{aligned} \tag{6}$$

In particular, the average number of queries to required to identify subordinated noise is given by $\Lambda^{N_1{}^L}(1) = \Lambda^L(H_{1/2})$.

To characterize the number of queries made until a non-transmitted element of the code-book is identified, which is the second part of the race, we assume that the code-book is uniformly random. For such code-books, the location of each of its elements in the guessing order of a received transmission is itself uniform in $\{1,\ldots,|\mathbb{A}|^n\}$. Thus the distribution of the number of guesses until any non-transmitted element of the code-book is hit upon is distributed as the minimum of $M_n = \lfloor|\mathbb{A}|^{nR}\rfloor$ such uniform random variables, where R is the code-book rate. We can, therefore, use the following result from [2].

Proposition 3 (LDP for Guessing a Non-transmitted Code–word [2])**.** *Assume that $M_n = \lfloor|\mathbb{A}|^{nR}\rfloor$ for some $R > 0$, and that $U^{n,1},\ldots,U^{n,M_n}$ are independent random variables, each uniformly distributed in $\{1,\ldots,|\mathbb{A}|^n\}$. Defining $U^n = \min_i U^{n,i}$, $\{1/n\log U^n\}$ satisfies a LDP with the lower semi-continuous rate-function*

$$I^U(u) = \begin{cases} 1-R-u & \text{if } u\in[0,1-R] \\ +\infty & \text{otherwise} \end{cases} \tag{7}$$

and $\lim_{n\to\infty} n^{-1}\log\mathbb{E}(U^n) = 1-R$.

A graphical representation of the probabilistic guessing race can be found in Fig. 1. When all symbols are subject to noise, the channel is within capacity so long as the zero of the rate-function for guessing noise, which occurs at the Shannon entropy rate of the noise H, is smaller than the zero of the rate-function for identifying a non-transmitted code-word, which occurs at $1-R$. As in all likelihood the correct decoding is identified after fewer queries than an incorrect decoding would be identified, the algorithm experiences concentration onto correct decodings, which leads to the Channel Coding

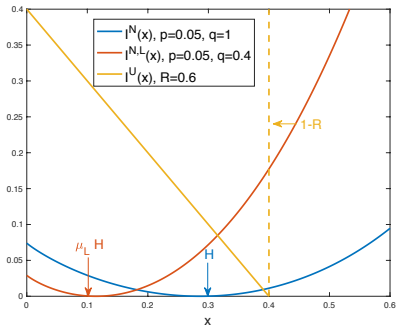

Fig. 1. Guesswork decoding race. $\mathbb{A} = \{0,1\}$, with a random code-book of rate R, with bits independently impacted by noise with probability q, whereupon they are flipped with probability p. For $p = 0.05$ and $R = 0.6$, plot shows large deviations rate function for: incorrectly identifying a non-transmitted element of the code-book, $I^U(x)$; guessing the true noise if $q = 1$, $I^{N_1}(x)$; with $q = 0.4$, guessing the true noise if a random set of locations are declared to be potentially noise-impacted, $I^{N_1{}^L}(x)$; With x being the value on the x-axis, when 2^{nx} noise guesses are made, the likelihood of success for each of these racing elements is approx. $2^{-n\inf_{y<x} I(y)}$ for the relevant $I(y)$.

Theorem, $R < 1-H$, in [2]. Here, the zero of the rate function for the subordinated noise-guessing with soft detection symbol reliability information occurs at μH. So long as μH is smaller than $1-R$, noise-guessing concentrates on identifying correct decodings, leading to any $R < 1-\mu H$ being achievable.

Theorem 1 (Soft Detection Channel Coding Theorem with SGRAND)**.** *Under Assumptions 1 and 2, and those of Proposition 3, if the code-book rate is less than capacity, $R < 1-\mu H$, then the probability that the SGRAND identified ML decoding is not the transmitted code-word decays exponentially in the block length n,*

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}\log P\left(U^n \le G\left(N_1^{L^n}\right)\right) \\ &= -\inf_{u\in[\mu H,1-R]}\{I^U(u) + I^{N_1{}^L}(u)\} < 0. \end{aligned} \tag{8}$$

If, in addition, g^* exists such that

$$\frac{d}{dg}I^{N_1}(g)|_{g=g^*} = 1, \tag{9}$$

which is analogous to one minus Gallager's critical rate, recalling $H_{1/2}$ is the Rényi entropy of the noise with parameter $1/2$, the SGRAND error rate is

$$\begin{aligned}\varepsilon(R) &= -\lim_{n\to\infty}\frac{1}{n}\log P\left(U^n \le G\left(N_1^{L^n}\right)\right) \\ &= \begin{cases} 1-R-\Lambda^L(H_{1/2}) & \text{if } R\in(0,1-\mu g^*) \\ I^{N_1{}^L}(1-R) & \text{if } R\in[1-\mu g^*, 1-\mu H) \\ 0 & \text{if } R\in(1-\mu H, 1]. \end{cases} \end{aligned} \tag{10}$$

Proof. As $\{U^n\}$ is independent of $\{(G(N_1^{L^n}), L^n)\}$, we have that $\left\{\left(n^{-1}\log U^n, n^{-1}\log G(N_1^{L^n}), L^n/n\right)\right\}$ satisfies an LDP with rate-function $I^U(u) + I^{N_1,L}(g,l)$. Now note the equivalence of the following two events:

$$\left\{U^n \le G\left(N_1^{L^n}\right)\right\} = \left\{\frac{1}{n}\log\left(U^n/G\left(N_1^{L^n}\right)\right) \le 0\right\}.$$

By the contraction principle, e.g. [17], with the continuous function $f(u,g,l) = (u-g,l)$, the process

$\left\{\left(n^{-1}\log\left(U^n/G\left(N_1^{L^n}\right)\right), n^{-1}L^n\right)\right\}$ satisfies a LDP with rate-function $\inf_{u\in[0,1-R]}\left\{I^U(u)+I^{N_1,L}(u-x,l)\right\}$. Consider

$$\begin{aligned}
&\varepsilon^L(R,l) := \\
&-\lim_{\delta\downarrow 0}\lim_{n\to\infty}\frac{1}{n}\log P\left(\frac{1}{n}\log\left(\frac{U^n}{G\left(N_1^{L^n}\right)}\right)\le 0, \frac{L^n}{n}\in(l-\delta,l+\delta)\right) \\
&= \inf_{x\le 0}\inf_{u\in[0,1-R]}\left\{I^U(u)+I^{N_1,L}(u-x,l)\right\} \\
&= \inf_{u\in[0,1-R]}\left\{I^U(u)+\inf_{g\ge u} lI^{N_1}\left(\frac{g}{l}\right)\right\}+I^L(l).
\end{aligned}$$

For $u\in[0,1-R]$, $I^U(u)=1-R-u$ is linearly decreasing, while $lI^{N_1}(g/l)$ is convex in g with minimum, zero, at $g=lH$. Thus if $R\ge 1-lH$, setting $u=1-R$ and $g=lH$, $\varepsilon^L(R,l)=I^L(l)$. If $R<1-lH$, then as both $I^U(u)$ and $lI^{N_1}(g/l)$, as a function of g, are strictly decreasing on $[0,lH]$, we have

$$\begin{aligned}
&\inf_{u\in[0,1-R]}\left\{I^U(u)+\inf_{g\ge u} lI^{N_1}\left(\frac{g}{l}\right)\right\} \\
&= \inf_{u\in[lH,1-R]}\left\{1-R-u+lI^{N_1}\left(\frac{u}{l}\right)\right\},
\end{aligned}$$

which is strictly positive as I^U is strictly decreasing to 0 on $[lH,1-R]$ while $lI^{N_1}(u/l)$ is strictly increasing in u on the same range. Assuming Gallager's critical rate, g^* defined in eq. (9), exists, as I^U is decreasing at rate 1 and

$$\frac{d}{dg}lI^{N_1}\left(\frac{g}{l}\right)|_{g=lg^*}=1,$$

then if $lg^*\le 1-R$, i.e. if $R\le 1-lg^*$,

$$\begin{aligned}
\inf_{u\in[lH,1-R]}\left\{1-R-u+lI^{N_1}\left(\frac{u}{l}\right)\right\} &= 1-R-lg^*+lI^{N_1}(g^*) \\
&= 1-R-lH_{1/2},
\end{aligned}$$

as $I^{N_1}(g^*)=g^*-H_{1/2}$. If, instead, $lg^*\ge 1-R$, then the infimum occurs at $u=1-R$ and

$$\inf_{u\in[lH,1-R]}\left\{1-R-u+lI^{N_1}\left(\frac{u}{l}\right)\right\} = lI^{N_1}\left(\frac{1-R}{l}\right)$$

if $R\in[1-lg^*,1-lH]$. The error exponent, $\varepsilon(R)$ in eq. (10), is obtained by the contraction principle, projecting out L^n/n and giving $\varepsilon(R)=\inf_{l\in[0,1]}\varepsilon^L(R,l)$. □

Combining Propositions 2 and 3 in a distinct way determines the asymptotic complexity of SGRAND in terms of the number of queries until a decoding, correct or incorrect, is identified: $D^n := \min\left(G\left(N^{L^n}\right), U^n\right)$. I.e SGRAND terminates at either identification of the noise that was in the channel or when a non-transmitted element of the code-book is identified, whichever occurs first. On the scale of large deviations, if the code-book is within capacity, $R<1-\mu H$, then the sole impact of the code-book is to curtail excessive guessing when unusual noise occurs.

Theorem 2 (Complexity of SGRAND)**.** *If $R<1-\mu H$, under Assumptions 1 and 2, and those of Proposition 3, $\{1/n\log D^n\}$ satisfies the LDP with a lower-semicontinuous convex rate-function*

$$I^D(d)=\begin{cases} I^{N_1{}^L}(d) & \text{if } d\in[0,1-R] \\ +\infty & \text{if } d>1-R \end{cases} \tag{11}$$

and the expected number of guesses until a ML decoding is found by SGRAND satisfies

$$\lim_{n\to\infty}\frac{1}{n}\log E(D^n)=\min\left(\Lambda^L(H_{1/2}),1-R\right).$$

Proof. Consider the process $\{n^{-1}\log D^n\}$. As $f(g,u)=\min(g,u)$ is a continuous function, by the contraction principle it satisfies a LDP with rate-function $I^{\mathrm{D}}(d)=\inf\{I^{N_1{}^L}(g)+I^U(u):\min(g,u)=d\}$. If $d>1-R$, $I^{\mathrm{D}}(d)=\infty$ as $I^U(d)=\infty$ for $d>1-R$. Alternatively, if $d\le 1-R$,

$$\begin{aligned}
I^{\mathrm{D}}(d) &= \min\left(I^{N_1{}^L}(d)+\inf_{x\ge d}I^U(x), \inf_{x\ge d}I^{N_1{}^L}(x)+I^U(d)\right) \\
&= \min\left(I^{N_1{}^L}(d), \inf_{x\ge d}I^{N_1{}^L}(x)+I^U(d)\right)
\end{aligned}$$

as $I^U(x)$ is decreasing for $x\in[0,1-R]$. If $R<1-\mu H$, then we make the following geometric considerations

$$I^{N_1{}^L}(0)=\inf_l\left\{I^L(l)+lI^{N_1}(0)\right\}=\inf_l\left\{I^L(l)+lH_{\min}\right\}\le\mu H_{\min},$$

where in the last inequality we have set $l=\mu$. As min-entropy is less than Shannon entropy $\mu H_{\min}\le\mu H<1-R$ and as $I^{N_1{}^L}$ is convex, $I^{N_1{}^L}(d)\le I^U(d)$ for all $d\in[0,H]$ while $I^{N_1{}^L}(d)$ is increasing on $[H,1-R]$ and so $I^{\mathrm{D}}(d)=I^{N_1{}^L}(d)$ for $d\in[0,1-R]$. For the scaling result for $E(D^n)$, we reverse the Legendre-Fenchel transform of I^{D} to obtain the sCGF of the process $\{n^{-1}\log D^n\}$ via Varadhan's Lemma. □

IV. Example: Soft Conditional BSC

Assume that data is transmitted in a binary alphabet, $\mathbb{A}=\{0,1\}$, noise is additive in $\mathbb{F}_2$, and each transmitted bit is impacted by noise independently with probability $p_S(1)=q\in[0,1]$ whereupon $p_{N|S}(1|1)=p\in[0,1]$. From eq. (5), the soft decoding channel's capacity is $C^{\mathrm{Soft}}=1-qh_2(p)$, where $h_2(p)$ is the binary Shannon entropy. The corresponding channel in which S^n is not observed is a binary symmetric channel (BSC) with probability $p_{N|S}(1|1)p_S(1)=pq$, whose channel capacity is $C^{\mathrm{Hard}}=1-h_2(qp)$. As h_2 is concave, $C^{\mathrm{Soft}}\ge C^{\mathrm{Hard}}$ for all q and p.

As the soft detection symbol reliability information $\{S^n\}$ is constructed of i.i.d. elements, the rate function governing the LDP for the proportion of noise impacted symbols, $\{L^n/n\}$, in Assumption 2 is the Kullback-Leibler divergence, $I^L(l)=-(1-l)\log((1-l)/(1-q))-l\log(l/q)$, which has the corresponding sCGF $\Lambda^L(\alpha)=\log(1-q+q2^\alpha)$. The rate function for LDP of the rescaled guesswork $\{1/n\log G(N_1^n)\}$ in eq. (4) is the Legendre-Fenchel transform, $I^{N_1}(g)=\sup_\alpha\left(\alpha g-\Lambda^{N_1}(\alpha)\right)$, of $\Lambda^{N_1}(\alpha)$

$$= \begin{cases} -\log\max(p,1-p) & \text{if } \alpha\le -1 \\ -p\log(p)-(1-p)\log(1-p) & \text{if } \alpha=1 \\ (1+\alpha)\log\left(p^{\frac{1}{1+\alpha}}+(1-p)^{\frac{1}{1+\alpha}}\right) & \text{if } \alpha\in(-1,1)\cup(1,\infty). \end{cases}$$

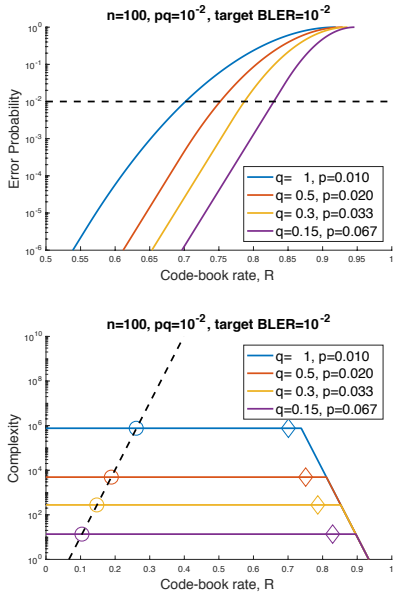

Fig. 2. BSC comparison, with $P(S=1)=q$ and $P(N=1|S=1)=p$), for a range of q and p such that $pq=10^{-2}$, $n=10^2$, a target block error rate of 10^{-2} and a code-rate of $\mathbb{R}$. In the upper plot, the horizontal dashed line is the target block error. In the lower plot the dashed black line gives approximate complexity of the brute force approach. Circles indicate complexity crossing points. Diamonds indicate the rate above which the target block error rate would be exceeded.

From eq. (6), the sCGF for the subordinated guesswork of true noise is $\Lambda^{N_1^L}(\alpha)=\Lambda^L(\Lambda^{N_1}(\alpha))$. The exponent of the average complexity required to identify the true noise in the channel with soft detection symbol reliability information is given by

$$\lim_{n\to\infty}\frac{1}{n}\log E\left(G\left(N_1^{L^n}\right)\right)=\log\left(1-q+q2^{2\log(p^{1/2}+(1-p)^{1/2})}\right),$$

while for the hard detection channel it is

$$\lim_{n\to\infty}\frac{1}{n}\log E(G(N^n))=2\log\left((qp)^{1/2}+(1-qp)^{1/2}\right).$$

While prefactors are not captured in the asymptotic analysis in Theorems 1 and 2, they allow approximations. For SGRAND's error probability we use $\approx 2^{-n\varepsilon(R)}$ for $R<1-qh_2(p)$, where the expression for $\varepsilon(R)$ can be found in eq. (10). For SGRAND, our measure of complexity is the average number of guesses per bit, $\approx 2^{n\min(1-R,\Lambda^L(H_{1/2}))}/n$. For comparison, let the complexity of the brute force computation of a ML decoding to be the number of conditional probabilities to be computed per bit before rank ordering and selecting the most likely code-book element, $2^{nR}/n$. Thus we are equating the work of one noise guess with computation of one conditional probability. As with SGRAND, this scheme results in ML decoding and so shares its error and success probabilities.

For $n=100$ and (q,p) pairs such that $pq=10^{-2}$ and so comparable with the hard detection channel, Fig. 2 plots approximate error probabilities and complexity as a function of code-book rate. The upper panel shows error probabilities, with a target block error rate indicated by the dashed horizontal line. Soft detection symbol reliability information greatly improves block error probability even though in this comparison the conditional probability of a bit flip given soft detection symbol reliability information increases as the soft detection probability decreases. The lower panel shows approximate complexity, with the dashed for the brute force approach, which grows exponentially in the code-book rate. The complexity of SGRAND is initially flat, corresponding to the average number of guesses until the true noise is identified. As R increases, eventually the SGRAND complexity drops, as encountering an erroneous element of the code-book clips the long guessing tail of true noise. Soft detection symbol reliability information dramatically reduces complexity.

REFERENCES

[1] K. R. Duffy, J. Li, and M. Médard, "Guessing noise, not code-words," in *IEEE Int. Symp. on Inf. Theory*, 2018.

[2] ——, "Capacity-achieving guessing random additive noise decoding (GRAND)," arXiv:1802.07010, Tech. Rep., 2018.

[3] N. Wernersson and M. Skoglund, "On source decoding based on finite-bandwidth soft information," in *IEEE Int. Symp. Inf. Theory*, 2005.

[4] R. A. Silverman and M. Balser, "Coding for constant data-rate systems - part 1 a new error-correcting code," Lincoln Laboratory, MIT, Tech. Rep., 1953.

[5] D. Chase, "Class of algorithms for decoding block codes with channel measurement information," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 170–182, January 1972.

[6] R. Koetter and A. Vardy, "Algebraic soft-decision decoding of Reed-Solomon codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 11, pp. 2809–2825, 2003.

[7] M. El-Khamy and R. J. McEliece, "Iterative algebraic soft-decision list decoding of Reed-Solomon codes," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 481–490, 2006.

[8] Y. S. Han, P.-N. Chen, and H.-B. Wu, "A maximum-likelihood soft-decision sequential decoding algorithm for binary convolutional codes," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 173–178, 2002.

[9] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. J.*, vol. 39, no. 5, pp. 1253–1265, 1960.

[10] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol. 42, no. 5, pp. 1977–1997, 1963.

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

[12] A. Beirami, R. Calderbank, M. Christiansen, K. R. Duffy, and M. Médard, "A characterization of guesswork on swiftly tilting curves," *IEEE Trans. Inform. Theory*, p. to appear, 2019.

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

[14] 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.

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

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

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

[18] M. M. Christiansen, K. R. Duffy, F. P. Calmon, and M. Médard, "Guessing a password over a wireless channel (on the effect of noise non-uniformity)," in *Asilomar Conf. Signals Syst. Comput.*, 2013.