Guessing random additive noise decoding with symbol reliability information (SRGRAND)

Guessing random additive noise decoding with symbol reliability information (SRGRAND)

Ken R. Duffy*, Muriel Médard† and Wei An†

*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, wei_an@mit.edu.

Abstract

The design and implementation of error correcting codes has long been informed by two fundamental results: Shannon's 1948 capacity theorem, which established that long codes use noisy channels most efficiently; and Berlekamp, McEliece, and Van Tilborg's 1978 theorem on the NP-hardness of decoding linear codes. These results shifted focus away from creating code-independent decoders, but recent low-latency communication applications necessitate relatively short codes, providing motivation to reconsider the development of universal decoders.

We introduce a scheme for employing binarized symbol soft information within Guessing Random Additive Noise Decoding, a universal hard detection decoder. We incorporate codebook-independent quantization of soft information to indicate demodulated symbols to be reliable or unreliable. We introduce two decoding algorithms: one identifies a conditional Maximum Likelihood (ML) decoding; the other either reports a conditional ML decoding or an error. For random codebooks, we present error exponents and asymptotic complexity, and show benefits over hard detection.

As empirical illustrations, we compare performance with majority logic decoding of Reed-Muller codes, with Berlekamp-Massey decoding of Bose-Chaudhuri-Hocquenghem codes, with CA-SCL decoding of CA-Polar codes, and establish the performance of Random Linear Codes, which require a universal decoder and offer a broader palette of code sizes and rates than traditional codes.

Keywords: Universal decoding, symbol reliability information, random codes.

I. Introduction

Since Shannon's 1948 opus [3] it has been known that channel capacity, the highest rate that an error correcting code can operate at while guaranteeing error-free communication over a noisy channel, is

A subset of these results was presented at the 2019 IEEE International Symposium on Information Theory, Paris, France, [1] and at the 2020 Annual Conference on Information Sciences and Systems, Princeton, USA [2]. In this article 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 we assume that $0 \in \mathbb{A}$ corresponds to no noise.

governed by the Shannon entropy of the channel's noise. By considering structureless random codes, his mathematical results proved that channel capacity is only achievable in the limit as the length of the error correcting code becomes large. By 1968, it was confirmed that his core theorems hold if structureless random codes are replaced with Random Linear Codes (RLCs) [4], which offer a more efficiently stored codebook description. In 1978, however, Berlekamp, McEliece, and Van Tilborg reported that maximum likelihood (ML) decoding of linear codes is an NP-complete problem [5], establishing that there exists a sequence of linear codes for which the decoding complexity is exponential as a function of block length. This feature, which underpins the McEliece cryptosystem [6], effectively halted practical consideration of universal decoding algorithms, with a couple of notable exceptions recounted in the Related Work.

The focus on long codes led to a working paradigm of pairing structured codes and code-specific decoders. Examples of such pairings are Reed-Muller (RM) codes [7], [8] with Majority Logic decoding, Reed-Solomon codes [9] with Berlekamp-Massey (BM) decoding [10], [11], [12], Low Density Parity Check Codes (LDPCs) [13] with belief propagation decoding [14], and, most recently, CRC-Assisted Polar (CA-Polar) codes, used in control channel communications in 5G New Radio (NR), with CRC-Assisted Successive Cancellation List (CA-SCL) decoding [15], [16], [17]. The structured nature of these codes leads to restrictions on lengths and rates. They are usually constructed based on the assumption of independent and identically distributed noise, which is then approximated through significant interleaving, with attendant delays. From an implementation point of view, distinct hardware is required for each code-decoder pair, and sometimes for different rates of the same code-decoder pair.

Many current communication systems require low-latency operation, where small bursts of data need efficient transmission [18], [19], [20]. Indeed, ultra-reliable low-latency communication (URLLC) is an important use-case in the 5G NR standard [21], [22]. Delivering URLLC necessitates efficient decoding of short, high-rate codes, motivating revisiting the possibility of high-accuracy universal decoders. Guessing Random Additive Noise Decoding (GRAND) [23], [24], [25], first proposed in 2018 for hard detection channels, is a class of decoding algorithms that can decode any code. GRAND's practical promise as a single efficient mechanism for any moderate redundancy code is such that circuit-based implementations have already been investigated [26], [27], [28] that avail of the inherent high level of parallelizability of the algorithm. That work demonstrates GRAND's performance credentials in hard detection channels, such as data storage system applications or communication systems with only hard detection demodulation.

GRAND's universal premise is that for a communication to be decodable the received signal must

faithfully contain information regarding the transmitted code-word *and* the error effect of the noise experienced on the channel. While most decoding algorithms utilize the codebook's structure to identify the transmitted code-word, GRAND endeavors to find the effect of the noise and so recover the transmitted code-word. To do this, it requires two devices: a method by which to query if a string is an element of the codebook; and a mechanism to sequentially create putative noise-effect sequences in decreasing order of their likelihood of occurrence on the channel. Armed with these, GRAND aims to produce an error corrected decoding for *any* block code, without restriction to binary or, indeed, linear codes.

Algorithm 1 Guessing Random Additive Noise Decoding. Given a demodulated channel output y^n and a function Φ such that $\Phi(y^n) = 0$ if and only if y^n is in the codebook, $c^{n,*}$ is the first codebook element identified and D^n is the number of codebook queries required to identify it, serving as a measure of confidence.

Inputs: y^n, Φ
Output: $c^{n,*}$, D^n
$d \leftarrow 0$, $D^n \leftarrow 0$.
$z^n \leftarrow$ next most likely noise effect sequence
$D^n \leftarrow D^n + 1$
if $\Phi(y^n \ominus z^n) = 0$ **then**
  $c^{n,*} \leftarrow y^n \ominus z^n$
  return $c^{n,*}$, D^n
end if

Pseudo-code for GRAND can be found in Algorithm 1, where the key step is "$z^n \leftarrow$ next most likely noise effect sequence". In the work that introduced GRAND the decoder only had access to a statistical description of the channel and hard-detection information. In that setting, GRAND provides ML decoding so long as the ordering of the putative noise effects matches the statistical description of the channel, even for channels with temporal noise correlations [24]. For standard models of hard detection noise effects from a highly interleaved channel or one subject to Markovian burst errors without an interleaver, dynamically creating putative noise effects is possible with simple logic [25]. That putative noise effects can be readily generated in parallel has been exploited in published hardware implementations of GRAND that perform multiple codebook membership queries per clock-cycle [26], [27], [28].

Incorporating soft information from per-realization measurements of received signals is known to be able to improve decoding significantly [29], but it is unclear how to do so with GRAND. It seems fraught at first blush as soft information seeks to represent continuous observations at the receiver, while GRAND searches over a collection of discrete noise effects on demodulated signals. A naïve approach, in which fine quantization of noise leads to guessing over a larger space of possible noise realizations, is inherently

undesirable from a complexity perspective.

Here we consider the problem of incorporating binary symbol reliability information to GRAND where soft information per received symbol is limited to a single bit to indicate whether a demodulated symbol has been demodulated with confidence or not. Analysis of the resulting schema, Symbol Reliability GRAND (SRGAND), results in: a universal ML decoder conditioned on one bit of symbol reliability information per received symbol; simulated performance for established linear codes and for RLCs, which exist at all lengths and rates and have theoretically desirable properties [30], but require a universal decoder; the mathematical evaluation of SRGRAND's complexity, showing an improvement vis-à-vis GRAND; error exponents for conditional ML decoding in the presence of a single bit of symbol reliability, and success exponents for the likelihood of correct decoding when the code-rate exceeds capacity.

II. Symbol reliability.

Essentially all digital communications involve taking discrete data, channel coding them to add robustness to noise, and then modulating those digital data into signals suitable for transmission and reception. For example, Phase Shift Keying (PSK), widely used in wireless communications systems, encodes groups of binary data into one of a finite set of phases of a carrier signal. To avail of better channel conditions in practice, not only is the codebook rate increased, but a modulation with a larger number of bits per modulated symbol is also employed. An illustration of Quadrature PSK (QPSK) is provided in Fig. 1. With transmitted symbols indicated by the red dots, assuming all symbol transmissions are equally likely and they are disturbed by independent additive Gaussian channel noise (AWGN), the probability density of a received signal being observed is indicated by the heat maps in Fig. 1 (a). Hard detection demodulation maps each received signal to the nearest potentially-transmitted symbol. One metric of confidence that a hard demodulated symbol corresponds to the transmitted one is the minimum across all possible alternate symbols of the Likelihood Ratio (LR) that a received signal was observed given the hard detection symbol was transmitted as compared with the alternate. The resulting LR surface is depicted in Fig. 1 (b). Instead of solely reporting the hard detection output, we envisage a further codebook independent quantization of the received signal into a symbol reliability indicator that separates reliably received symbols from unreliable ones. The principle behind the approach is illustrated in Fig. 1 (c) where a thresholding of the LR results in a masked region such that if a signal is received within that region, the hard detection demodulated symbol is flagged as being unreliable. The probability density of receiving a signal conditional on being in the masked region of potentially noise-impacted symbols is shown in Fig.

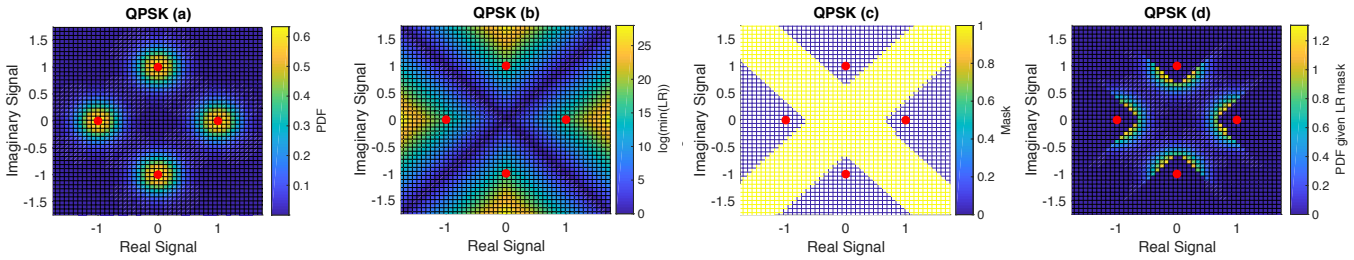

Fig. 1. QPSK subject to uncorrelated bivariate AWGN. Each pair of bits is coded into one of four symbols, indicated by the red dots. (a) displays a heat map of the probability density that the received signal is at a given location. When hard detection is employed, received signals are demodulated to the symbol in the quadrant where the received signal is measured and that is provided to the decoder. (b) shows the minimum Likelihood Ratio (LR) between each hard detection symbol and all others as heat maps, providing a measure of confidence in the hard demodulation. (c) displays a mask of the LR surfaces in (b) where within the hatched area the LR is greater than a threshold, and in the yellow mask area it is less than the threshold, identifying the potentially noise-impacted symbols. In the symbol reliability quantization, symbols received in the yellow masked region are demodulated but are also marked as being potentially noise-impacted. (d) provides heat map views of the probability density function of a received signal, conditioned on it being observed in the yellow masked area of uncertainty.

1 (d). Thus, the uncertainty region corresponding to the potentially noise-impacted symbols serves as a mask that labels received symbols whose values are questionable, enabling the decoder to focus on them. As the GRAND approach is codebook independent and noise-centric, we establish that we can incorporate symbol reliability information in a way that results in reduced complexity.

Our mathematical abstraction of this reliability information assumes that symbols received from the channel have been accurately indicated to be error free or to have possibly been subjected to independent additive random noise. SRGRAND then provides an ML decoding conditioned on the mask that was provided. In practice, creation of this symbol reliability information corresponds to a situation where soft information, such as instantaneous Signal to Interference plus Noise Ratios (SINR), has been thresholded so as to provide false negatives with a sufficient small likelihood that poor masking, i.e. incorrect identification of potentially noise-impacted symbols, does not dominate the block error probability. In effect, this symbol reliability information is a codebook-independent quantization of soft information [31]. In Section VII, using a simple threshold rule, for an additive white Gaussian noise channel we empirically find that the provision of symbol reliability information results in a 0.75 to 1 dB gain over optimal ML hard detection decoding, even when the symbol reliability information is potentially erroneous.

III. Guessing Random Additive Noise Decoding

The contribution of the current article is to identify how to incorporate symbol reliability information into the GRAND approach, and the ensuant increased capacity, reduced block error probability, and decreased complexity. We assume that, as well as being in receipt of a 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. This model is similar in spirit to the well-known Gilbert-Elliott model [32], [33], although our results will hold for channel state process

$\{S^n\}$ that have more involved correlation structures than Markovian. The core idea is that the vector S^n be used as a mask that separates symbols that require guessing, since they are potentially noise-impacted, from those that do not.

A. GRAND

Consider a hard-detection channel with inputs, X^n, and outputs, Y^n, consisting of blocks of n symbols from a finite alphabet $\mathbb{A} = \{0, \ldots, |\mathbb{A}| - 1\}$. Assume that channel input is altered by random noise effects, N^n, that are independent of the channel input and also take values in $\mathbb{A}^n$. Assume the function, $\oplus$, describing the channel's action, is invertible so that knowing the output and input the noise can be recovered:

$$Y^n = X^n \oplus N^n \text{ and } N^n = Y^n \ominus X^n. \tag{1}$$

In the hard detection setting, the receiver is solely provided with the discrete channel output Y^n.

Assuming code-words are selected uniformly at random, to implement ML decoding, the sender and receiver first share a codebook $\mathscr{C}_n = \{c^{n,1}, \ldots, c^{n,M_n}\}$ 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 $i \in \{1, \ldots, M_n\}$. The ML decoding is then

$$c^{n,*} \in \arg\max \left\{ p_{Y^n|C^n}(y^n|c^{n,i}) : c^{n,i} \in \mathscr{C}_n \right\}. \tag{2}$$

For hard detection, the principle underlying the algorithms in [23], [24] is to focus on identifying the noise that was experienced in the channel rather than directly trying to identify the transmitted code-word. Based on a statistical model of the symbol-level channel, the receiver achieves this by first rank-ordering noise sequences from most likely to least likely, breaking ties arbitrarily. In that order, the decoder sequentially queries whether the sequence that remains when the effect of the putative noise is removed from the received signal is an element of the codebook. The first instance where the answer is in the affirmative is the decoded element. To see that GRAND corresponds to ML decoding for channels described in Eq. (1), note that, owing to the definition of $c^{n,*}$ in Eq. (2),

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

Irrespective of how the codebook is constructed, by sequentially subtracting noise sequence effects from the received sequence in order from the most likely to least likely and querying if it is in the codebook,

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 codebook [24]. The second algorithm considered in [24], GRANDAB (GRAND with ABandonment), follows the same procedure as GRAND, but abandons noise guessing and declares an error if more than $|\mathbb{A}|^{n(H+\delta)}$ queries have been made, where H is the Shannon entropy rate of the noise and $\delta > 0$ is arbitrary. If more than $|\mathbb{A}|^{n(H+\delta)}$ queries are needed to identify a ML decoding, then the noise has been sufficiently unusual that, in query number terms, it is beyond the Shannon typical set. As a result, the block-error rate cost of abandoning is asymptotically negligible. Note the conditional likelihood that a ML decoding is in error increases as the number of queries made before identification of a codebook element increases, so one is abandoning a less certain decoding.

B. SRGRAND

The adaptation of this noise guessing principle to the symbol reliability setting results in a ML decoder conditioned on the veracity of that symbol reliability information. SRGRAND that proceeds as follows:

- Given channel output y^n and symbol reliability information $s^n = (s_1^n, s_2^n, \ldots, s_n^n)$, initialize $i = 1$, set the non-noise-impacted symbol locations of guessed noise sequence z^n to 0, and set the masked (i.e. potentially noise-impacted) entries of z^n to be the most likely noise effect 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 potentially noise-impacted symbols z^n to be the next most likely noise effect sequence of length l^n.
- The x^n that results from this while loop is the decoded element.

Note that SRGRAND can directly co-opt sequential noise pattern generators that were developed for GRAND by restricting their application to masked symbols alone.

Based on the same logic as for GRAND, which has only hard detection information, this procedure identifies a conditional ML decoding in this setting, but, depending on s^n, it will have performed fewer queries than GRAND and the output element is more likely to be the transmitted one, owing to the targeted nature of the querying. While SRGRAND always returns an element of the codebook that is a ML decoding conditioned on the symbol reliability information, the version with abandonment, SRGRANDAB, either provides a conditional ML decoding or returns an erasure. Without impacting the capacity-achieving nature

of the decoder, several distinct abandonment thresholds, which can be used in combination, are possible and result in reduced decoding complexity. We comment on two other possibilities in Section VIII, and prove results for one representative rule:

- With $L^n = \sum_{i=1}^{n} S_i^n$ being the random number of potentially noise-impacted symbols, assuming it exists, let $\lim_n E(L^n/n) = \mu^L > 0$ be the long run average proportion of potentially noise-impacted symbols. SRGRANDAB proceeds as SRGRAND, but abandons and declares an error without providing an element of the codebook if more than $|\mathbb{A}|^{n(\mu^L H+\delta)}$ queries are made, where H is the Shannon entropy of the noise for a potentially noise-impacted symbol, and $\delta > 0$ is arbitrary.

This is similar to the GRANDAB abandonment rule, but where enough queries are made to cover the typical set of the average number of potentially noise-impacted symbols.

In Section V we mathematically determine the gain in capacity, reduction in block error rate, and decrease in complexity that can be obtained by leveraging this symbol reliability information within the GRAND approach. The desirable features of GRAND stem from its focus on the noise rather than on the codebook as transmissions that are subject to light noise are quickly decoded, irrespective of the codebook construction or its rate, and these properties are preserved as we incorporate the symbol reliability information. We illustrate the gains to be obtained by considering a worked mathematical example in Section VI and, in Section VII, simulated performance evaluation with Reed-Muller (RM), Bose-Chaudhuri-Hocquenghem (BCH), CA-Polar, and RLC that also treats the possibility of decoding errors due to erroneous masks.

IV. Related Work

While the vast majority of codes and decoding systems, including all those currently used in practice, are co-designed, a few universal decoders have been developed. The original ML decoder works by computing the conditional probability of the received signal for every element of the not-necessarily-linear codebook and selecting the most likely. This approach means it can be used with structureless codes stored in a dictionary, and for channels with memory so long as the decoder has an accurate statistical description of it. This brute force evaluation requires an enormous number of real-valued computations for every received code-word, rendering the approach infeasible for all but the shortest of codes [34]. It is, however, amenable to mathematical analysis and remains of theoretical importance in the provision of performance bounds for an optimal decoder.

Restricting to binary linear $[n,k]$ codes, universal decoders have been studied for both cryptographical and communications purposes. Finding its roots in Prange's seminal research [35], Information Set Decoding (ISD) and its variants [36], [37], [38], [39], [40], [41] are randomized algorithms used to assess mathematically the security provided by code-based cryptosystems as the code becomes long. The core cryptographic scenario essentially maps to memoryless hard detection channels. Given a binary linear code-word and a received hard detection communication that has been subject to a known number of flipped bits, for each demodulated binary output the basic version of ISD works in two iterated steps until a decoding is found. The first is a transformation where the columns of the binary parity check matrix are randomly permuted and Gaussian elimination is performed to rewrite the code in systematic format. In the second step, for a number of columns that is less than the code's correction capability, the difference between the syndrome and all linear combinations of that number of columns is evaluated. Once this difference is found to be the zero vector, the Gaussian elimination transformation is inverted to identify the decoded code-word. Probabilistic analysis of the algorithm provides worse case bounds for decoding any code, and later tweaks to the algorithm serve to reduce the exponent in the complexity as a function of code-length. To use ISD for communications requires some adaptation owing to the assumption of an *a priori* known number of flipped bits.

In communications, soft information has been exploited to produce approximate ML decoders for binary linear codes. In 1974 [42] Dorsch introduced the idea of the Most Reliable Basis (MRB), and developments on that theme have led to Ordered Statistics Decoding (OSD) [43] and its variants [44], [45], [46], [47], [48], [49], [50]. The principle underlying OSD is to approximate ML decoding by computing conditional probabilities of the received signal for a substantially smaller list than the whole codebook, which one hopes contains the ML decoding. The number of real-valued conditional probabilities that must then be computed per received signal is determined by the size of the list. As with ISD, in OSD the linear code's column order is re-arranged and Gaussian elimination used to systematize it, but rather than using repeated random permutations the columns are ordered once in terms of decreasing bit reliability of the received transmission as determined from the soft information. The most reliable k bits are hard demodulated and a list of all binary sequences within a fixed Hamming distance, t, of that sequence is created. Each of those $\sum_{i=0}^{t}\binom{k}{i}$ sequences are multiplied by the revised code generator to create putative code-words in the MRB. The conditional probabilities of the received sequences for these code-words, rather than all code-words, is then computed. The most likely one is identified and converted back to the original basis

as the decoding. OSD's approximation relies on the principle that if one takes a hard decision on the k most reliable channel observations, depending on channel conditions, only few errors are expected within them, with the majority of the errors introduced by the channel instead being contained within the least reliable channel outputs, which are essentially ignored for decoding purposes. The larger the list, the better the approximation to ML decoding.

The original hard detection GRAND algorithm [23], [24], [25] is a true ML block-code decoder for hard detection channels subject to noise with or without memory. GRAND's operation requires a method to query a string's membership of a codebook. If the code is unstructured and stored in a dictionary, each query corresponds to a tree-search with a complexity that is logarithmic in the code-length. If the code was a Cyclic Redundancy Check (CRC) code, which is traditionally only used for error detection, checking for codebook membership requires only a simple polynomial calculation. If the code is linear in any field, codebook membership can be determined by a single matrix multiplication and comparison. The matrix multiplication results in the evaluation of a syndrome, but GRAND is not a syndrome decoder. No syndrome table is kept and, if channel conditions change, GRAND naturally adapts its decoding without recomputing an entire syndrome table. This latter point is particularly significant in the presence of soft information, which effectively serves to provide a distinct channel for each communication.

V. Mathematical Analysis

As in [24], for the analysis of SRGRAND and SRGRANDAB we exploit the fact that the algorithm is a race between sequential queries either identifying the noise in the channel, which results in a correct decoding, or encountering a non-transmitted element of the codebook, which results in an error. The difference with SRGRAND is that the decoder is faster and more precise than GRAND because it only asks questions of the sub-string that has been potentially impacted by noise. While the analysis is more involved, the results obtained are, possibly surprisingly, as clean as in the hard detection setting. Our mathematical treatment relies on techniques from Large Deviation Theory. While we endeavour to provide guiding heuristics, to follow the arguments in detail requires familiarity with that theory [51], [52], [53].

To analyze the algorithm, we recall notions of guesswork [54], [55]. Given the receiver is told that n symbols have been potentially impacted by noise, it creates a list of noise sequences, $G : \mathbb{A}^n \mapsto \{1, \ldots, |\mathbb{A}|^n\}$, ordered from most likely to least likely, with ties broken arbitrarily: $G(z^{n,i}) \leq G(z^{n,j})$ iff $P(N^n = z^{n,i}) \geq P(N^n = z^{n,j})$. For a sequence, $z^n \in \mathbb{A}^n$, its guesswork is the integer $G(z^n)$. For example, if the channel were binary, $\mathbb{A} = \{0,1\}$, and noise was Bernoulli for some $p < 0.5$, then the guesswork order follows

Hamming weight. For independent and identically distributed noise on more general alphabets, the family of measures that share the same guesswork order are described by an exponential family [56].

Assumption 1 (Noise distribution)**.** *When noise occurs, it is independent and identically 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 [55] first established how the non-negative moments of guesswork, $E(G(N^n)^\alpha)$ for $\alpha > 0$, scale in n in terms of Rényi entropies of order α. Building on those and subsequent results that treated negative moments, [57] for $\alpha > -1$ and for $\alpha \leq -1$, and more general noise sources, it was established [58] that the logarithm of guesswork satisfies a Large Deviation Principle (LDP) [51]. The LDP provides estimates on the distribution of the number of queries required to correctly identify a noise-string and was used as the basis to analyze one side of the decoding race in the hard detection setting [23], [24]. Recall that all logarithms are base $|\mathbb{A}|$.

Proposition 1 (Guesswork Moments and Large Deviation Principle [55], [57], [58])**.** *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:*

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

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

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

Moreover, given $S^n = 1^n$*, the process* $\{n^{-1}\log G(N^n)\}$ *satisfies a LDP (e.g. [51]) with convex rate-function*

$$I^N(x) = \sup_{\alpha\in\mathbb{R}}(x\alpha - \Lambda^N(\alpha)), \textit{where}\ I^N(0) = H_{\min}\ \textit{and}\ I^N(H) = 0. \tag{4}$$

Heuristically, Eq. (4) implies that $P(G(N^n) \approx |\mathbb{A}|^{nx}) = |\mathbb{A}|^{-nI^N(x)}$ for large n. As $I^N(H) = 0$, with high probability the number of queries until N^n is identified concentrates at $|\mathbb{A}|^{nH}$. The probability that N^n is identified in either fewer or more queries decays exponentially in n with a rate governed by the convex function I^N defined in Eq. (4). 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 greater 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 but has little probability.

In the symbol reliability setting, it is not necessary to guess a noise-string of length n. Instead, one must guess a random number of symbols corresponding to those inside the mask that are potentially noise-impacted. To that end, we have the following assumption on the size of the mask, which is the number of potentially noise-impacted symbols per transmission.

Assumption 2 (Number of potentially noise-impacted symbols - size of mask)**.** *With $L^n = \sum_{i=1}^n S_i^n$ being the mask size, i.e. 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 $I^L(l) = \infty$ if $l \notin [0,1]$ and $I^L(\mu^L) = 0$, where $\lim_n E(L^n/n) = \mu^L > 0$. Define the sCGF for $\alpha \in \mathbb{R}$ to be $\Lambda^L(\alpha) = \lim_{n\to\infty} \frac{1}{n} \log E\left(|\mathbb{A}|^{\alpha L^n}\right) = \sup_{l\in[0,1]} \left(\alpha x - I^L(l)\right)$, which exists in the extended reals owing to Varadhan's Lemma (e.g. [51][Theorem 4.3.1]).*

Roughly, Assumption 2, which is true for a broad class of processes $\{S^n\}$ including i.i.d., Markov and general mixing, e.g. [51], says the probability of having nl potentially noise impacted symbols decays exponentially in n with a rate, $I^L(l)$, that is positive unless l is the mean μ^L, i.e. $P(L^n \approx nl) \approx |\mathbb{A}|^{-nI^L(l)}$.

With some abuse of notation for Shannon entropy, under Assumptions 1 and 2, recalling that we define all logarithms as base $|\mathbb{A}|$, the symbol reliability decoding channel's capacity, $C^{\text{Sym. Rel.}}$ is upper bounded by

$$\begin{aligned} C^{\text{Sym. Rel.}} &\leq \limsup_{n\to\infty} \frac{1}{n} \sup I(X^n; (Y^n, S^n)) \leq 1 - \limsup_{n\to\infty} \frac{1}{n} H\left(N_1^{L^n}\right) = 1 - \limsup_{n\to\infty} \frac{E(L^n)}{n} H(N_1) \\ &= 1 - \mu^L h(p_{N|S}(\cdot|1)), \end{aligned} \tag{5}$$

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 $p_{N|S}(\cdot|1)$, we have upper-bounded the entropy of the input by 1, and used the fact that the channel is invertible (i.e. Eq. (1)). Through constructing SRGRAND and SRGRANDAB, we will show $C^{\text{Sym. Rel.}}$ is attainable.

Under Assumptions 1 and 2, in a distinct context and for a distinct purpose, it was established in [59] that with a random number of characters to be guessed one has the following LDP.

Proposition 2 (LDP for guessing subordinated noise [59])**.** *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

$$I^{N,L}(g,l) = lI^N\left(\frac{g}{l}\right) + I^L(l), \tag{6}$$

where I^N is the guesswork rate-function defined in Eq. (4) *and I^L is the length rate-function defined in Assumption 2. Note that $I^{N,L}(H,\mu^L) = 0$, where H is Shannon entropy of a noise-impacted symbol and μ^L is the average number of potentially noise-impacted symbols.*

The subordinated guesswork process $\{1/n \log G(N_1^{L^n})\}$ satisfies a LDP with the convex rate function

$$I^{N^L}(g) = \inf_{l\in[0,1]} \left(lI^N\left(\frac{g}{l}\right) + I^L(l)\right), \text{ where } I^{N^L}(\mu^L H) = I^{N,L}(H,\mu^L) = 0. \tag{7}$$

The sCGF for $\{1/n \log G(N_1^{L^n})\}$, the Legendre-Fenchel transform of I^{N^L}, is given by the composition of the sCGF for the length with the sCGF for the guesswork of non-subordinated noise

$$\Lambda^{N^L}(\alpha) = \lim_{n\to\infty} \frac{1}{n} \log E\left(G\left(N_1^{L^n}\right)^\alpha\right) = \Lambda^L(\Lambda^N(\alpha)) = \sup_g \left(g\alpha - I^{N^L}(g)\right) \text{ for } \alpha \in \mathbb{R}. \tag{8}$$

In particular, the average number of queries required to identify subordinated noise is given by

$$\Lambda^{N^L}(1) = \lim_{n\to\infty} \frac{1}{n} \log E\left(G\left(N_1^{L^n}\right)\right) = \Lambda^L(H_{1/2}). \tag{9}$$

Roughly speaking, the joint LDP indicates that, for large n, $P\left(\left(\frac{1}{n}\log G\left(N_1^{L^n}\right), \frac{L^n}{n}\right) \approx (g,l)\right) \approx |\mathbb{A}|^{-nI^{N,L}(g,l)}$, and $I^{N,L}(g,l)$ in Eq. (6) can be interpreted as follows: if the number of potentially noise-impacted symbols is $L^n \approx nl$, which is exponentially unlikely with rate $I^L(l)$, then having the logarithm of the subordinated guesswork be $\log G\left(N_1^{L^n}\right) \approx ng$ has essentially the same likelihood as $\log G\left(N_1^{ln}\right) \approx ng$, which has rate $lI^N(g/l)$ as a total deviation of g must be accrued over a smaller proportion of potentially noise-impacted symbols. The unconditioned LDP follows from the large deviations mantra that rare events occur in the most likely way, so that the rate-function I^{N^L} is determined from the proportion of potentially noise-impacted symbols that gives the smallest decay rate for the probability.

Results on the subordinated guesswork process $\{1/n \log G(N_1^{L^n})\}$ governed by the rate-function in Eq. (7) are sufficient to enable us to prove a Channel Coding Theorem for the symbol reliability channel. Finer-grained results on error exponents that depend on the proportion of symbols that were noise-impacted, however, follow from the LDP for the joint subordinated guesswork and length process $\{(1/n \log G(N_1^{L^n}), L^n/n)\}$ governed by the rate-function given in Eq. (6).

We note that Λ^L is a convex function whose derivative at the origin is μ^L, the mean number of potentially

noise-impacted symbols, so that $\Lambda^L(H_{1/2}) \geq \mu^L H_{1/2}$. Hence, from Eq. (9), the average number of queries until the true channel-noise is identified grows exponentially in n at a potentially larger rate than the guesswork required for the average proportion of potentially noise-impacted symbols. Despite that, the zero of the rate-function in Eq. (7) occurs at $\mu^L H$, so that the majority of the likelihood of identifying the true subordinated noise occurs by the Shannon entropy of the typical set of average number of potentially noise-impacted symbols. Thus, while stochastic fluctuations in the number of potentially noise-impacted symbols has relevance to complexity and error exponents, that variability has no impact on capacity. In a manner akin to GRANDAB, without loss of capacity, complexity can be ameliorated by abandoning guessing after a suitable number of queries.

To mathematically characterize the number of queries made until a non-transmitted code-word is identified, which is the second part of the guesswork decoding race, we assume that the codebook is created uniformly at random. For uniformly distributed codebooks, the location of each element in the guessing order of a received transmission is itself uniform in $\{1,\ldots,|\mathbb{A}|^n\}$. The distribution of the number of guesses until any non-transmitted element of the codebook is hit upon is thus distributed as the minimum of M_n uniform random variables. We can, therefore, use the following result from [23], [24], again recalling that our logarithm is base $|\mathbb{A}|$.

Proposition 3 (LDP for Guessing a Non-transmitted Code-word [23], [24])**.** *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 $U^n = \min_i U^{n,i}$, $\{1/n \log U^n\}$ then $\lim_{n\to\infty} \frac{1}{n}\log \mathbb{E}(U^n) = 1-R$. and U^n satisfies a LDP with the lower semi-continuous rate-function*

$$I^U(u) = \begin{cases} 1-R-u & \text{if } u \in [0,1-R] \\ +\infty & \text{otherwise.} \end{cases} \tag{10}$$

A graphical representation of the rate-functions that determine the asymptotic likelihoods of outcomes of this guessing race can be found in Fig. 2. When all symbols are subject to noise, as in [23], [24], 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 $1-R$, where R is the normalized codebook rate. As in all likelihood the correct decoding is identified after fewer queries than an incorrect element of the codebook would be identified, the algorithm experiences concentration onto a correct decoding, which leads to the

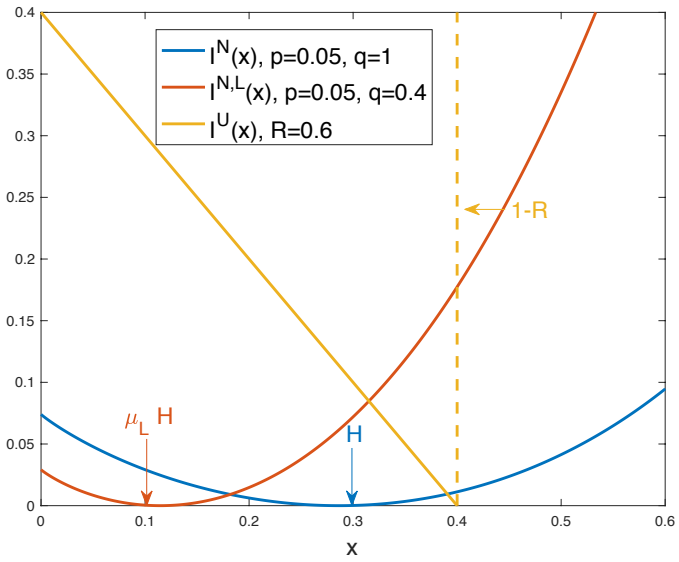

Fig. 2. Probabilistic guesswork decoding race in a SR-BSC. With $p = 0.05$ and codebook rate $R = 0.6$, the large deviations rate function for: incorrectly identifying a non-transmitted element of the codebook, $I^U(x)$; guessing the true noise if $q = 1$ and all bits are potentially noise-impacted, $I^N(x)$; with $q = 0.4$, guessing the true noise if a random set of locations are potentially noise-impacted, $I^{N^L}(x)$. With x being the value on the x-axis, when 2^{nx} noise guesses are made the likelihood of success for each of these three racing elements is approximately $2^{-n\inf_{y<x} I(y)}$ for the relevant rate function, $I(y)$.

proof of the classical hard detection Channel Coding Theorem, $R < 1 - H$, in [23], [24]. In the present paper, the zero of the rate function for the subordinated noise-guessing occurs at $\mu^L H$, the average number of potentially noise-impacted symbols times the Shannon entropy of the noise. So long as $\mu^L H$ is smaller than $1 - R$, noise-guessing concentrates on identifying correct decodings before erroneous ones, leading to the Symbol Reliability Channel Coding Theorem, proved below, where any $R < 1 - \mu^L H$ is achievable.

The proportion of potentially noise-impacted symbols is available to the receiver and so it is reasonable to consider error exponents subject to its knowledge. We characterize these error exponents in terms of R and the rate-function $I^{N,L}$ given in Eq. (6).In particular, define

$$\varepsilon^L(R,l) = -\lim_{\delta \downarrow 0} \lim_{n\to\infty} \frac{1}{n} \log P\left(\frac{L^n}{n} \in (l-\delta, l+\delta), U^n \leq G\left(N_1^{L^n}\right)\right) \tag{11}$$

to be the probability exponent that the proportion of potentially noise-impacted symbols, representing the size of the mask, is l, and that there is an error, as the number of queries required to identify a non-transmitted code-word is smaller than the number of queries required to identify the true noise.

Theorem 1 (Symbol Reliability Channel Coding Theorem)**.** *Assuming $R < 1 - \mu^L H$, under Assumptions 1 and 2, and those of Proposition 3, we have that the probability that the conditional ML decoding of SRGRAND is incorrect decays exponentially in n,*

$$\lim_{n\to\infty} \frac{1}{n} \log P\left(U^n \leq G\left(N_1^{L^n}\right)\right) = -\inf_{u\in[\mu^L H, 1-R]} \{I^U(u) + I^{N^L}(u)\} < 0. \tag{12}$$

If g^* exists such that

$$\frac{d}{dg}I^N(g)|_{g=g^*}=1, \tag{13}$$

which is analogous to one minus Gallager's critical rate, then the joint error exponent of (11)*, subject to a given proportion of potentially noise-impacted symbols satisfies*

$$\varepsilon^L(R,l)=\begin{cases} I^L(l)+1-R-lH_{1/2} & \text{if } R\in(0,1-lg^*] \\ I^L(l)+lI^N\left(\dfrac{1-R}{l}\right) & \text{if } R\in[1-lg^*,1-lH] \\ I^L(l) & \text{if } R\in(1-lH,1]. \end{cases} \tag{14}$$

The unconditioned SRGRAND error rate is

$$\varepsilon(R)=\inf_{l\in[0,1]}\varepsilon^L(R,l)=-\lim_{n\to\infty}\frac{1}{n}\log P\left(U^n\leq G\left(N_1^{L^n}\right)\right)=\begin{cases} 1-R-\Lambda^L(H_{1/2}) & \text{if } R\in(0,1-\mu^Lg^*) \\ I^{N^L}(1-R) & \text{if } R\in[1-\mu^Lg^*,1-\mu^LH) \\ 0 & \text{if } R\in(1-\mu^LH,1]. \end{cases} \tag{15}$$

With $\delta>0$, abandoning guessing if $|\mathbb{A}|^{n(\mu^LH+\delta)}$ queries have been made without identifying an element of the codebook, the SRGRANDAB error rate is also negative,

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}\log P\left(\left\{U^n\leq G(N_1^{L^n})\right\}\cup\left\{G(N_1^{L^n})\geq|\mathbb{A}|^{n(\mu^LH+\delta)}\right\}\right)\\&=-\min\left(\inf_{u\in[\mu^LH,1-R]}\{I^U(u)+I^{N^L}(u)\},I^{N^L}(\mu^LH+\delta)\right)<0.\end{aligned} \tag{16}$$

If, in addition, g^* defined in Eq. (13) *exists then the expression simplifies to $\varepsilon^{AB}(R)=\min\left(\varepsilon(R),I^{N^L}(H+\delta)\right)<0$ where $\varepsilon(R)$ is the conditional ML decoding error rate in Eq.* (15)*.*

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 a LDP with rate-function $I^U(u)+I^{N,L}(g,l)$. Noting the equivalence of the following two events,

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

By the contraction principle (e.g. [51][Theorem 4.2.1]) with the continuous function $f(u,g,l)=(u-g,l)$, the process $\left\{\left(\frac{1}{n}\log\left(U^n/G\left(N_1^{L^n}\right)\right),\frac{L^n}{n}\right)\right\}$ satisfies a LDP with rate-function $\inf_{u\in[0,1-R]}\left\{I^U(u)+I^{N,L}(u-x,l)\right\}$.

Consider $\varepsilon^L(R,l)$ defined in Eq. (14), where the limits exist as the rate-functions are convex and so

continuous on the interior of where they are finite,

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

This final expression essentially encapsulates that the error exponent is the exponent for the likelihood that the proportion of potentially noise-impacted symbols, or size of the mask, is l, plus the smallest exponent (corresponding to the most likely event) for the minimum of the scaled uniforms being at u, while the scaled sub-ordinated guesswork occurs at any value g at least as large as u.

For $u\in[0,1-R]$, $I^U(u)=1-R-u$ is linearly decreasing, while $lI^N(g/l)$ 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, alternatively, $R<1-lH$, then as both $I^U(u)$ and $lI^N(g/l)$, as a function of g, are strictly decreasing on $[0,lH]$,

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

which is strictly positive as I^U is strictly decreasing to 0 on $[lH,1-R]$ while $lI^N(u/l)$ is strictly increasing in u on the same range. Assuming g^* defined in Eq. (13), exists, as I^U is decreasing at rate 1 and $\frac{d}{dg}lI^N\left(\frac{g}{l}\right)|_{g=lg^*}=1$, then if $lg^*\le 1-R$, i.e. if $R\le 1-lg^*$,

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

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

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

and the expression in (14) follows. The unconditional error exponent, $\varepsilon(R)$ in Eq. (15), is obtained from that in (14) by the contraction principle, projecting out L^n/n, giving $\varepsilon(R)=\inf_{l\in[0,1]}\varepsilon^L(R,l)$. If $R\ge 1-\mu^L H$, then $\varepsilon(R)=\varepsilon^L(R,\mu^L)=0$. If $R\in[1-lg^*,1-lH]$, then $\varepsilon(R)=\inf_l\left\{I^L(l)+lI^N\left(\frac{1-R}{l}\right)\right\}=I^{N^L}(1-R)$. Finally, if $R\in(0,1-l]$, then $\varepsilon(R)=\inf_l\left\{I^L(l)+1-R-lH_{1/2}\right\}=(1-R)-\inf_l\left\{lH_{1/2}-I^L(l)\right\}=1-R-\Lambda^L(H_{1/2})$, inverting the Legendre-Fenchel transform in the last step.

To determine the error exponent of SRGRANDAB, by the Principle of the Largest Term [51, Lemma 1.2.15] it suffices to consider only the smallest of the two exponential rates in Eq. (16). The first term is the error rate for GRAND. The second term is the exponent of the probability of error due to abandonment

of guessing. Note that $P\left(G(N_1^{L^n}) \geq |\mathbb{A}|^{n(\mu^L H+\delta)}\right) = P\left(\frac{1}{n}\log G(N_1^{L^n}) \geq \mu^L H+\delta\right)$ and the result follows from the LDP as $I^{N^L}(x)$ is convex and increasing for $x > \mu^L H$. □

Consider the error exponent for the conditional ML decoding via SRGRAND, $\varepsilon^L(R,l)$ in (14). The exponent for the likelihood that the proportion of potentially noise-impacted symbols, L^n/n, which is approximately l, is $I^L(l)$. The error-exponent is then as in a channel where only a proportion l of transmitted symbols are in the mask of symbols subject to noise [24]. The unconditional equivalent, $\varepsilon(R)$ in Eq. (15) identifies the most likely proportion of noise-impacted symbols that may give rise to an error for a given codebook rate. For SRGRANDAB, an error occurs either if the identified conditional ML decoding is in error or if abandonment occurs. The more likely of these two events dominates in the limit.

Combining Propositions 2 and 3 in a distinct way enables us to determine the asymptotic complexity of the SRGRAND and SRGRANDAB in terms of the number of queries until a decoding, correct or incorrect, is identified: $D^n := \min\left(G\left(N^{L^n}\right), U^n\right)$ .That is, the algorithm terminates when the channel noise or a non-transmitted element of the codebook are identified, whichever occurs first. On the scale of large deviations, if the codebook is within capacity, $R < 1-\mu^L H$, then it becomes apparent that the sole impact of the codebook is to curtail excessive guessing when unusual noise occurs. The number of guesses SRGRANDAB makes until terminating is $D^n_{\text{AB}} := \min\left(G\left(N^{L^n}\right), U^n, |\mathbb{A}|^{n(\mu^L H+\delta)}\right)$. The final term corresponds to the abandonment threshold, curtailing guessing shortly after the Shannon typical set for an average number of potentially noise impacted symbols.

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

$$I^D(d) = \begin{cases} I^{N^L}(d) & \text{if } d \in [0, 1-R] \\ +\infty & \text{if } d > 1-R \end{cases} \tag{17}$$

and the expected number of guesses for SRGRAND to find a conditional ML decoding satisfies $\lim_{n\to\infty} \frac{1}{n}\log E(D^n) = \min\left(\Lambda^L(H_{1/2}), 1-R\right)$. With $\delta > 0$, the complexity of SRGRANDAB, $\{1/n \log D^n_{AB}\}$, satisfies a LDP with a convex rate function

$$I^{D\text{-}AB}(d) = \begin{cases} I^{N^L}(d) & \text{if } d \in [0, \min(1-R, \mu^L H)] \\ +\infty & \text{if } d > \min(1-R, \mu^L H) \end{cases} \tag{18}$$

and the expected number of guesses until SRGRANDAB terminates, $\{D_{AB}^n\}$*, satisfies* $\lim_{n\to\infty}\frac{1}{n}\log E(D_{AB}^n) = \min\left(\Lambda^L(H_{1/2}), 1-R, \mu H+\delta\right)$.

Proof. Consider the process $\{n^{-1}\log D^n\}$, following [24][Proposition 2], as $f(g,u)=\min(g,u)$ is a continuous function, by the contraction principle it satisfies a LDP with rate-function $I^{\mathrm{D}}(d)=\inf\{I^{N^L}(g)+I^U(u): \min(g,u)=d\}$. If $d>1-R$, $I^{\mathrm{D}}(d)=\infty$ as $I^U(d)=\infty$ for $d>1-R$. Alternatively, if $d\leq 1-R$,

$$I^{\mathrm{D}}(d)=\min\left(I^{N^L}(d)+\inf_{x\geq d}I^U(x), \inf_{x\geq d}I^{N^L}(x)+I^U(d)\right)=\min\left(I^{N^L}(d), \inf_{x\geq d}I^{N^L}(x)+I^U(d)\right)$$

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

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

where in the last inequality we have set $l=\mu^L$. As min-entropy is less than Shannon entropy $\mu^L H_{\min}\leq\mu^L H<1-R$ and as I^{N^L} is convex, $I^{N^L}(d)\leq I^U(d)$ for all $d\in[0,H]$ while $I^{N^L}(d)$ is increasing on $[H,1-R]$ and so $I^{\mathrm{D}}(d)=I^{N^L}(d)$ for $d\in[0,1-R]$.

To obtain the scaling result for $E(D^n)$ we invert the transformation from the rate function I^{D} to its Legendre-Fenchel transform, the sCGF of the process $\{n^{-1}\log D^n\}$ via Varadhan's Theorem [51][Theorem 4.3.1]. In particular, note that, regardless of whether I^{D} is convex or not,

$$\lim_{n\to\infty}\frac{1}{n}\log E(D^n)=\lim_{n\to\infty}\frac{1}{n}\log E\left(|\mathbb{A}|^{\log D^n}\right)=\sup_{d\in\mathbb{R}}\{d-I^{\mathrm{D}}(d)\}=\min\left(\Lambda^L(H_{1/2}),1-R\right).$$

The final component of the minimum satisfies an LDP with a rate function 0 if $d=\mu+\delta$ and $+\infty$ if $d\neq\mu+\delta$. and, again, as minimum is continuous by the contraction principle the LDP with a rate-function given in Eq. (18) and the scaling of $E(D_{\mathrm{AB}}^n)$ follows from similar considerations. □

Theorem 2 effectively says that in SRGRAND the algorithm terminates with a correct decoding so long as the number of queries made before identifying an element of the codebook is less than $|\mathbb{A}|^{n(1-R-\varepsilon)}$ for some $\varepsilon>0$. If more queries than that are made, the conditional ML decoding will be erroneous. SRGRAND queries until it identifies the true noise or until an erroneous identification, whichever comes first. In this realization of SRGRANDAB, querying is abandoned for noise sequences beyond the typical set of the average number of potentially noise impacted symbols, curtailing complexity.

VI. Mathematical Example: Symbol Reliability Binary Symmetric Channel (SR-BSC)

We consider a setting where it is possible to mathematically compare channels with and without knowledge of the symbol reliability information vector S^n, the Symbol Reliability Binary Symmetric Channel (SR-BSC). For the SR-BSC, we assume that each transmitted symbol is potentially impacted independently by noise with probability $p_S(1) = q \in [0,1]$. Code-book and noise symbols take values in a binary alphabet $\mathbb{A} = \{0,1\}$, $\oplus$ is addition in $\mathbb{F}_2$, and thus 0 represents the no-noise character. Given a symbol has been potentially noise-impacted, the conditional probability that the corresponding bit has been flipped is $p_{N|S}(1|1) = p \in [0,1]$, $p_{N|S}(0|1) = 1-p$ and $p_{N|S}(0|0) = 1$. The overall bit-flip probability of the SR-BSC is thus pq. We consider capacity and error exponents, which are properties of ML decoding no matter whether it is identified by the noise-guessing methodology or by brute force, as well as complexity, which is a feature of the noise-guessing approach. From Eq. (5), the capacity of the symbol reliability channel is $C^{\text{Sym. Rel.}}(q,p) = 1 - qh_2(p)$, where $h_2(p) = -(1-p)\log_2(1-p) - p\log_2(p)$ is the binary Shannon entropy. The corresponding hard detection channel is a BSC with probability $P(N=1) = P(N=1|S=1)P(S=1) = pq$ and so the hard detection channel capacity is $C^{\text{Hard}}(q,p) = 1 - h_2(pq)$. As h_2 is concave, $C^{\text{Sym. Rel.}}(q,p) \geq C^{\text{Hard}}(q,p)$ for all q and p, and so the capacity of the channel with symbol reliability information is necessarily higher. Depending on the parametrization, the symbol reliability channel's capacity can be several orders of magnitude larger than the hard detection capacity.

As the symbol reliability information 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, $I^L(l) = -(1-l)\log_2\left(\frac{1-l}{1-q}\right) - l\log_2\left(\frac{l}{q}\right)$, which has the corresponding sCGF

$$\Lambda^L(\alpha) = \log_2\left(1 - q + q2^{\alpha}\right). \tag{19}$$

The rate function for LDP of the rescaled guesswork $\{1/n\log_2 G(N_1^n)\}$ in Eq. (4) is the Legendre-Fenchel transform, $I^N(g) = \sup_\alpha\left(\alpha g - \Lambda^N(\alpha)\right)$, of

$$\Lambda^N(\alpha) = \begin{cases} -\log_2\max(p,1-p) & \text{if } \alpha \leq -1 \\ -p\log_2(p) - (1-p)\log_2(1-p) & \text{if } \alpha = 1 \\ (1+\alpha)\log_2\left(p^{1/(1+\alpha)} + (1-p)^{1/(1+\alpha)}\right) & \text{if } \alpha \in (-1,1)\cup(1,\infty). \end{cases} \tag{20}$$

From Eq. (8), the sCGF for the subordinated guesswork of true noise is $\Lambda^{N^L}(\alpha) = \Lambda^L(\Lambda^N(\alpha))$, where Λ^L

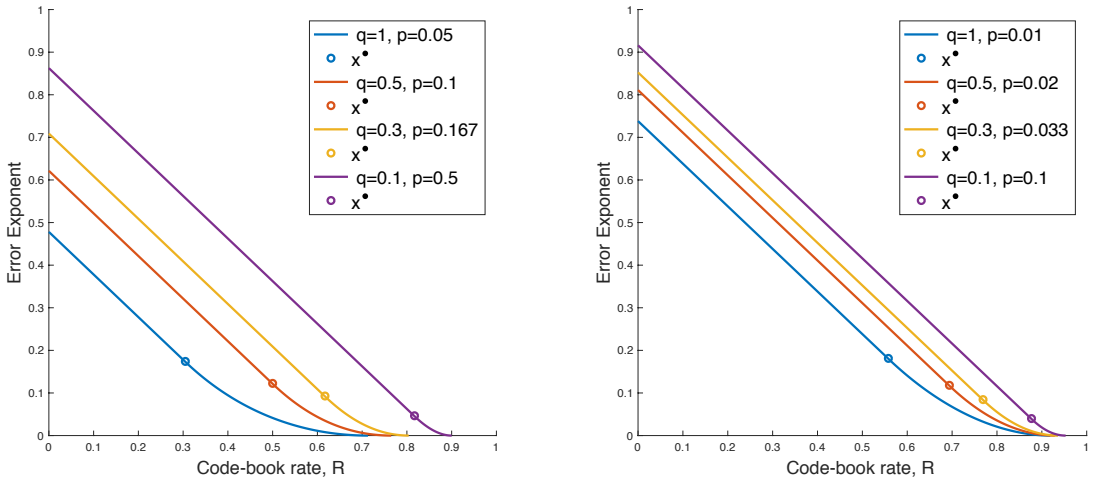

Fig. 3. Block error exponent comparison between BSCs with and without symbol reliability information. pq, the overall bit-flip probability, is constant. Error exponents are plotted as a function of codebook rate R. In the left hand side plot $pq = 0.05$, and on the right-hand side $pq = 0.01$. Circles indicate Gallager's critical rate. The lowest line has $q = 1$ and is the error exponent of the hard detection channel. Higher lines correspond to different (q, p) combinations and have larger error exponents, meaning decoding errors are less likely.

and Λ^N are given by Eq. (19) and Eq. (20), respectively. The exponent of the average complexity required to identify the true noise in the symbol reliability channel is given by $\lim_{n\to\infty} n^{-1}\log_2 E\left(G\left(N_1^{L^n}\right)\right) = \Lambda^{N^L}(1) = \Lambda^L(\Lambda^N(1)) = \log_2\left(1-q+q2^{2\log_2(p^{1/2}+(1-p)^{1/2})}\right)$, while for the hard detection channel it is $\lim_{n\to\infty} n^{-1}\log_2 E(G(N^n)) = 2\log_2\left((pq)^{1/2}+(1-pq)^{1/2}\right)$.

Armed with the sCGFs for the proportion of potentially noise impacted bits and for the rescaled logarithm of the guesswork of potentially noise impacted bits, the asymptotic error exponent given in (15) is readily computable numerically. Recall that, as a function of the codebook rate R, this is the exponent in the decay rate in the likelihood than a conditional ML decoding is in error as the block length increases.

While prefactors are not captured in the asymptotic analysis of Theorems 1 and 2, they allow the following approximations. The conditional ML probability of error is approximately $2^{-n\varepsilon(R)}$ for $R < 1 - qh_2(p)$, which holds true regardless of whether it is identified by SRGRAND or brute force, where the expression for $\varepsilon(R)$ can be found in Eq. (15). For SRGRAND decoding, our measure of complexity is the average number of guesses per bit per decoding, approximately $2^{n\min(1-R,\Lambda^L(H_{1/2}))}/n$. For comparison, we define the complexity of the computation of the ML decoding in Eq. (2) by the method described in [60] to be the number of conditional probabilities that must be computed per bit before rank ordering and determining the most likely codebook element, equal to $2^{n\min(R,1-R)}/n$, where we are equating the work performed in one noise guess, which amounts to checking if a string is an element of the codebook, with the computation of one conditional probability.

For two values of block size, $n = 100$ and $n = 1000$, and (q, p) pairs such that pq is constant and so comparable with the hard detection channel, Fig. 4 plots the approximate error probabilities and complexity

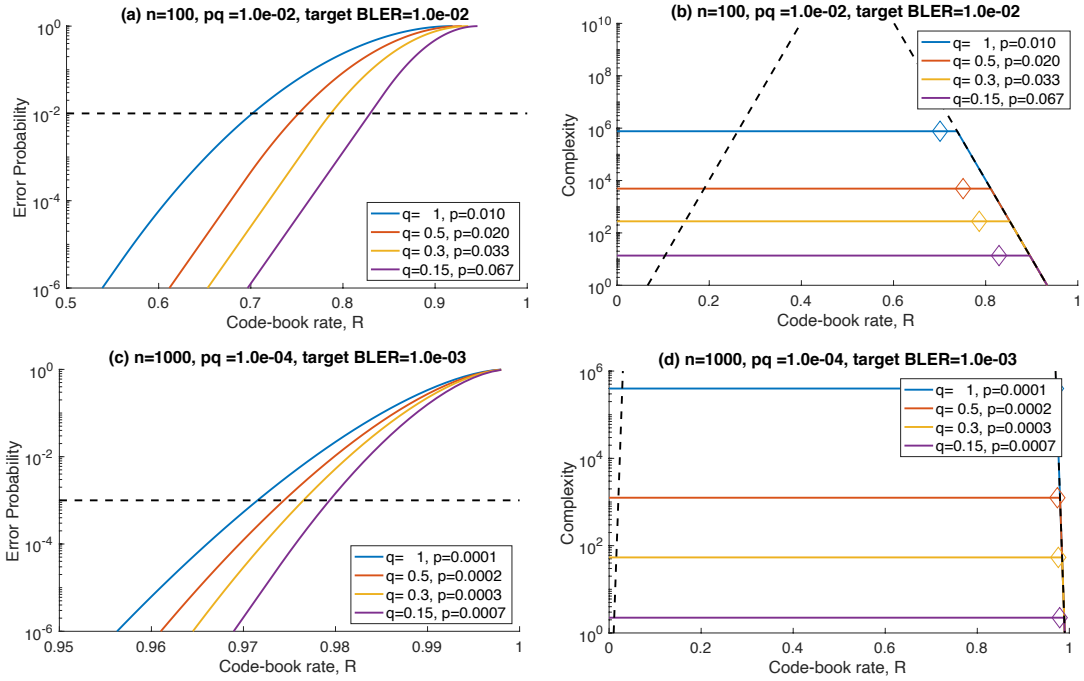

Fig. 4. Approximate block error probability and complexity for BSCs with and without symbol reliability information. The BSC without symbol reliability corresponds to the SR-BSC with $q = 1$, and the overall bit-flip probability, pq, is constant. (a-b) Show results for $n = 100$, $pq = 10^{-2}$ and a target block error of 10^{-2}. In (a), the horizontal dashed line is the target block error and approximate block error probabilities are shown as a function of codebook rate, R, for a selection of (q, p) pairs. (b) shows the approximate complexity, in terms of average number of guesses per-bit to identification of a codebook element, which decreases with q, even though p is increasing. The dashed black line gives complexity for the approach of [60], Diamonds indicate the rate above which the target block error rate would be exceeded, while the inflection point occurs at cut-off rate. (c-d) show corresponding results for $n = 1000$, $pq = 10^{-4}$ and a target block error of 10^{-3}.

as a function of codebook rate. The upper panels show the error probabilities with a target block error rate indicated by the dashed horizontal line. The provision of symbol reliability information greatly improves the block error probability, even though in this comparison the conditional probability of a bit flip given symbol reliability information increases as the symbol reliability probability decreases.

The lower two panels show the approximate complexity. The dashed line gives the approximate complexity for the approach in [60], which grows exponentially in R, when computing a conditional probability for every codeword. In contrast, the complexity of the SRGRAND approach is initially flat. As the rate, R, increases, eventually the SRGRAND complexity drops, as encountering an erroneous element of the codebook clips the long guessing tail of true noise. The diamonds indicate the rate above which the target block error rate would be violated.

VII. Empirical Performance Evaluation

A distinctive aspect of the GRAND approach is that it is readily implemented and can be used with any block code construction. While the theoretical results in Section VI are for uniform-at-random codebooks, in practice nearly all error correction codebooks are linear in a finite field with k input bits and n coded

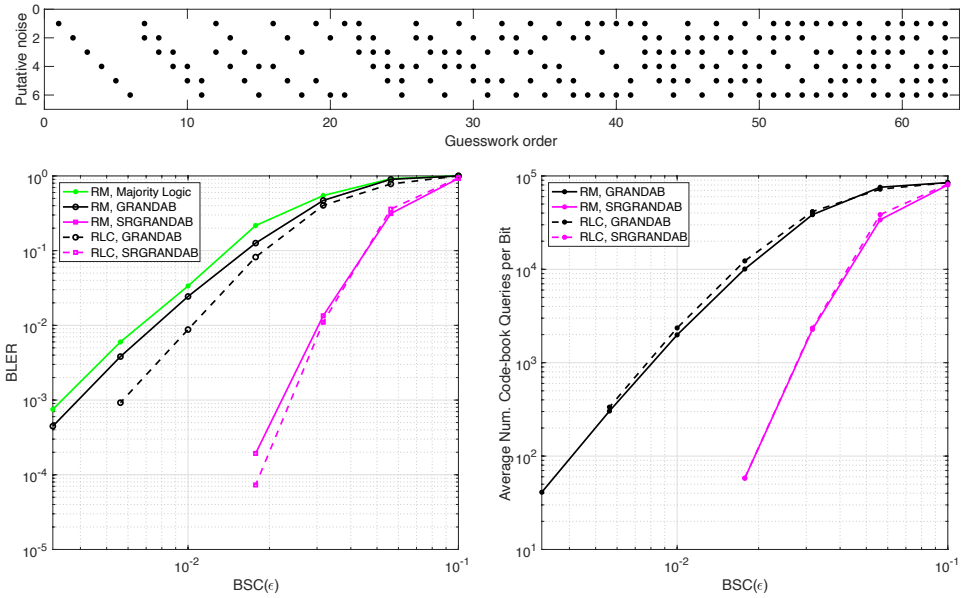

Fig. 5. Performance evaluation with RM and RLC $[128,99]$, rate 0.77, codes in an SR-BSC. Upper panel illustrates GRAND and SRGRAND guessing order on an $n=7$ code, where each column is a putative noise sequence with a dot indicating a 1, and sequences are queried in order from left to right. Left panel gives BLER performance for majority logic decoding, GRANDAB and SRGRANDAB decoding of a RM code as well as GRANDAB and SRGRANDAB decoding of RLCs. Lower right panel gives average number of codebook queries per bit per decoding for GRANDAB and SRGRANDAB.

bits. Associated with each code is a check matrix $H^{n\times n-k}\in\{0,1\}^{n\times n-k}$ and to test if a string, y^n, is in the codebook a single matrix multiplication and comparison, $H^{n\times n-k}(y^n)^T \stackrel{?}{=} (0^{n-k})^T$, suffice, in the appropriate field. Here we compare the decoding performance of GRANDAB, SRGRANDAB for four types of binary linear codes.

A. The Symbol Reliability Binary Symmetric Channel

In the context of the SR-BSC introduced in Section VI, when the unconditional bit flip probability is ε, we set the probability that a bit is marked as unreliable to be $q=\sqrt{\varepsilon}$ and the bit flip probability conditioned on unreliability to be $p=\sqrt{\varepsilon}$. For hard detection GRANDAB, putative noise strings are queried in order of Hamming weight. Within each set of strings with the same Hamming weight, the ordering is arbitrary and we do so in the order illustrated in Fig. 5, first panel. For SRGRANDAB, we assume that the channel state is known and use the same search pattern, but confined to querying only bits for which the channel state was marked as unreliable for any given communication. For GRANDAB, we set the abandonment threshold to check for up to four bit-flips. For SRGRANDAB, we allow the same number of codebook queries as GRANDAB before abandoning and reporting a decoding error.

RM codes, which only exist for some $[n,k]$ pairs, are broadly used in wireless communications and have a well-established hard detection decoder, majority logic decoding [12]. Fig. 5 reports Block Error Rates (BLER) as a function of the bit flip probability ε for a rate 0.77, $[128,99]$, RM code. As majority

logic decoding is tailored to a BSC and is known to be accurate in that setting, its performance is only slightly degraded from the ML BLER that GRANDAB provides. The provision of reliability information to SRGRANDAB gives it a distinct advantage, resulting in significantly enhanced BLER. The right panel reports the average number of codebook queries per received bit that GRANDAB and SRGRANDAB make. As each query solely requires a matrix multiplication by a sparse vector, for typical target BLER of 10^{-2} or lower, the complexity requirements of GRANDAB and SRGRANDAB are modest.

Since the 1960s, RLCs have been known to be capacity-achieving if twined with ML decoding [4] with the same error exponents as those for uniform-at-random codebooks [61]. Those results hinge on a proof that at high rates the average RLC is a good one. The lack of an efficient decoder that can accurately decode any linear, high-rate codebook has meant, however, that this avenue is little explored. Here we consider the application of GRANDAB and SRGRANDAB for decoding RLCs. For any $[n,k]$ pair we can construct systematic binary RLCs by making a random generator matrix $\left[I^{k\times k}|C^{k\times n-k}\right]$, where $I^{k\times k}$ is the identity matrix and the entries of the random check matrix $C^{k\times n-k}$ are independent Bernoulli 1/2 random variables. To check if y^n is a member of the codebook, one can test if $y^k C^{k\times n-k} \stackrel{?}{=} (y_{k+1},\ldots,y_n)$, obviating the need for the receiver to determine the associated check matrix. Consistently with theoretical results, in an empirical evaluation codes are re-randomized after each use. In practice, the sender and receiver could share a seed for the random number generator from which the check matrix is produced.

Fig. 5 also reports the BLER and complexity performance of GRANDAB and SRGRANDAB for [128,99] RLCs, so that the results are directly comparable to those for RM codes. With hard detection ML decoding by GRANDAB, it can be seen that RLCs slightly outperform RM codes, leading to better BLER and comparable decoding complexity. This result is potentially surprising as the re-randomization in the RLC would lead one to suspect that some codes are poor performers, but is consistent with theory that says that RLCs are typically good. The provision of symbol reliability information changes matters and SRGRANDAB gets equally good performance from both RM and RLCs. The use of RLCs, which necessitates a universal decoder, holds appeal as changing codebooks may provide enhanced security, and our results suggest there is no loss in terms of error performance in using them.

For a rate 0.83, BCH $[127,106]$ Fig. 6 reports BLER as a function of ε, as well as RLCs of the same rate. The results mirror those found for RM codes, where the dedicated hard detection decoder provides similar performance to the universal GRANDAB and the provision of symbol reliability information leads SRGRANDAB to significantly outperform both. As with RM codes, RLCs, which can only be efficiently

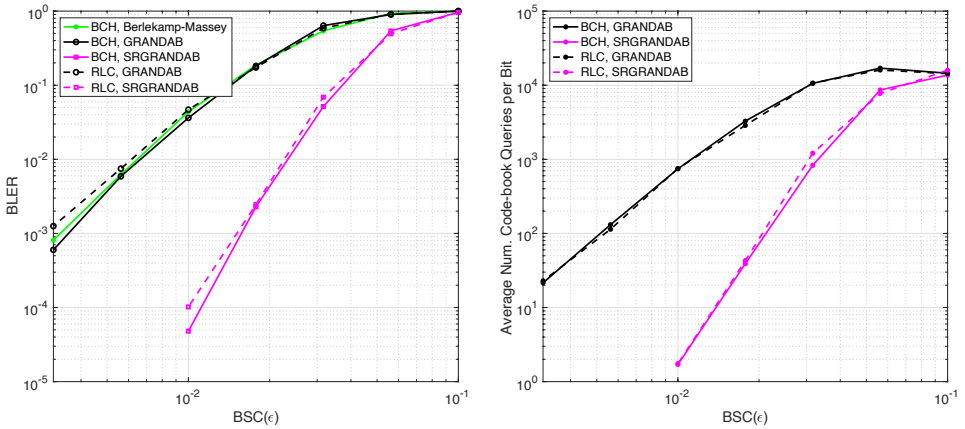

Fig. 6. Performance with BCH and RLC $[127,106]$, rate 0.83, codes in an SR-BSC. Guessing order as in the upper panel of Fig. 5. Left panel gives BLER performance for BM, GRANDAB and SRGRANDAB decoding of a BCH code, as well as GRANDAB and SRGRANDAB decoding of RLCs. Right panel gives average number of codebook queries per bit per decoding for GRANDAB and SRGRANDAB.

decoded with the GRAND approach, lead to similar BLERs as the BCH code with essentially identical complexity for both. The latter is not surprising as the complexity of GRANDAB and SRGRANDAB is largely dominated by properties of the noise rather than those of the codebook. These results suggest that for BLER performance of moderate-redundancy codes, the accuracy of the decoding mechanism is more important than the codebook structure, opening up a rich palette of code sizes and rates for URLLC in a single algorithmic instantiation.

B. Quantizing Soft Information to Create Symbol Reliability Information

The mathematical analysis assumes that the mask provided to the decoder, s^n, is correct, with symbols accurately tagged as reliable or unreliable. In practice, that requires binary quantization of soft information. Should quantization result in a symbol being reliable when it is not, that would necessarily result in an erroneous decoding or abandonment. Here we illustrate simple means by which mask creation can be achieved such that the frequency of provision of erroneous masks, the Mask Error Rate (MER), does not dominate the BLER. The masking rule is a function of the SNR, the length and redundancy of the code.

Consider an AWGN with noise variance is σ^2 and a transmitter-receiver pair employing BPSK with transmitted the binary symbols corresponding to ± 1. We wish to identify a threshold, τ, such that if the absolute value of a received signal is beyond τ it is likely to be reliable. Given τ, the probability an individual bit is erroneously labeled as reliable when it is incorrect is $P(\sigma\mathcal{N} > 1+\tau)$, where $\mathcal{N}$ is a Gaussian with mean zero and variance one. Thus, for a code of length n we have that the likelihood one or more bits are erroneously marked as reliable, resulting in a mask error, is $\text{MER} = 1 - P(\sigma\mathcal{N} \leq 1+\tau)^n$. Hence, by setting a target MER as a function of the code length and SNR such that the MER will not dominate the BLER, the receiver determines the static signal threshold by $\tau = \sigma F_{\mathcal{N}}^{-1}((1-\text{MER})^{1/n}) - 1$,

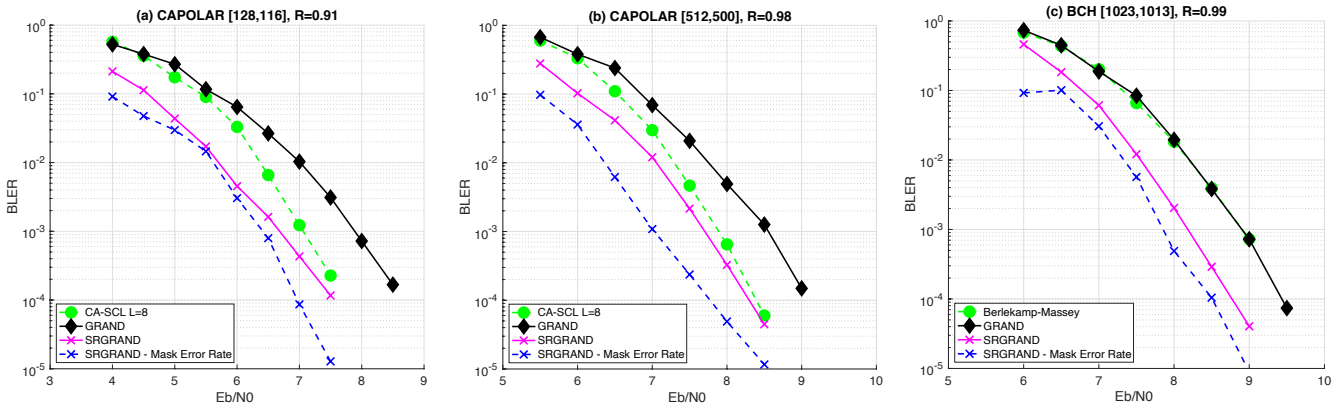

Fig. 7. Performance evaluation in an AWGN with a threshold mask for CA-Polar [128,116], CA-Polar [512,500], and BCH [1023,1013] codes. Green lines correspond to full soft detection CA-SCL decoding of CA-Polar codes with a list size of eight, and hard detection BM decoding for the BCH code. Black lines correspond to ML GRANDAB decoding, magenta to SRGRAND, with blue lines indicating the contribution to SGRAND's BLER that comes from mask errors due to erroneous quantization of soft information.

where $F_{\mathcal{N}}^{-1}$ is the inverse of a Normal distribution. In addition, we set the mask so it always includes the $n-k$ least reliable bits, which allows SRGRAND to do a small amount of corrective work if necessary. In particular, that avoids circumstances where the reception is indicated to be error free, but the received demodulated signal is not in the codebook.

Figure 7 shows results of this system for two CA-Polar codes, [128,116] and [512,500], and a BCH [1023,1013] code. Green lines correspond to full soft decoding with CA-SCL of CA-Polar codes [62] using a list-size of 8 [63], [64], [65], and BM decoding of the BCH code. Black lines indicate optimal ML decoding with hard detection GRAND. The blue lines indicate the contribution to BLER that comes from the masks being in error, MER, where static target mask error rates are determined in advance and used to identify the marking threshold τ. The magenta lines report the overall SRGRAND BLER, inclusive of the MER. At a BLER of 10^{-3}, these results demonstrate a BLER gain of 0.75 to 1dB can be obtained with SRGRAND over ML hard detection decoding, irrespective of code-length.

VIII. Discussion

We have introduced SRGRAND and SRGRANDAB, two noise-centric decoding algorithms using symbol reliability information. By using the symbol reliability information to mask symbols that are reliable and guessing noise only on unreliable symbols, these algorithms can realize higher rates, with lower error probabilities, and less complexity, than without symbol reliability information.

All of the GRAND algorithms are suitable for use with any codebook so long as testing membership of the codebook for a string of symbols is efficient. For linear codes, such testing requires only a matrix multiplication over a finite field. CRC codes, CA-Polar Codes and RLCs are all linear. Moreover,

guesswork orders are known to be robust to mismatch [66], and so decoding precision should not be sensitive to minor imprecision in the channel noise model.

We empirically compared SRGRAND and SRGRANDAB with the well established majority logic decoding of RM codes and BM decoding of a BCH code. The provision of symbol reliability information to SRGRANDAB results in substantially better performance. As the algorithms are universal, they enable us to empirically consider decoding RLCs, which is little explored outside of theory. The BLER performance is comparable with the highly structured RM and BCH codes of the same rate. This opens the possibility of using SRGRANDAB for security, based on a principle of having the sender and receiver use a distinct linear code drawn using a cryptographically secure random number generator for each transmission.

While we presented results for one SRGRANDAB abandonment rule that reduces average algorithmic complexity without sacrificing channel capacity, others are possible and, indeed, can be used in combination. Here we mention two more. The first is a natural extension to the rule of abandoning guessing when coverage of the typical set for the average number of potentially noise impacted symbols. In the symbol reliability model, the specific number of potentially noise-impacted symbols, L^n, for each received transmission, Y^n, is known to the algorithm and querying is abandoned after $|\mathbb{A}|^{L^n(H+\delta)}$ guesses. Analysis of the impact of this rule on error exponents and complexity follows the same line of argument as presented in the paper, though the resulting expressions are less elegant. A distinct alternative is not to guess at all if too many symbols are reported to be potentially noise impacted; i.e. if $L^n > n(\mu^L + \delta)$. It is straight forward to show this rule does not impact capacity, but an analysis of complexity, which would now be conditional on $L^n \leq n(\mu^L + \delta)$, would not follow immediately from the large deviation arguments presented here. The analysis in this paper for codes of fixed length could, however, be readily extended to decoding with symbol reliability information for variable length codes [67], [68] and rateless codes.

SRGRAND avails of symbol reliability information, which is the most succinct form of soft information, and lends itself to both mathematical analysis and implementation in hardware, seeing a 0.5 to 0.75dB gain over hard detection GRAND. A natural question is how to use more fine-grained soft information in a GRAND algorithm, what the additional algorithmic complexity would be, and what performance gains would be available. By creating a bespoke noise effect query order for each reception, it is possible to use one real-valued piece of soft information per bit to identify soft-detection ML decodings [69]. Although the resulting algorithm does not lend itself to theoretical determination of error exponents or to efficient implementation in hardware due to the need for dynamic memory, a software implementation enables the

empirical evaluation of a bound on the achievable performance for a given code. A heuristic algorithm that uses $\log_2(n)$ bits of soft information per received bit has been reported that appears to empirically approximate the performance available from full soft information with a simpler algorithm [70], [71]. Again, its construction does not lend itself to mathematical identification of error exponents, but it is more suitable for implementation in hardware and an architecture for it has been proposed [72], albeit one that is significantly more complex in terms of energy and area than is the case for GRAND [26], [27] or would be for SRGRAND. The question of whether SRGRAND could be augmented to avail of more finely quantized soft information while retaining the simplicity of its operation remains outstanding.

The GRAND algorithms can themselves provide, in addition to a decoding, soft information through the number of noise queries. A lower number of guesses corresponds to a higher likelihood of correct decoding. Such soft information can be of use, for example, for component codes in a concatenated code or Turbo code [73], [74], [75], [76]. Thus one may envisage using the information on decoding reliability of SRGRAND and SRGRANDAB in a manner akin to the reliability information provided by the Soft-Output Viterbi Algorithm [73], [74], [77], [75], [78], by the operation of Turbo decoding [76], [79], [80], [81], [82], [83], [84], by the syndrome information used in Ordered Statistics Decoding (OSD) [43], [85], [86], or other soft-input, soft-output schemes [87], [88], [68], [89]. In general, we can envisage in future work systems that meld equalization and decoding as in [90] or soft information originating from other decoding processes, [91], [92], [93], [94], [95].

REFERENCES

[1] K. R. Duffy and M. Médard, "Guessing random additive noise decoding with soft detection symbol reliability information," in *IEEE Int. Symp. on Inf. Theory*, 2019.

[2] K. R. Duffy, A. Solomon, K. M. Konwar, and M. Médard, "5G NR CA-Polar maximum likelihood decoding by GRAND," in *Annual Conference on Information Sciences and Systems*, 2020.

[3] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

[4] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: John Wiley & Sons, Inc., 1968.

[5] E. Berlekamp, R. McEliece, and H. Van Tilborg, "On the inherent intractability of certain coding problems (corresp.)," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 384–386, 1978.

[6] R. J. McEliece, "A public-key cryptosystem based on algebraic," *Deep Space Network Progress Report*, vol. 42-44, pp. 114–116, 1978.

[7] I. Reed, "A class of multiple-error-correcting codes and the decoding scheme," *Transactions of the IRE Professional Group on Information Theory*, vol. 4, no. 4, pp. 38–49, 1954.

[8] D. E. Muller, "Application of boolean algebra to switching circuit design and to error detection," *Transactions of the IRE professional group on electronic computers*, no. 3, pp. 6–12, 1954.

[9] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *Journal of the society for industrial and applied mathematics*, vol. 8, no. 2, pp. 300–304, 1960.

[10] E. Berlekamp, *Algebraic coding theory*. World Scientific, 1968.

[11] J. Massey, "Shift-register synthesis and bch decoding," *IEEE Trans. Inf Theory*, vol. 15, no. 1, pp. 122–127, 1969.

[12] S. Lin and D. J. Costello, *Error control coding: fundamentals and applications*. Pearson-Prentice Hall, 2004.

[13] R. G. Gallager, "Low density parity check codes," 1963.

[14] M. P. Fossorier, M. Mihaljevic, and H. Imai, "Reduced complexity iterative decoding of low-density parity check codes based on belief propagation," *IEEE Trans. Commun.*, vol. 47, no. 5, pp. 673–680, 1999.

[15] K. Niu and K. Chen, "CRC-aided decoding of Polar codes," *IEEE Commun. Letters*, vol. 16, no. 10, pp. 1668–1671, 2012.

[16] I. Tal and A. Vardy, "List decoding of Polar codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2213–2226, 2015.

[17] A. Balatsoukas-Stimming, M. B. Parizi, and A. Burg, "LLR-based successive cancellation list decoding of Polar codes," *IEEE Trans. Signal Process.*, vol. 63, no. 19, pp. 5165–5179, 2015.

[18] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, 2016.

[19] C. She, C. Yang, and T. Q. Quek, "Radio resource management for ultra-reliable and low-latency communications," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 72–78, 2017.

[20] H. Chen, R. Abbas, P. Cheng, M. Shirvanimoghaddam, W. Hardjawana, W. Bao, Y. Li, and B. Vucetic, "Ultra-reliable low latency cellular networks: Use cases, challenges and approaches," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 119–125, 2018.

[21] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai, "A survey on low latency towards 5G: RAN, core network and caching solutions," *IEEE Commun. Surv.*, vol. 20, no. 4, pp. 3098–3130, 2018.

[22] M. Médard, "Is 5 just what comes after 4?" *Nature Electronics*, vol. 3, no. 1, pp. 2–4, 2020.

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

[24] ——, "Capacity-achieving guessing random additive noise decoding (GRAND)," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4023–4040, 2019.

[25] W. An, M. Médard, and K. R. Duffy, "Keep the bursts and ditch the interleavers," in *IEEE GLOBECOM*, 2020.

[26] S. M. Abbas, T. Tonnellier, F. Ercan, and W. J. Gross, "High-throughput VLSI architecture for GRAND," in *IEEE SiPS*, 2020.

[27] A. Riaz, V. Bansal, A. Solomon, W. An, Q. Liu, K. Galligan, K. R. Duffy, M. Médard, and R. T. Yazicigil, "Multi-code multi-rate universal maximum likelihood decoder using GRAND," in *IEEE ESSCIRC*, 2021.

[28] S. M. Abbas, M. Jalaleddine, and W. J. Gross, "High-throughput VLSI architecture for GRAND Markov order," in *IEEE SiPS*, 2021.

[29] A. Goldsmith, *Wireless communications*. Cambridge University Press, 2005.

[30] J. T. Coffey and R. M. Goodman, "Any code of which we cannot think is good," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1453–1461, 1990.

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

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

[33] 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.

[34] L. D'Alessio, L. Liu, K. R. Duffy, Y. C. Eldar, M. Médard, and M. Babadi, "A coding theory perspective on multiplexed molecular profiling of biological tissues," in *IEEE Int. Symp. Inf. Theory Appl.*, 2020.

[35] E. Prange, "The use of information sets in decoding cyclic codes," *IRE Trans. Inf. Theory*, vol. 8, no. 5, pp. 5–9, 1962.

[36] J. Stern, "A method for finding codewords of small weight," in *Int. Colloquium Coding Theory Appl.* Springer, 1988, pp. 106–113.

[37] P. J. Lee and E. F. Brickell, "An observation on the security of McEliece's public-key cryptosystem," in *Workshop on the Theory and Application of of Cryptographic Techniques*. Springer, 1988, pp. 275–280.

[38] J. S. Leon, "A probabilistic algorithm for computing minimum weights of large error-correcting codes," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 1354–1359, 1988.

[39] C. Peters, "Information-set decoding for linear codes over $\mathbb{F}_q$," in *International Workshop on Post-Quantum Cryptography*. Springer, 2010, pp. 81–94.

[40] D. J. Bernstein, T. Lange, and C. Peters, "Smaller decoding exponents: ball-collision decoding," in *Annual Cryptology Conference*. Springer, 2011, pp. 743–760.

[41] A. Becker, A. Joux, A. May, and A. Meurer, "Decoding random binary linear codes in $2^{n/20}$: How 1+ 1= 0 improves information set decoding," in *EUROCRYPT*. Springer, 2012, pp. 520–536.

[42] B. Dorsch, "A decoding algorithm for binary block codes and J-ary output channels (corresp.)," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 391–394, 1974.

[43] M. P. C. Fossorier and S. Lin, "Soft-decision decoding of linear block codes based on ordered statistics," *IEEE Trans. Inf. Theory*, vol. 41, no. 5, pp. 1379–1396, 1995.

[44] D. Gazelle and J. Snyders, "Reliability-based code-search algorithms for maximum-likelihood decoding of block codes," *IEEE Tran. Inf. Theory*, vol. 43, no. 1, pp. 239–249, 1997.

[45] A. Valembois and M. Fossorier, "Box and match techniques applied to soft-decision decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 796–810, 2004.

[46] Y. Wu and C. N. Hadjicostis, "Soft-decision decoding of linear block codes using preprocessing and diversification," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 378–393, 2006.

[47] ——, "Soft-decision decoding using ordered recodings on the most reliable basis," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 829–836, 2007.

[48] M. Baldi, N. Maturo, E. Paolini, and F. Chiaraluce, "On the use of ordered statistics decoders for low-density parity-check codes in space telecommand links," *EURASIP J Wirel. Comm.*, no. 1, p. 272, 2016.

[49] Q. Guo, T. Johansson, E. Mårtensson, and P. S. Wagner, "Some cryptanalytic and coding-theoretic applications of a soft Stern algorithm," *Adv. Math. Commun.*, vol. 13, no. 4, 2019.

[50] C. Choi and J. Jeong, "Fast and scalable soft decision decoding of linear block codes," *IEEE Commun. Lett.*, vol. 23, no. 10, pp. 1753–1756, 2019.

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

[52] S. R. S. Varadhan, "Large deviations," *Ann. Appl. Probab.*, vol. 36, no. 2, pp. 397 – 419, 2008.

[53] A. Weiss and A. Shwartz, *Large Deviations For Performance Analysis: Queues, Communication and Computing*. Routledge, 2019.

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

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

[56] A. Beirami, R. Calderbank, M. Christiansen, K. R. Duffy, and M. Médard, "A characterization of guesswork on swiftly tilting curves," *IEEE Trans. Inform. Theory*, vol. 65, no. 5, pp. 2850–2871, 2019.

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

[58] 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.

[59] 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.

[60] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate (corresp.)," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284–287, 1974.

[61] Y. Domb, R. Zamir, and M. Feder, "The random coding bound is tight for the average linear code or lattice," *IEEE Trans. Inform. Theory*, vol. 62, no. 1, pp. 121–130, 2016.

[62] A. Cassagne *et al.*, "Aff3ct: A fast forward error correction toolbox!" *Elsevier SoftwareX*, vol. 10, p. 100345, Oct. 2019.

[63] H. Zhang, R. Li, J. Wang, S. Dai, G. Zhang, Y. Chen, H. Luo, and J. Wang, "Parity-check polar coding for 5G and beyond," in *IEEE ICC*, 2018, pp. 1–7.

[64] P. Chen, B. Bai, Z. Ren, J. Wang, and S. Sun, "Hash-polar codes with application to 5G," *IEEE Access*, vol. 7, pp. 12 441–12 455, 2019.

[65] C. Kestel, L. Johannsen, O. Griebel, J. Jimenez, T. Vogt, T. Lehnigk-Emden, and N. Wehn, "A 506Gbit/s polar successive cancellation list decoder with CRC," in *IEEE PIMRC*, 2020.

[66] R. Sundaresan, "Guessing based on length functions," in *IEEE Int. Symp. Inf. Theory*, 2007.

[67] J. Wen and J. Villasenor, "Soft-input soft-output decoding of variable length codes," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 689–692, 2002.

[68] C. Weidmann, "Reduced-complexity soft-in-soft-out decoding of variable-length codes," in *IEEE Int. Symp. Inf. Theory*, 2003, pp. 201–201.

[69] A. Solomon, K. R. Duffy, and M. Médard, "Soft maximum likelihood decoding using GRAND," in *IEEE Int. Commun. Conf.*, 2020.

[70] K. R. Duffy, "Ordered reliability bits guessing random additive noise decoding," in *IEEE IEEE Int Conf Acoust Speech Signal Process*, 2021.

[71] W. An, M. Médard, and K. R. Duffy, "CRC codes as error correcting codes," in *IEEE ICC*, 2021.

[72] S. M. Abbas, T. Tonnellier, F. Ercan, M. Jalaleddine, and W. J. Gross, "High-throughput VLSI architecture for soft-decision decoding with ORBGRAND," in *IEEE IEEE Int Conf Acoust Speech Signal Process*, 2021.

[73] J. Hagenauer and P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications," in *IEEE Global Tele. Conf. Exhib.*, vol. 3, 1989, pp. 1680–1686.

[74] C. Berrou, P. Adde, E. Angui, and S. Faudeil, "A low complexity soft-output Viterbi decoder architecture," in *IEEE Int. Conf. Commun.*, vol. 2, 1993, pp. 737–740.

[75] M. P. C. Fossorier, F. Burkert, S. Lin, and J. Hagenauer, "On the equivalence between SOVA and max-log-MAP decodings," *IEEE Commun. Lett.*, vol. 2, no. 5, pp. 137–139, 1998.

[76] J. P. Woodard and L. Hanzo, "Comparative study of Turbo decoding techniques: an overview," *IEEE Trans. Veh. Technol.*, vol. 49, no. 6, pp. 2208–2233, 2000.

[77] L. Papke, P. Robertson, and E. Villebrun, "Improved decoding with the SOVA in a parallel concatenated (Turbo-code) scheme," in *IEEE Int. Conf. Commun.*, vol. 1, 1996, pp. 102–106.

[78] V. Guruswami and M. Sudan, "Decoding concatenated codes using soft information," in *17th IEEE CCC*, 2002, pp. 148–157.

[79] G. Colavolpe, G. Ferrari, and R. Raheli, "Noncoherent iterative (Turbo) decoding," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1488–1498, 2000.

[80] J. Hagenauer and L. Papke, "Decoding "Turbo"-codes with the soft output Viterbi algorithm (SOVA)," in *IEEE Int. Symp. Inf. Theory*, 1994, p. 164.

[81] K. Kim, J. W. Choi, A. C. Singer, and K. Kim, "A new adaptive Turbo equalizer with soft information classification," in *IEEE Int Conf Acoust Speech Signal Process*, 2010.

[82] X. Wang and H. V. Poor, "Iterative (Turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, 1999.

[83] S. Song, A. C. Singer, and K.-M. Sung, "Soft input channel estimation for Turbo equalization," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2885–2894, 2004.

[84] R. Johannesson and K. S. Zigangirov, *Fundamentals of convolutional coding*. John Wiley & Sons, 2015.

[85] Y. Kaji and D. Ikegami, "Decoding linear block codes using the ordered-statistics and the MLD techniques," in *IEEE Int. Symp. Inf. Theory*, 2002.

[86] P. A. Martin, A. Valembois, M. P. C. Fossorier, and D. P. Taylor, "Reduced complexity soft-input soft-output "box and match" decoding," in *IEEE Int. Symp. Inf. Theory*, 2003, pp. 202–202.

[87] T. K. Moon and J. H. Gunther, "Decoding by iterative detection (decidet): Soft-in/soft-out decoding of arbitrary linear block codes over arbitrary finite fields," in *Asilomar Conf. on Sig. Sys. Comp.*, 2011, pp. 674–681.

[88] S. Fujimoto, T. Kusaka, and S. Ueda, "A study on soft-out of soft-in/soft-out decoding algorithms for binary linear codes," in *Int. Symp. Inf. Theory and Appl.*, 2016, pp. 300–304.

[89] C. Studer and H. Bolcskei, "Soft—input soft—output single tree-search sphere decoding," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4827–4842, 2010.

[90] G. Ungerboeck, "Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems," *IEEE Trans. Commun.*, vol. 22, no. 5, pp. 624–636, 1974.

[91] G. D. Forney, "Concatenated codes," Research Laboratory for Electronics, Massachusetts Institute of Technology, Tech. Rep., 1965.

[92] S. Benedetto and G. Montorsi, "Soft-input soft-output building blocks to construct and iteratively decode code networks," in *IEEE Int. Symp. Inf. Theory*, 1997, p. 7.

[93] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "A soft-input soft-output app module for iterative decoding of concatenated codes," *IEEE Commun. Letters*, vol. 1, no. 1, pp. 22–24, 1997.

[94] R. Sivasankaran and S. W. McLaughlin, "Performance of soft-in soft-out stack decoding," in *IEEE Int. Symp. Inf. Theory*, 2001, p. 234.

[95] J. Bellorado, A. Kavcic, M. Marrow, and L. Ping, "Low-complexity soft-decoding algorithms for Reed—Solomon codes -—Part II: Soft-input soft-output iterative decoding," *IEEE Tran. Inf. Theory*, vol. 56, no. 3, pp. 960–967, 2010.