Nonorthogonal Multiple Access With Guessing Random Additive Noise Decoding-Aided Macrosymbol (GRAND-AM)

Kathleen Yang, *Graduate Student Member, IEEE*, Muriel Médard, *Fellow, IEEE*, and Ken R. Duffy, *Senior Member, IEEE*

***Abstract*—We propose guessing random additive noise decoding-aided macrosymbols (GRAND-AMs) as a nonorthogonal multiple access (NOMA) method that can detect, error correct, and decode multiple users with imperfect channel estimation, asynchronous transmission, and interference, which are all topics of concern for Internet of Things. GRAND-AM is a NOMA method that uses both joint multiuser detection and joint error correction decoding to handle multiple access interference (MAI). For the joint multiuser detector, we introduce the concept of a macrosymbol, which is constructed from the combination of all user symbols. For the error correction decoding component, we introduce multiple access channel (MAC) codes, which are codes that are used to split the channel rate between users and correct errors due to MAI. In this scheme, each user has their information bits encoded with independent MAC codes. We use a soft detection variant of GRAND, an efficient and practical decoding method that inverts noise effect sequences from a sequence of symbols to arrive at a codeword, to correct a sequence of macrosymbols, ensuring that all user codebooks are simultaneously satisfied. The joint detection and decoding of GRAND-AM can outperform time division multiple access (TDMA) by 10 dB with perfect channel estimation, and by 6 dB with imperfect channel estimation. Considering a more complete communication chain, when additional forward error correction is used along with the MAC code, the GRAND-AM method performs similarly to a same rate low-density parity-check-coded TDMA system.**

***Index Terms*—Error correction, joint detection, multiple access, multiuser detection, nonorthogonal multiple access (NOMA).**

Manuscript received 14 March 2024; revised 28 May 2024; accepted 11 June 2024. Date of publication 24 June 2024; date of current version 23 August 2024. This work was supported in part by the National Science Foundation under Grant EECS-2128555, and in part by Defense Advanced Research Projects Agency under Grant HR00112120008. This article was presented in part at International Conference on Communications 2023 [DOI: 10.1109/ICC45041.2023.10278938] and Military Communications Conference 2023 [DOI: 10.1109/MILCOM58377.2023.10356355]. We extend our prior work by considering the addition of nonidealities, such as channel estimation error and asynchronicity to the detection, as well as comparing the performance of GRAND-AM versus an orthogonal multiple access method such as TDMA. *(Corresponding author: Kathleen Yang.)*

Kathleen Yang and Muriel Médard are with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: klyang@mit.edu; medard@mit.edu).

Ken R. Duffy is with the Electrical and Computer Engineering Department, Northeastern University, Boston, MA 02115 USA (e-mail: k.duffy@northeastern.edu).

Digital Object Identifier 10.1109/JIOT.2024.3418211

I. Introduction

THE GROWING number of users from sources such as machine-type communications (MTCs) for the Internet of Things (IoT) and global connectivity as well as increasing data rates [3], [4], [5], [6] has led to a shift away from traditionally used orthogonal multiple access (OMA) techniques to nonorthogonal multiple access (NOMA) techniques. In OMA, multiple users share a channel resource such that there is no overlap in access, such as in the case for time-division multiple access (TDMA) where users access the channel at different assigned times, frequency-division multiple access (FDMA) where users access the channel with different assigned frequencies, or code-division multiple access (CDMA) where users are assigned orthogonal spreading codes [7]. In contrast to OMA methods, NOMA methods allow multiple users share a channel resource simultaneously which increases spectral efficiency and helps address increasing connectivity demands [8], [9].

As an example of this, consider the scenario where there are 30 000 MTC devices connected to the network, with a uniform arrival distribution for access requests over 60 s [10]. For a frame of length 10 ms, there will be on average five devices that attempt to access the network over this period. With a traditionally used OMA method such as TDMA, this will lead to the devices experiencing long delays as only one device can access the network at a time. This is detrimental for IoT, especially when it is used for mission critical and ultrareliable low-latency communication (URLLC) applications. In order to reduce the latency, NOMA methods must be used to simultaneously service the five requests while still achieving target error rates such as 10^{-4} for frame error rates (FERs) for mission critical services [11].

In order for NOMA methods to achieve these desired error rates while handling the multiple users, they must first address the multiple access interference (MAI) of the users, which arises from the overlapping user signals. A combination of both multiuser detection and encoding/decoding methods at the receiver must be considered to achieve these error rates. Some commonly investigated methods that consider both of these aspects are power-domain NOMA (PD-NOMA) and code-domain NOMA (CD-NOMA).

PD-NOMA handles multiple users through power control and interference cancellation methods such as successive

interference cancellation (SIC) [8], [9], [12]. Each user is assigned a power or experiences different channel gains, and at the receiver, is treated as an interferer to the other users accessing the same channel resource. SIC takes advantage of the different powered users by detecting the highest powered user first while treating others users as noise, then removing its contribution from the channel output so that the next highest powered user can be detected with the same process. This process iterates through all users until only noise is left in the remaining channel output. While SIC is less complex than an optimal maximum-likelihood (ML) detector, its iterative nature and the treatment of other users as interferers can lead to error propagation if an error occurs when detecting earlier users, as well as asymmetric error rates when users are of similar powers [13], [14], [15].

The other commonly investigated NOMA method, CD-NOMA, handles multiple users through the usage of nonorthogonal codes. In particular, in low-density CDMA or sparse code multiple access (SCMA), which are variants of CD-NOMA, the nonorthogonal codes are sparse in nature, which helps limit the number of overlapping users per chip of the spreading sequence used or channel resources, respectively [9], [16], [17], [18], [19], [20]. Limiting the number of users per chip or resource reduces the MAI that each user experiences. A factor graph can be employed to represent the structure of the error-correcting codes in low-density CDMA and SCMA, which then allows for a message-passing algorithm (MPA) to be used, which results in an iterative, near-optimal detection method [17], [18], [19]. While MPA is near optimal and less complex than a maximum a posteriori (MAP) detector, the low density or sparse codebooks requires a design based on the number of users, as well as the desired load per resource, which would be challenging to orchestrate in a dynamic environment.

PD-NOMA and CD-NOMA techniques can be used to address the required increase in spectral efficiency necessary to support the growing number of users and the problem of achieving target error rates, but there are other important aspects in multiple access channel (MAC) that should be considered for IoT applications. Factors such as asynchronicity for grant-free access such that transmitters can transmit without going through contention-based access, interference handling from sources, such as heterogeneous networks and channel reuse, and the lack of perfect channel estimation must also be addressed [21], [22], [23].

Taking into account these additional requirements, we propose guessing random additive noise decoding-aided macrosymbol (GRAND-AM), which uses joint MUD and joint decoding to handle MAI, and can handle interference and symbol-wise asynchronicity. Our proposed method introduces short error-correcting codes called MAC codes that split the channel rate to handle MAI, as shown in Fig. 1. Furthermore, the joint MUD avoids the error propagation and asymmetry issues that interference cancellation methods such as PD-NOMA face, while the MAC codes do not need to be designed for the expected load and for each user such as in CD-NOMA.

For the detection component of GRAND-AM, we base the MUD on the jointly optimal ML detector for multiple

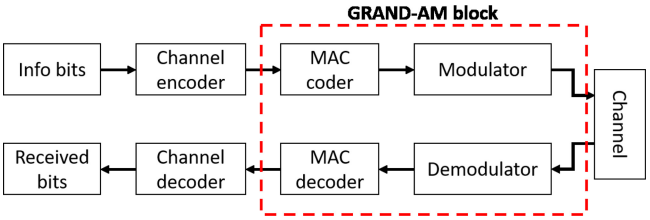

Fig. 1. Block diagram showing how GRAND-AM is incorporated into the communication process. A MAC code is a short error-correcting code used to rate split and correct errors that arise from simultaneous transmission. This is a separate code from the forward error-correcting code.

users, which performs similarly to the individually optimal ML detector in uncoded scenarios [13], [14], [24], [25], [26], [27], [28]. For the joint detector, we introduce the concept of a macrosymbol, which is a symbol that is formed from the combination of all available user symbols during that symbol time with their channel gains applied. An example of a macrosymbol can be seen in Fig. 2. This allows the joint MUD to act as a single-user detector for a macrosymbol, leading to a less complex detector compared to the per user MUD when evaluating log likelihoods.

Then, for the joint decoding component of GRAND-AM, we use a soft information variant of GRAND that can provide near ML decodings for any moderate redundancy code of any structure in order to recover the sequence of macrosymbols [29], [30], [31]. GRAND is suited for decoding the sequence of macrosymbols, as it queries most likely to least likely noise sequences before removing the noise sequence that corresponds with a sequence of symbols contained within the codebook. By inverting the digital noise effect from the sequence of macrosymbols, and ensuring that the sequence of macrosymbols satisfy all user codebooks simultaneously, the joint nature of the decoding process can be preserved. Through the joint MUD and joint decoding process in GRAND-AM, the MAC code can be used to more effectively remove the MAI effects from the users compared to a more conventional per user MUD and decoding process.

This work is organized as follows. In Section II, we introduce the MAC model, with and without an interferer, as well as the assumptions made for the detection and decoding. In Section III, we define the concept of a macrosymbol, and its usage in a joint MUD. In addition, we also discuss the per user MUD to compare the two detectors. For Section IV, we give an overview of the GRAND algorithm used for the joint decoding. We then discuss our results comparing GRAND-AM versus a more conventional individual, per user detection and decoding process and TDMA as an OMA method, as well as show GRAND-AM's performance in the presence of symbol-wise asynchronicity, channel estimation error, and interference in Section V. We show that the usage of joint detection and decoding in GRAND-AM leads to large gains such as 10 dB over TDMA given perfect channel estimation, which can allow GRAND-AM to achieve desired FERs at lower E_b/N_0 compared to a turbo-coded TDMA user. We also consider how GRAND-AM fits into a communication chain with additional forward error correction (FEC) and show that GRAND-AM

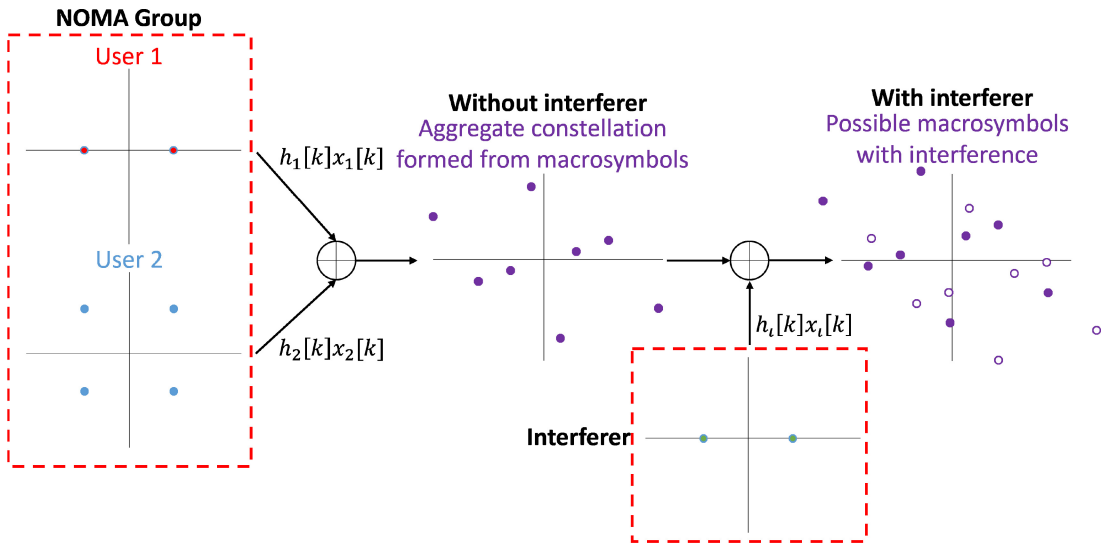

Fig. 2. Aggregate constellation is formed from macrosymbols, which are unique combinations of all users that have transmitted within the same resource block. In this example, there are two users transmitting, which are modulated with BPSK and 4QAM, respectively. Without an interferer, there are only eight valid macrosymbols to detect. With an interferer of size 2 modulation, there are 16 potential macrosymbols to detect, which depend on what the interferer symbol is. Only 8 of these are the true macrosymbols.

can outperform a turbo-coded TDMA user while taking into account overall rates and block length. Finally, we give our conclusions in Section VI.

II. System Model

For the system model, we consider only the GRAND-AM block outlined in Fig. 1. That is, we assume that the bits provided are already encoded with FEC and apply the MAC code upon the channel encoded bits. Consider u users simultaneously accessing the same channel resource, with the possibility that the users are asynchronous, or an interferer is present due co-channel interference, heterogeneous networks, or adversarial interference [23], [32], [33]. Note that this interference is different compared to MAI due to the receiver not attempting to recover the interferer's message.

Each user, $i \in [1, u]$, transmits n_i MAC coded bits, of which k_i bits are information bits. The (n_i, k_i) MAC codes are independent from user to user. The n_i bits for the ith user are modulated with an m_i size discrete modulation, where the modulations are independent from user to user. We denote the symbols associated with the ith user's modulation as being contained within the set $\mathcal{S}_i = \{x_{i,1}, x_{i,2}, \ldots, x_{i,m_i}\}$. Thus, the length of the sequence of symbols of the ith user transmits is $l_i = \lceil n_i / \log_2(m_i) \rceil$. In Fig. 3, an example can be seen where there are $u = 2$ users, and each user transmits $n_i = 8$ MAC coded bits, which are modulated with 4 quadrature amplitude modulation (QAM), leading to $l_i = 4$. When an interferer is present within the system, it is denoted with subscript q, and is modulated with an m_q size discrete modulation with possible symbols $\mathcal{S}_q = \{x_{q,1}, x_{q,2}, \ldots, x_{q,m_q}\}$.

Assuming a rich multipath channel such that the channel gain can be represented with Rayleigh fading, the received signal at time t is

$$y[t] = \sum_{i=1}^{u} h_i[t] x_i[d_i + o_i] + h_q[t] x_q[t] + w[t] \tag{1}$$

where $x_i[d_i + o_i]$ is the ith user's transmitted symbol that belongs to the set $\mathcal{S}_i$ and $d_i \in [1, l_i]$, o_i indicates the offset of the ith user and without loss of generality $o_1 = 0$, $x_q[t]$ is the interferer's transmitted symbol that is randomly sampled from the set $\mathcal{S}_q$, $h_i[t]$ and $h_q[t]$ are the channel gains experienced by the users and the interferer and distributed such that $h_i[t], h_q[t] \overset{iid}{\sim} \mathcal{CN}(0, 1)$, and $w[t]$ is the complex additive white Gaussian noise (AWGN) distributed as $w[t] \overset{iid}{\sim} \mathcal{CN}(0, 1)$. Note that due to the independent MAC codes and modulations, the index symbol t ranges from $t \in [1, \max(l_i + o_i)] \forall i \in [1, u]$. For the ith user, in the case where $t \notin [1 + o_i, l_i + o_i]$, $x_i[t] = 0$, indicating that the ith user has not transmitted for time t. The transmit powers of the users and interferer are defined as $P_i = \mathbb{E}[|x_i[d_i + o_i]|^2]$ and $P_q = \mathbb{E}[|x_q[t]|^2]$, respectively. The absence of an interferer can be represented by setting $P_q = 0$. An example of the asynchronous transmission can be seen in Fig. 3 where there is a symbol offset between the two users of $o_2 - o_1 = 1$. Given this offset, the 1st and 5th received macrosymbols are formed from the $d_1 = 1$ symbol of user 1 and $d_2 = 4$ symbol from user 2, respectively, while the 2nd through 4th macrosymbols are formed from the combination of the $d_1 \in [2, 4]$ symbols of user 1 and $d_2 \in [1, 3]$ symbols of user 2.

For this work, we make the following assumptions: The receiver has knowledge of the users' modulations $\mathcal{S}_i$, the (n_i, k_i) MAC codes used, the symbol offsets o_i, the transmit powers P_i, and an estimate of the channel gains $\hat{h}_i[t]$, which we will shortly define. In addition, if there is an interferer present, and the receiver has knowledge of it, then we assume that the receiver knows the interferer modulation $\mathcal{S}_q$, its power P_q, and has an estimate of its channel gain $\hat{h}_q[t]$. In this case, the receiver is an interferer aware receiver. If the receiver does not have the above knowledge of the interferer, we define it as an interferer ignorant receiver.

We define the estimate of the channel gain such that there is AWGN added to the true channel gain. For generality, we use

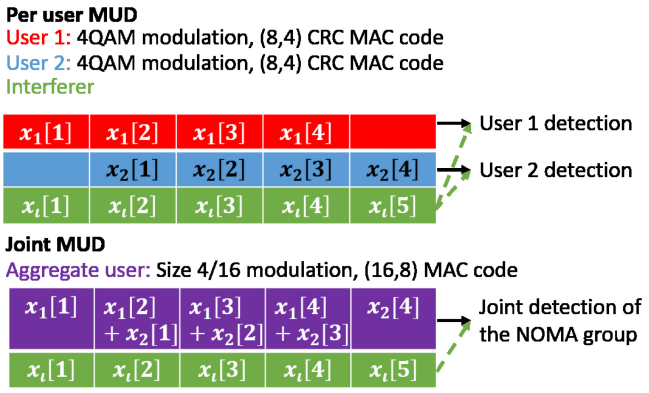

Fig. 3. High-level depiction of the per user and joint MUD, when the receiver is either aware or ignorant of the interferer composition. Red corresponds to user 1, blue corresponds to user 2, green corresponds to the interferer, and purple corresponds to the aggregate user with the resulting macrosymbol. When the users are asynchronous and do not overlap, the macrosymbol is composed of one symbol, and when the users simultaneously access the same channel resource, the macrosymbol is composed of two symbols. The presence of green text indicates whether or not the receiver has knowledge of the interferer if there is an interferer present. In this example, it is assumed the channel gains are unitary.

$\hat{h}[t]$ and $h[t]$ without subscripts, as this channel gain estimate is applicable to both the users and the interferer at all times, The estimate is

$$\hat{h}[t] = h[t] + e[t] \tag{2}$$

where $e[t] \sim \mathcal{CN}(0, \sigma^2)$ is the noise added to the channel gain, and $\sigma^2 = P^{-\alpha}$ [34], [35], [36]. Note that P is used as a general variable for the power of the users and interferer. By defining the noisy channel estimate as in (2), the estimation error is related to the power of the transmitted signal–a stronger transmit power will lead to less estimation error, and a weaker transmit power will lead to more estimation error. In addition, the tuning parameter α can be used to control the quality of the noisy estimate. When $\alpha \to \infty$, there will be no channel estimation error, and when $\alpha = 0$, the error is independent from the transmit power.

III. Multiuser Detection

In this section, we discuss the multiuser detection component of GRAND-AM. Previously, in Section I, we stated that there are two types of MUDs, an individually optimal per user detector, and a jointly optimal detector. Here, we discuss in greater detail what these two detectors are, and the differences between them. In addition, we incorporate the channel gain estimate into the detection, which we have previously not done so before in [1] and [2]. For the following equations, we assume that all u users have transmitted. When a user has not transmitted due to asynchronicity, the following estimators can be modified by removing the user that has not transmitted.

A. Per User MUD

The individually optimal per user MUD optimizes the probability that each user is correct, independent from the other users [13]. Here, we define the estimator for the per user MUD in a system where there is an interferer, and the receiver is interferer aware. Without loss of generality, the estimate for user 1 at time t is

$$\hat{x_1}[t] = \underset{x_1[t]}{\operatorname{argmax}} f_{Y|X_1}(y[t]|x_1[t]) \tag{3}$$

where

$$\begin{aligned} &f_{Y|X_1}(y[t]|x_1[t]) \\ &\quad= \sum_{v_q=1}^{m_q} f_{Y|X_1,X_q}(y[t]|x_1[t], x_{q,v_q}) p(X_q = x_{q,v_q}) \\ &\quad= \frac{1}{m_q} \sum_{v_q=1}^{m_q} f_{Y|X_1,X_q}(y[t]|x_1[t], x_{q,v_q}) \\ &\quad= \frac{1}{m_q} \frac{1}{m_2, \ldots m_u} \sum_{v_q=1}^{m_q} \sum_{v_2=1}^{m_2}, \ldots \sum_{v_u=1}^{m_u} \\ &\qquad f_W\left(y[t] - \hat{h}_1[t]x_1[t] - \sum_{i=2}^{u} \hat{h}_i[t]x_{i,v_i} - \hat{h}_q[t]x_{q,v_q}\right) \end{aligned} \tag{4}$$

where $f_W(\cdot)$ is the probability distribution function (PDF) of the AWGN from the channel, and $\hat{h}_i[t]$ are the estimated channel gains. Note that under the condition that the receiver is interferer aware and has knowledge about its modulation and an estimate of its channel gains, the effect of the interferer can be accounted for through marginalization. This method is also used when accounting for the effect of the other users, in order to maximize the probability that user 1 is correctly detected.

Recall that the absence of an interferer can be represented by setting $P_q = 0$, in which case, the estimator for the per user MUD is simplified to

$$\begin{aligned} f_{Y|X_1}(y[t]|x_1[t]) = \frac{1}{m_2 \cdots m_u} \sum_{v_2=1}^{m_2} \cdots \sum_{v_u=1}^{m_u} \\ f_W\left(y[t] - \hat{h}_1[t]x_1[t] - \sum_{i=2}^{u} \hat{h}_i[t]x_{i,v_i}\right). \end{aligned} \tag{5}$$

In addition, in the case of an interferer ignorant receiver, (5) is used at the receiver. This is due to the receiver lacking knowledge about the interferer modulation and channel gains, and treating the interferer as part of the AWGN.

The per user MUD requires the summation of $u - 1$ PDFs when there is not an interferer present. Within each of these PDFs, there are u additions and u multiplications. When the noise is assumed to be AWGN, this leads to the summation of $\prod_{i=2}^{u} m_i$ exponential functions, which leads to high complexity when evaluating the likelihoods of each possible estimate for user 1. In addition, user 1 will require a total of m_1 estimates generated in this fashion. While approximations can be made to reduce the complexity of evaluating the estimates [27], the approximations remove the optimality of this detector.

B. Macrosymbols

Before discussing the jointly optimal MUD discussed in [13], we first define the concept of a macrosymbol. The jointly optimal MUD maximizes the probability that all users

are simultaneously detected correctly. We formulate the joint of all the users accessing the channel as a single aggregate user, whose constellation is formed from the combination of all user symbols and channel gains. This allows for the visualization of the joint MUD as a single-user detector for the aggregate user's macrosymbols, which leads to simplifications when calculating the log likelihood compared to the individually optimal per user MUD. The set of possible macrosymbols is formally defined as all unique combinations of

$$\mathcal{S}_\mu[t] = \{\mu[t]\} = \left\{ \sum_{i=1}^{u} h_i[t] x_{i,j_i} \right\} \tag{6}$$

where j_i are chosen from $j_i \in [1, m_i]$ and $i \in [1, u]$. Compared to the ith user having a constellation of size m_i, the aggregate user instead has a constellation of size $\prod_{i=1}^{u} m_i$.

An example of the set of macrosymbols can be seen in Fig. 2, where there are two users modulated with binary phase-shift keying (BPSK) and 4QAM, respectively. Before an interferer contributes to the channel output, there are eight macrosymbols in the aggregate constellation. After an interferer contributes to the channel output, there are still eight possible macrosymbols, but they have been corrupted by the interferer such that the receiver can observe 16 possible macrosymbols in the case where the interferer is modulated with BPSK. If the receiver is interferer aware, it must then take into account the interferer contribution in order to more accurately detect the users' transmissions.

C. Joint MUD

With the macrosymbol defined as such in (6), the estimator for the jointly optimal MUD in a system where there is an interferer, and the receiver is interferer aware is

$$\hat{\mu}[t] = \underset{\mu[t]}{\operatorname{argmax}} f_{Y|\mathcal{M}}(y[t]|\mu[t]) \tag{7}$$

where

$$\begin{aligned} &f_{Y|\mathcal{M}}(y[t]|\mu[t]) \\ &\quad = \sum_{v_q=1}^{m_q} f_{Y|\mathcal{M},X_q}(y[t]|\mu[t], x_{q,v_q}) p(X_q = x_{q,v_q}) \\ &\quad = \frac{1}{m_q} \sum_{v_q=1}^{m_q} f_{Y|\mathcal{M},X_q}(y[t]|\mu[t], x_{q,v_q}) \\ &\quad = \frac{1}{m_q} \sum_{v_q=1}^{m_q} f_N\left(y[t] - \hat{\mu}[t] - \hat{h_q}[t] x_{q,v_q}\right) \\ &\quad = \frac{1}{m_q} \sum_{v_q=1}^{m_q} f_W\left(y[t] - \sum_{i=1}^{u} \hat{h}_i[t] x_{i,j_i} - \hat{h_q}[t] x_{q,v_q}\right) \end{aligned} \tag{8}$$

where $f_N(\cdot)$ is the PDF of the AWGN from the channel, $\hat{h}_i[t]$ are the estimated channel gains, and j_i are chosen from $j_i \in [1, m_i]$ and $i \in [1, u]$.

In the case of an interferer ignorant receiver, or a system where there is no interferer, the estimator for the jointly optimal MUD is simplified to

$$f_{Y|\mathcal{M}}(y[t]|\mu[t]) = f_W\left(y[t] - \sum_{i=1}^{u} \hat{h}_i[t] x_{i,j_i}\right) \tag{9}$$

which is the estimator corresponding to a single-user detector.

Unlike the per user MUD, the joint MUD can be simplified when the noise is AWGN and there is not an interferer present. A logarithmic operation can be performed, which leads to the estimator only requiring u additions and u multiplications per estimate, of which $\prod_{i=1}^{u} m_i$ are required. There is necessary processing required when generating the macrosymbols, however, they also only require additions and multiplications, unlike the estimator for the per user MUD. To generate a macrosymbol, u sums and u multiplications are required, leading to a total number of operations needed for the set of macrosymbols being $2u \prod_{i=1}^{u} m_i$.

IV. Multiuser Decoding

In this section, we discuss the multiuser decoding component of GRAND-AM. In particular, we go into further details of the decoding algorithm used, and its ability to handle joint decoding across all users. We compare the process of a per user versus a joint decoding and highlight their difference. We also discuss ways of reducing the complexity of the joint decoding process if there are many points in the aggregate constellation and a long aggregate codebook.

A. Decoding With GRAND

Recall that each user $i \in [1, u]$ is coded with an (n_i, k_i) MAC code that is used for correcting errors due to MAI. In addition, there is the requirement that the users must be decoded jointly, in order to maintain the properties associated with the joint MUD process. Considering that the joint of the users results in a joint codebook that has a different structure from a single code, we use a decoding algorithm based on GRAND, which is a universal decoder that queries noise sequences from most to least likely and removes them from the received sequence of symbols or bits [30], [37].

In particular, we use symbol-level ordered reliability bit GRAND (ORBGRAND), which is a near ML variant of GRAND that queries noise sequences based on symbol-level reliabilities [30]. Below, we give a brief overview of symbol-level ORBGRAND. A list of symbol-level reliabilities is generated from the estimator when detecting user symbols. The reliabilities of the detected symbols are removed from the list, leaving a list of potential symbol-level reliabilities, which will be used for correcting the detected sequence of symbols if the sequence is not contained within the codebook. A list of potential symbol swaps is generated using the symbol-level reliabilities. The list is rank-ordered using a logistic weight principle, which is defined as the sum of the indices associated with the rank-ordered symbols that will be swapped. Note that when ranking the possible symbol swaps, any symbol swaps that duplicates the symbol to be swapped is excluded from the list. Symbols are then swapped with the originally

detected symbols based on the ordering of the alternative symbol list. The resulting sequence of symbols is then checked for membership within the codebook. If it is not contained within the codebook, the process continues until a sequence of symbols satisfies the codebook.

Here, we give a simple example of the logistic weight ranking process. Consider a symbol sequence of length 4, which we will denote as (s_1, s_2, s_3, s_4), and the symbol modulation is binary. Assume that the order from least to most reliable is (s_3, s_4, s_1, s_2), with corresponding weights of $(1, 2, 3, 4)$. Then, the ranking of symbol swaps based on logistic weights is $\{s_3, s_4, s_1, (s_3, s_4), s_2, (s_3, s_1), \ldots\}$. Note that both (s_3, s_4) and s_1 swaps have weights of 3, where the weight for (s_3, s_4) comes from the sum of the original weights of s_3 and s_4, and s_1 is originally weighted as a 3. Thus, both symbol swaps have equal priority, and order does not matter. In addition, observe that the (s_4, s_4) swap is removed from the list despite being logistic weight rank 4 due to the requirement that the same symbol cannot be swapped multiple times.

Symbol-level ORBGRAND is well suited for both individual and joint multiuser decoding. As it works on a symbol-level basis, it will also work for the macrosymbol, allowing for joint decoding in addition to the joint MUD. The process differs slightly between a per user and a joint decoding. For the receiver that utilizes per user MUD and decoding, symbol-level ORBGRAND will be used to correct each user's associated symbol sequence that has been encoded with the MAC code. Thus, symbol-level ORBGRAND must be used u times, and will directly give each user's corrected codeword. For the receiver that utilizes joint MUD and decoding, the symbol-level ORBGRAND process will differ slightly. Symbol-level ORBGRAND will be used once, in order to correct the macrosymbol sequence. The requirement is that for a corrected sequence of macrosymbols to be found, all u MAC codebooks must be simultaneously satisfied. Then, the resulting sequence of macrosymbols will be decomposed to each of the users' symbol sequence. An example of the difference between the individual and joint error correction processes is illustrated in Fig. 4, which shows how each user has its own errors to correct with its own MAC code for per user GRAND, while the aggregate user has the combination of errors across both users with the combined MAC code for GRAND-AM. Pseudocode for the joint MUD and decoding process in GRAND-AM is found in Algorithm 1, and further outlines this process.

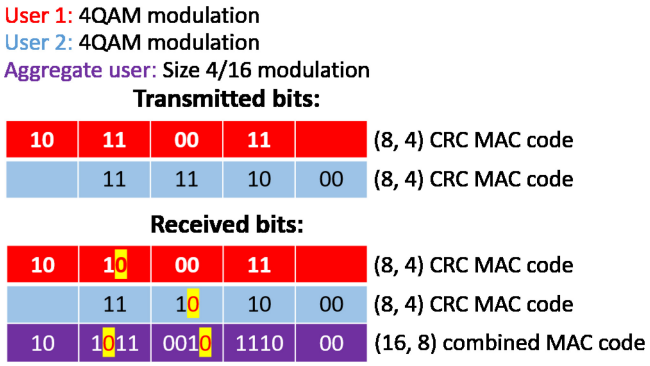

Fig. 4. Example of bit errors due to the detection component of GRAND-AM. In a per user decoding, each user independently corrects their own errors, while in a joint decoding, the users' bits are combined into the aggregate user's bits which are corrected and decoded according to combined codebook across all users. Correction and decoding with the combined codebook requires all users be simultaneously satisfied. Note that when there is asynchronicity, the received bits are composed of a single user's bits, and when both users transmit simultaneously, the received bits are the combination of the two users.

Algorithm 1 GRAND-AM

Input: Received signals $y[t]$, macrosymbols set $\mathcal{S}_\mu[t]$, MAC codes (n_i, k_i) with codewords $\{C_i\}$ for $i \in [1, u], k \in [1, \lceil n_i / \log_2(m_i) \rceil]$

Output: Corrected codewords $\hat{c}_i$

$a = 1$
for $a \leq \max(\lceil n_i / \log_2(m_i) \rceil)$ **do**
  Detect $\hat{\mu}[t]$ from $y[t]$
  Save list of likelihoods per macrosymbol excluding $\hat{\mu}[t]$ as $\{\mathcal{LL}[t]\}$
end for
Construct macrosymbol sequence $\vec{m} \leftarrow (\hat{\mu}[1], \hat{\mu}[2], \cdots)$
Separate macrosymbol sequence $\vec{m}$ into user symbol sequences $\vec{x_i} \forall i \in [1, u]$
Generate symbol swap list $\mathcal{Q}$ according to logistic weight principle from lists $(\{\mathcal{LL}[1]\}, \{\mathcal{LL}[2]\}, ...)$
$b = 1$, $\vec{r} = \vec{m}$
while $\vec{x_i}$ does not satisfy all $\{C_i\}$ simultaneously **do**
  Swap symbols of $\vec{d}$ according to $\mathcal{Q}[l]$
  Separate $\vec{r}$ into user symbol sequences $\vec{x_i}$
  Check $\vec{x_i} \in \{C_i\} \forall i$
  $b = b + 1$, $\vec{r} = \vec{m}$
end while
return corrected symbol sequences $\vec{c_i} = \vec{x_i}$

B. Reducing the Complexity of Decoding

As the numbers of users in the MAC increase, both the size of the aggregate constellation and the size of the aggregate codebook increase. There are two aspects of complexity associated with the decoding—generating the symbol swap lists based on the likelihoods associated with each possible aggregate constellation point and the number of queries it takes until all user codebooks are simultaneously satisfied [29]. We will discuss methods of reducing the complexity of both aspects if in an SNR regime that necessitates it.

In [29], it is stated that the landslide algorithm used to generate the symbol swap lists is efficient and will not bottleneck the decoding process. However, considering the multiple users combining to form an aggregate constellation and codebook, we will discuss how to reduce the length of the likelihood lists, which will reduce the complexity necessary for a large number of symbol swaps. Recall that the aggregate constellation contains $\prod_{i=1}^{U} m_i$ points. For simplicity of analysis, assume that the number of symbols per codeword for each user, l_i, are the same, which we will simplify to l. Then, the length of the likelihood list used for corrected the aggregate user will equal $l \prod_{i=1}^{U} m_i$. Indeed, as either

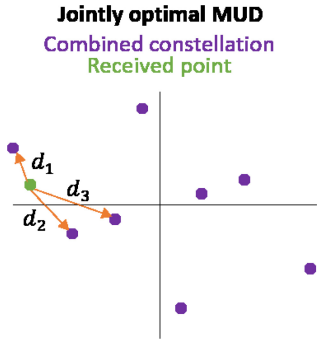

Fig. 5. Length of the likelihood lists can be reduced by only taking into account the q nearest neighbors, where $d = 3$ for the above figure, instead of all $\prod_{i=1}^{U} m_i$ points in the aggregate constellation.

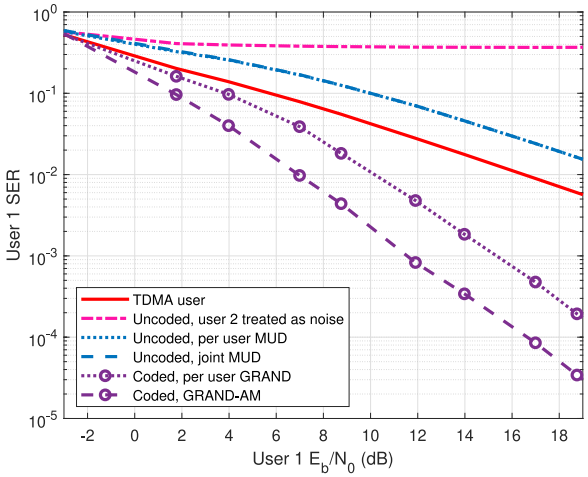

Fig. 6. Performance of a single user modulated with 4QAM and without an $(8, 4)$ CRC MAC code when TDMA is used versus two users simultaneously accessing the channel with or without MAC codes. There are no interferers, and the users are perfectly synchronized.

the number of users, the size of the users' constellation, or number of symbols increases, the length of the likelihood list will grow large. When multiple symbol swaps are required for large logistic weight ranks, which tends to occur in low SNR regimes, generating the symbol swap list may become intensive.

One method that can be used to reduce the length of the likelihood list is to consider the μ nearest neighbors to a received point, instead of considering all $\prod_{i=1}^{U} m_i$ aggregate constellation points. This is shown in Fig. 5, where $\mu = 3$ and only the three nearest neighbors will be used for the symbol swaps. These constellation points are far more likely to be the original transmitted symbol than the other five aggregate constellation points. Using this technique will allow for the overall likelihood list to be reduced from length $l \prod_{i=1}^{U} m_i$ to length $l\mu$. In addition, using the μ nearest neighbors will prevent the likelihood list from growing exponentially with the users in the MAC and the size of the users' constellations.

The other aspect of decoding complexity that should be considered is the number of queries symbol wise ORBGRAND requires until a codeword that satisfied the aggregate codebook is reached. For a codebook of size (n, k), on average, at most 2^{n-k} queries are required before there is an erroneous decoding for the codeword [37]. For the aggregate codebook, then $n = \sum_{i=1}^{U} n_i$ and $k = \sum_{i=1}^{U} k_i$. The erroneous decodings happen most often at low SNR regimes, implying that decoding in these regimes will be more complex. Thus, it can be seen that the number of users increases or the as the number of parity bits per user codebook increases, the decoding complexity will increase. One method of controlling the number of queries required is to use an abandonment threshold [37]. Once the number of queries has passed this threshold, decoding will halt. This will help prevent scenarios where the number of queries to reach a codeword that satisfies all codebooks is exceedingly high. While this method will increase the error rates in low SNR regimes, it should be noted that the regimes of interest, such as FERs of 10^{-4} for mission critical services will rarely reach this abandonment threshold.

V. Results and Discussion

We first consider the performance of GRAND-AM in comparison to other methods, such as treating the other user as noise as the worst case scenario, TDMA as an example of an OMA method, using only MUD as a NOMA method, and using per user MUD and GRAND when a MAC code is used to split the channel for NOMA. The results for two users modulated with 4QAM and with equivalent powers and perfect synchronization are shown in Fig. 6, with independent $(8, 4)$ CRC codes with hexcode 0x9 used as MAC codes when splitting the channel in $1/2$ for each user. It is assumed that there are no additional interferers other than users, and in the scenario where the users are not treated as noise, there is perfect channel estimation available at the receiver.

The worst case scenario for NOMA, where one user is recovered while treating the other as noise when the two users are equal power, leads to high symbol error rates (SERs). This is due to the high signal to interference and noise ratio (SINR $= P_u/(P_q + N_0)$)—as the powers of the two users increase, the SINR approaches 0 dB, leading to poor recovery of the user of interest. Thus, signals should not be treated as noise, especially if information can be obtained about them due to them having high power. This is further discussed later, where an interferer has been added to the system, and where both low- and high-powered interfering signals are considered. This detection method is the upper bound on the SER, as it assumes no information about other users sharing the channel other than their power.

While it is important to discuss this worst case scenario upper bound, in MACs, this method will not be used unless in combination with other techniques such as SIC, which we have already discussed in the introduction. Thus, we will primarily focus on comparing GRAND-AM's results with the results when TDMA, only MUD, and per user MUD and GRAND are used for MACs. For these methods, (5) and (9) are used as the estimators for the detection component, with the NOMA methods computing the estimators with two users, and TDMA computing the estimator with one user. When only MUD is used, both the joint MUD and the per user MUD perform very similarly, as noted in [13] and as seen in

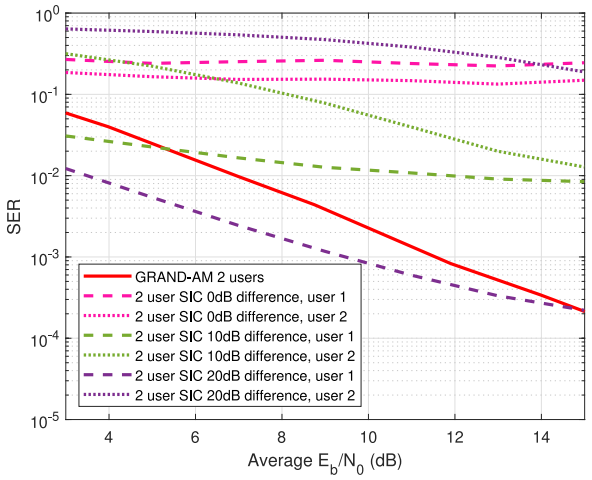

Fig. 7. Comparison between GRAND-AM versus NOMA when SIC is used for detection when there are two users modulated with 4QAM and coded with (8, 4) CRC codes as MAC codes.

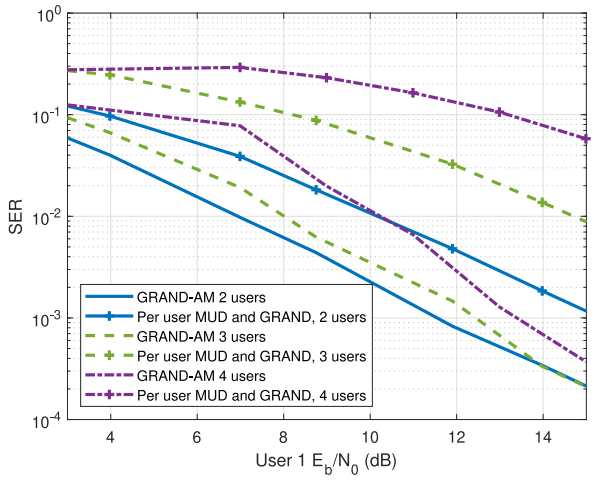

Fig. 8. Performance of per user GRAND and GRAND-AM when recovering 2, 3, or 4 users modulated with 4QAM and with (8, 4) CRC codes as MAC codes without an interferer present and with perfect channel estimation and synchronicity.

Fig. 6, where both versions of the blue dotted lines overlap. The MUD detection without a MAC code "overloads" the channel, as both users share the channel without splitting the channel nonorthogonally, thus leading to the MUD only method performing worse than TDMA by ∼3 dB. However, once the channel is split, using rate 1/2 MAC codes for both users to have the same channel "occupancy" as TDMA, the combined MUD and detection methods outperform TDMA.

In comparison to TDMA outperforming the MUD only method by ∼3 dB, incorporating the MAC codes for the channel splitting results in per user MUD and GRAND outperforming TDMA by ∼7 dB and GRAND-AM outperforming TDMA by ∼10 dB. The addition of the MAC code used for handling MAI from the simultaneous channel usage and splitting the channel nonorthogonally leads to great gains over an OMA method. In particular, the GRAND-AM method, where joint MUD and joint GRAND are used, leads to greater gains compared to the per user method.

The difference in performance between the joint and per user methods is due to the requirement that all users must be simultaneously satisfied when the users are jointly decoded when using GRAND-AM. For the individual decoding process, the GRAND algorithm separately decodes each user such that the resulting sequence of symbols for each user has no impact on the results of the other users. Due to this independence, the scenario where the recombination of all users' symbol sequences is far away from the original received signal may arise. In contrast, GRAND-AM maintains the jointness of the decoding process by ensuring that the sequence of macrosymbols being tested against the users' codebooks must simultaneously satisfy all codebooks before ending the algorithm, which is similar to the sequence of macrosymbols being required to satisfy a $(\sum_{i=1}^{u} n_i, \sum_{i=1}^{u} k_i)$ length code. The larger number of parity bits helps improve the SER. Furthermore, unlike the case with the per user MUD and decoding process, GRAND-AM minimizes the distance between the resulting sequence of macrosymbols, and the received sequence of macrosymbols, leading to a lower error rate for all users once the separation of the macrosymbols into user symbols is completed.

While we have shown that GRAND-AM outperforms TDMA with the usage of the MAC codes, the question arises how it compares when other NOMA methods are used. We compare GRAND-AM's results versus a method that uses an iterative detection method such as SIC, such as in PD-NOMA. We consider the scenario where the multiple users accessing the MAC are provided a total power budget, and for SIC-based detection and decoding methods, the users are allowed to have disparate transmit powers. Fig. 7 shows the case when there are two users, each modulated with 4QAM, and given (8, 4) CRC codes for error correction. When SIC detection is used, the user powers differ by 0, 10, or 20 dB, while for GRAND-AM, the user powers are equal, that is, with a 0-dB difference. It can be seen that regardless of the power difference between the users, the usage of SIC as a detector leads to disparate error rates between users 1 and 2. One user will have better error rates than the other, and as the power difference increases, the difference in performance between user 1 and user 2 increase further. While one user with the SIC detection may outperform GRAND-AM when the power difference is set to 20 dB, the other user greatly suffers in comparison. In addition, note that SIC-based methods must have some power difference between the users—when the users have similar transmit powers, all user error rates suffer as a result. These downsides show that iterative detection and decoding methods such as SIC, which are vital to PD-NOMA methods, should be avoided.

A concern with IoT scenarios and NOMA is how many users can be simultaneously supported, and how much load can the channel support. We consider the what happens when the number of users is increased while the MAC code rate remains the same, which will increase the channel load. Fig. 8 shows this scenario, where there are 3 or 4 users with the same powers in the NOMA group, each modulated with 4QAM and with independent (8, 4) CRC codes as MAC codes. Indeed, as MAI increases due to the increase in the number of users,

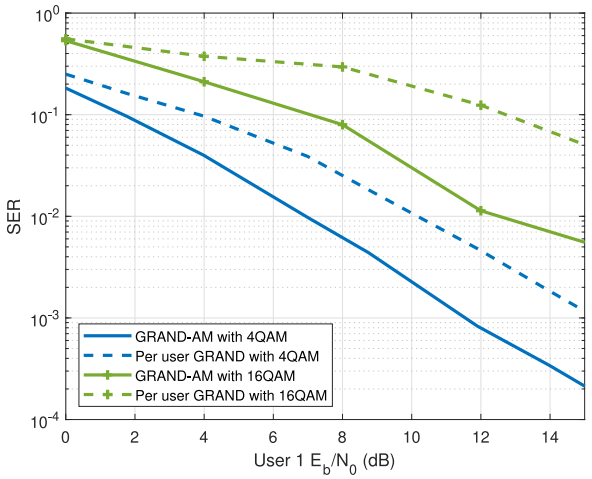

Fig. 9. Comparison between GRAND-AM and per user MUD and GRAND when there are two users modulated with 4QAM versus two users modulated with 16QAM and coded with (8, 4) CRC codes as MAC codes.

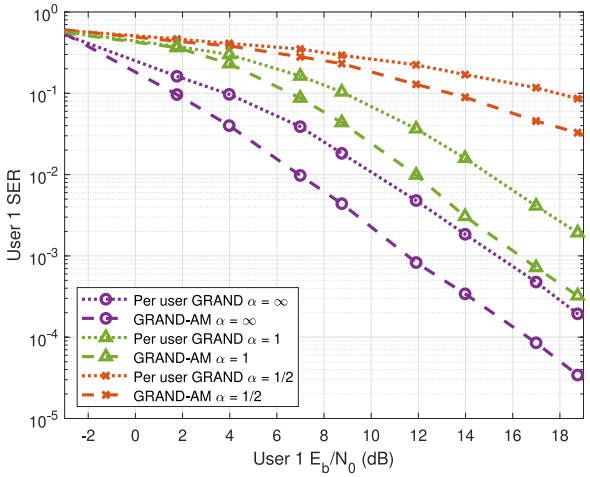

Fig. 10. Impact of imperfect channel estimation on the performance of per user GRAND and GRAND-AM when recovering two users modulated with 4QAM and with (8, 4) CRC codes as MAC codes without an interferer present and with perfect synchronicity.

performance degrades. When per user MUD and GRAND are used, the error rates greatly increase, while for GRAND-AM, the error rates slightly increase. For three users, GRAND-AM outperforms per user MUD and GRAND by ∼7 dB, while for four users, GRAND-AM outperforms per user MUD and GRAND by ∼9 dB. Despite the increased MAI and size of the aggregate constellation, the aggregate codebook of size $(\sum_{i=1}^{u} n_i, \sum_{i=1}^{u} k_i)$ helps to offset these factors, leading to the slight degradation in performance for GRAND-AM compared to the large degradation in performance for per user MUD and GRAND.

While the TDMA curve is not shown in Fig. 8, note that due to the orthogonality of TDMA, it will result in the same performance as the curve in Fig. 6. However to account for three users transmitting information, each user can only transmit $1/3$ of the time, leading to a reduction in throughput per user. Similarly, with four users, the throughput per user is reduced even more. Meanwhile, with GRAND-AM, all users can transmit simultaneously, even with the channel overloaded with 3 and 4 users. While there are increases in error rates, GRAND-AM can still outperform TDMA with 3 or 4 users accessing the MAC. Thus, GRAND-AM shows great promise as a NOMA method, as even when the number of users grows and the sum of the codebook rates of the users is greater than 1, it can still reliably recover user information.

For the previous results, we have considered the case where all the users are modulated with 4QAM. While the aggregate constellation size does increase exponentially with the number of users, as in the case with Fig. 8, it can be seen that the increase in size of the aggregate codebook helps offset the increased number of errors due to the exponentially growing aggregate constellation. However, when the number of users remains fixed while the size of each users' constellations increases, there is no corresponding growth in aggregate codebook size. Fig. 9 shows the impact of increasing the user constellation size from 4QAM to 16QAM for two users while the MAC code remains fixed. With this, the aggregate constellation sizes are 16 and 256, respectively, while the aggregate codebook is of size $(16, 8)$ for both cases. As the user modulation goes from 4QAM to 16QAM, there is ∼6–7 dB loss when GRAND-AM is used, versus the ∼9-dB loss when only per user MUD and GRAND is used. This indicates that while the error rates do increase as expected when higher order constellations are used, the joint MUD and decoding process may help to mitigate some of the losses compared to the per user MUD and decoding process.

Note that in the case of the two users modulated with 16QAM in Fig. 9 versus the four users modulated with 4QAM in Fig. 8, the overall size of the aggregate constellation is equal at 256 points in both cases. The primary difference is that for the 2 user scenario, the aggregate codebook is only size $(16, 8)$ versus for the 4 user scenario, the aggregate codebook is size $(32, 16)$. As a result of the increased number of parity bits, despite having the same aggregate constellation, the scenario where there are four users with 4QAM modulation performs better. The implication is that for higher order aggregate constellations, aggregate codebooks with a larger number of parity bits are required, whether these parity bits are obtained through the combination of more users with short MAC codes versus less users with longer MAC codes.

While GRAND-AM performs well when there is perfect channel estimation, it is an impractical requirement, especially in IoT scenarios. Thus, we consider the performance of our proposed method when there is channel estimation error that scales with the transmit power of the signal, as described in (2). Fig. 10 shows how GRAND-AM and per user MUD and decoding perform once channel estimation error is introduced when there are two users modulated with 4QAM and with $(8, 4)$ CRC codes as MAC codes. In this figure, two α terms are included for comparison—$\alpha = 1$, which indicates that the power of the channel estimation error is inversely proportional to the user power, and $\alpha = 1/2$, which will lead to a higher channel estimation error. Here, we see that given a large enough α term, GRAND-AM is robust against channel estimation error. For $\alpha = 1$, the performance of GRAND-AM degrades by 4 dB, and when $\alpha = 1/2$, the performance

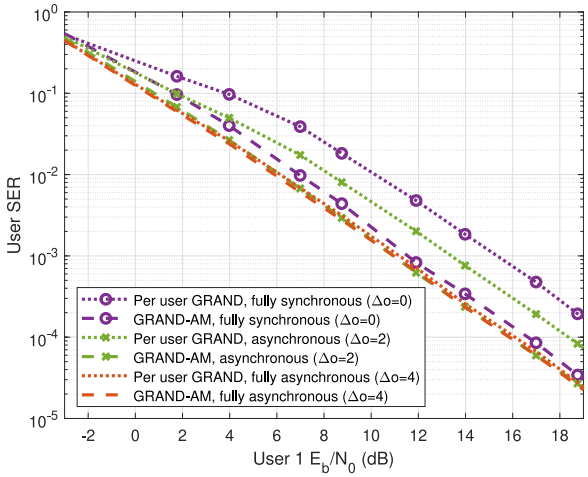

Fig. 11. Impact of symbol-level asynchronicity when recovering two users modulated with 4QAM and with $(8, 4)$ CRC codes as MAC codes without an interferer present and with perfect channel estimation. Note that $\Delta o = o_2 - o_1$ is the symbol wise offset value between user 1 and user 2.

degrades by 15 dB at larger E_b/N_0. Thus, it is important that some accurate channel estimate is acquired for GRAND-AM to work. However, note that when $\alpha = 1$, even with the degradation of 4 dB due to the channel estimation error, GRAND-AM still outperforms TDMA by $\sim$6 dB at larger E_b/N_0. This shows that GRAND-AM is a powerful technique, as even with channel estimation error, it can still outperform an OMA method with perfect channel estimation. In addition, it is reasonable to assume $\alpha = 1$, considering that this corresponds with the scenario where the channel is reciprocal, that is, the uplink and downlink channels can use the same channel state information, though some error will be introduced due to the channel not remaining completely static [35].

Other than target error rates and channel estimation error, another aspect that is important to consider for IoT applications is the issue of synchronization and grant-free access. Considering that IoT transmissions may be interrupt-based due to power saving modes being used, not every user can be synchronized such that they always simultaneously transmit their frames together. In addition, synchronization requires coordination between the transmitters and receiver, which would further add to the overhead required. Therefore, we should consider asynchronicity and explore how it impacts the performance of GRAND-AM, with the assumption that a timing lock has been obtained or maintained such that there is only symbol-wise asynchronicity, as exhibited in Fig. 3, where there is a single symbol offset asynchronicity.

Fig. 11 shows how symbol-wise asynchronicity impacts the error rates of per user MUD and GRAND and GRAND-AM, using the same parameters as in the previous figures with two users. We consider three cases of symbol-wise offsets—0 offset, also known as, full synchronicity; 2 offset for partial asynchronicity; and 4 offset, also known as full asynchronicity. The case of full asynchronicity is similar to TDMA, as the user transmissions are orthogonal, though there is still the $(8, 4)$ CRC MAC code applied on each user. The symbol-wise asynchronicity improves the SERs for both the per user MUD and GRAND method and GRAND-AM, though there is a greater impact on the per user MUD and GRAND method. The offset leads to user symbols being detected by themselves, which leads to lower SER for the offset symbol compared to the case when two user symbols overlap, which leads to an overall decrease in SER. In particular, this is useful for the per user MUD and GRAND method, as it is heavily dependent per user—when the error rate associated with the per user MUD decreases, the overall SER will decrease. In comparison, GRAND-AM does not experience such gains with the asynchronicity. While the asynchronicity does improve GRAND-AM's SERs, the improvement is minor. Unlike the per user MUD and GRAND method, as GRAND-AM relies on jointly detecting and decoding the users, it can effectively correct errors that arise from the macrosymbols. Thus, there is less to gain from the lower error rate associated with individual symbols. This behavior can be seen when comparing the two extremes of full synchronicity and full asynchronicity. For per user MUD and GRAND, there is a $\sim$5 dB improvement, while for GRAND-AM, there is a < 1 dB improvement. This shows that assuming that a timing lock is obtained or maintained such that there is only symbol-wise asynchronicity, SERs will improve, whether or not per user MUD and GRAND or GRAND-AM is used. Regardless of the synchronicity, GRAND-AM will perform well.

For the previous results, we have considered the case where there is only MAI from the multiple users simultaneously accessing the same channel resource. However, there can be other interferers in the channel that arise from transmissions from sources not of interest. We consider similar parameters as in the previous parts of the discussion, but with an additional interferer in the system. To handle the interferer, we consider two types of receivers, the interferer aware receiver, which corresponds to the estimators in (4) and (8), and the interferer ignorant receiver, which corresponds to the estimators in (5) and (9). We consider the case where there is and is not channel estimation error for both the users and the interferer, as well as what happens when the interferer has a higher channel estimation error compared to the receiver.

Fig. 12 shows how interference aware and ignorant receivers perform with GRAND-AM when recovering two users modulated with 4QAM and coded with $(8, 4)$ CRC MAC codes, as well as the performance of an interference-aware TDMA user, when in the presence of an interferer modulated with 16QAM. We assume that there is perfect channel estimation available for both the users and the interferer. The users have a fixed power of 21.8 dB, while the interferer has a power ranging from the AWGN power to the user power. This range allows for a look into the performance of these NOMA methods in low to high interference regimes. Note that for the MUD only methods, only the joint MUD is shown, as the per user and joint MUDs will perform similarly as discussed earlier.

Similar to the case of Fig. 6, TDMA outperforms the MUD only NOMA methods, while performing worse than the interference-aware per user MUD and GRAND method and GRAND-AM. At larger SINRs, the TDMA user outperforms the MUD only NOMA methods by 5 dB, which is more than in the case without an interferer. The TDMA user only

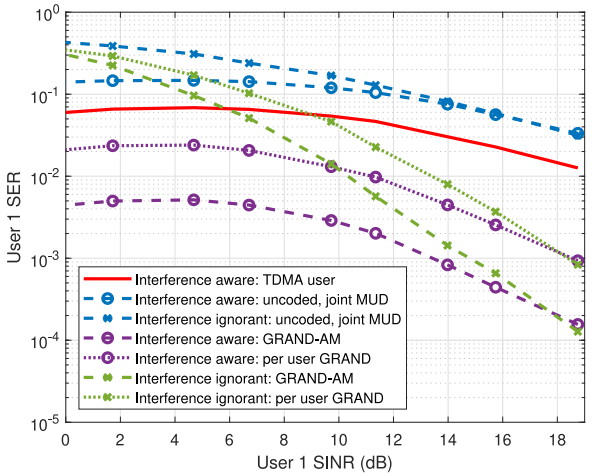

Fig. 12. Comparison between interference ignorant and interference-aware receivers when using per user GRAND or GRAND-AM to recover two users modulated with 4QAM and with $(8, 4)$ CRC codes as MAC codes with an interferer present, perfect synchronization, and perfect channel estimation.

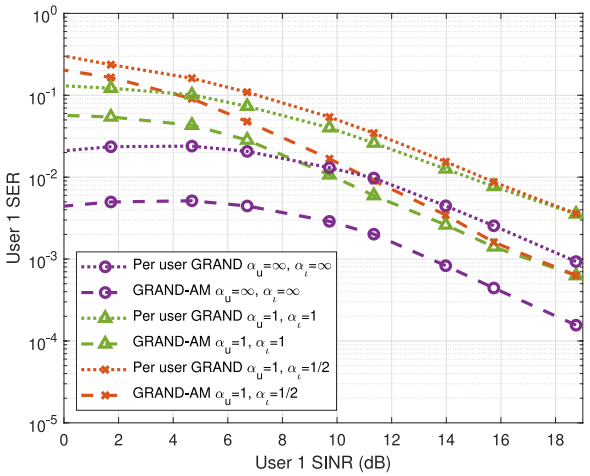

Fig. 13. Impact of imperfect channel estimation on the performance of per user GRAND or GRAND-AM when the receiver is interferer aware.

has a single interference source to account for, while the MUD only NOMA methods must account for both the MAI and the interferer, which leads to the larger difference. Thus, a MAC code should be used to handle the MAI. With a MAC code, the interference-aware per user MUD and GRAND method outperforms TDMA by ∼8 dB, while the GRAND-AM outperforms TDMA by a factor that is too large, in terms of dB, to illustrate in the figure. Even when the receiver is interference ignorant and treats the interferer signal as noise, the usage of MAC codes can still allow for good performance. While in low SINR regimes, the interference ignorant per user MUD and GRAND method and GRAND-AM perform worse than the interference-aware TDMA user, as the SINR increases, the addition of the MAC code allows the interference ignorant methods to outperform TDMA. The interference ignorant per user MUD and GRAND method begins performing better than TDMA at ∼10 dB, while interference ignorant GRAND-AM begins performing better at ∼6 dB. This shows that when there is both MAI and an interferer present, MAC codes and GRAND-AM are powerful tools that can handle both of these effects.

In the beginning of this section, we briefly mentioned that the worst case scenario for NOMA is when one user is recovered while treating the other user as noise. This is analogous to the interference ignorant receiver being used under the assumption there is only a single user accessing the channel when there are actually two users present. As seen in these results, interference ignorant receivers perform poorly at low SINRs, which was exactly the case discussed earlier. Given that the power of the interferer is on par with the user power, there should be obtainable information about the interferer that can be used to improve the error rates. However, when the SINR is large, that is, the power of the interferer is small relative to the power of the users and on par with the AWGN power, having information about the interferer is less crucial for good performance. Indeed, in Fig. 12, as the SINRs increase, the performance of the interference ignorant and interference-aware receivers approach each other. Thus, interference aware or ignorant receivers should be chosen based on the expected interference, if this information is available beforehand, and the receiver can afford to use interference-aware detection.

Similar to the case where there isn't an interferer, it is impractical to assume perfect channel estimation for the users and interferers. In addition, because the receiver must obtain information about the interferer, instead of being given information as in the case of the users, it is reasonable to assume that the channel estimation for the interferer will at most be as good as the channel estimation for the users. We consider the scenario where $\alpha_u = 1$, which corresponds to a reciprocal channel for the users, and $\alpha_q = 1/2$ and $\alpha_q = 1$ for the interferers. Fig. 13 shows how the channel estimation error for both the users and the interferer impacts the error rates. In the high SINR range, the SERs converge, indicating that the error introduced by the channel estimation originates from the channel estimation error for the users, with less impact from the change in α_q for the interferers due to the power of the interferer being similar to the power of the noise. However, when the SINR grows small and approaches 0, the change in α_q impacts the SERs of the users more heavily, as it is important to have accurate channel estimation for the interferer when its power is large. However, even with inaccurate channel estimation, the interference-aware receiver can still outperform the interference ignorant receiver when the SINR is small, especially with GRAND-AM being used. The interference-aware receiver using GRAND-AM with inaccurate channel estimation for both $\alpha_q = 1$ and $\alpha_q = 1/2$ can outperform the interference ignorant receiver using GRAND-AM up to an SINR of 9 dB. In contrast, the interference-aware receiver using per user MUD and decoding can outperform the interference ignorant receiver up to an SINR of 7 dB for both $\alpha_q = 1$ and $\alpha_q = 1/2$. This shows that a receiver using GRAND-AM in a scenario with both user and interferer channel estimation error is more robust against errors compared to a more conventional per user MUD and decoding process.

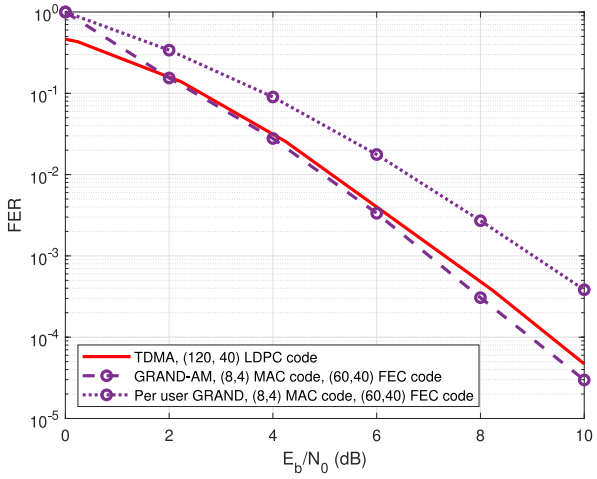

Fig. 14. Comparison of FERs when overall code rates of $\sim 1/3$ are used for both a TDMA user with LDPC coding, and per user GRAND and GRAND-AM with CRC codes as inner and outer codes.

Now that we have shown that GRAND-AM performs well and is robust in the presence of channel estimation error, asynchronicity, and interferers when it comes to NOMA, we should also consider how well GRAND-AM can be incorporated to the overall system, when we take into account the FEC block. Typically, only a single code such as a low-density parity-check (LDPC) code is used for error correction. However, recall from the block diagram in Fig. 1 that there are two codes being used for our proposed method, the MAC code used to deal with MAI from NOMA, and the FEC code used to deal with other errors. For a comparison with the traditionally used FEC, the product of the rates of the MAC code and FEC code used in GRAND-AM should be considered the overall coding rate. If rate $1/3$ LDPC codes are used for IoT applications, the outer FEC code for GRAND-AM should be rate $2/3$ if a rate $1/2$ inner code for MAI is used. Considering the importance of URLLC in many IoT applications, we use a payload size of 40 information bits, leading to a $(120, 40)$ LDPC code generated using the ETSI published standards [38]. While CRC codes are not used for FEC in the standards due to the lack of an error correction decoder, let us consider them, as they have flexible codeword lengths and the recent development of the GRAND algorithm allows the usage of them for error correction [39], [40]. For GRAND-AM and per user MUD and GRAND, we consider a $(60, 40)$ CRC code for the FEC. In combination with the $(8, 4)$ MAC code that has been used for the previous results, this will lead to an overall code of size $(120, 40)$ with rate $1/3$.

Fig. 14 shows a comparison between the FERs when there is a TDMA user with the $(120, 40)$ LDPC code, versus per user MUD and GRAND and GRAND-AM where a $(8, 4)$ CRC code is used as the MAC code and a $(60, 40)$ CRC code with hexcode 0xd41cf is used as the FEC code. Both users are modulated with BPSK modulation. For simplicity, we consider only the case where there is perfect channel estimation available to the receiver. The LDPC coded TDMA user outperforms GRAND-AM with two users at low E_b/N_0, but as the E_b/N_0 increases, GRAND-AM with two users begins to outperform by $\sim$0.5 dB. Both the TDMA user and the GRAND-AM users reach an FER of 10^{-4} at 9 dB, showing that using the LDPC code for the TDMA user and the GRAND-AM method of a MAC code and an FEC code are similarly reliable. However, recall that as shown in Fig. 8, when the number of users increases, GRAND-AM can mitigate the number of errors through the aggregate codebook, which grows in size with the number of users. As a result, increasing the number of NOMA users will slightly increase the FER, allowing for all NOMA users to freely transmit. In contrast, while the TDMA users will maintain the same FER due to the orthogonality of TDMA, as the number of users increases, the transmit duration per user decreases. Thus, GRAND-AM as a NOMA method has the potential to support higher throughput for each user compared to OMA method.

VI. Conclusion

In this work, we have proposed GRAND-AM, which uses joint ML MUD and joint decoding, as a method for handling NOMA. For GRAND-AM, there are three crucial components involved with the joint MUD and decoding—the macrosymbol, the MAC code, and the GRAND decoding algorithm. The macrosymbol is a concept that combines all user symbols and channel gains into a single macrosymbol that is then detected with the ML receiver. The detection of the macrosymbol directly corresponds with detecting the joint of all users. The MAC codes are short codes used to split the channel rate between users, and correct errors that arise due to MAI. In order to jointly decode across all users, we use the macrosymbol, and the combination of all users' MAC codes to form a joint codebook that a GRAND algorithm then uses to jointly decode across all users. For the joint decoding process, the algorithm does not halt until all user codebooks are simultaneously satisfied, unlike individual decoding methods, where each user codebook can be independently satisfied. This leads to the aggregate user, which is formed from all possible combinations of the individual users, having a codebook that has a larger number of parity checks compared to an individual user, resulting in lower error rates.

We have considered GRAND-AM under various circumstances, such as with imperfect channel estimation, symbol-wise asynchronous transmissions, and interference. The use of a MAC code for handling MAI allows for GRAND-AM to outperform TDMA by 10 dB in the ideal scenario, and even with imperfect channel estimation, the use of the MAC code can lead to better performance by 6 dB. In particular, the combination of the joint MUD and joint decoding via GRAND for GRAND-AM can outperform both the per user MUD and GRAND method as well as conventional OMA techniques such as TDMA in these circumstances, even when the channel is overloaded. When FEC is taken into account, GRAND-AM as a NOMA method results in FERs similar to a TDMA user for a given rate. The ability of GRAND-AM

to handle these scenarios shows GRAND-AM's potential in IoT applications, such as in grant-free multiple access or for URLLC applications.

ACKNOWLEDGMENT

The authors would like to thank Dr. Peihong Yuan for providing the MATLAB code that generates error rate curves for LDPC codes in Rayleigh fading channels.

REFERENCES

[1] K. Yang, M. Médard, and K. R. Duffy, "Multiuser detection using GRAND-aided macrosymbols," in *Proc. IEEE ICC*, 2023, pp. 4646–4651.
[2] K. Yang, M. Médard, and K. R. Duffy, "Separating interferers from multiple users in interference aware guessing random additive noise decoding aided Macrosymbol," in *Proc. IEEE MILCOM*, 2023, pp. 643–648.
[3] *Global Connectivity Report 2022*, Int. Telecommun. Union, Geneva, Switzerland, 2022.
[4] *IMT Traffic Estimates for the Years 2020 to 2030*, ITU-Rec. M. 2370-0, Int. Telecommun. Union, Geneva, Switzerland, 2015.
[5] H. Chen et al., "Ultra-reliable low latency cellular networks: Use cases, challenges and approaches," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 119–125, Dec. 2018.
[6] 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, Sep. 2016.
[7] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[8] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE 77th VTC*, 2013, pp. 1–5.
[9] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-lin, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.
[10] "Technical specification group radio access network; study on RAN improvements for machine-type communications; Version 11.0.0," 3GPP, Sophia Antipolis, France, Rep. TR 37.868, 2011.
[11] "Next generation protocols (NGP); next generation protocol requirements; Version 1.1.1," ETSI, Sophia Antipolis, France, Rep. GS NGP 005, 2017.
[12] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.
[13] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[14] M. L. Honig, *Advances in Multiuser Detection*. Hoboken, NJ, USA: Wiley, 2008.
[15] J. G. Andrews and T. H. Y. Meng, "Performance of Multicarrier CDMA with successive interference cancellation in a multipath fading channel," *IEEE Trans. Commun.*, vol. 52, no. 5, pp. 811–822, May 2004.
[16] Y. Cai, Z. Qin, F. Cui, G. Y. Li, and J. A. McCann, "Modulation and multiple access for 5G networks," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 629–646, 1st Quart., 2018.
[17] R. Hoshyar, F. P. Wathan, and R. Tafazolli, "Novel low-density signature for synchronous CDMA systems over AWGN channel," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1616–1626, Apr. 2008.
[18] S. Chaturvedi, Z. Liu, V. A. Bohara, A. Srivastava, and P. Xiao, "A tutorial on decoding techniques of sparse code multiple access," *IEEE Access*, vol. 10, pp. 58503–58524, 2022.
[19] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. IEEE 24th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, 2013, pp. 332–336.
[20] M. Rebhi, K. Hassan, K. Raoof, and P. Chargé, "Sparse code multiple access: Potentials and challenges," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1205–1238, 2021.
[21] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1805–1838, 3rd Quart., 2020.
[22] D. C. Nguyen et al., "6G Internet of Things: A comprehensive survey," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 359–383, Jan. 2022.
[23] D. Lopez-Perez, I. Guvenc, G. De La Roche, M. Kountouris, T. Q. Quek, and J. Zhang, "Enhanced intercell interference coordination challenges in heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 22–30, Jun. 2011.
[24] J. Lee, D. Toumpakaris, and W. Yu, "Interference mitigation via joint detection," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 6, pp. 1172–1184, Jun. 2011.
[25] J. Bao, Z. Ma, G. K. Karagiannidis, M. Xiao, and Z. Zhu, "Joint multiuser detection of multidimensional constellations over fading channels," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 161–172, Jan. 2017.
[26] W.-K. Ma and P. Ching, "Asymptotic minimum bit error rate property of maximum-likelihood multiuser detection," *IEEE Commun. Lett.*, vol. 7, no. 9, pp. 425–427, Sep. 2003.
[27] D. J. Jakubisin and R. M. Buehrer, "Approximate joint MAP detection of co-channel signals in non-gaussian noise," *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4224–4237, Oct. 2016.
[28] D. Bai, J. Lee, S. Kim, and I. Kang, "Near ML modulation classification," in *Proc. IEEE Veh. Technol. Conf.*, 2012, pp. 1–5.
[29] K. R. Duffy, W. An, and M. Médard, "Ordered reliability bits guessing random additive noise decoding," *IEEE Trans. Signal Process.*, vol. 70, pp. 4528–4542, Aug. 2022.
[30] W. An, M. Médard, and K. R. Duffy, "Soft decoding without soft demapping with ORBGRAND," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2023, pp. 1080–1084.
[31] A. Riaz et al., "A sub-0.8pJ/b 16.3Gbps/mm^2 universal soft-detection decoder using ORBGRAND in 40nm CMOS," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, 2023, pp. 432–434.
[32] A. Ghosh, J. Zhang, J. G. Andrews, and R. Muhamed, *Fundamentals of LTE*, 1st ed. Upper Saddle River, NJ, USA: Prentice Hall Press, 2010.
[33] R. A. Poisel, *Introduction to Communication Electronic Warfare Systems*, 2nd ed. Boston, MA, USA: Artech House, 2008.
[34] S. M. Razavi and T. Ratnarajah, "Performance analysis of interference alignment under CSI mismatch," *IEEE Trans. Veh. Technol.*, vol. 63, no. 9, pp. 4740–4748, Nov. 2014.
[35] P. Aquilina and T. Ratnarajah, "Performance analysis of IA techniques in the MIMO IBC with imperfect CSI," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1259–1270, Apr. 2015.
[36] T. Yoo and A. Goldsmith, "Capacity of fading MIMO channels with channel estimation error," in *Proc. IEEE ICC*, 2004, pp. 808–813.
[37] K. R. Duffy, J. Li, and M. Médard, "Capacity-achieving guessing random additive noise decoding," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4023–4040, Jul. 2019.
[38] "NR; multiplexing and channel coding," 3GPP, Sophia Antipolis, France, Rep. TS 38.212, 2018.
[39] W. An, M. Médard, and K. R. Duffy, "CRC codes as error correction codes," in *Proc. IEEE ICC*, 2021, pp. 1–6.
[40] P. Koopman. "Best CRC polynomials." Accessed: May 2023. [Online]. Available: http://users.ece.cmu.edu/koopman/crc/

Kathleen Yang (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering from California Institute of Technology, Pasadena, CA, USA, in 2019, and the M.S. degree from Massachusetts Institute of Technology, Cambridge, MA, USA, in 2022, where she is currently pursuing the Ph.D. degree.

Her research interests lie at the intersection of digital communications and signal processing.

Muriel Médard (Fellow, IEEE) received the first bachelor's degree in EECS, the second bachelor's degree in mathematics, the third bachelor's degree in humanities, and the M.S. and Sc.D. degrees from Massachusetts Institute of Technology, Cambridge, MA, USA, 1989, 1989, 1991, 1991, and 1995, respectively.

She is the NEC Professor of Software Science and Engineering with the Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology, Cambridge, MA, USA, where she leads the Network Coding and Reliable Communications Group, Research Laboratory for Electronics. She is also the Chief Scientist with Steinwurf, Aalborg, Denmark, which she has co-founded.

Dr. Médard received the 2017 IEEE Communications Society Edwin Howard Armstrong Achievement Award and the 2016 IEEE Vehicular Technology James Evans Avant Garde Award. She was awarded the 2022 IEEE Kobayashi Computers and Communications Award. She received the 2019 Best Paper Award for IEEE Transactions on Network Science and Engineering, the 2018 ACM SIGCOMM Test of Time Paper Award, the 2009 IEEE Communication Society and Information Theory Society Joint Paper Award, the 2009 William R. Bennett Prize in the Field of Communications Networking, the 2002 IEEE Leon K. Kirchmayer Prize Paper Award, as well as nine conference paper awards. Most of her prize papers are co-authored with students from her group. She has served as the Technical Program Committee Co-Chair of ISIT (twice), CoNext, WiOpt, WCNC, and many workshops. She has chaired the IEEE Medals Committee, and served as a member and a chair of many committees, including the Inaugural Chair of the Millie Dresselhaus Medal. She was the Editor-in-Chief of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and has served as the Editor or a Guest Editor of many IEEE publications, including the IEEE TRANSACTIONS ON INFORMATION THEORY, the IEEE JOURNAL OF LIGHTWAVE TECHNOLOGY, and the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY. She was a member of the inaugural steering committees for the IEEE TRANSACTIONS ON NETWORK SCIENCE and the IEEE JOURNAL ON SELECTED AREAS IN INFORMATION THEORY. She served as the Editor-in-Chief for the IEEE TRANSACTIONS ON INFORMATION THEORY. She was the Elected President of the IEEE Information Theory Society in 2012 and served on its board of governors for a dozen years. She is a member of the German National Academy of Sciences Leopoldina (elected 2022), the American Academy of Arts and Sciences (elected 2021), the U.S. National Academy of Engineering (elected 2020), a Fellow of the U.S. National Academy of Inventors (elected 2018), and a Fellow of the Institute of Electrical and Electronics Engineers (elected 2008). She holds Honorary Doctorates from the Technical University of Munich in 2020, from The University of Aalborg in 2022, and from the Budapest Institute of Technology and Economics (BME) in 2023.

Ken R. Duffy (Senior Member, IEEE) received the B.A. degree in mathematics and the Ph.D. degree in probability theory from Trinity College Dublin, Dublin, Ireland, in 1996 and 2000, respectively.

He is a Professor of Electrical and Computer Engineering and a Professor of Mathematics with Northeastern University, Boston, MA, USA, where he is affiliated to the Institute for the Wireless Internet of Things. He is a Co-Founder of the Royal Statistical Society's Applied Probability Section in 2011 and co-authored a cover article of Trends in Cell Biology in 2012. He works in works collaborative multidisciplinary teams to design, analyze, and realize algorithms using tools from probability, statistics, and machine learning. Algorithms he has developed have been implemented in digital circuits and in DNA.

Prof. Duffy is a winner of the Best Paper Award at the IEEE International Conference on Communications in 2015, the Best Paper Award from IEEE Transactions on Network Science and Engineering in 2019, and the Best Research Demo Award from COMSNETS in 2022.