FASURA: A Scheme for Quasi-Static Fading Unsourced Random Access Channels

Michail Gkagkos, *Graduate Student Member, IEEE,* Krishna R. Narayanan, *Fellow, IEEE,* Jean-Francois Chamberland, *Senior Member, IEEE,* Costas N. Georghiades, *Fellow, IEEE,* Department of Electrical and Computer Engineering, Texas A&M University

***Abstract*—Unsourced random access emerged as a novel wireless paradigm enabling massive device connectivity on the uplink. We consider quasi-static Rayleigh fading wherein the access point has multiple receive antennas and every mobile device a single transmit antenna. The objective is to construct a coding scheme that minimizes the energy-per-bit subject to a maximum probability of error given a fixed message length and a prescribed number of channel uses. Every message is partitioned into two parts: the first determines pilot values and spreading sequences; the remaining bits are encoded using a polar code. The transmitted signal contains two distinct sections. The first features pilots and the second is composed of spread modulated symbols. The receiver has three modules: an energy detector, tasked with recovering the set of active pilot sequences; a bank of Minimum Mean Square Error (MMSE) estimators acting on measurements at the receiver; and a polar list-decoder, which seeks to retrieve the coded information bits. A successive cancellation step is applied to subtract recovered codewords, before the residual signal is fed back to the decoder. Empirical evidence suggests that an appropriate combination of these ideas can outperform state-of-the-art coding techniques when the number of active users exceeds one hundred.**

***Index Terms*—Unsourced random access, massive multi-user MIMO, machine-type communication, polar codes, spread spectrum.**

I. Introduction

UNSOURCED random access (URA) has recently emerged as a wireless paradigm to accommodate the type of traffic generated by unattended devices [2]–[4]. A motivation behind this new perspective is that, as device density and machine-type traffic grow, the allocation of spectral resources based on sustained connections or coordinated requests becomes impractical. This is especially true when devices sporadically transmit short packets. A pragmatic alternative to fine-grain scheduling is to have all active devices share the same codebook. This way, the system can operate irrespective of the total user population, and focus on the number of active devices instead. In such situations, the decoder aims to recover the set of sent messages, without regard for user identities. If a device wishes to reveal its identity, it can embed this information in the payload of its own message. Many conceptual approaches and candidate solutions attuned to URA have appeared in recent years, with a majority of them confined to access points equipped with a single antenna. These schemes are informed by the natural connection between URA and sparse support recovery in large dimensions. Two noteworthy lines of work have emerged. A first group of publications are inspired by compressed sensing (CS) solvers [5]–[13]. These schemes also leverage notions from forward error correction to enable the application of CS solvers to very large spaces. A second group of results draw inspiration from more traditional multiple access techniques, including notions from multi-user detection [14] and random access [15]–[17].

This material is based upon work supported, in part, by the National Science Foundation (NSF) under Grants CCF-2131106 & CNS-2148354, and by Qualcomm Technologies, Inc., through their University Relations Program. This work was presented in part at the IEEE International Workshop on Signal Processing Advances in Wireless Communications, 2022 [1].

Manuscript received April 19, 2021; revised August 16, 2021.

A. Related Work

Following the success of URA schemes tailored to single-antenna Gaussian channels, researchers started to extend this access philosophy to SISO channels with fading. Fundamental limits are provided by Kowshik and Polyanskiy in [18], where they consider the many-user asymptotics of Chen et al. [19], and then derive bounds on the optimal required energy-per-bit for the reliable quasi-static multiple access channel. A converse bound for multiple-access channels with Rayleigh fading can be found in [20]. Furthermore, the authors in [20] construct a practical scheme for this problem based on LDPC codes and T-fold ALOHA, including a characterization of the gap between the fundamental bound and the performance of their scheme. The aforementioned results are derived, largely, under the assumption that received signals are coarsely synchronous, helped perhaps by the presence of a beacon. This assumption has been relaxed in [21], a contribution which offers a treatment of the asynchronous case. The authors therein exploit the fact that a scheme operating in the frequency domain allows the conversion of the synchronization problem to a phase-shift estimation problem. Andreev et al. [22] examine a quasi-static fading URA channel, and they propose a scheme that blends polar codes and expectation-maximization (EM) clustering. Andreev et al. [13] consider a coded compressed sensing approach by replacing the outer tree code proposed by Amalladinne et al. [5] with a list-recoverable code capable of correcting t errors. They show that the list-recoverable codes can improve the performance of the coded compressed sensing scheme significantly compared to the original tree code, when applied to certain fading environments. Many recent develop-

ments by the URA research community center on practical aspects of wireless communications, such as MIMO models. Liu and Wang [23] proposed an iterative receiver based on a sparse Tanner graph; therein, every iteration can recover at most three codewords. In their system model, the access point is equipped with four antennas, and there are around 3000 complex channel uses. On the other hand, Andreev and Frolov in [24] consider eight antennas and ten times more channel uses. The performance of a coded compressed sensing and T-fold ALOHA with a polar codes schemes is evaluated.

An early contribution to massive MIMO URA can be found in [25], where signals from different devices across slots are stitched together using channel properties rather than an outer code. Fengler et al. in [11] combine a non-Bayesian sparse recovery algorithm with an outer code aimed at message disambiguation to facilitate MIMO URA; the performance of this algorithm can be improved using successive cancellation list-decoding, as shown in [26]. Liu and Wang borrow ideas from slot-based transmissions and propose a receiver that merges simultaneous orthogonal matching pursuit (S-OMP) and blind channel estimation [27]. In [28], Srivatsa and Murthy analyze the throughput of irregular repetition slotted aloha (IRSA) and derive channel estimates for three schemes. Decurninge et al. offer an efficient solution based on a tensor construction, with good performance in some regimes [29]. Massive MIMO channels with correlated fading and a line-of-sight component are considered in [30]. Xie et al. design a covariance-based recovery algorithm to detect active users and their channels. They show that the proposed scheme suffers a $0.9-1.3$dB loss when used for correlated channels. The same authors exploit the angular domain sparsity and construct a scheme based on an EM-aided generalized approximate message passing algorithm. The idea is to remove the redundancies required by the tree encoder/decoder in the Coded Compressed Sensing (CCS) framework [31].

Among published work, two schemes put forth by Fengler et al. [32] and by Ahmadi et al. [33] offer the best performance in a different range of parameters. The construction in [32] relies on preamble-selected pilot sequences to estimate channels, and message payloads are encoded using a polar code. The receiver exploits the fact that the recovery of pilot sequences can be modeled as an MMV problem, and they use approximate message passing (MMV-AMP) to recover the active set. Then, channel coefficients are evaluated based on MMSE estimation, and maximum ratio combining (MRC) is used to combine the signals from different antennas. The outputs of the MRC serve as symbol estimates, and a polar decoder attempts to recover the most likely message. Finally, a successive interference canceller subtracts the contribution of the decoded users, and the process starts anew with pilot recovery applied to the residual signal. On the other hand, authors in [33] design a slot-based scheme. The main idea is the construction of orthogonal pilots by concatenating rows from a Hadamard matrix. A user divides its message into multiple parts, and each subset of the payload selects one of the orthogonal pilots. Then, all information bits are encoded, QPSK modulated, and appended after the pilots. Finally, a user transmits its codeword in one of the slots selected at random. An iterative receiver is designed that consists of a Neyman-Pearson test to identify the active pilots, an MRC for channel estimation, and symbol estimations. Furthermore, the soft estimates at the output of the MRC are passed to a list polar decoder. If a codeword satisfies the CRC constraint, its energy is subtracted from the received vector, and the process continues until the decoder is unable to output a new decoded user.

B. Main Contributions

The objective of this paper is to design a scheme for a quasi-static, massive MIMO, unsourced random access system. Since previous schemes, such as [3], [15], show outstanding performance in the AWGN case, we extend them to the MIMO URA setting. We call the proposed scheme FASURA (fading spread unsourced random access) since one of the main characteristics is the spreading sequences and because it is designed for fading channels. Instead of dividing the message into a multiple parts (more than two), like the CCS based schemes and recently the scheme in [33], we divide messages into two parts. The operational parameters, such as pilot, spreading sequences, etc, are selected based on a preamble. The payload is encoded using a single-user polar code. This approach is similar to the architectures of [15], [34] for SISO systems, and [32] in the MIMO case. The differences between [32] and FASURA are the introduction of spreading sequences, the detection of active sequences, and distinct channel/symbol estimation techniques. In [32], the devices send modulated polar codewords directly. On the other hand, FASURA spreads every coded symbol before transmission. The hope is that the spreading sequences can mitigate interference during the decoding process, especially when the number of active devices is large.

Furthermore, we borrow the idea from the past, of using estimates of the data as pilots to estimate the channel. We call this noisy pilot channel estimation (NOPICE). The idea is to estimate the channel using both pilots and preliminary decisions about coded symbols. Although some preliminary symbol decisions are erroneous, the overall impact of this approach is better performance. This estimation technique can be embedded in an iterative loop and, when parameters are picked judiciously, it leads to progressively better channel estimates. It should be noted that NOPICE can be removed from the FASURA algorithm at the expense of some performance loss, yet with a corresponding reduction in computational complexity. The proposed architecture outperforms the scheme put forth by Fengler et al. [32] and by Ahmadi et al. [33] when the number of antennas at the base station is 50 and 100. For example, for 500 active devices and 50 antennas, the difference in E_{b}/N_0 is around 3dB and 1dB, respectively. This makes FASURA the state-of-the-art for MIMO URA in certain practical regimes. Also, to characterize its possibilities under more practical scenarios, we show the performance of our scheme when the number of antennas is small.

C. Notation

Throughout, $\mathbb{Z}_+$ and $\mathbb{C}$ refer to the non-negative integers and complex numbers, respectively. We use $[n]$ to denote

$\{1, 2, \ldots, n\}$. We employ boldface lowercase $\mathbf{a}$ and boldface uppercase letters $\mathbf{A}$ to indicate vectors and matrices. The complex conjugate transpose of $\mathbf{A}$ is denoted by $\mathbf{A}^*$, and the transpose of a matrix $\mathbf{A}$ by $\mathbf{A}^{\mathrm{T}}$. Sets are labeled with calligraphic letters, e.g., $\mathcal{A}$. We also adopt programming style notation with $\mathbf{A}[:, t]$ and $\mathbf{A}[k, :]$ representing the tth column and kth row of $\mathbf{A}$.

II. System Model

In the wireless network considered, K_{tot} devices are presented, but only a small subset of K users ($K \ll K_{\mathrm{tot}}$) are active. We assume that K is known at the receiver. Without loss of generality, we label active users using integers from the set $\mathcal{K} = [K]$. Although slot-based schemes reduce the inter-user interference and the decoding complexity, it necessitates coordination. Thus, we allow the users to use the same total n complex channel uses. The objective of each user is to transmit a B-bit message to a common destination. Also, every user is equipped with a single antenna, whereas the access point features M antennas. The channel between each user and the base station is modeled as quasi-static Rayleigh fading, whereby channel coefficients remain fixed during the transmission of an entire codeword. In other words, the coherence time is longer than n symbol transmissions. Furthermore, the receive antennas at the access point are located far enough from one another, and so are devices, to create KM independent channel realizations.

Let the B-bit message of user k be denoted by $\mathbf{m}_k$. The encoded and modulated signal (input to the channel) corresponding to $\mathbf{m}_k$ is defined as $\mathbf{x}_k = \mathcal{E}(\mathbf{m}_k) \in \mathbb{C}^n$, where $\mathcal{E}(\cdot)$ is the encoding function described in Section III-A. Note that the encoding function is independent of a user's identity, which is one of the defining characteristics of a URA system. Then, the received signal at the M antennas takes the form

$$\mathbf{Y} = \sum_{k=1}^{K_{\mathrm{tot}}} \delta_k \mathbf{x}_k \mathbf{h}_k^{\mathrm{T}} + \mathbf{Z} = \sum_{k \in \mathcal{K}} \mathbf{x}_k \mathbf{h}_k^{\mathrm{T}} + \mathbf{Z}, \tag{1}$$

where the indicator function δ_k is introduced and takes value one when user k is active, and zero otherwise. Vector $\mathbf{x}_k = \mathcal{E}(\mathbf{m}_k)$ is the channel input of the kth user. The channel coefficients between user k and the M received antennas are captured by $\mathbf{h}_k \in \mathbb{C}^M$. Since we consider Rayleigh fading with no correlation between the KM channels, the elements of each $\mathbf{h}_k$ are independent, circularly symmetric complex Gaussian random variables with mean zero and unit variance. The additive noise component $\mathbf{Z} \in \mathbb{C}^{n \times M}$ is a matrix with independent and identically distributed (i.i.d.) entries, each drawn from a circularly symmetric complex Gaussian distribution $\mathcal{CN}(0, \sigma_z^2)$. The maximum power that a user can transmit is equal to 1, and thus every transmit signal must satisfy power constraint $\frac{1}{n}\|\mathbf{x}(\mathbf{m}_k)\|^2 \leq 1, \ \forall k \in [K]$. As a result, we can define the energy per bit to noise power spectral density ratio of the system by

$$\frac{E_{\mathbf{b}}}{N_0} = \frac{\|\mathbf{x}(\mathbf{m}_k)\|^2}{B\sigma_z^2} \leq \frac{1}{R\sigma_z^2}.$$

where $R = B/n$ is the rate of the coding scheme. The access point receives the transmitted signals corrupted by the fading channel and additive noise, i.e., $\mathbf{Y}$, and aims to produce a set $\hat{\mathcal{K}}$ of candidate messages with cardinality at most K. In some of the related works, system performance is evaluated in terms of the per-user probability of error (PUPE) [2]. We note that PUPE is the same as the missed detection probability in a detection problem. In this article, in a manner akin to related papers, e.g., [32], we consider the probability of missed detection P_{md} and the probability of false alarm P_{fa}. This makes our results stronger. For the problem we are interested in, these two error probabilities are given by

$$\mathrm{P}_{\mathrm{md}} = \mathbb{E}\left[\frac{1}{|\mathcal{K}|}\sum_{\mathbf{m}_k \in \mathcal{K}} \Pr\left[\mathbf{m}_k \notin \hat{\mathcal{K}}\right]\right],$$

$$\mathrm{P}_{\mathrm{fa}} = \mathbb{E}\left[\mathbf{1}_{\{|\hat{\mathcal{K}}| \neq 0\}}\frac{1}{|\hat{\mathcal{K}}|}\sum_{\hat{\mathbf{m}}_k \in \hat{\mathcal{K}}} \Pr\left[\hat{\mathbf{m}}_k \notin \mathcal{K}\right]\right]$$

where $\mathcal{K}$ and $\hat{\mathcal{K}}$ denote the set of messages and the set of the recovered messages , respectively, and the expectation is taken over the randomness of the fading process, the noise process, and the relevant algorithmic components. The number of messages declared by the decoder is captured by random variable $\hat{K}$. We want to point out that the receiver includes some constraints that drive the access point to output $\hat{K}$ messages instead of K, where $\hat{K} \leq K$. To evaluate the performance of FASURA, we define the error rate P_{e} to be the sum of the two types of probability introduced above,

$$\mathrm{P}_{\mathrm{e}} = \mathrm{P}_{\mathrm{md}} + \mathrm{P}_{\mathrm{fa}}.$$

We fix system parameters B, n, K, M, and target error ε. The objective of this work is to construct a communication scheme that minimizes E_{b}/N_0 while also satisfying the constraint $\mathrm{P}_{\mathrm{e}} \leq \varepsilon$.

III. FASURA

The encoding process and the decoding strategy are described in this section. Simple examples are provided to give some intuition on the operation of FASURA.

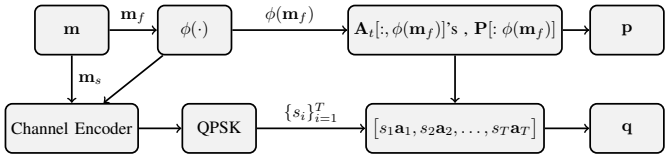

Fig. 1. A synopsis of the encoding process.

A. The Encoder

Each user k in the set $\mathcal{K}$ aims to transmit a message $\mathbf{m}_k$ to the access point. Following the URA framework, the encoding process is independent of the identity of the user, and the signal to the channel is only associated with the message $\mathbf{m}_k$. Since the encoder for each user is identical, we describe the functionality of the encoder for a generic message $\mathbf{m}$. The same steps are repeated by all the active devices. We follow the approach found in [35], [15], [32] and we split the B-bit message into two parts $\mathbf{m}_f$ and $\mathbf{m}_s$ of lengths B_f bits

and B_s bits, respectively. Since the encoding process of $\mathbf{m}_s$ depends on $\mathbf{m}_f$, FASURA encodes $\mathbf{m}_f$ first. The two parts $\mathbf{m}_f$ and $\mathbf{m}_s$ are encoded as follows.

1) Encoding $\mathbf{m}_f$: The encoding process of the first B_f bits is a mapping from $\{0,1\}^{B_f}$ to a set of random vectors. Let $\mathcal{A} = \{\mathbf{A}_t\}_{t=1}^{T}$ denote a collection of $T = n_c/2$ spreading matrices indexed by time t, where $\mathbf{A}_t \in \mathbb{C}^{L\times J}$. Note that n_c is the length of the code and L and J are the length and the number of spreading sequences, respectively. Each coefficient of the matrices in $\mathcal{A}$ is drawn from a complex Gaussian distribution with zero mean and variance one, i.e. $a_t^{i,j} \sim \mathcal{CN}(0,1)$, where $a_t^{i,j}$ is the ith, jth element of $\mathbf{A}_t$. Then the columns of matrix $\mathbf{A}_t$ are normalized to have second norm equal to $\sqrt{L}$, since they will be used in conjunction with Quadrature Phase Shift Keying (QPSK) symbols chosen from $\left\{\pm\frac{1}{\sqrt{2}} \pm \frac{j}{\sqrt{2}}\right\}$. Then the product has energy L. We view a column of $\mathbf{A}_t$ as spreading sequences of length L, which can be utilized at time t. This implies that the spreading sequences are time-varying, which helps to decrease the correlation between the users. In other words, assume that the spreading sequences are time-invariant, and imagine that Users 1 and 2 pick $\mathbf{a}(1)$ and $\mathbf{a}(2)$, where $\mathbf{a}(i)$ is the i-th column of $\mathbf{A}$. If the correlation $\mathbf{a}(1)^*\mathbf{a}(2)$ is high, then these two users are not separated well. On the other hand, by using time-variant spreading sequences, the probability that all pairs of spreading sequence have high correlation is small. Note that for a fixed t, there are $J = 2^{B_f}$ possible spreading sequences. Similarly, we define $\mathbf{P} \in \left\{\pm\frac{1}{\sqrt{2}} \pm \frac{j}{\sqrt{2}}\right\}^{n_p\times J}$ to be a matrix whose columns are possible pilot sequences. The selection of spreading sequences and pilots is performed using a function $\phi : \{0,1\}^{B_f} \to [J]$. The function $\phi(\cdot)$ is a bijection that maps $\mathbf{m}_f$ to an index in $[J]$. This is done by computing the decimal representation of $\mathbf{m}_f$. An important aspect of the proposed scheme is that $J \ll K_{\text{tot}}$, which is critical in limiting the complexity at the decoder. We emphasize that there is no guarantee that every active user will select a distinct sequence from an orthogonal set. Instead of assigning unique spreading sequences and a pilot sequence to each user in the network, we apportion only J different combinations of spreading sequences and pilot sequences. The active users pick sequences at random from a non-orthogonal set and there is a positive probability that two users pick the same sequences. To minimize the probability of this event, we can increase the length of $\mathbf{m}_f$.

To summarize, a user employs the series of spreading sequences $\{\mathbf{A}_t[:,\phi(\mathbf{m}_f)]\}_{t=1}^{T}$, along with pilot sequence $\mathbf{P}[:,\phi(\mathbf{m}_f)]$ to send the first part of the message $\mathbf{m}_f$. Thus, the overall encoding function for $\mathbf{m}_f$ can be described as follows,

$$g(\mathbf{m}_f) \to \{\mathbf{A}_t[:,\phi(\mathbf{m}_f)]\}_{t=1}^{T} \cup \{\mathbf{P}[:,\phi(\mathbf{m}_f)]\}.$$

The example below illustrates the encoding procedure.

Example 1. *Consider the encoding process of the first part of the message. Let this message be $\mathbf{m}_f = [0\ 1\ 1]$, then there are $J = 2^3 = 8$ possible pilot and spreading sequences in each set $\mathbf{A}_t$. Furthermore, let $T = 4$, thus there are 4 matrices in $\mathcal{A}$. By applying $\phi(\cdot)$ to $\mathbf{m}_f$ the 3th column of $\mathbf{P}$, $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$ and $\mathbf{A}_4$ is selected. As a result, $\mathbf{m}_f$ is mapped to a set of vectors as follows*

$$g(\mathbf{m}_f) \to \{\mathbf{P}[:,3], \mathbf{A}_1[:,3], \mathbf{A}_2[:,3], \mathbf{A}_3[:,3], \mathbf{A}_4[:,3]\}.$$

The encoding procedure for this example is illustrated in Fig. 2.

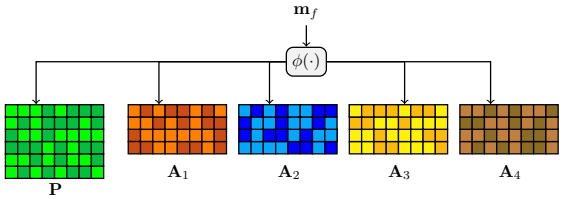

Fig. 2. Encoding procedure of $\mathbf{m}_f$ using a toy example. The decimal representation of $\mathbf{m}_f$ is three, and the third column of all matrices is selected.

2) Encoding $\mathbf{m}_s$: The second part of the message, $\mathbf{m}_s$, is first encoded using a cyclic redundancy check (CRC) code in order to leverage a list polar decoder at the receiver side. Then the CRC codeword[1] of length $B_c = B_s + B_{crc}$, is mapped to an n_c binary vector by using an (n_c, B_c) polar code with $n_c - B_c$ frozen bit positions. Instead of fixing the values of the frozen positions to be all zeros, we construct at random a matrix $\mathbf{F} \in \{0,1\}^{n_c-B_c\times J}$. Then an active user chooses the $\phi(\mathbf{m}_f)$th column of $\mathbf{F}$ to act as the frozen bits during the polar encoder. Note that by choosing different values for the frozen positions, two users with different first part of the message, i.e. $\mathbf{m}_{k,f} \neq \mathbf{m}_{j,f},\ k \neq j$ select two different cosets of a polar code. This is an additional level of error detection to identify false alarms when estimating the active spreading sequences at the receiver. Let $\mathbf{c} \in \{0,1\}^{n_c}$ be the output of the polar encoder, and define $\mathcal{I} = \{\pi_1(\cdot), \ldots, \pi_J(\cdot)\}$ be a set of J interleavers. Then another way to increase the probability of identifying false alarms during decoding is obtained by permuting the resulting polar codeword $\mathbf{c}$ using $\pi_{\phi(\mathbf{m}_f)}(\cdot)$. The ensuing vector $\tilde{\mathbf{c}} = \pi_{\phi(\mathbf{m}_f)}(\mathbf{c})$ is modulated using QPSK to obtain vector $\mathbf{s}$ of length $T = n_c/2$. Finally, to reduce the interference between the users during the decoding procedure, the tth symbol of $\mathbf{s}$ is spread using the $\phi(\mathbf{m}_f)$th column of $\mathbf{A}_t$. We express the codeword of $\mathbf{m}_s$ as $\mathbf{q}$, given by

$$\mathbf{q}(\mathbf{m}_f, \mathbf{m}_s) = \begin{bmatrix} s_1\mathbf{a}_1^{\mathrm{T}} & s_2\mathbf{a}_2^{\mathrm{T}} & \cdots & s_T\mathbf{a}_T^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \tag{2}$$

where $\mathbf{a}_t = \mathbf{A}_t[:,\phi(\mathbf{m}_f)]$ is the $\phi(\mathbf{m}_f)$th column of $\mathbf{A}_t$, and s_t is the QPSK symbol at time t. With this procedure, the message $\mathbf{m}$ is mapped to the channel input signal as the concatenation of the pilot sequence $\mathbf{p}(\mathbf{m}_{k,f})$ and spread codeword $\mathbf{q}(\mathbf{m}_{k,f}, \mathbf{m}_{k,s})$. In a nutshell, when the message of user k is $\mathbf{m}_k = (\mathbf{m}_{k,f}, \mathbf{m}_{k,s})$, the signal sent by this user becomes

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{p}^{\mathrm{T}}(\mathbf{m}_{k,f}) & \mathbf{q}^{\mathrm{T}}(\mathbf{m}_{k,f}, \mathbf{m}_{k,s}) \end{bmatrix}^{\mathrm{T}}$$

where $\mathbf{p}(\mathbf{m}_{k,f}) = \mathbf{P}[:,\phi(\mathbf{m}_{k,f})]$. Note that the parameters are chosen such that $n_p + L\frac{n_c}{2} = n$. Also, we observe that $\|\mathbf{x}\|^2 = n$.

[1]We refer to the concatenation of sequence $\mathbf{m}_s$ and the CRC bits, i.e. the output of the CRC encoder, as the CRC codeword.

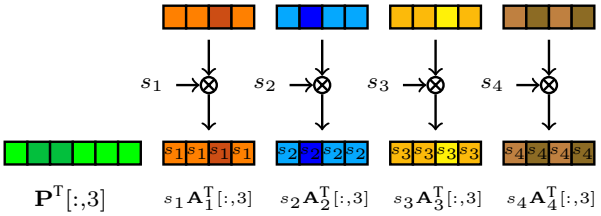

Fig. 3. Construction of the channel input, $\mathbf{x}$, using the information of the toy Example 1 and Example 2.

Example 2. *Consider the same parameters as in Example 1, and let* $\mathbf{s}$ *be the output of the QPSK modulator. Then the four QPSK symbols are spread using the* 3*rd column of* $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ *and* $\mathbf{A}_4$*, respectively, to create*

$$\mathbf{q} = \begin{bmatrix} s_1\mathbf{A}_1^T[:,3] & s_2\mathbf{A}_2^T[:,3] & s_3\mathbf{A}_3^T[:,3] & s_4\mathbf{A}_4^T[:,3] \end{bmatrix}^T.$$

Then the input signal to the channel is given as

$$\mathbf{x} = \begin{bmatrix} \mathbf{P}^T[:,3] & s_1\mathbf{A}_1^T[:,3] & s_2\mathbf{A}_2^T[:,3] & s_3\mathbf{A}_3^T[:,3] & s_4\mathbf{A}_4^T[:,3] \end{bmatrix}^T$$

Figure 3 shows the procedure to construct the channel input $\mathbf{x}_k$.

The encoding procedure is illustrated in Fig. 1. The system model of (1) can be written as the concatenation of

$$\mathbf{Y}_p = \mathbf{P}_a\mathbf{H}_a + \mathbf{Z}_p \qquad \text{and} \qquad \mathbf{Y}_q = \mathbf{Q}_a\mathbf{H}_a + \mathbf{Z}_q \tag{3}$$

or,

$$\begin{bmatrix} \mathbf{Y}_p \\ \mathbf{Y}_q \end{bmatrix} = \begin{bmatrix} \mathbf{P}_a \\ \mathbf{Q}_a \end{bmatrix} \mathbf{H}_a + \begin{bmatrix} \mathbf{Z}_p \\ \mathbf{Z}_q \end{bmatrix}$$

where subscript a indicates sub-matrices with active columns only. For example if there are only 5 active users with decimal representation of their $\mathbf{m}_f$'s to be 2, 123, 20, 98 and 7, then $\mathbf{P}_a = \mathbf{P}[:, \{2, 123, 20, 98, 7\}]$. That is, the kth column of $\mathbf{P}_a$ is $\mathbf{P}[:, \phi(\mathbf{m}_{k,f})]$ and the kth column of $\mathbf{Q}_a$ is $\mathbf{q}_k$.

B. The Decoder

Our objective here is to construct an iterative receiver that aims to recover the messages transmitted by the active users. We use techniques from estimation theory, such as MMSE, LMMSE, and energy detection (order statistics), to estimate and detect various parameters that are unknown to the receiver. Additionally, a list-polar decoder performs channel decoding. We provide a big-picture of the decoding process, and a more detailed description of each part in the following subsections.

Recall that in a traditional multiple access channel, the base station assigns a pilot sequence to every user. As a result, the first step of the decoding process is to estimate the channel. On the other hand, in the URA setting, the pilot sequences cannot be assigned to the users beforehand which means active columns of $\mathbf{P}$ must be identified first. We use energy detection to accomplish this. Then an MMSE estimator is utilized to estimate the channel coefficients. The channel estimates, combined with the recovered spreading sequences, are then used to form an LMMSE to estimate the QPSK symbols transmitted by the active users. We introduce the NOPICE block, which aims to improve the channel estimates and thus the symbol estimates. This idea has been used in the past literature, i.e. [36], but also recently [37] in the fifth generation of mobile communications (5G) channel and symbol estimation. Afterward, a polar list-decoder outputs the most likely messages. In the end, a successive interference canceller subtracts the estimated contribution of the decoded messages, and the procedure starts anew on the residual signal, iterating until the decoder cannot output a new message. The main building blocks of the algorithm are highlighted in Fig. 4. We elaborate on individual components below.

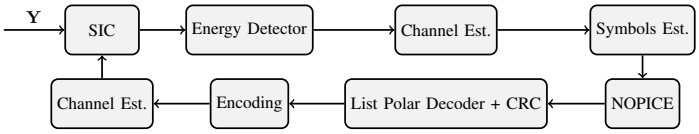

Fig. 4. This notional diagram outlines the message recovery process at the receiver. This iterative scheme includes the identification of selected spreading sequence, channel/symbol estimation, and polar decoding.

1) The Energy Detector: An energy detector is used to determine the active spreading sequences. This can be obtained by correlating the received signal $\mathbf{Y}$ with the corresponding columns in $\mathbf{P}$ and $\{\mathbf{A}_t\}_{t=1}^T$ and computing the statistic λ_j given by

$$\lambda_j = \|\mathbf{P}^*[:,j]\mathbf{Y}_p\|^2 + \sum_{t=1}^{T} \|\mathbf{A}_t^*[:,j]\mathbf{Y}_q[\mathbf{n}_t,:]\|^2, \forall j \in [J], \tag{4}$$

where $\mathbf{n}_t = [(t-1)L + 1 : tL]$. The statistic λ_j is computed for all $j \in [J]$; then, the energy detector outputs the indices corresponding to the largest K values. Let $\hat{\mathcal{M}}_f$ denote the set of $\mathbf{m}_f$'s that correspond to the K largest values of λ_j's. After this step, the estimate of the first part of the message of users is recovered. This implies that, the active columns of $\mathbf{P}$ and $\{\mathbf{A}_t\}_{t=1}^T$, the values of the frozen positions, and the interleavers are considered known for the purpose of the algorithmic progression. The next step is to recover the second part of every message.

2) Channel Estimation: To recover the symbols, the channel coefficients must be estimated first. Since the channel is complex Gaussian, an MMSE filter can be derived to estimate the SIMO channels of the candidate users included in $\hat{\mathcal{K}}$ by minimizing the mean square error. Using active pilots obtained by the energy detector step, the MMSE filter is given as

$$\mathbf{W}_1 = \left(\mathbf{I}_{\hat{K}} + \frac{\hat{\mathbf{P}}^*\hat{\mathbf{P}}}{\sigma_z^2} \right)^{-1} \frac{\hat{\mathbf{P}}^*}{\sigma_z^2}, \tag{5}$$

where $\hat{\mathbf{P}} = \mathbf{P}[:, \phi(\hat{\mathcal{M}}_f)]$. It should be noted that the filter is independent of the antenna index, and the estimated channel coefficients between the $\hat{K} = |\hat{\mathcal{K}}|$ users and the M antennas are taken to be

$$\hat{\mathbf{H}} = \mathbf{W}_1\mathbf{Y}_p, \tag{6}$$

where $\hat{\mathbf{H}}$ is a $\hat{K} \times M$ matrix containing all the estimated coefficients. The next step is to estimate the QPSK symbols.

3) Symbol Estimation: After the energy detector and channel estimation, the information needed for symbol estimation is available. Recall that each symbol is spread with a different

spreading sequence, to reduce the interference between users. The received signal (of length L) for each symbol t is

$$\begin{aligned}\mathbf{Y}_q[\mathbf{n}_t, m] &= \mathbf{A}_t \,\mathrm{diag}(\mathbf{r}_t)\mathbf{H}[:, m] + \mathbf{Z}[\mathbf{n}_t, m] \\ &= \mathbf{A}_t \,\mathrm{diag}\left(\mathbf{H}[:, m]\right)\mathbf{r}_t + \mathbf{Z}[\mathbf{n}_t, m]\end{aligned}$$

where $\mathbf{r}_t = (s_{t,1}, s_{t,2}, \ldots, s_{t,K})$ are the symbols of the users at time t. Note that noisy observations of the same QPSK symbols $\mathbf{r}_t$ are received from M different antennas. As a result, to improve the symbol estimates, we stack the columns of $\mathbf{Y}_q$, which corresponds to different antennas, to obtain

$$\underbrace{\begin{bmatrix}\mathbf{Y}_q[\mathbf{n}_t, 1] \\ \vdots \\ \mathbf{Y}_q[\mathbf{n}_t, M]\end{bmatrix}}_{\mathtt{y}_t \in \mathbb{C}^{LM\times 1}} = \underbrace{\begin{bmatrix}\mathbf{A}_t \,\mathrm{diag}\left(\mathbf{H}[:, 1]\right) \\ \vdots \\ \mathbf{A}_t \,\mathrm{diag}\left(\mathbf{H}[:, M]\right)\end{bmatrix}}_{\mathtt{B}_t \in \mathbb{C}^{LM\times K}} \underbrace{\mathbf{r}_t}_{K\times 1} + \underbrace{\begin{bmatrix}\mathbf{Z}_q[\mathbf{n}_t, 1] \\ \vdots \\ \mathbf{Z}_q[\mathbf{n}_t, M]\end{bmatrix}}_{\mathtt{z}_t \in \mathbb{C}^{LM\times 1}}.$$

As a consequence, we can apply an LMMSE estimator to the vectorized received signal

$$\mathtt{y}_t = \mathtt{B}_t\mathbf{r}_t + \mathtt{z}_t. \tag{7}$$

Note that active spreading sequences and channel coefficients are not known a priori and the estimation error is not zero. We can modify (7) to account for the estimation inaccuracies by writing,

$$\mathtt{y}_t = \hat{\mathtt{B}}_t\mathbf{r}_t + \left(\mathtt{B}_t - \hat{\mathtt{B}}_t\right)\mathbf{r}_t + \mathtt{z}_t. \tag{8}$$

In view of the last equation, one could take into consideration the interference term and increase the effective noise variance. However, for the simulation parameters we use, the performance of our scheme essentially remains unaffected when the second term is neglected. Disregarding the interference term, the LMMSE filter for (8) takes the form

$$\mathtt{W}_t = \left(\mathtt{I}_{\hat{K}} + \frac{\hat{\mathtt{B}}_t^*\hat{\mathtt{B}}_t}{\sigma_z^2}\right)^{-1}\frac{\hat{\mathtt{B}}_t^*}{\sigma_z^2}. \tag{9}$$

One of the disadvantages of using time-varying spreading sequences is that the LMMSE filter must be computed T times. This introduces complexity to our decoder. Still, this extra cost improves the overall performance. By computing the LMMSE filter in (9), one can estimate the QPSK symbols at time t as follows

$$\hat{\mathbf{r}}_t = \mathtt{W}_t\mathtt{y}_t, \;\; \forall \quad t \in [T].$$

4) NOPICE: One of the salient features of our scheme is the estimation of the channel using, not only the original pilots, but also the temporary coded decisions. This idea has been developed before, known as *data-aided channel estimation* and for example [36], use this method to increase the performance of their schemes. Here we provide our view of this idea by defining the term *Temporary Coded Decisions*.

Definition 3. *(Temporary Coded Decisions) After the third step above is completed, the symbols across time corresponding to the same user are aggregated in the form of noisy codewords. These soft estimates are then passed to a polar list-decoder. (The channel decoder block is explained in the next section.) The decoder produces a set of most likely messages. After the CRC step, one of them is re-encoded and QPSK modulated. The outcome of this process is called* temporary coded decisions.

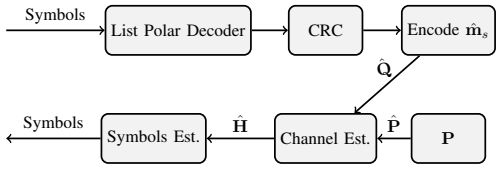

Fig. 5. The main functionalities of NOPICE. The channel is estimated using pilots and estimated symbols.

The main blocks in NOPICE to produce the temporary coded decisions for all users are a polar decoder, an encoder, and a modulator. Also, the channel and the symbol estimates computed in (9) are used to produce the new estimates.

Let $\hat{\mathbf{s}}_k = [\hat{s}_{1,k} \;\; \hat{s}_{2,k} \;\; \ldots \;\; \hat{s}_{T,k}]$ be the temporary coded decisions of user k, where $\hat{s}_{t,k}$ is the tth symbol of user k. Note that the pilots, the spreading sequences, and the symbol candidates for this user are available, thus we can construct its channel input,

$$\hat{\mathbf{x}}_k = \left[\mathbf{p}^{\mathrm{T}}(\hat{\mathbf{m}}_{k,f}) \quad \mathbf{q}^{\mathrm{T}}(\hat{\mathbf{m}}_{k,f}, \hat{\mathbf{m}}_{k,s})\right]^{\mathrm{T}}$$

where $\hat{\mathbf{q}}_k$ can be constructed as (2) by replacing the true values with the candidate symbols. The idea behind the NOPICE is the following. Assuming that $\hat{\mathbf{x}}_k$'s are decoded correctly, we use these vectors as known signals while re-estimating the channel coefficients. Note that some of the symbols in $\hat{\mathbf{s}}_k$ may be erroneous. Nevertheless, the hope is that most candidate symbols are correct and the estimates of the channel can be improved. Consider the received signal at antenna m.

$$\mathbf{Y}[:, m] = \mathbf{X}\mathbf{H}[:, m] + \mathbf{Z}[:, m]$$

Due to the errors that can occur in the symbols $\hat{\mathbf{s}}_k$, we can express the receive signal as follows

$$\mathbf{Y}[:, m] = \hat{\mathbf{X}}\mathbf{H}[:, m] + \left(\mathbf{X} - \hat{\mathbf{X}}\right)\mathbf{H}[:, m] + \mathbf{Z}[:, m].$$

Once again, by ignoring the interference term, we can apply the following MMSE filter to estimate the channel coefficients,

$$\mathbf{W}_2 = \hat{\mathbf{X}}^*\left(\hat{\mathbf{X}}^*\hat{\mathbf{X}} + \mathbf{I}_{\hat{K}}\right)^{-1}. \tag{10}$$

As before, the MMSE filter in (10) is independent of the antenna index and, hence can be applied directly to observation $\mathbf{Y}$ to obtain the channel estimates as follows:

$$\hat{\mathbf{H}} = \mathbf{W}_2\mathbf{Y}. \tag{11}$$

Then the algorithm proceeds with the re-estimation of the QPSK symbols by the procedure described before, by utilizing the updated channel estimates. Figure 5 shows a block diagram of this process. Note that the input and the output of NOPICE are the symbols, hence one can remove this block without changing any other part of FASURA to reduce complexity, at the cost of some performance loss.

5) The Channel Decoder: A single-user list polar decoder, and a CRC validation step are the two main blocks of the channel decoder. Since this block is independent of the user index, we describe the process for user k, with the understanding that this step is repeated for all candidate users in $\hat{\mathcal{K}}$.

Also, it is easy to see that this procedure can be parallelized in a straightforward manner, thereby reducing execution time. Let $\hat{\mathbf{s}}_k$ be the estimates of the kth user, i.e., the output of the symbol estimation block, and construct the estimates of the bits as follows:

$$\hat{\mathbf{c}}_k = \left[\mathrm{Re}(\hat{\mathbf{s}}_k)\ \mathrm{Im}(\hat{\mathbf{s}}_k)\right], \tag{12}$$

where $\mathrm{Re}(\hat{\mathbf{s}}_k)$ and $\mathrm{Im}(\hat{\mathbf{s}}_k)$ are the real and imaginary part of $\hat{\mathbf{s}}_k$, respectively. Then the input to the channel decoder is $\hat{\mathbf{c}}_k$ and the frozen values $\mathbf{F}[:, \phi(\hat{\mathbf{m}}_f)]$. We consider a sum-product polar list decoder with a list of length n_L. As a result, after completion, the decoder outputs a list of the n_L most likely messages. The next step is to use the CRC decoder to pick the messages that meet the constraint. If more than one possible message satisfies the CRC constraint, the most likely message within the pruned list is returned by this block. Let $\hat{\mathbf{m}}_{k,s}$ be this message. The final step in the channel decoder block is called *hard decision decoder*. We use this block to reduce the number of false alarms as follows. First compute the hamming distance $\mathrm{d}_H(\cdot,\cdot)$ between the codeword corresponding to $\hat{\mathbf{m}}_{k,s}$ and hard decisions (HD) of $\hat{\mathbf{c}}_k$, i.e.

$$g = \mathrm{d}_H\Big(\mathrm{encode}(\hat{\mathbf{m}}_{k,s}), \mathrm{HD}(\hat{\mathbf{c}}_k)\Big) \tag{13}$$

where $\mathrm{encode}(\cdot)$ is the polar encoder function and $\mathrm{HD}(\cdot)$ is the hard decoder defined by

$$x = \begin{cases} 1, & \text{if } \hat{\mathbf{c}}_k(i) \le 0 \\ 0, & \text{if } \hat{\mathbf{c}}_k(i) > 0 \end{cases}$$

with $\hat{\mathbf{c}}_k(i)$ the ith estimated value. When $g > \gamma$, the recovered message $\hat{\mathbf{m}}_{k,s}$, is discarded. Note that γ is a hyperparameter which can be tuned to achieve a trade-off between the missed-detection and false-alarm probabilities.

Remark 4. *(List-Decoder and NOPICE) Under normal operation, the polar/CRC decoder returns the most likely consistent message. Yet, it is possible that the list contains no codewords that fulfill their CRC constraints. When the list-decoder is used during the NOPICE block, the most likely and, necessarily, inconsistent message is returned. The hope is that, although the message is guaranteed to be wrong, a portion of the encoded symbols can still be correct.*

6) SIC: An iteration of this composite iterative algorithm is done by subtracting the interference of the decoded users through successive interference cancellation (SIC) step. For this part, the estimated channel input $\hat{\mathbf{x}}$ for all the users whose most likely messages met the CRC validation process is calculated and channel estimation similar to (10) is applied to obtain better estimates. Let $\hat{\mathbf{X}}$ be the matrix with columns the channel inputs of the decoded users, and let $\hat{\mathbf{H}}$ be the matrix each row of which presents the SIMO channels between a user and the access point. Then, we subtract their contribution from the received signal and obtain the residual as

$$\mathbf{Y}_r = \mathbf{Y} - \hat{\mathbf{X}}\hat{\mathbf{H}}, \tag{14}$$

where $\mathbf{Y}_r$ is the residual. It should be mentioned that the SIC step is crucial for the performance of FASURA. If the codewords and the channel estimates are completely wrong, severe error propagation is introduced. To avoid that, a series of constraints, e.g. CRC, to discard the codeword have been used. Nevertheless, even if the codewords are correct, the channel estimates can not be the true channel. As a result, error propagation exists but does not affect overall performance.

The algorithm progresses as described above with the first block to be the energy detector and $\mathbf{Y}_r$ as the input of this block. Note that if a total of $\hat{K} < K$ users are decoded at the end of an iteration, then in the next iteration the output of the energy detector is $K - \hat{K}$ indices. This process continues until the output list contains K messages or there is no improvement between two consecutive rounds of iterations.

Remark 5. *(Residual) After the first iteration, all the building blocks of the receiver use the residual. However, for all iterations, SIC uses the received signal and produces a new residual for the next iteration.*

C. Time Complexity

Before diving into simulation results, we dedicate this section to discuss the time complexity of our scheme. Let us recall that the number of spreading sequence J is a linear function of K, similarly $L = O(K)$, $T = O(B)$ and $n_L = O(B)$. We analyze only the most time consuming parts of the algorithm. Let us begin with the first step of FASURA, which is the energy detector. It is clear, from equation (4), that the time complexity is $\mathrm{TC}_{\mathrm{ED}} = O(MJ(n_p + TL))$, or equivalently $\mathrm{TC}_{\mathrm{ED}} = O(MKn_p + MK^2B)$. For all channel estimation steps, (6) and (11), the MMSE filter needs $O(K^3)$ operation to be computed, and the calculation of the channel estimates requires $O(Kn_pM)$. Similarly, the LMMSE filter in (9) requires $O(K^3)$, for each time t. As a result the total time complexity of symbol estimation is $\mathrm{TC}_{\mathrm{SE}} = O(K^3B) + O(BK^2M)$. It is easy to show that the NOPICE step is computed in $O(K^3 + KM(n_p + BK))$ steps. After the channel and symbol estimation, the list-polar decoder runs in $\mathrm{TC}_{\mathrm{PD}} = O(n_L n_c \log n_c) = O(B^2 \log B)$ for each user. Finally the SIC step needs $O(K^3 + KM(n_p + BK))$ steps. If $n_p = O(K \log K)$ and by combining everything together, the time complexity of our scheme is

$$\begin{aligned}\mathrm{TC}_{\mathrm{FASURA}} &= O(\rho(K^3B + MK^2(\log K + B) + KB^2 \log B)) \\ &= O(\rho(K^3 + MK^2 \log K)),\ \text{if } B \text{ is fixed}\end{aligned}$$

where ρ is the number of iteration of the receiver algorithm. In our simulation ρ is between 1–4.

IV. Simulation Results

In this section we investigate the performance of our scheme[2] by looking at some parts of the decoder. Also, at the end of this section, a comparison of our scheme with other schemes in the literature is provided.

[2]The source code for the FASURA communication scheme is available at https://github.com/EngProjects/mMTC

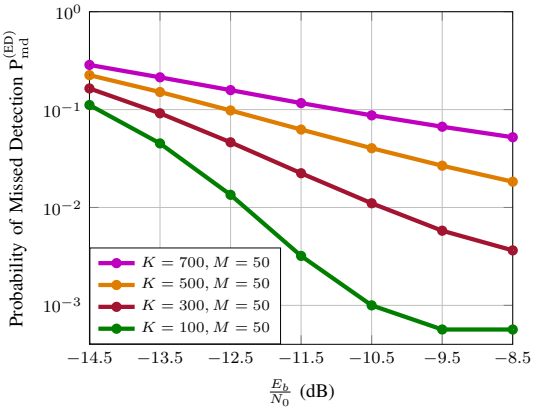

Fig. 6. Probability of Miss of the Energy Detector Algorithm at the first iteration of FASURA.

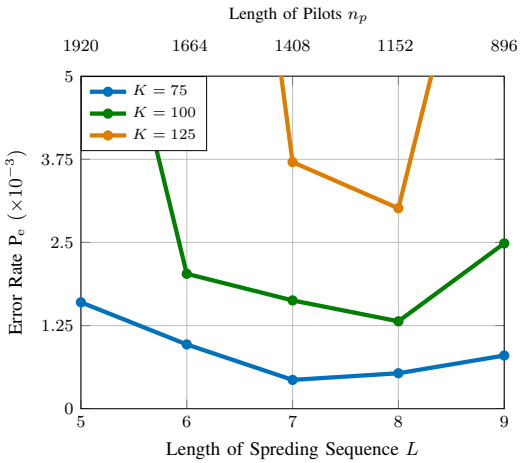

Fig. 7. Error Rate as a function of the length of the spreading sequences L, for $M = 4$ and SNR= 0dB.

A. The Energy Detector

As already noted, one of the most critical functions of any URA scheme is activity detection. It is a block that can introduce error propagation. In this part, the encoder aims to detect active users. We want to investigate the performance of the energy detector for different values $\frac{E_b}{N_0}$ and number of users. The performance metric is the missed detection and false alarm probabilities of the energy detector, i.e.,

$$\mathrm{P}_{\mathrm{md}}^{(ED)} = \frac{\mathbb{E}[n_{\mathrm{ms}}]}{K}, \quad \mathrm{P}_{\mathrm{fa}}^{(ED)} = \mathbb{E}\left[\frac{n_{\mathrm{fa}}}{\hat{K}}\right],$$

where n_{ms} and n_{fa} denote the number of misses and false alarms, respectively. Since, the energy detector outputs the largest K values, the two probabilities are equal ($\hat{K} = K$ is not random, and if n_{d}, indicates the number of detections, then $K - n_{\mathrm{d}} = n_{\mathrm{ms}} = n_{\mathrm{fa}}$)[3]. We consider the number of antennas to be $M = 50$, and the number of users to vary from 100 to 700. Figure 6 shows how the missed detection probability behaves for different values of $\frac{E_b}{N_0}$ and K. Since the number of spreading sequences, J, is fixed for all K, the probability of miss saturates as the SNR increases (see the case $K = 100$) because of the collisions.

B. Pilot Length - Length of Spreading Sequence Trade-off

Once the number of channel uses n and the length of the code n_c are fixed, then a designer can choose either the length of the pilots or the length of the spreading sequence, since $n = n_p + L\frac{n_c}{2}$. The trade-off between n_p and L is important, especially when the number of antennas is small and there is a need to separate the users in the code domain. As a result, we set the number of antennas to be $M = 4$. Note that for all simulations, where the number of antennas is four, we follow the same parameters as Liu and Wang [23] in order to compare FASURA with the proposed scheme in [23] later. This means the number of channel uses is $n = 2840$, the number of information bits is $B = 70$ and the SNR is defined as,

$$\mathrm{SNR} = \frac{\mathbb{E}\left[\mathbf{h}\mathbf{x}^{\mathrm{T}}\right]^2}{Mn\sigma^2} = \frac{1}{\sigma^2}$$

as adopted to our notation. Figure 7 demonstrates the trade-off between the length of the pilots, n_p and the length of the spreading sequences L as a function of the number of users. The SNR value is 0dB and the number of users take values from $\{75, 100, 125\}$. Note that, as the length of the spreading sequences increases, the number of pilots decreases. It is clear from the figure that as the number of users increases, the length of the spreading sequences, L, which achieves minimum error-rate increases, as expected. For example for 75 and 100 users the length of the spreading sequences is 7 and 8, respectively. This result implies that the degrees of freedom, from the antennas, are not enough, and there is a need of code-domain separation.

C. Small Number of Antennas

In this section, we want to evaluate the performance of FASURA when the number of antennas at the base station is relatively small. Particularly, we consider an access point with $M = 4$ antennas and we stick with the same parameters as Section IV-B, i.e. $B = 70$, $n = 2840$ and SNR = 0dB. We compare the performance of FASURA in the small number of antennas regime with the slot-based scheme proposed by Liu and Wang [23].

Figure 8 illustrates the performance of FASURA with and without NOPICE to the Slot-Based scheme by Liu and Wang [23]. It is clear that our scheme outperforms their approach. Interestingly, the performance of FASURA with and without NOPICE is roughly the same. This is expected since the scheme operates in a high SNR regime and the channel estimates using only the pilots are good enough to recover the QPSK symbols.

D. Massive MIMO

Next, we move to the Massive MIMO regime. We compare FASURA with the state-of-the-art schemes proposed by

[3]We define the two probabilities assuming that the probability of collision is zero. However, in our case this probability is greater than zero but very small. As a result, Fig. 6 illustrates an approximation of the missed detection probability.

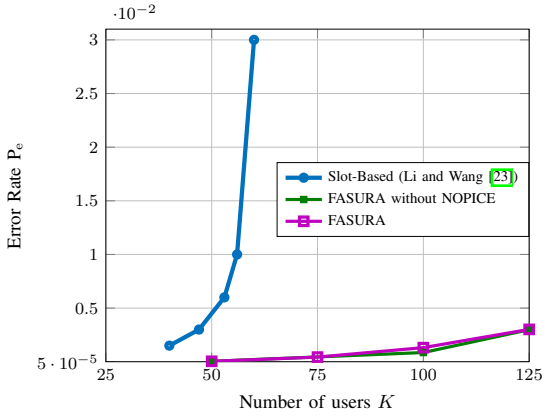

Fig. 8. Comparison between the FASURA and the scheme proposed in [23], for $B = 70$, $n = 2840$ and SNR = 0dB.

Fengler et al. [32] and Ahmadi et al. [33]. To ensure a fair comparison between the three URA communication schemes, we pick parameters for our system that match their reported implementation. Specifically, we choose $B = 100$ message bits, $n = 3200$ complex channel uses, and a target probability of error $\mathrm{P_e} \leq 0.05$. Beyond these constraints, the other parameters for FASURA are $n_p = 896$, $L = 9$, $n_c = 512$, $n_L = 64$ and $J = 2^{16}$. We randomly generate $T = \frac{n_c}{2} = 256$ spreading sequence matrices and one pilot matrix. Elements of the spreading sequence matrices are drawn independently and uniformly from the set $\left\{ \pm \frac{1}{2\sqrt{n}} \pm \frac{j}{2\sqrt{n}} \right\}$. Similarly, elements of the pilot matrix are generated independently and with equal probability from $\left\{ \pm \frac{1}{\sqrt{2n}} \pm \frac{j}{\sqrt{2n}} \right\}$. We employ 16 CRC bits, and the number of antennas at the base station takes values from the set $M \in \{50, 100\}$.

When the number of antennas is 50, Figure 9 plots the performance of FASURA, along with the communication scheme in [32] and [33]. Also, to motivate the use of the NOPICE block, we report the performance of FASURA with and without the NOPICE channel estimation technique. For the operational parameters studied and a user population exceeding 100 active devices, FASURA outperforms the scheme proposed in [32]. Interestingly, as the number of active users grows, the gap between the FASURA and the Pilot-Based scheme widens. For example, when the number of users goes from 100 to 500, the gap increases from 0.33 dB to 3.05 dB. On the other hand, the gain of FASURA compared to the scheme by Ahmadi and Duman remains roughly the same as the number of users increases. Furthermore, the presence of the NOPICE block seems to uniformly improve the performance of FASURA. Our proposed scheme also substantially outperforms the tensor-based modulation scheme in [29]. We should mention, however, that these benefits in terms of E_b/N_0 come at the expense of additional computations.

Figure 10 illustrates the performance of the three schemes, when the number of antennas is 100. Interestingly, the performance of the scheme in [33] is worse than the performance of the rest. For example, the gap between FASURA and the Orthogonal Pilots scheme ranges from 1.4–8.4dB. Furthermore, the Pilot-Based scheme outperforms both alternatives when the number of users is less than 200. Finally, the gain in terms of performance between FASURA with and without NOPICE is

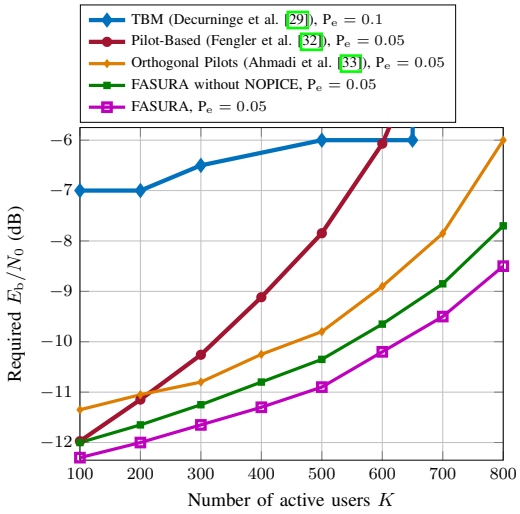

Fig. 9. Performance comparison between FASURA and other previously published schemes. The number of antennas at the base station is $M = 50$, users transmit $B = 100$ information bits each, and the total number of channel uses is $n = 3200$, and $\mathrm{P_e} \leq 0.05$.

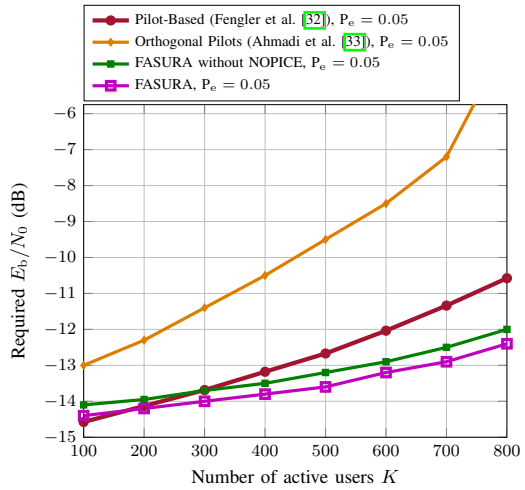

Fig. 10. Performance comparison of FASURA and other previously published schemes. The number of antennas at the base station is $M = 100$, users transmit $B = 100$ information bits each, and the total number of channel uses is $n = 3200$, and $\mathrm{P_e} \leq 0.05$.

clear.

Remark 6. *(Finite blocklength bounds for Massive Random Access) Gao et al. [38] recently derive achievability and converse bounds on the minimum energy-per-bit required for each active user to transmit B bits with blocklength n under a PUPE constraint, for the massive random access problem in MIMO quasi-static Rayleigh fading channels. In Fig. 1 in [38] the achievability and converse bounds on the minimum required energy-per-bit under the PUPE for the same parameters as our simulation are presented. Interestingly, FASURA performs near the achivability bound [38, Theorem 1] when $K < 400$.*

V. Conclusion

This article considers the MIMO unsourced random access problem on a quasi-static Rayleigh fading channel. We pro-

pose a novel communication scheme called FASURA that outperforms existing schemes in the massive MIMO setup for number of users exceeding 100 and various numbers of antennas at the base-station, albeit at the cost of increased complexity. FASURA splits the payload into two parts and encodes the first to a set of randomly generated pilots and spreading sequences. The remainder of the information bits are encoded using a polar code. Then the modulated coded symbols are spread using the spreading sequences. The receiver is equipped with multiple antennas, and the objective is to recover the transmitted messages. The decoding process includes pilot detection, channel and symbol estimation, polar list decoding, and successive interference cancellation. We explore a different way to perform channel estimation, which we call NOPICE, whereby after temporary decoding decisions have been made (Definition 3), the channel is re-estimated assuming that the decoded messages are accurate. This scheme is somewhat reminiscent of the *certainty equivalent principle* in control theory [39]. Also, in the communication literature is known as *data-aided channel estimation*. With NOPICE, the proposed scheme outperforms the pilot-based scheme recently published in [32], and the orthogonal pilot scheme in [33] when the number of users is more than 100 and $M = 50$. On the other hand, when the number of antennas is 100, the proposed scheme outperforms the scheme in [32] when $K > 200$. Also, through numerical simulations, we conducted a comparative study of FASURA with and without NOPICE. It seems that the approach adopted within the NOPICE block leads to uniformly better performance, with gains of 0.25 dB or more. Moreover, the behaviour of the scheme for different parameters was investigated through exhaustive simulations. For example, the trade-off between the length of the spreading sequence.

REFERENCES

[1] M. Gkagkos, K. R. Narayanan, J.-F. Chamberland, and C. N. Georghiades, "FASURA: A scheme for quasi-static massive MIMO unsourced random access channels," in *2022 IEEE 23rd International Workshop on Signal Processing Advances in Wireless Communication (SPAWC)*, Oulu, Finland, 04-06 July 2022, pp. 1–5.

[2] Y. Polyanskiy, "A perspective on massive random-access," in *IEEE International Symposium on Information Theory (ISIT)*, Aachen, Germany, 25-30 June 2017, pp. 2523–2527.

[3] A. Vem, K. R. Narayanan, J. Cheng, and J.-F. Chamberland, "A user-independent serial interference cancellation based coding scheme for the unsourced random access Gaussian channel," in *IEEE Information Theory Workshop (ITW)*, Kaohsiung, Taiwan, 06-10 November 2017, pp. 121–125.

[4] K.-H. Ngo, A. Lancho, G. Durisi, and A. G. i. Amat, "Massive uncoordinated access with random user activity," in *IEEE International Symposium on Information Theory (ISIT)*, Melbourne, Australia, 12-20 July 2021, pp. 3014–3019.

[5] V. K. Amalladinne, J.-F. Chamberland, and K. R. Narayanan, "A coded compressed sensing scheme for unsourced multiple access," *IEEE Transactions on Information Theory*, vol. 66, no. 10, pp. 6509–6533, July 2020.

[6] R. Calderbank and A. Thompson, "CHIRRUP: A practical algorithm for unsourced multiple access," *Information and Inference*, December 2019.

[7] A. Fengler, P. Jung, and G. Caire, "SPARCs for unsourced random access," vol. 67, no. 10, pp. 6894–6915, May 2021.

[8] V. K. Amalladinne, A. K. Pradhan, C. Rush, J.-F. Chamberland, and K. R. Narayanan, "Unsourced random access with coded compressed sensing: Integrating AMP and belief propagation," vol. 68, no. 4, pp. 2384–2409, April 2022. [Online]. Available: https://dx.doi.org/10.1109/TIT.2021.3136437

[9] V. K. Amalladinne, J. R. Ebert, S. Rini, J.-F. Chamberland, and K. R. Narayanan, "Coded demixing for unsourced random access," vol. 70, pp. 2972–2984, June 2022. [Online]. Available: https://dx.doi.org/10.1109/TSP.2022.3182224

[10] V. Shyianov, F. Bellili, A. Mezghani, and E. Hossain, "Massive unsourced random access based on uncoupled compressive sensing: Another blessing of massive MIMO," vol. 39, no. 3, pp. 820–834, August 2020. [Online]. Available: https://doi.org/10.1109/JSAC.2020.3019722

[11] A. Fengler, S. Haghighatshoar, P. Jung, and G. Caire, "Non-Bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver," vol. 67, no. 5, pp. 2925–2951, March 2021.

[12] V. K. Amalladinne, J. R. Ebert, J.-F. Chamberland, and K. R. Narayanan, "An enhanced decoding algorithm for coded compressed sensing with applications to unsourced random access," *Sensors*, vol. 22, no. 2, p. 676, January 2022.

[13] K. Andreev, P. Rybin, and A. Frolov, "Unsourced random access based on list recoverable codes correcting t errors," in *IEEE Information Theory Workshop (ITW)*, Kanazawa, Japan, 17-21 October 2021, pp. 1–6.

[14] S. Verdu, *Multiuser Detection*, 1st ed. USA: Cambridge University Press, 1998.

[15] A. K. Pradhan, V. K. Amalladinne, K. R. Narayanan, and J.-F. Chamberland, "Polar coding and random spreading for unsourced multiple access," in *IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[16] Z. Han, X. Yuan, C. Xu, S. Jiang, and X. Wang, "Sparse kronecker-product coding for unsourced multiple access," *IEEE Wireless Communications Letters*, vol. 10, no. 10, pp. 2274–2278, July 2021.

[17] A. Pradhan, V. Amalladinne, A. Vem, K. R. Narayanan, and J.-F. Chamberland, "A joint graph based coding scheme for the unsourced random access Gaussian channel," in *2019 IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, 09-13 December 2019, pp. 1–6.

[18] S. S. Kowshik and Y. Polyanskiy, "Quasi-static fading MAC with many users and finite payload," in *2019 IEEE International Symposium on Information Theory (ISIT)*, Paris, France, 07-12 July 2019, pp. 440–444.

[19] X. Chen, T.-Y. Chen, and D. Guo, "Capacity of Gaussian many-access channels," *IEEE Transactions on Information Theory*, vol. 63, no. 6, pp. 3516–3539, February 2017.

[20] S. S. Kowshik, K. Andreev, A. Frolov, and Y. Polyanskiy, "Energy efficient random access for the quasi-static fading MAC," in *2019 IEEE International Symposium on Information Theory (ISIT)*, Paris, France, 07-12 July 2019, pp. 2768–2772.

[21] ——, "Short-packet low-power coded access for massive MAC," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, 03-06 November 2019, pp. 827–832.

[22] K. Andreev, E. Marshakov, and A. Frolov, "A polar code based TIN-SIC scheme for the unsourced random access in the quasi-static fading MAC," in *IEEE International Symposium on Information Theory (ISIT)*, Los Angeles, CA, USA, 21-26 June 2020, pp. 3019–3024.

[23] J. Liu and X. Wang, "Unsourced multiple access based on sparse tanner graph – efficient decoding, analysis and optimization," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 5, pp. 1509–1521, January 2022.

[24] K. Andreev and A. Frolov, "On unsourced random access for the MIMO channel," in *2021 XVII International Symposium "Problems of Redundancy in Information and Control Systems" (REDUNDANCY)*, Moscow, Russian Federation, 25-29 October 2021, pp. 17–21.

[25] V. Shyianov, F. Bellili, A. Mezghani, and E. Hossain, "Massive unsourced random access based on uncoupled compressive sensing: Another blessing of massive MIMO," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 3, pp. 820–834, March 2021.

[26] V. K. Amalladinne, J.-F. Chamberland, and K. R. Narayanan, "Coded compressed sensing with successive cancellation list decoding for unsourced random access with massive MIMO," *arXiv preprint arXiv:2105.02185*, 2021.

[27] J. Liu and X. Wang, "Sparsity-exploiting blind receiver algorithms for unsourced multiple access in MIMO and massive MIMO channels," *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8055–8067, December 2021.

[28] C. R. Srivatsa and C. R. Murthy, "On the impact of channel estimation on the design and analysis of IRSA based systems," *arXiv preprint arXiv:2112.07242*, December 2021.

[29] A. Decurninge, I. Land, and M. Guillaud, "Tensor-based modulation for unsourced massive random access," *IEEE Wireless Communications Letters*, vol. 10, no. 3, pp. 552–556, March 2021.

[30] X. Xie, Y. Wu, J. Gao, and W. Zhang, "Massive unsourced random access for massive MIMO correlated channels," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, Taipei, Taiwan, 07-11 December 2020, pp. 1–6.
[31] X. Xie, Y. Wu, J. An, J. Gao, W. Zhang, C. Xing, K.-K. Wong, and C. Xiao, "Massive unsourced random access: Exploiting angular domain sparsity," *IEEE Transactions on Communications*, vol. 70, no. 4, pp. 2480–2498, April 2022.
[32] A. Fengler, O. Musa, P. Jung, and G. Caire, "Pilot-based unsourced random access with a massive MIMO receiver, interference cancellation, and power control," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 5, pp. 1522–1534, May 2022.
[33] M. J. Ahmadi and T. M. Duman, "Unsourced random access with a massive MIMO receiver using multiple stages of orthogonal pilots," in *2022 IEEE International Symposium on Information Theory (ISIT)*, Espoo, Finland, 26 June 2022 - 01 July 2022, pp. 2880–2885.
[34] E. Marshakov, G. Balitskiy, K. Andreev, and A. Frolov, "A polar code based unsourced random access for the Gaussian MAC," in *IEEE Vehicular Technology Conference (VTC)*, Honolulu, HI, USA, 22-25 September 2019, pp. 1–5.
[35] A. K. Pradhan, V. K. Amalladinne, K. R. Narayanan, and J.-F. Chamberland, "LDPC codes with soft interference cancellation for uncoordinated unsourced multiple access," *arXiv preprint arXiv:2105.13985*, May 2021.
[36] J. Ma and L. Ping, "Data-aided channel estimation in large antenna systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3111–3124, June 2014.
[37] C. S. K. Valmeekam and K. R. Narayanan, "A semi-blind decision directed iterative channel estimation and decoding for LDPC coded OFDM systems," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, Austin, TX, USA, 10-13 April 2022, pp. 686–691.
[38] J. Gao, Y. Wu, T. Li, and W. Zhang, "Energy efficiency of MIMO massive unsourced random access with finite blocklength," *arXiv preprint arXiv:2302.02048*, February 2023. [Online]. Available: https://arxiv.org/abs/2302.02048
[39] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Belmont, MA, USA: Athena Scientific, 2005, vol. I.

Michail Gkagkos (Graduate Student Member, IEEE) received the B.Sc. degree in Electrical Engineering from University of Cyprus in 2019. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Texas A&M University at College Station, College Station, TX, USA. From May 2022 to August 2022, he was employed as a Communication Systems Intern at MaxLinear, and from May 2023 to August 2023 as a Wireless R&D Intern at Qualcomm. His research interests include wireless communications, signal processing, error control coding and information theory.

Krishna R. Narayanan (Fellow, IEEE) received the B.E. degree from Coimbatore Institute of Technology, M.S. degree from Iowa State University and Ph.D. degree in Electrical Engineering from Georgia Institute of Technology in 1998. Since 1998, he has been with the Department of Electrical and Computer Engineering at Texas A&M University, where he is currently the Eric D. Rubin '06 professor. His current research interests are in the design of massive uncoordinated multiple access schemes, coded distributed computing, exploring connections between coding theory and sparse signal recovery and in analyzing data defined on graphs. He received the 2022 IEEE Joint Communications Society and Information Theory Society best paper award. He also received the 2006 and 2020 best paper awards from the IEEE technical committee for signal processing for data storage. In 2014, he received the Professional Progress in Engineering award given to an outstanding alumnus of Iowa State University. He was awarded a university level distinguished teaching award in 2018 at Texas A&M University. He has served as an associate editor for coding techniques for the IEEE Transactions on Information Theory and as the area editor (and editor) for the coding theory and applications area of the IEEE Transactions on Communications. He was elected as a Fellow of the IEEE for contributions to coding for wireless communications and data storage.

Jean-Francois Chamberland (Senior Member, IEEE) is a professor in the Department of Electrical and Computer Engineering at Texas A&M University. He completed the B.Eng. degree at McGill University, the M.S. degree at Cornell University, and the Ph.D. degree at the University of Illinois, Urbana-Champaign. His research interests are in the areas of communications and information theory, statistical inference, computer systems and networks, and learning. Recently, he has been studying the efficient design of wireless systems and the fundamental limits of communication networks. His contributions have been recognized through an IEEE Young Author Best Paper Award from the IEEE Signal Processing Society, a Faculty Early Career Development (CAREER) Award from the National Science Foundation (NSF), and an IEEE Communications Society & Information Theory Society Joint Paper Award. He currently serves as a senior area editor for the IEEE Open Journal of Signal Processing, and he was an associate editor for the IEEE Transactions on Information Theory.

Costas N. Georghiades (Fellow, IEEE) received the B.E. degree with distinction from the American University of Beirut in 1980, and M.Sc. and D.Sc. degrees from Washington University, St. Louis, MO, in 1983 and 1985, respectively, all in Electrical Engineering. He joined the Electrical Engineering Department at Texas A& M in 1985 where he is professor and holder of the Delbert A. Whitaker Chair. He served as Department Head 2005-2012, as Associate Dean for Research and Texas A&M Experiment Station Assistant and Associate Director 2012-2016 and as Senior Associate Vice President for Research at Texas A&M 2016-2022. His portfolio included multidisciplinary research engagement, centers and institutes and core facilities. He currently serves as Interim Director of the Global Cyber Research Institute endowed by Ray Rothrock '77 and Anthony Wood '90 and as Interim Department Head of Electrical & Computer Engineering.

His personal research interests are in the application of statistical communication, information, and estimation theories to the study of telecommunication systems, and specifically wireless and optical systems. He served in various editorial positions with the Institute of Electrical and Electronic Engineers (IEEE), as well as in executive level positions with several conference organizations. In other professional service, he chaired several IEEE awards committees, the Information Theory Society's Fellows Evaluation Committee, the Communication Theory Technical Committee, and the Wireless Communication Letters Steering committee. He was elected Fellow of the IEEE in 1998 and in 2012 he received the IEEE Communication Society's Communication Theory Technical Committee Service Award. He recently served as a member of the IEEE Fellows Committee, and he is currently a member of the IEEE Eric E. Sumner Award Committee. In 2021 he received the Texas A&M College of Engineering Excellence Award for Service.