OTFS Signaling for Uplink NOMA of Heterogeneous Mobility Users

Yao Ge, *Graduate Student Member, IEEE*, Qinwen Deng, *Graduate Student Member, IEEE*,
P. C. Ching, *Life Fellow, IEEE*, and Zhi Ding, *Fellow, IEEE*

Abstract— We investigate a coded uplink non-orthogonal multiple access (NOMA) configuration in which groups of co-channel users are modulated in accordance with orthogonal time frequency space (OTFS). We take advantage of OTFS characteristics to achieve NOMA spectrum sharing in the delay-Doppler domain between stationary and mobile users. We develop an efficient iterative turbo receiver based on the principle of successive interference cancellation (SIC) to overcome the co-channel interference (CCI). We propose two turbo detector algorithms: orthogonal approximate message passing with linear minimum mean squared error (OAMP-LMMSE) and Gaussian approximate message passing with expectation propagation (GAMP-EP). The interactive OAMP-LMMSE detector and GAMP-EP detector are respectively assigned for the reception of the stationary and mobile users. We analyze the convergence performance of our proposed iterative SIC turbo receiver by utilizing a customized extrinsic information transfer (EXIT) chart and simplify the corresponding detector algorithms to further reduce receiver complexity. Our proposed iterative SIC turbo receiver demonstrates performance improvement over existing receivers and robustness against imperfect SIC process and channel state information uncertainty.

Index Terms— Co-channel interference, NOMA, OTFS, resource allocation, turbo receiver.

I. INTRODUCTION

THE rapid advancement of wireless transmission and mobile communication techniques has led to an explosive rise in data traffics of wireless networks. In order to support such tremendous needs, high spectrum efficiency techniques such as non-orthogonal multiple access (NOMA) have been considered as promising solutions for improving spectrum utilization and user connectivity [1]–[3]. Unlike conventional orthogonal multiple access (OMA), NOMA allows multiple users to access the same spectrum simultaneously at different

power levels [4] or with the help of low-density spreading codes [5]. To overcome the inevitable co-channel interference (CCI), advanced receivers such as successive interference cancellation (SIC) are required as effective multi-user detection for NOMA systems. Note that the NOMA spectrum sharing is common among users with different channel conditions or quality of service (QoS) levels [2], [4].

Broadband mobile communications in high-mobility environments such as high-speed railways and autonomous vehicles represent another strong arena of growth. High mobility communications are particularly challenging because of the well-known fast channel fading and distortions due to the large Doppler spread. Recently, the advent of orthogonal time frequency space (OTFS) [6] modulation shows strong promise as an effective PHY-layer alternative to traditional orthogonal frequency division multiplexing (OFDM) in high-mobility environments. OTFS exhibits performance advantages since it can exploit the diversity gain coming from both the delay and Doppler channel domains. OTFS can effectively convert a rapidly time-varying channel in time-frequency domain into a quasi-stationary channel model in delay-Doppler domain. This quasi-stationary channel model simplifies channel estimation [7]–[9] and symbol detection [10]–[14] for wireless receivers in high-mobility scenarios. OTFS achieves diversity gain for stationary multipath channels and doubly-selective channels, as shown in [15] and [16], [17], respectively. Other related works considered OTFS modulation in multiple-input multiple-output (MIMO) systems [18] and millimeter wave (mmWave) communication systems [19].

Recognizing the superior performance of OTFS in delay-Doppler channel domain, several works studied multiple user access based on OTFS framework in high-mobility scenarios [20]–[25]. For OTFS-OMA, the authors in [20], [21] proposed to allocate different time-frequency resources to different users. This orthogonal resource allocation can be either in contiguous [20] or interleaved [21] fashions when the ideal bi-orthogonal pulses are available. Practically, however, such ideal pulses are not realizable in view of the Heisenberg uncertainty principle [26]. Allocation of different delay-Doppler resources to different users [22] has shown significant CCI that requires complex receivers. For massive MIMO-OTFS networks, a new path division multiple access (PDMA) [23] can assign angle-domain resources to different users so as to eliminate CCI at the receiver in the angle-delay-Doppler domain. To further improve the spectral efficiency and support

the massive connectivity, the OTFS-NOMA schemes were proposed in [24], [25], where multiple mobile users are sharing the same delay-Doppler resources and distinguished by either different power levels [24] or sparse codewords [25].

A recent work [27], [28] suggested a new form of OTFS-NOMA in which a single high-mobility user using OTFS in the delay-Doppler domain is paired with a group of low-mobility OFDM users for non-orthogonal channel sharing. The improved spectrum efficiency shown by [27], [28], however, relies on the ideal bi-orthogonal OTFS pulses have been elusive to practitioners and may not even exist physically. These ideal bi-orthogonal OTFS pulses are also essential to the proposed simple equalizations in [27], [28] for OTFS-NOMA system. In addition, the performance analysis of OTFS-NOMA scheme in [27], [28] only considers the perfect SIC process and requires mobile channels to exhibit on-the-grid delays and Doppler shifts, which are still unrealistic assumptions in practical OTFS-NOMA system deployment.

To alleviate the dependency on the assumptions of ideal bi-orthogonal OTFS pulses and on-the-grid channel delays and/or Doppler shifts, we investigate a more general coded uplink OTFS-NOMA scenario in this work. Without loss of generality, we focus on a simple scenario in which the stationary users and mobile users are grouped for NOMA. Unlike [27], [28], both users utilize OTFS modulation. We design an efficient NOMA protocol by grouping users with different mobility profiles. With the use of OTFS, co-channel users of different mobility profiles can take advantage of different resource allocations in the delay-Doppler domain to mitigate their CCI and simplify the receiver complexity. We eliminate the unrealistic assumptions of ideal bi-orthogonal pulses, perfect SIC process and on-the-grid channel delay/Doppler shifts. By using the practical pulses such as rectangular pulses, the block circulant matrices in [27], [28] no longer apply. To this end, we develop a novel receiver architecture to effectively mitigate CCI and recover the signal for each user. Our contributions in this work are summarized as follows:

1) We propose an OTFS-based NOMA (OBNOMA) configuration that groups users of different mobility profiles by only utilizing OTFS modulation. Without loss of generality, we consider groups of stationary and mobile users that occupy different sub-vector resources of delay-Doppler domain via OTFS. This OBNOMA framework can effectively tackle the CCI and is amenable to effective receiver algorithms.

2) We design an iterative turbo receiver for multi-user detection and decoding that leverages the SIC principle. The proposed joint SIC detector and individual user decoders exchange the extrinsic information iteratively to improve receiver performance. In particular, we develop an orthogonal approximate message passing with linear minimum mean squared error (OAMP-LMMSE) algorithm for detecting the OBNOMA stationary users' signal and a Gaussian approximate message passing with expectation propagation (GAMP-EP) algorithm for detecting the signal of OBNOMA mobile users.

3) We develop a novel customized extrinsic information transfer (EXIT) chart framework to analyze the convergence property of our proposed iterative SIC turbo receiver. More importantly, we further propose reduced complexity variants for both OAMP-LMMSE and GAMP-EP detectors without significant performance drop.

4) We demonstrate that our proposed iterative SIC turbo receiver for OBNOMA system outperforms the existing methods and has robustness to the imperfect SIC process and channel state information (CSI) uncertainty.

We organize the remaining sections of the paper as follows. Section II summarizes the basics of OTFS transmission. Section III proposes the novel coded uplink OBNOMA system model and describes the resource allocation in delay-Doppler domain for OBNOMA users. In Section IV, we propose our iterative SIC turbo receiver and the two component detectors respectively for signal detection of stationary and mobile users. We further analyze the convergence behavior of the proposed iterative SIC turbo receiver through a novel customized EXIT chart and simplify the two detector algorithms to reduce receiver complexity in Section V. We provide simulation results in Section VI under different scenarios. Our conclusions are finally drawn in Section VII. The Appendix contains some detailed proofs at the end of the paper.

II. PRELIMINARIES

In this section, we briefly introduce basic OTFS concepts and system transmission model. We also provide the mathematical descriptions of OTFS used for both mobile and stationary users.

A. Basic Concepts of OTFS

Unlike conventional OFDM, OTFS multiplexes and processes each information symbol in the roughly constant delay-Doppler domain rather than time-frequency domain. A lattice in time-frequency plane is sampled by intervals T (seconds) and $\Delta f = 1/T$ (Hz) along the time and frequency axes, i.e.,

$$\Lambda = \{(m\Delta f, nT), m = 0, \cdots, M-1; n = 0, \cdots, N-1\},$$

where $M \in \mathcal{Z}$ and $N \in \mathcal{Z}$ represent the total available numbers of subcarriers and time intervals, respectively. According to the channel characteristics, T and Δf are chosen, respectively, larger than the maximal channel delay spread and maximum Doppler frequency shift.

The corresponding lattice in delay-Doppler plane is described by

$$\Gamma = \left\{\left(\frac{\ell}{M\Delta f}, \frac{k}{NT}\right), \ell = 0, \cdots, M-1; k = 0, \cdots, N-1\right\},$$

where $1/M\Delta f$ and $1/NT$ denote the resolutions of delay dimension and Doppler dimension, respectively. Note that signals placed on delay-Doppler grids in a given packet burst can be transformed into time-frequency samples with time duration $T_f = NT$ and bandwidth $B = M\Delta f$. Additional details on OTFS can be found in existing works such as [6], [13].

B. OTFS Signal Model for Mobile Users

At OTFS transmitter, the MN random information symbols (e.g., QAM) are generated from a complex alphabet $\mathbb{A} = \{a_1, a_2, \cdots, a_Q\}$ and placed on the delay-Doppler plane Γ. These delay-Doppler symbols $\mathbf{X} \in \mathbb{C}^{M \times N}$ are mapped into a lattice in time-frequency domain $\bar{\mathbf{X}} \in \mathbb{C}^{M \times N}$ through the inverse symplectic finite Fourier transform (ISFFT) [29],

$$\bar{\mathbf{X}} = \mathbf{F}_M \mathbf{X} \mathbf{F}_N^H, \tag{1}$$

where $\mathbf{F}_M \in \mathbb{C}^{M \times M}$ and $\mathbf{F}_N \in \mathbb{C}^{N \times N}$ are the normalized M-point and N-point discrete Fourier transform (DFT) matrices, respectively. Next, the Heisenberg transform is adopted to the time-frequency signal $\bar{\mathbf{X}}$ with a transmit pulse $g_{tx}(t)$ to generate the time domain signal $\mathbf{s} \in \mathbb{C}^{MN \times 1}$,

$$s[c] = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \bar{X}[m,n] g_{tx}(cT_s - nT) e^{j2\pi m \Delta f (cT_s - nT)},$$
$$c = 0, \cdots, MN - 1, \tag{2}$$

where $T_s = 1/M\Delta f$ is the symbol spaced sampling interval.

To overcome the inter-frame interference, we append a cyclic prefix (CP) of length no shorter than the maximal channel delay spread to signal $\mathbf{s}$. After passing a transmit filter, the resulted time domain signal enters the multipath fading channels characterized by sampled response of

$$h[c,p] = \sum_{i=1}^{L} h_i e^{j2\pi \nu_i (cT_s - pT_s)} \mathrm{P_{rc}}(pT_s - \tau_i),$$
$$c = 0, \cdots, MN - 1; \quad p = 0, \cdots, P - 1, \tag{3}$$

where L denotes the number of multipaths; h_i, τ_i and ν_i are the complex gain, delay and Doppler shift associated with the i-th path, respectively. The channel tap P is determined by the maximal channel delay spread as well as the duration of the overall filter response.

In (3), $\mathrm{P_{rc}}(pT_s - \tau_i)$ is the sampled overall filter response that comprises a pair of bandlimiting matched filters adopted by the transmitter and receiver to control signal transmission bandwidth and to achieve maximum signal-to-noise ratio (SNR) at the receiver. In practice, the most common implemented pulse shaping filters at the transmitter and receiver are the root raised-cosine (RRC) filters, leading to a raised-cosine (RC) rolloff pulse for $\mathrm{P_{rc}}(\tau)$. In addition, the Doppler frequency shift of the i-th path can be written as $\nu_i = (k_{\nu_i} + \beta_{\nu_i})/NT$, where the integer k_{ν_i} and real $\beta_{\nu_i} \in (-0.5, 0.5]$ are respectively represent the index and fractional Doppler shift of ν_i.

Consider the baseband model. At OTFS receiver, the channel output signal enters a user-defined receive filter. After removing CP, we can obtain the received signal $\mathbf{r} \in \mathbb{C}^{MN \times 1}$ as

$$r[c] = \sum_{p=0}^{P-1} h[c,p] s\left[[c-p]_{MN}\right] + n[c], \quad c = 0, \cdots, MN - 1, \tag{4}$$

where $\mathbf{n}$ represents the filtered noise and the notation $[\cdot]_m$ denotes mod-m operation. The received time domain signal

$\mathbf{r}$ is then processed by Wigner transform (i.e., the inverse of Heisenberg transform) using a receive pulse $g_{rx}(t)$ to produce the time-frequency domain signal

$$\bar{Y}[m,n] = \sum_{c=0}^{MN-1} g_{rx}^*(cT_s - nT) r[c] e^{-j2\pi m \Delta f (cT_s - nT)},$$
$$m = 0, \cdots, M - 1; \ n = 0, \cdots, N - 1. \tag{5}$$

Finally, the signal matrix $\bar{\mathbf{Y}} \in \mathbb{C}^{M \times N}$ in the time-frequency domain are transformed back to the delay-Doppler domain via symplectic finite Fourier transform (SFFT) as described below [29]:

$$\mathbf{Y} = \mathbf{F}_M^H \bar{\mathbf{Y}} \mathbf{F}_N. \tag{6}$$

For simplicity, we utilize a rectangular pulse for $g_{tx}(t)$ and $g_{rx}(t)$ in the above steps, for which the baseband OTFS input-output relationship in delay-Doppler domain is given by [30]

$$Y[\ell, k] = \sum_{p=0}^{P-1} \sum_{i=1}^{L} \sum_{q=0}^{N-1} h_i \mathrm{P_{rc}}(pT_s - \tau_i) \gamma(k, \ell, p, q, k_{\nu_i}, \beta_{\nu_i})$$
$$\times X\left[[\ell - p]_M, [k - k_{\nu_i} + q]_N\right] + \omega[\ell, k], \tag{7}$$

where $\boldsymbol{\omega} \in \mathbb{C}^{M \times N}$ is the noise at the output of the SFFT. We also use the following definitions:

$$\gamma(k, \ell, p, q, k_{\nu_i}, \beta_{\nu_i})$$
$$= \begin{cases} \frac{1}{N} \xi(\ell, p, k_{\nu_i}, \beta_{\nu_i}) \theta(q, \beta_{\nu_i}), \\ \qquad p \leq \ell < M, \\ \frac{1}{N} \xi(\ell, p, k_{\nu_i}, \beta_{\nu_i}) \theta(q, \beta_{\nu_i}) \phi(k, q, k_{\nu_i}), \\ \qquad 0 \leq \ell < p, \end{cases} \tag{8a}$$

$$\xi(\ell, p, k_{\nu_i}, \beta_{\nu_i}) = e^{j2\pi \left(\frac{\ell-p}{M}\right) \left(\frac{k_{\nu_i} + \beta_{\nu_i}}{N}\right)}, \tag{8b}$$

$$\theta(q, \beta_{\nu_i}) = \frac{e^{-j2\pi(-q-\beta_{\nu_i})} - 1}{e^{-j\frac{2\pi}{N}(-q-\beta_{\nu_i})} - 1}, \tag{8c}$$

$$\phi(k, q, k_{\nu_i}) = e^{-j2\pi \frac{[k - k_{\nu_i} + q]_N}{N}}. \tag{8d}$$

To summarize, the input-output relationship in (7) can be vectorized column-wise into

$$\tilde{\mathbf{y}}_{\mathcal{M}} = \tilde{\mathbf{H}}_{\mathcal{M}} \tilde{\mathbf{x}}_{\mathcal{M}} + \boldsymbol{\omega}_{\mathcal{M}}, \tag{9}$$

where $\tilde{\mathbf{x}}_{\mathcal{M}}, \tilde{\mathbf{y}}_{\mathcal{M}}, \boldsymbol{\omega}_{\mathcal{M}} \in \mathbb{C}^{MN \times 1}$ and $\tilde{\mathbf{H}}_{\mathcal{M}} \in \mathbb{C}^{MN \times MN}$ is a sparse matrix.

C. OTFS Signal Model for Stationary Users

Since the stationary user do not experience Doppler shifts (i.e., $\nu_i = 0, \forall i$), the baseband channel impulse response in (3) can be simplified into

$$h[p] = \sum_{i=1}^{L} h_i \mathrm{P_{rc}}(pT_s - \tau_i), \quad p = 0, \cdots, P - 1. \tag{10}$$

In this case, the channel input-output relationship of OTFS in (7) reduces to

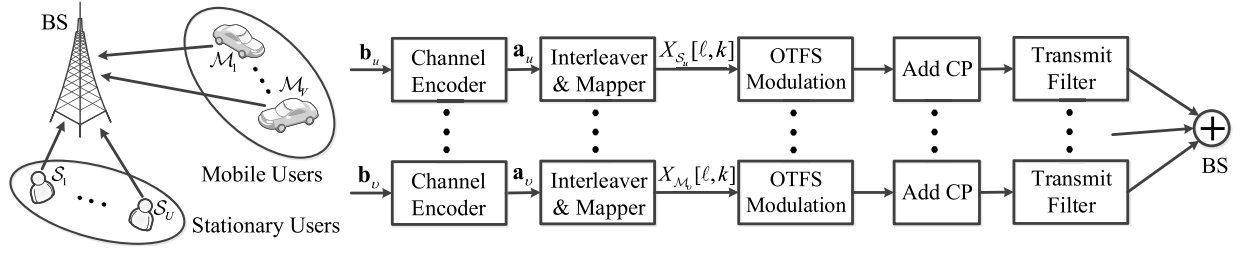$$Y[\ell, k] = \sum_{p=0}^{P-1} h[p] \bar{\gamma}(k, \ell, p) X\left[[\ell - p]_M, k\right] + \omega[\ell, k], \tag{11}$$

Fig. 1. System model for coded uplink OBNOMA scheme.

where

$$\bar{\gamma}(k,\ell,p) = \begin{cases} 1, & p \le \ell < M, \\ e^{-j2\pi\frac{k}{N}}, & 0 \le \ell < p. \end{cases} \quad (12)$$

Let us denote the frequency domain channel response

$$H[c] = \sum_{p=0}^{P-1} h[p]e^{-j\frac{2\pi cp}{MN}}, \quad c = 0, \cdots, MN-1, \quad (13)$$

and define the diagonal matrix $\bar{\mathbf{H}}_k \in \mathbb{C}^{M \times M}$ via

$$\bar{\mathbf{H}}_k = \text{diag}\{H[k], H[k+N], \cdots, H[k+(M-1)N]\}. \quad (14)$$

The following proposition summarizes the findings:

Proposition 1: For stationary user, the OTFS input-output relationship in (11) is equivalent to

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \boldsymbol{\omega}_k, \quad k = 0, \cdots, N-1, \quad (15)$$

where $\mathbf{x}_k \in \mathbb{C}^{M \times 1}$ and $\mathbf{y}_k \in \mathbb{C}^{M \times 1}$ are the k-th column of $\mathbf{X}$ and $\mathbf{Y}$, respectively. $\boldsymbol{\omega}_k \in \mathbb{C}^{M \times 1}$ is the k-th noise vector and the equivalent channel matrix $\mathbf{H}_k \in \mathbb{C}^{M \times M}$ is given by

$$\mathbf{H}_k = \mathbf{U}_k^H \bar{\mathbf{H}}_k \mathbf{U}_k, \quad (16)$$

where $\mathbf{U}_k = \mathbf{F}_M \boldsymbol{\Lambda}_k$ is a unitary matrix with

$$\boldsymbol{\Lambda}_k = diag\left\{1, e^{-j\frac{2\pi k}{MN}}, e^{-j\frac{2\pi 2k}{MN}}, \cdots, e^{-j\frac{2\pi(M-1)k}{MN}}\right\}. \quad (17)$$

Proof: See Appendix. ∎

The input-output relationship in (15) can be rewritten as

$$\tilde{\mathbf{y}}_S = \tilde{\mathbf{H}}_S\tilde{\mathbf{x}}_S + \boldsymbol{\omega}_S, \quad (18)$$

where $\tilde{\mathbf{x}}_S = \left[\mathbf{x}_0^T, \mathbf{x}_1^T, \cdots, \mathbf{x}_{N-1}^T\right]^T$, $\tilde{\mathbf{y}}_S = \left[\mathbf{y}_0^T, \mathbf{y}_1^T, \cdots, \mathbf{y}_{N-1}^T\right]^T$, $\boldsymbol{\omega}_S = \left[\boldsymbol{\omega}_0^T, \boldsymbol{\omega}_1^T, \cdots, \boldsymbol{\omega}_{N-1}^T\right]^T$ and $\tilde{\mathbf{H}}_S = \text{diag}\{\mathbf{H}_0, \mathbf{H}_1, \cdots, \mathbf{H}_{N-1}\}$.

Remark 1: Note that the mathematical description of OTFS transmission with rectangular pulses for stationary user in Proposition 1 is equivalent to the Vector OFDM [31]–[33], which is proposed to as a bridge between conventional OFDM and single carrier modulations. We also notice that similar result has been mentioned in [15] according to the equivalent system structures of OTFS and Vector OFDM for stationary user. Such intrinsic equivalence between OTFS and Vector OFDM offers a new insight for better understanding of Vector OFDM. They share the same characteristics and properties directly.

Remark 2: From OTFS signal models, we observe that the received signal is only affected by a quasi-stationary

channel in the delay-Doppler domain as shown in (7) for the mobile users. For the stationary users, OTFS converts an inter-symbol interference (ISI) channel into multiple "ISI-free" vector channels as in (15). Evidently, OTFS helps simplify the system models for both the mobile and stationary users.

III. OBNOMA SYSTEM MODEL AND RESOURCE ALLOCATION

A. Mobility-Profile Based User Grouping in OBNOMA

Consider a coded uplink multiuser system with $(U + V)$ users communicating with a base station (BS) simultaneously as shown in Fig. 1, where $\mathcal{U} = \{1, 2, \cdots, U\}$ denotes the set of stationary users and $\mathcal{V} = \{U + 1, U + 2, \cdots, U + V\}$ represents the set of mobile users. For simplicity, we use $u = \{1, 2, \cdots, U\} \in \mathcal{U}$ and $v = \{1, 2, \cdots, V\} \in \mathcal{V}$ to denote the u-th stationary user and v-th mobile user, respectively. To avoid unnecessary confusion, we use a simple model in which the user terminals and the BS receiver are equipped with a single transmit antenna and receive antenna. Each user utilizes OTFS for uplink transmission to take advantage of its benefits (Remark 2). Naturally, our model also applies to the cases involving multiple transmit and receive antennas, with expected diversity gain.

In practice, the stationary and mobile users experience different Doppler shifts and fading rates. This channel difference allows us to develop a special OTFS-based NOMA (OBNOMA) by grouping users with different mobility profiles. Note that laissez faire resource allocation may lead to severe CCI. Considering characteristics of stationary and mobile users in terms of Doppler shifts and channel delay spreads, we propose a novel resource allocation in OBNOMA scheme as shown in Fig. 2. Specifically, the non-overlapping bins along the Doppler axis are assigned to stationary users orthogonally without mutual interference among these stationary users. On the other hand, disjoint and contiguous bins along the delay axis are allocated to mobile users to mitigate their mutual interference.

B. Proposed OBNOMA Signal Models

Consider K_g binary information bits $\mathbf{b}_g$ for user $g \in \{\mathcal{U}, \mathcal{V}\}$ which are encoded into a codeword $\mathbf{a}_g$ of length N_g, where the code rate equals to K_g/N_g. The codeword $\mathbf{a}_g$ is interleaved to give a data block $\mathbf{d}_g$ before being modulated into Gray-mapped symbols $\mathbf{e}_g$ drawn from a complex alphabet $\mathbb{A}$. Thereby, the resulting transmit symbols for the u-th stationary
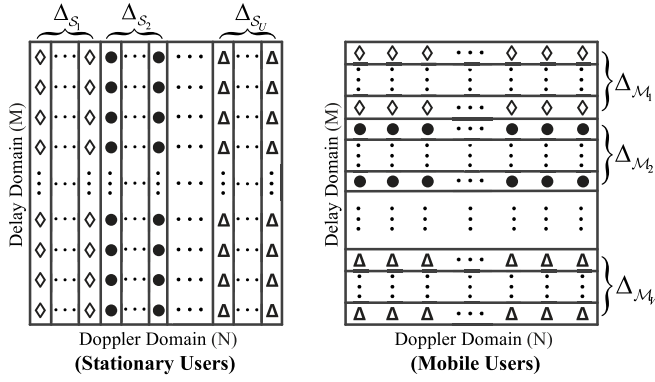
Fig. 2. Resource allocation in OBNOMA.

user and v-th mobile user are placed in the delay-Doppler plane Γ, and denoted respectively as (19a) and (19b), as shown at the bottom of the page, where $\Delta_{\mathcal{S}_u}$ is the set of bins along the Doppler axis assigned to the u-th stationary user of cardinality $|\Delta_{\mathcal{S}_u}|$ and $\Delta_{\mathcal{M}_v}$ is the set of bins along the delay axis assigned to the v-th mobile user of cardinality $|\Delta_{\mathcal{M}_v}|$, respectively. Without loss of generality, the mean symbol energies are denoted as $P_{\mathcal{S}}$ and $P_{\mathcal{M}}$ for stationary and mobile users, respectively, with the corresponding bit energies $E_{\mathcal{S}}$ and $E_{\mathcal{M}}$. In addition, the sets assigned to different users with the same mobility profile are mutually independent, i.e., $\cup_{u=1}^{U}\Delta_{\mathcal{S}_u} = \{0, 1, \cdots, N-1\}$, $\cup_{v=1}^{V}\Delta_{\mathcal{M}_v} = \{0, 1, \cdots, M-1\}$, $\Delta_{\mathcal{S}_u} \cap \Delta_{\mathcal{S}_{u'}} = \emptyset$ and $\Delta_{\mathcal{M}_v} \cap \Delta_{\mathcal{M}_{v'}} = \emptyset$ when $u \neq u'$ and $v \neq v'$, respectively.

Each user employs OTFS modulation (i.e., ISFFT and Heisenberg transform) and adds a CP in front of the generated time domain signal. After passing through the transmit filter, each stationary user signal is sent out over channel response

$$h_{\mathcal{S}_u}[p] = \sum_{i=1}^{L_{\mathcal{S}_u}} h_{\mathcal{S}_u,i} \mathrm{P}_{\mathrm{rc}}(pT_s - t_{\mathcal{S}_u} - \tau_{\mathcal{S}_u,i}),$$
$$p = 0, \cdots, P_{\mathcal{S}_u} - 1, \quad (20)$$

whereas each mobile user signal is transmitted over channel response

$$h_{\mathcal{M}_v}[c, p]$$
$$= \sum_{i=1}^{L_{\mathcal{M}_v}} h_{\mathcal{M}_v,i} e^{j2\pi \nu_{\mathcal{M}_v,i}(cT_s - pT_s)} \mathrm{P}_{\mathrm{rc}}(pT_s - t_{\mathcal{M}_v} - \tau_{\mathcal{M}_v,i}),$$
$$c = 0, \cdots, MN - 1, \ p = 0, \cdots, P_{\mathcal{M}_v} - 1. \quad (21)$$

Recall that $L_{\mathcal{S}_u}$ and $t_{\mathcal{S}_u}$ are the number of multipaths and the amount of timing offset experienced by the u-th stationary user; $h_{\mathcal{S}_u,i}$ and $\tau_{\mathcal{S}_u,i}$ represent the gain and delay associated with the i-th path of the u-th stationary user's channel.

Similarly, $L_{\mathcal{M}_v}$ and $t_{\mathcal{M}_v}$ denote the number of multipaths and the timing offset experienced by the v-th mobile user; $h_{\mathcal{M}_v,i}$, $\tau_{\mathcal{M}_v,i}$ and $\nu_{\mathcal{M}_v,i}$ stand for the complex gain, delay and Doppler frequency shift associated with the i-th path of the v-th mobile user's channel, respectively. $P_{\mathcal{S}_u}$ and $P_{\mathcal{M}_v}$ represent, respectively, the channel taps of the u-th stationary user and v-th mobile user.

We assume that the CP is sufficiently long to accommodate both the maximum timing offset and the maximal channel delay spread. Hence, there is no inter-frame interference. At the receiver, the CP is removed after the received filter. We apply the standard Wigner transform and SFFT structure to demodulate OTFS signals in the delay-Doppler domain, where the input-output relationship can be expressed as

$$\mathbf{y} = \sum_{u=1}^{U} \tilde{\mathbf{H}}_{\mathcal{S}_u} \tilde{\mathbf{x}}_{\mathcal{S}_u} + \sum_{v=1}^{V} \tilde{\mathbf{H}}_{\mathcal{M}_v} \tilde{\mathbf{x}}_{\mathcal{M}_v} + \boldsymbol{\omega} \quad (22a)$$
$$= \bar{\mathbf{H}}_{\mathcal{S}} \bar{\mathbf{x}}_{\mathcal{S}} + \bar{\mathbf{H}}_{\mathcal{M}} \bar{\mathbf{x}}_{\mathcal{M}} + \boldsymbol{\omega}. \quad (22b)$$

Here, we have used the following notations

$$\bar{\mathbf{H}}_{\mathcal{S}} = \left[\tilde{\mathbf{H}}_{\mathcal{S}_1}, \tilde{\mathbf{H}}_{\mathcal{S}_2}, \cdots, \tilde{\mathbf{H}}_{\mathcal{S}_U} \right] \in \mathbb{C}^{MN \times UMN},$$
$$\bar{\mathbf{x}}_{\mathcal{S}} = \left[\tilde{\mathbf{x}}_{\mathcal{S}_1}^T, \tilde{\mathbf{x}}_{\mathcal{S}_2}^T, \cdots, \tilde{\mathbf{x}}_{\mathcal{S}_U}^T \right]^T \in \mathbb{C}^{UMN \times 1},$$
$$\bar{\mathbf{H}}_{\mathcal{M}} = \left[\tilde{\mathbf{H}}_{\mathcal{M}_1}, \tilde{\mathbf{H}}_{\mathcal{M}_2}, \cdots, \tilde{\mathbf{H}}_{\mathcal{M}_V} \right] \in \mathbb{C}^{MN \times VMN},$$
$$\bar{\mathbf{x}}_{\mathcal{M}} = \left[\tilde{\mathbf{x}}_{\mathcal{M}_1}^T, \tilde{\mathbf{x}}_{\mathcal{M}_2}^T, \cdots, \tilde{\mathbf{x}}_{\mathcal{M}_V}^T \right]^T \in \mathbb{C}^{VMN \times 1}.$$

Also, $\mathbf{y} \in \mathbb{C}^{MN \times 1}$ represents the received signal and $\boldsymbol{\omega} \in \mathbb{C}^{MN \times 1} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}})$ denotes the noise vector. $\tilde{\mathbf{x}}_{\mathcal{S}_u} \in \mathbb{C}^{MN \times 1}$ and $\tilde{\mathbf{x}}_{\mathcal{M}_v} \in \mathbb{C}^{MN \times 1}$ contain the transmitted symbols from u-th stationary user and v-th mobile user, respectively. The equivalent channels $\tilde{\mathbf{H}}_{\mathcal{S}_u}$ and $\tilde{\mathbf{H}}_{\mathcal{M}_v}$ have the similar structures as $\tilde{\mathbf{H}}_{\mathcal{S}}$ in (18) and $\tilde{\mathbf{H}}_{\mathcal{M}}$ in (9), respectively.

Note that $\bar{\mathbf{x}}_{\mathcal{S}}$ and $\bar{\mathbf{x}}_{\mathcal{M}}$ are sparse vectors due to the resource allocations of (19a) and (19b). The numbers of non-zero elements in $\bar{\mathbf{x}}_{\mathcal{S}}$ and $\bar{\mathbf{x}}_{\mathcal{M}}$ are identically MN. Let $\mathbf{x}_{\mathcal{S}} \in \mathbb{C}^{MN \times 1}$ and $\mathbf{x}_{\mathcal{M}} \in \mathbb{C}^{MN \times 1}$ denote the effective input vectors after removing the zeros in $\bar{\mathbf{x}}_{\mathcal{S}}$ and $\bar{\mathbf{x}}_{\mathcal{M}}$; Let $\mathbf{H}_{\mathcal{S}} \in \mathbb{C}^{MN \times MN}$ and $\mathbf{H}_{\mathcal{M}} \in \mathbb{C}^{MN \times MN}$ represent the effective matrices after deleting the columns corresponding to the indices of the zeros in $\bar{\mathbf{x}}_{\mathcal{S}}$ and $\bar{\mathbf{x}}_{\mathcal{M}}$, respectively. We can then simplify the relationship of (22) to

$$\mathbf{y} = \mathbf{H}_{\mathcal{S}} \mathbf{x}_{\mathcal{S}} + \mathbf{H}_{\mathcal{M}} \mathbf{x}_{\mathcal{M}} + \boldsymbol{\omega}. \quad (23)$$

From (23), we observe that the conventional single user detection [12]–[14] or multi-user detection with OMA [21], [22] cannot be directly applied to recover the signal due to the strong presence of CCI at the receiver. Therefore, additional processing such as iterative SIC techniques should be employed to mitigate CCI effect.

$$X_{\mathcal{S}_u}[\ell, k] = \begin{cases} \mathbf{e}_u[\ell + k_u N], & \ell \in \{0, 1, \cdots, M-1\} \ \& \ k \in \Delta_{\mathcal{S}_u}, k_u = \{0, 1, \cdots, |\Delta_{\mathcal{S}_u}| - 1\} \\ 0, & \text{otherwise}, \end{cases} \quad (19a)$$

$$X_{\mathcal{M}_v}[\ell, k] = \begin{cases} \mathbf{e}_v[\ell_v M + k], & \ell \in \Delta_{\mathcal{M}_v} \ \& \ k \in \{0, 1, \cdots, N-1\}, \ell_v = \{0, 1, \cdots, |\Delta_{\mathcal{M}_v}| - 1\} \\ 0, & \text{otherwise}, \end{cases} \quad (19b)$$

IV. ITERATIVE SIC TURBO RECEIVER FOR OBNOMA

We now investigate the recovery of the signal for each user from the received aggregated signal at the BS. Here, we propose an iterative SIC turbo receiver to overcome the CCI and self-interference in delay-Doppler domain.

A. Receiver Structure

The structure of iterative SIC turbo receiver is shown in Fig. 3. The key point is the iterative exchange of information between the SIC detector and the co-channel individual user channel decoders. The extrinsic information generated from the SIC detector is treated as *a priori* information by the individual channel decoders. Next, the channel decoders generate extrinsic information to be used by the SIC detector as its *a priori* information to form a soft-input-soft-output turbo processing loop. Note that the concept of turbo equalization has been extensively studied for stationary communication systems [34] such as single-user [35] or multi-user scenarios [36], [37], which can achieve excellent performance.

Without loss of generality, we let $\mathcal{R} = \log_2 Q$ be the number of bits in each symbol. In this receiver, the SIC detector generates extrinsic probabilities $P_E (\mathbf{e}_g) , \forall g \in \{\mathcal{U}, \mathcal{V}\}$ by taking the feedback information $P_D (\mathbf{e}_g)$ from the channel decoders as the input *a priori* probabilities. To initialize, the SIC detector starts with equiprobable symbols without prior information from channel decoders. These extrinsic probabilities $P_E (\mathbf{e}_g)$ are then demapped and their bit log-likelihood ratios (LLRs) can be expressed as

$$L_E (\mathbf{d}_g[c\mathcal{R} + j]) = \log \frac{\sum_{\mathbf{e}_g[c] \in \mathbb{A} | \mathbf{d}_g[c\mathcal{R}+j]=0} P_E (\mathbf{e}_g[c])}{\sum_{\mathbf{e}_g[c] \in \mathbb{A} | \mathbf{d}_g[c\mathcal{R}+j]=1} P_E (\mathbf{e}_g[c])},$$
$$c = 0, \cdots, \frac{N_g}{\mathcal{R}} - 1, \quad (24)$$

where $\mathbf{d}_g[c\mathcal{R} + j]$ denoting the j-th bit associated to the c-th symbol of the g-th user. The resulted LLRs are deinterleaved as $L_E (\mathbf{a}_g)$ before passing to the channel decoders. The channel decoder computes an estimation of the information bits $\hat{\mathbf{b}}_g$, along with the extrinsic LLRs on the coded bits $\mathbf{a}_g$ according to

$$L_D (\mathbf{a}_g[i]) = \log \frac{\Pr (\mathbf{a}_g[i] = 0 | L_E (\mathbf{a}_g))}{\Pr (\mathbf{a}_g[i] = 1 | L_E (\mathbf{a}_g))} - L_E (\mathbf{a}_g[i]),$$
$$i = 1, 2, \cdots, N_g. \quad (25)$$

These extrinsic LLRs are interleaved as $L_D (\mathbf{d}_g)$ and mapped back to update the input *a priori* probabilities of SIC detector

$$P_D (\mathbf{e}_g[c] = \chi) \propto \prod_{j=1}^{\mathcal{R}} e^{-\varphi_j^{-1}(\chi) L_D(\mathbf{d}_g[c\mathcal{R}+j])}, \quad (26)$$

where $\varphi_j^{-1}(\chi)$ denotes the value of the j-th bit labelling the symbol $\chi \in \mathbb{A}$.

This "turbo" message passing process is repeated iteratively before terminating at a maximum iteration number n_t or upon meeting other preset stopping criteria. A detailed implementation of the iterative SIC turbo receiver is summarized in **Algorithm 1**. Here the SIC detector is adopted, where we detect stationary users' signal first before removing their

---

Algorithm 1 Iterative SIC Turbo Receiver

Initialization: $P_D (\mathbf{e}_g[c] = \chi) = 1/Q, c = 0, \cdots, \frac{N_g}{\mathcal{R}} - 1, \forall g \in \{\mathcal{U}, \mathcal{V}\}, \chi \in \mathbb{A}$.
for $\mathcal{T} = 0, 1, \cdots, n_t$ **do**
  SIC detector:
  1) Obtain $P (\mathbf{x}_{\mathcal{S}})$ and $P_E (\mathbf{e}_u) , \forall u \in \mathcal{U}$ by employing OAMP-LMMSE detector in **Algorithm 2**;
  2) Obtain $P_E (\mathbf{e}_v) , \forall v \in \mathcal{V}$ by employing GAMP-EP detector in **Algorithm 3**;
  3) Demap the extrinsic probabilities $P_E (\mathbf{e}_g) , \forall g \in \{\mathcal{U}, \mathcal{V}\}$ and compute their extrinsic bit LLRs $L_E (\mathbf{d}_g) , \forall g \in \{\mathcal{U}, \mathcal{V}\}$ in (24);
  4) Deinterleave $L_E (\mathbf{d}_g) , \forall g \in \{\mathcal{U}, \mathcal{V}\}$ and deliver them to the channel decoders;
  Channel decoders:
  5) Run each channel decoder to output $L_D (\mathbf{a}_g) , \forall g \in \{\mathcal{U}, \mathcal{V}\}$ in (25);
  6) Interleave the extrinsic LLRs $L_D (\mathbf{a}_g) , \forall g \in \{\mathcal{U}, \mathcal{V}\}$ and map them to obtain $P_D (\mathbf{e}_g) , \forall g \in \{\mathcal{U}, \mathcal{V}\}$ in (26);
end for
Output: The decisions of the information bits $\hat{\mathbf{b}}_g, \forall g \in \{\mathcal{U}, \mathcal{V}\}$ from the channel decoders.

---

contribution to the aggregated received signal via SIC. The mobile users' signal are then detected upon the removal of stationary users' signal. Based on different system models, we propose two detection algorithms next.

B. OAMP-LMMSE Detector for Stationary Users in OBNOMA

After obtaining the feedback information $P_D (\mathbf{e}_g) , \forall g \in \{\mathcal{U}, \mathcal{V}\}$ from the co-channel individual user channel decoders, we first initialize the *a priori* probabilities of $P_D (\mathbf{x}_{\mathcal{S}})$ and $P_D (\mathbf{x}_{\mathcal{M}})$ according to the indices of symbols for each user corresponding to $P_D (\mathbf{e}_g) , \forall g \in \{\mathcal{U}, \mathcal{V}\}$. We then project each entry of probabilities $P_D (\mathbf{x}_{\mathcal{M}})$ into Gaussian distribution, with respective mean and variance given by

$$\mu_{\mathcal{M}_c} = \sum_{\chi \in \mathbb{A}} \chi P_D (x_{\mathcal{M}_c} = \chi), \quad (27a)$$

$$\eta_{\mathcal{M}_c} = \sum_{\chi \in \mathbb{A}} |\chi|^2 P_D (x_{\mathcal{M}_c} = \chi) - |\mu_{\mathcal{M}_c}|^2, \quad (27b)$$

for $c = 0, 1, \cdots, MN - 1$. Next, we can approximately rewrite (23) as

$$\mathbf{y}_{\mathcal{S}} \simeq \mathbf{H}_{\mathcal{S}} \mathbf{x}_{\mathcal{S}} + \mathbf{z}_{\mathcal{S}}, \quad (28)$$

where $\mathbf{y}_{\mathcal{S}} = \mathbf{y} - \mathbf{H}_{\mathcal{M}} \boldsymbol{\mu}_{\mathcal{M}}$ and $\mathbf{z}_{\mathcal{S}}$ is modeled as $\mathcal{CN} (\mathbf{0}, \boldsymbol{\Sigma}_{\mathcal{S}})$ with covariance matrix $\boldsymbol{\Sigma}_{\mathcal{S}} = \boldsymbol{\Sigma}_{\boldsymbol{\omega}} + \mathbf{H}_{\mathcal{M}} \text{diag} \{\boldsymbol{\eta}_{\mathcal{M}}\} \mathbf{H}_{\mathcal{M}}^H$.

For $m = 0, 1, \cdots, M - 1$, we define the following notations

$$\mathbf{x}_{\mathcal{S}_k} = \mathbf{x}_{\mathcal{S}} [kN + m], \quad \mathbf{H}_{\mathcal{S}_k} = \mathbf{H}_{\mathcal{S}} [kN + m, kN + m],$$
$$\mathbf{y}_{\mathcal{S}_k} = \mathbf{y}_{\mathcal{S}} [kN + m], \quad \mathbf{z}_{\mathcal{S}_k} \sim \mathcal{CN} (\mathbf{0}, \boldsymbol{\Sigma}_{\mathcal{S}_k}),$$
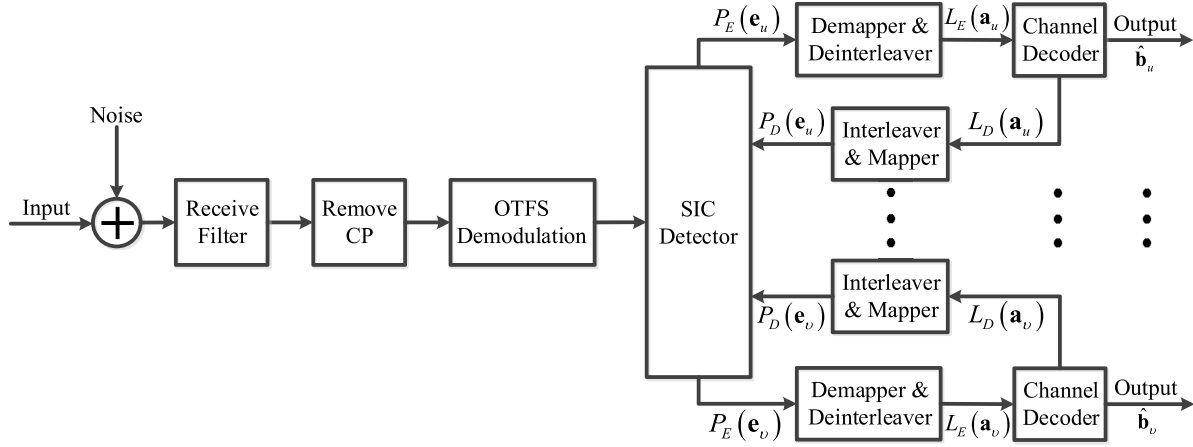$$\boldsymbol{\Sigma}_{\mathcal{S}_k} = \boldsymbol{\Sigma}_{\mathcal{S}} [kN + m, kN + m].$$

Fig. 3. Iterative SIC turbo receiver structure.

As a result, we can rewrite (28) as multiple "ISI-free" vector channels as in (15),

$$\mathbf{y}_{\mathcal{S}_k} \simeq \mathbf{H}_{\mathcal{S}_k} \mathbf{x}_{\mathcal{S}_k} + \mathbf{z}_{\mathcal{S}_k}, \quad k = 0, 1, \cdots, N-1. \quad (29)$$

Through this process, the stationary users' signal can be detected block by block as in (29). One can apply conventional linear receivers such as zero-forcing (ZF) and LMMSE [10], [12], [33] or an efficient message passing (MP) algorithm [13], [15] for symbol detection.

Recently, an OAMP algorithm is proposed in [38], where the extrinsic messages passed iteratively in the factor graph are only required to be orthogonal rather than stringent independent in the original MP. The successful performance improvement of OAMP motivates us to detect stationary users' signal by combining OAMP with LMMSE.

To describe the detail steps of the OAMP-LMMSE receiver, we focus on the k-th block without loss of generality. Specifically, we use $x_{\mathcal{S}_{k,m}}$ to denote the m-th symbol in the k-th block and assume its *a priori* distribution to be Gaussian, modeled as $q_D\left(x_{\mathcal{S}_{k,m}}\right) \sim \mathcal{CN}\left(\mu_{\mathcal{S}_{k,m}}, \eta_{\mathcal{S}_{k,m}}\right)$. Hence, the joint posteriori distribution can be decomposed as follows:

$$p\left(\mathbf{x}_{\mathcal{S}_k} \mid \mathbf{y}_{\mathcal{S}_k}, \mathbf{H}_{\mathcal{S}_k}\right) \propto p\left(\mathbf{y}_{\mathcal{S}_k} \mid \mathbf{x}_{\mathcal{S}_k}, \mathbf{H}_{\mathcal{S}_k}\right) \prod_{m=0}^{M-1} q_D\left(x_{\mathcal{S}_{k,m}}\right)$$
$$\sim \mathcal{CN}\left(\mathbf{A}_{\mathcal{S}_k}, \mathbf{B}_{\mathcal{S}_k}\right). \quad (30)$$

Here, we can represent (30) by using a factor graph, where a factor node $\mathbf{y}_{\mathcal{S}_k}$ is connected to multiple variable nodes $x_{\mathcal{S}_{k,m}}, m = 0, 1, \cdots, M-1$. We approximate the *a posteriori* distribution by computing and passing messages between the factor node $\mathbf{y}_{\mathcal{S}_k}$ and variable nodes $x_{\mathcal{S}_{k,m}}, m = 0, 1, \cdots, M-1$ in this factor graph iteratively. The OAMP-LMMSE detector is summarized in **Algorithm 2**. We now describe its detailed steps in iteration ι:

1) From Factor Node $\mathbf{y}_{\mathcal{S}_k}$ to Variable Nodes $x_{\mathcal{S}_{k,m}}, m = 0, 1, \cdots, M-1$: For simplicity, we can apply the LMMSE criterion at the factor node to compute the *a posteriori*

Algorithm 2 OAMP-LMMSE Detector

Input: $P_D\left(\mathbf{e}_g\right), \forall g \in \{\mathcal{U}, \mathcal{V}\}$, $\mathbf{y}$, $\mathbf{H}_{\mathcal{S}}$, $\mathbf{H}_{\mathcal{M}}$ and $n_{\mathcal{S}}$.
Initialization: $P_D\left(\mathbf{x}_{\mathcal{S}}\right)$, $P_D\left(\mathbf{x}_{\mathcal{M}}\right)$, $\boldsymbol{\mu}_{\mathcal{M}}$, $\boldsymbol{\eta}_{\mathcal{M}}$, $\mathbf{y}_{\mathcal{S}}$, $\boldsymbol{\mu}_{\mathcal{S}}^{(0)}$, $\boldsymbol{\eta}_{\mathcal{S}}^{(0)}$ and $\alpha_{\mathcal{S}}^{(0)} = 0$.
for $k = 0, 1, \cdots, N-1$ **do**
  Set iteration count $\iota = 1$.
  repeat
    1) Factor node $\mathbf{y}_{\mathcal{S}_k}$ generates the extrinsic mean $C_{\mathcal{S}_{k,m}}^{(\iota)}$ and variance $D_{\mathcal{S}_{k,m}}^{(\iota)}$ in (33), then sends them to the variable nodes $x_{\mathcal{S}_{k,m}}, m = 0, 1, \cdots, M-1$;
    2) Each variable node $x_{\mathcal{S}_{k,m}}$ computes the mean $\mu_{\mathcal{S}_{k,m}}^{(\iota)}$ and variance $\eta_{\mathcal{S}_{k,m}}^{(\iota)}$ in (36), and passes them back to the factor node $\mathbf{y}_{\mathcal{S}_k}$;
    3) Calculate the convergence indicator $\alpha_{\mathcal{S}}^{(\iota)}$ in (37);
    4) Update $P\left(\mathbf{x}_{\mathcal{S}_k}\right) = \bar{P}^{(\iota)}\left(\mathbf{x}_{\mathcal{S}_k}\right)$ and $P_E\left(\mathbf{x}_{\mathcal{S}_k}\right) = \bar{P}_E^{(\iota)}\left(\mathbf{x}_{\mathcal{S}_k}\right)$ if $\alpha_{\mathcal{S}}^{(\iota)} > \alpha_{\mathcal{S}}^{(\iota-1)}$;
    5) $\iota := \iota + 1$;
  until $\alpha_{\mathcal{S}}^{(\iota)} = 1$ or $\iota = n_{\mathcal{S}}$.
end for
Output: $P\left(\mathbf{x}_{\mathcal{S}}\right)$ and $P_E\left(\mathbf{e}_u\right), \forall u \in \mathcal{U}$.

---

distribution [12], [33], [39]:

$$\mathbf{B}_{\mathcal{S}_k}^{(\iota)} = \left(\mathbf{H}_{\mathcal{S}_k}^H \boldsymbol{\Sigma}_{\mathcal{S}_k}^{-1} \mathbf{H}_{\mathcal{S}_k} + \text{diag}\left\{\boldsymbol{\eta}_{\mathcal{S}_k}^{(\iota-1)}\right\}^{-1}\right)^{-1}, \quad (31)$$

$$\mathbf{A}_{\mathcal{S}_k}^{(\iota)} = \mathbf{B}_{\mathcal{S}_k}^{(\iota)} \left(\mathbf{H}_{\mathcal{S}_k}^H \boldsymbol{\Sigma}_{\mathcal{S}_k}^{-1} \mathbf{y}_{\mathcal{S}_k} + \text{diag}\left\{\boldsymbol{\eta}_{\mathcal{S}_k}^{(\iota-1)}\right\}^{-1} \boldsymbol{\mu}_{\mathcal{S}_k}^{(\iota-1)}\right), \quad (32)$$

where $\boldsymbol{\mu}_{\mathcal{S}_k}^{(\iota-1)}$ and $\boldsymbol{\eta}_{\mathcal{S}_k}^{(\iota-1)}$ are the mean and variance vectors for the symbols of k-th block, which are acquired from the variable nodes in the $(\iota - 1)$-th iteration and can be initialized in the first iteration by projecting the probabilities $P_D\left(\mathbf{x}_{\mathcal{S}_k}\right)$ from the channel decoders into Gaussian distributions. Following the Gaussian message combining rule [40], we then update the extrinsic marginal

distribution $q_E^{(\iota)}\left(x_{\mathcal{S}_{k,m}}\right) \sim \mathcal{CN}\left(C_{\mathcal{S}_{k,m}}^{(\iota)}, D_{\mathcal{S}_{k,m}}^{(\iota)}\right)$, with

$$D_{\mathcal{S}_{k,m}}^{(\iota)} = \left[\left(B_{\mathcal{S}_{k,m}}^{(\iota)}\right)^{-1} - \left(\eta_{\mathcal{S}_{k,m}}^{(\iota-1)}\right)^{-1}\right]^{-1}, \quad (33\text{a})$$

$$C_{\mathcal{S}_{k,m}}^{(\iota)} = D_{\mathcal{S}_{k,m}}^{(\iota)}\left[\frac{A_{\mathcal{S}_{k,m}}^{(\iota)}}{B_{\mathcal{S}_{k,m}}^{(\iota)}} - \frac{\mu_{\mathcal{S}_{k,m}}^{(\iota-1)}}{\eta_{\mathcal{S}_{k,m}}^{(\iota-1)}}\right], \quad (33\text{b})$$

where $B_{\mathcal{S}_{k,m}}^{(\iota)}$ denotes the m-th diagonal element of $\mathbf{B}_{\mathcal{S}_k}^{(\iota)}$. At last, the factor node sends the mean $C_{\mathcal{S}_{k,m}}^{(\iota)}$ and variance $D_{\mathcal{S}_{k,m}}^{(\iota)}$ to the variable nodes.

2) From Variable Nodes $x_{\mathcal{S}_{k,m}}, m = 0, 1, \cdots, M-1$ *to Factor Node* $\mathbf{y}_{\mathcal{S}_k}$: The *a posteriori* probability can be decomposed as follows at each variable node

$$\bar{P}^{(\iota)}\left(x_{\mathcal{S}_{k,m}} = \chi\right) \propto P_D\left(x_{\mathcal{S}_{k,m}} = \chi\right)$$
$$\times \exp\left(-\frac{\left|\chi - C_{\mathcal{S}_{k,m}}^{(\iota)}\right|^2}{D_{\mathcal{S}_{k,m}}^{(\iota)}}\right), \quad \forall \chi \in \mathbb{A}, \quad (34)$$

and then projected into a Gaussian distribution $\mathcal{CN}\left(E_{\mathcal{S}_{k,m}}^{(\iota)}, F_{\mathcal{S}_{k,m}}^{(\iota)}\right)$. In order to avoid numerical instabilities, we set a minimum allowed variance ε, i.e., $F_{\mathcal{S}_{k,m}}^{(\iota)} = \max\{\varepsilon, F_{\mathcal{S}_{k,m}}^{(\iota)}\}$. Then, we can update the extrinsic distribution $\bar{q}_E^{(\iota)}\left(x_{\mathcal{S}_{k,m}}\right) \sim \mathcal{CN}\left(\bar{\mu}_{\mathcal{S}_{k,m}}^{(\iota)}, \bar{\eta}_{\mathcal{S}_{k,m}}^{(\iota)}\right)$, where

$$\bar{\eta}_{\mathcal{S}_{k,m}}^{(\iota)} = \left[\left(F_{\mathcal{S}_{k,m}}^{(\iota)}\right)^{-1} - \left(D_{\mathcal{S}_{k,m}}^{(\iota)}\right)^{-1}\right]^{-1}, \quad (35\text{a})$$

$$\bar{\mu}_{\mathcal{S}_{k,m}}^{(\iota)} = \bar{\eta}_{\mathcal{S}_{k,m}}^{(\iota)}\left[\frac{E_{\mathcal{S}_{k,m}}^{(\iota)}}{F_{\mathcal{S}_{k,m}}^{(\iota)}} - \frac{C_{\mathcal{S}_{k,m}}^{(\iota)}}{D_{\mathcal{S}_{k,m}}^{(\iota)}}\right]. \quad (35\text{b})$$

To improve the performance and control the convergence speed of the algorithm, we apply a damping factor $\delta_{\mathcal{S}} \in (0, 1]$ [41]–[43], i.e.,

$$\eta_{\mathcal{S}_{k,m}}^{(\iota)} = \left[\frac{\delta_{\mathcal{S}}}{\bar{\eta}_{\mathcal{S}_{k,m}}^{(\iota)}} + \frac{(1-\delta_{\mathcal{S}})}{\eta_{\mathcal{S}_{k,m}}^{(\iota-1)}}\right]^{-1}, \quad (36\text{a})$$

$$\mu_{\mathcal{S}_{k,m}}^{(\iota)} = \eta_{\mathcal{S}_{k,m}}^{(\iota)}\left[\delta_{\mathcal{S}}\frac{\bar{\mu}_{\mathcal{S}_{k,m}}^{(\iota)}}{\bar{\eta}_{\mathcal{S}_{k,m}}^{(\iota)}} + (1-\delta_{\mathcal{S}})\frac{\mu_{\mathcal{S}_{k,m}}^{(\iota-1)}}{\eta_{\mathcal{S}_{k,m}}^{(\iota-1)}}\right]. \quad (36\text{b})$$

If the renewed variance $\eta_{\mathcal{S}_{k,m}}^{(\iota)}$ becomes negative, we would skip this update. Finally, $\mu_{\mathcal{S}_{k,m}}^{(\iota)}$ and $\eta_{\mathcal{S}_{k,m}}^{(\iota)}$ are passed back to the factor node.

3) Convergence Indicator: We define a convergence indicator $\alpha_{\mathcal{S}}^{(\iota)}$ for stationary users as

$$\alpha_{\mathcal{S}}^{(\iota)} = \frac{1}{M}\sum_{m=0}^{M-1}\mathbb{I}\left(\max_{\chi \in \mathbb{A}}\bar{P}^{(\iota)}\left(x_{\mathcal{S}_{k,m}} = \chi\right) \geq 1 - \varrho\right), \quad (37)$$

where $\mathbb{I}(\cdot)$ represents the indicator function and $\varrho > 0$ is a small value.

4) Update Criterion: If $\alpha_{\mathcal{S}}^{(\iota)} > \alpha_{\mathcal{S}}^{(\iota-1)}$, we update

$$P\left(\mathbf{x}_{\mathcal{S}_k}\right) = \bar{P}^{(\iota)}\left(\mathbf{x}_{\mathcal{S}_k}\right), \quad P_E\left(\mathbf{x}_{\mathcal{S}_k}\right) = \bar{P}_E^{(\iota)}\left(\mathbf{x}_{\mathcal{S}_k}\right), \quad (38)$$

where $\bar{P}_E^{(\iota)}\left(x_{\mathcal{S}_{k,m}} = \chi\right) \propto \exp\left(-\frac{\left|\chi - C_{\mathcal{S}_{k,m}}^{(\iota)}\right|^2}{D_{\mathcal{S}_{k,m}}^{(\iota)}}\right), \forall \chi \in \mathbb{A}$.

5) Stopping Criterion: The OAMP-LMMSE detector terminates when either $\alpha_{\mathcal{S}}^{(\iota)} = 1$ or the maximum iteration number $n_{\mathcal{S}}$ is reached.

Note that the OAMP-LMMSE can be used to detect each block's symbols parallelly, thus, the delay to detect the whole stationary users' signal is manageable. Finally, we obtain the extrinsic probabilities of $P_E\left(\mathbf{e}_u\right), \forall u \in \mathcal{U}$ according to the indices of each stationary user's symbols corresponding to $P_E\left(\mathbf{x}_{\mathcal{S}}\right)$, and output $P\left(\mathbf{x}_{\mathcal{S}}\right)$ and $P_E\left(\mathbf{e}_u\right), \forall u \in \mathcal{U}$.

C. GAMP-EP Detector for Mobile Users in OBNOMA

With the *a posteriori* probabilities $P\left(\mathbf{x}_{\mathcal{S}}\right)$ of the stationary users' symbols from OAMP-LMMSE detector, we first project each entry of these probabilities into Gaussian distribution, denoted as $\hat{q}\left(x_{\mathcal{S}_c}\right) \sim \mathcal{CN}\left(\hat{\mu}_{\mathcal{S}_c}, \hat{\eta}_{\mathcal{S}_c}\right), c = 0, 1, \cdots, MN-1$. From (23), we can approximate

$$\mathbf{y}_{\mathcal{M}} \simeq \mathbf{H}_{\mathcal{M}}\mathbf{x}_{\mathcal{M}} + \mathbf{z}_{\mathcal{M}}, \quad (39)$$

where $\mathbf{y}_{\mathcal{M}} = \mathbf{y} - \mathbf{H}_{\mathcal{S}}\hat{\boldsymbol{\mu}}_{\mathcal{S}}$ and $\mathbf{z}_{\mathcal{M}}$ is modeled as $\mathcal{CN}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\mathcal{M}}\right)$ with covariance matrix $\boldsymbol{\Sigma}_{\mathcal{M}} = \boldsymbol{\Sigma}_{\boldsymbol{\omega}} + \mathbf{H}_{\mathcal{S}}\text{diag}\left\{\hat{\boldsymbol{\eta}}_{\mathcal{S}}\right\}\mathbf{H}_{\mathcal{S}}^H$.

Direct solution of (39) by employing OAMP-LMMSE detector could be computationally costly since it involves a large matrix inverse while the typical value of MN can be in the order of thousands or even larger in OTFS system. Fortunately, $\mathbf{H}_{\mathcal{M}}$ is a sparse matrix and the index sets of non-zero components in the d-th row and c-th column can be denoted as $\mathcal{I}(d)$, $d = 0, 1, \cdots, MN-1$ and $\mathcal{J}(c)$, $c = 0, 1, \cdots, MN-1$, respectively. We also represent the corresponding numbers of non-zero components in the d-th row and c-th column as $|\mathcal{I}(d)|$ and $|\mathcal{J}(c)|$. Hence, we can use a sparsely connected factor graph to describe the system model of (39), where the entries of $\mathbf{y}_{\mathcal{M}}$ and $\mathbf{x}_{\mathcal{M}}$ are regarded as factor nodes and variable nodes, respectively.

Unlike the existing works in [13], [18], [25], [30], which use MP for symbol detection. Here, we propose a GAMP-EP detector for performance improvement. Note that EP algorithm is a Bayesian inference technique developed to approximate the true posterior. It has been already successfully applied for symbol detection in the stationary communication systems [42], [43] with modest complexity. We approximate the messages updated and passed between the factor nodes and variable nodes on the factor graph as Gaussian. **Algorithm 3** contains a detailed description of GAMP-EP detector, and the steps of the κ-th iteration are introduced below:

1) From Factor Node $y_{\mathcal{M}_d}$ *to Variable Nodes* $x_{\mathcal{M}_c}, c \in \mathcal{I}(d)$: At each factor node, we can represent the received signal $y_{\mathcal{M}_d}$ as

$$y_{\mathcal{M}_d} = H_{\mathcal{M}_{d,c}}x_{\mathcal{M}_c} + \sum_{e \in \mathcal{I}(d), e \neq c} H_{\mathcal{M}_{d,e}}x_{\mathcal{M}_e} + z_{\mathcal{M}_d}. \quad (40)$$

Algorithm 3 GAMP-EP Detector

Input: $P_D\left(\mathbf{e}_v\right), \forall v \in \mathcal{V}$, $P\left(\mathbf{x}_{\mathcal{S}}\right)$, $\mathbf{y}$, $\mathbf{H}_{\mathcal{S}}$, $\mathbf{H}_{\mathcal{M}}$ and $n_{\mathcal{M}}$.

Initialization: $P_D\left(\mathbf{x}_{\mathcal{M}}\right)$, $\hat{\boldsymbol{\mu}}_{\mathcal{S}}$, $\hat{\boldsymbol{\eta}}_{\mathcal{S}}$, $\mathbf{y}_{\mathcal{M}}$, $\mu_{\mathcal{M}_{d,c}}^{(0)} = \mu_{\mathcal{M}_c}$, $\eta_{\mathcal{M}_{d,c}}^{(0)} = \eta_{\mathcal{M}_c}$, $c = 0, 1, \cdots, MN - 1, d \in \mathcal{J}(c)$, $\alpha_{\mathcal{M}}^{(0)} = 0$ and iteration count $\kappa = 1$.

repeat

  1) Each factor node $y_{\mathcal{M}_d}$ generates the mean $C_{\mathcal{M}_{d,c}}^{(\kappa)}$ and variance $D_{\mathcal{M}_{d,c}}^{(\kappa)}$ in (41) and (42), then delivers them to the connected variable nodes $x_{\mathcal{M}_c}, c \in \mathcal{I}(d)$;

  2) Each variable node $x_{\mathcal{M}_c}$ computes the mean $\mu_{\mathcal{M}_{d,c}}^{(\kappa)}$ and variance $\eta_{\mathcal{M}_{d,c}}^{(\kappa)}$ in (45), and sends them back to the connected factor nodes $y_{\mathcal{M}_d}, d \in \mathcal{J}(c)$;

  3) Calculate the convergence indicator $\alpha_{\mathcal{M}}^{(\kappa)}$ in (46);

  4) Update $P\left(\mathbf{x}_{\mathcal{M}}\right) = \bar{P}^{(\kappa)}\left(\mathbf{x}_{\mathcal{M}}\right)$ and $P_E\left(\mathbf{x}_{\mathcal{M}}\right) = \bar{P}_E^{(\kappa)}\left(\mathbf{x}_{\mathcal{M}}\right)$ if $\alpha_{\mathcal{M}}^{(\kappa)} > \alpha_{\mathcal{M}}^{(\kappa-1)}$;

  5) $\kappa := \kappa + 1$;

until $\alpha_{\mathcal{M}}^{(\kappa)} = 1$ or $\kappa = n_{\mathcal{M}}$.

Output: $P_E\left(\mathbf{e}_v\right), \forall v \in \mathcal{V}$.

The messages passed from the factor node $y_{\mathcal{M}_d}$ to variable node $x_{\mathcal{M}_c}$ are the mean $C_{\mathcal{M}_{d,c}}^{(\kappa)}$ and variance $D_{\mathcal{M}_{d,c}}^{(\kappa)}$, respectively, given by

$$C_{\mathcal{M}_{d,c}}^{(\kappa)} = \left[y_{\mathcal{M}_d} - \sum_{e \in \mathcal{I}(d), e \neq c} H_{\mathcal{M}_{d,e}} \mu_{\mathcal{M}_{d,e}}^{(\kappa-1)} \right] \Big/ H_{\mathcal{M}_{d,c}}, \quad (41)$$

$$D_{\mathcal{M}_{d,c}}^{(\kappa)} = \left[\sum_{e \in \mathcal{I}(d), e \neq c} \left|H_{\mathcal{M}_{d,e}}\right|^2 \eta_{\mathcal{M}_{d,e}}^{(\kappa-1)} + \sigma_{\mathcal{M}_d} \right] \Big/ \left|H_{\mathcal{M}_{d,c}}\right|^2, \tag{42}$$

where $\mu_{\mathcal{M}_{d,e}}^{(\kappa-1)}$ and $\eta_{\mathcal{M}_{d,e}}^{(\kappa-1)}$ are the mean and variance received from variable node $x_{\mathcal{M}_e}$ in the $(\kappa-1)$-th iteration. They can be initialized in the first iteration according to (27). $\sigma_{\mathcal{M}_d}$ is the d-th diagonal element of $\boldsymbol{\Sigma}_{\mathcal{M}}$.

2) From Variable Node $x_{\mathcal{M}_c}$ to Factor Nodes $y_{\mathcal{M}_d}, d \in \mathcal{J}(c)$: The *a posteriori* probability at each variable node is given by

$$\bar{P}^{(\kappa)}\left(x_{\mathcal{M}_c} = \chi\right) \propto P_D\left(x_{\mathcal{M}_c} = \chi\right) \prod_{e \in \mathcal{J}(c)}$$

$$\times \exp\left(-\frac{\left|\chi - C_{\mathcal{M}_{e,c}}^{(\kappa)}\right|^2}{D_{\mathcal{M}_{e,c}}^{(\kappa)}}\right), \quad \forall \chi \in \mathbb{A}. \quad (43)$$

We again project this probability into a Gaussian distribution $\mathcal{CN}\left(E_{\mathcal{M}_c}^{(\kappa)}, F_{\mathcal{M}_c}^{(\kappa)}\right)$ and set a minimum allowed variance ε, i.e., $F_{\mathcal{M}_c}^{(\kappa)} = \max\{\varepsilon, F_{\mathcal{M}_c}^{(\kappa)}\}$ to avoid numerical instabilities. We then update the extrinsic distribution

$$\bar{q}_E^{(\kappa)}\left(x_{\mathcal{M}_{d,c}}\right) \sim \mathcal{CN}\left(\bar{\mu}_{\mathcal{M}_{d,c}}^{(\kappa)}, \bar{\eta}_{\mathcal{M}_{d,c}}^{(\kappa)}\right) \text{ in which}$$

$$\bar{\eta}_{\mathcal{M}_{d,c}}^{(\kappa)} = \left[\left(F_{\mathcal{M}_c}^{(\kappa)}\right)^{-1} - \left(D_{\mathcal{M}_{d,c}}^{(\kappa)}\right)^{-1} \right]^{-1}, \quad (44a)$$

$$\bar{\mu}_{\mathcal{M}_{d,c}}^{(\kappa)} = \bar{\eta}_{\mathcal{M}_{d,c}}^{(\kappa)} \left[\frac{E_{\mathcal{M}_c}^{(\kappa)}}{F_{\mathcal{M}_c}^{(\kappa)}} - \frac{C_{\mathcal{M}_{d,c}}^{(\kappa)}}{D_{\mathcal{M}_{d,c}}^{(\kappa)}} \right]. \quad (44b)$$

Finally, the variable node $x_{\mathcal{M}_c}$ updates the mean $\mu_{\mathcal{M}_{d,c}}^{(\kappa)}$ and variance $\eta_{\mathcal{M}_{d,c}}^{(\kappa)}$ as follows and delivers them to the factor node $y_{\mathcal{M}_d}$.

$$\eta_{\mathcal{M}_{d,c}}^{(\kappa)} = \left[\frac{\delta_{\mathcal{M}}}{\bar{\eta}_{\mathcal{M}_{d,c}}^{(\kappa)}} + \frac{(1 - \delta_{\mathcal{M}})}{\eta_{\mathcal{M}_{d,c}}^{(\kappa-1)}} \right]^{-1}, \quad (45a)$$

$$\mu_{\mathcal{M}_{d,c}}^{(\kappa)} = \eta_{\mathcal{M}_{d,c}}^{(\kappa)} \left[\delta_{\mathcal{M}} \frac{\bar{\mu}_{\mathcal{M}_{d,c}}^{(\kappa)}}{\bar{\eta}_{\mathcal{M}_{d,c}}^{(\kappa)}} + (1 - \delta_{\mathcal{M}}) \frac{\mu_{\mathcal{M}_{d,c}}^{(\kappa-1)}}{\eta_{\mathcal{M}_{d,c}}^{(\kappa-1)}} \right], \quad (45b)$$

where $\delta_{\mathcal{M}} \in (0, 1]$ is a damping factor applied to improve the accuracy and convergence. Similar to OAMP-LMMSE detector, we ignore the update if the variance $\eta_{\mathcal{M}_{d,c}}^{(\kappa)}$ is negative.

3) Convergence Indicator: The convergence indicator $\alpha_{\mathcal{M}}^{(\kappa)}$ for mobile users is defined as

$$\alpha_{\mathcal{M}}^{(\kappa)} = \frac{1}{MN} \sum_{c=0}^{MN-1} \mathbb{I}\left(\max_{\chi \in \mathbb{A}} \bar{P}^{(\kappa)}\left(x_{\mathcal{M}_c} = \chi\right) \geq 1 - \varrho\right). \quad (46)$$

4) Update Criterion: If $\alpha_{\mathcal{M}}^{(\kappa)} > \alpha_{\mathcal{M}}^{(\kappa-1)}$, we update

$$P\left(\mathbf{x}_{\mathcal{M}}\right) = \bar{P}^{(\kappa)}\left(\mathbf{x}_{\mathcal{M}}\right), \quad P_E\left(\mathbf{x}_{\mathcal{M}}\right) = \bar{P}_E^{(\kappa)}\left(\mathbf{x}_{\mathcal{M}}\right), \quad (47)$$

where $\bar{P}_E^{(\kappa)}\left(x_{\mathcal{M}_c} = \chi\right) \propto \prod_{e \in \mathcal{J}(c)} \exp\left(-\frac{\left|\chi - C_{\mathcal{M}_{e,c}}^{(\kappa)}\right|^2}{D_{\mathcal{M}_{e,c}}^{(\kappa)}}\right), \forall \chi \in \mathbb{A}$.

5) Stopping Criterion: The GAMP-EP detector terminates when either $\alpha_{\mathcal{M}}^{(\kappa)} = 1$ or the maximum iteration number $n_{\mathcal{M}}$ is reached.

Once the stopping criterion is satisfied, we obtain the extrinsic probabilities $P_E\left(\mathbf{e}_v\right), \forall v \in \mathcal{V}$ according to the indices of each mobile user's symbols corresponding to $P_E\left(\mathbf{x}_{\mathcal{M}}\right)$.

V. PERFORMANCE ANALYSIS AND COMPLEXITY REDUCTION

We now analyze the performance property of our proposed iterative SIC turbo receiver and develop reduced complexity variants for both the OAMP-LMMSE and GAMP-EP detectors.

A. Performance Analysis With EXIT Chart

Based on the main idea of EXIT chart [37], [44], [45], we develop a novel customized variant to analyze the convergence behavior of our proposed iterative SIC turbo receiver. An EXIT chart is a semi-analytical tool to study the transfer characteristics of mutual information (MI) between transmitted bits and their LLRs computed by receiver components through iterations in turbo detector. It has been widely adopted for convergence behavior analysis and prediction of turbo processing in stationary communication systems involving single

user [45] or multiple users [37]. Specifically, the receiver components are modeled as devices mapping the *a priori* MI I_i at the input to a new extrinsic MI I_e at the output. Based on EXIT charts, extrinsic information exchanges between the detector and the channel decoder can be visualized as a decoding trajectory. This yields an asymptotic convergence analysis for turbo receivers.

Unfortunately, EXIT chart cannot be directly applied to convergence analysis of NOMA systems because of user asymmetry in our proposed OBNOMA framework. To this end, we develop a customized EXIT chart for OBNOMA to analyze the convergence of our proposed iterative SIC turbo receiver.

Note that the input *a priori* information of mobile users will affect the output extrinsic information of stationary users in the SIC detector, and vice versa. Hence, we need to separately depict EXIT charts for stationary and mobile users. Specifically, we fix the input *a priori* MI of OBNOMA mobile users to several different values (i.e., different $I_i^{(\mathcal{M})}$) in the SIC detector and generate the corresponding EXIT chart for OBNOMA stationary users. The system decoding trajectory path will follow the transfer curves of the channel decoders and the detectors based on different $I_i^{(\mathcal{M})}$, and finally approach the desired operating point.

We can similarly generate EXIT chart for OBNOMA mobile users by fixing the input *a priori* MI of OBNOMA stationary users in the SIC detector. Compared with the traditional EXIT chart, the newly customized EXIT chart can provide more insights into the iterative behavior of the proposed SIC turbo receiver and graphically anticipate its convergence better.

To illustrate how the customized EXIT chart works for OBNOMA, Fig. 4 shows an example of the proposed iterative SIC turbo receiver in system with $U = 4$ stationary users and $V = 4$ mobile users. All user symbols are QPSK and channel decoders are low-density parity-check (LDPC) as introduced in Section VI. We also set the relative signal energy ratio between stationary and mobile users to $E_{\mathcal{S}}/E_{\mathcal{M}} = 5$ dB and fix the mobile users' SNR to $E_{\mathcal{M}}/N_0 = 3.5$ dB. Here, a typical urban channel model [46] is adopted for each user and the channel response for each mobile user is generated by utilizing Jakes formulation [10], [13], [30] with maximum Doppler spread equals to 1111 Hz. The stationary users' channels are generated with 0 Doppler shift.

From Fig. 4(a) and Fig. 4(b), we observe that the system trajectories are staircase traces between the transfer curves of the detector and decoder components for both the stationary and mobile users. In addition, the convergence region and average required number of iterations for the proposed iterative SIC turbo receiver can be predicted by EXIT charts. By checking the quantity of staircase projections (steps) through the trajectory curves in Fig. 4, we notice that four iterations are already enough to achieve the expected performance. This convergence results corroborate the average bit error rate (BER) of the receiver output for both stationary and mobile users in Fig. 5 with $E_{\mathcal{S}}/E_{\mathcal{M}} = 5$ dB. We found that the average BER drop becomes negligible beyond four iterations for both group of OBNOMA users.
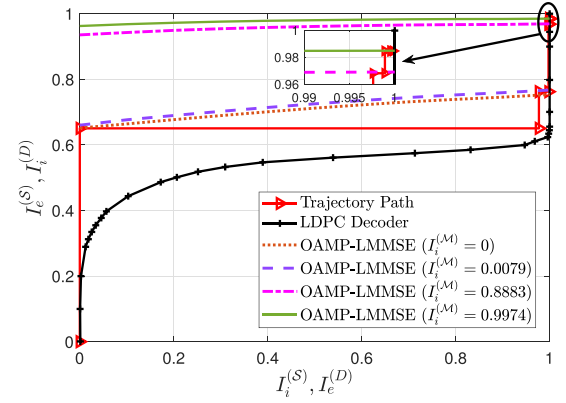
(a) EXIT chart for stationary users.
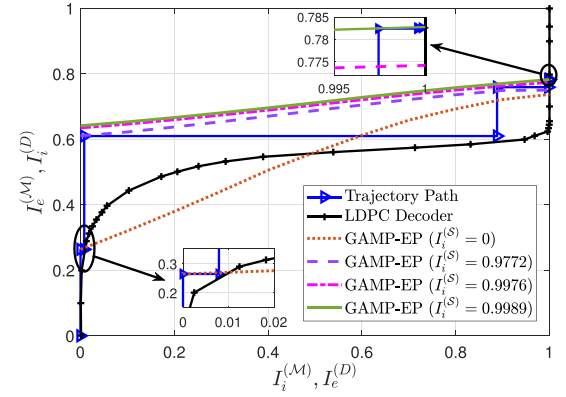
(b) EXIT chart for mobile users.

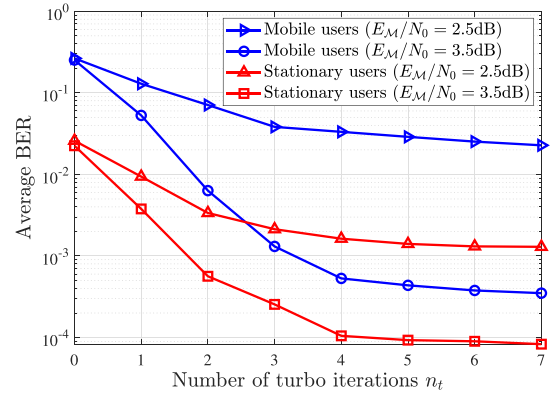Fig. 4. EXIT charts for the iterative SIC turbo receiver with $E_{\mathcal{S}}/E_{\mathcal{M}} = 5$ dB and $E_{\mathcal{M}}/N_0 = 3.5$ dB.

Fig. 5. Average BER convergence of OBNOMA users under $E_{\mathcal{S}}/E_{\mathcal{M}} = 5$ dB.

B. Complexity Reduction

The complexity of the proposed iterative SIC turbo receiver is mainly dominated by the advanced SIC detector. TABLE I summarizes the implementation complexity for each iteration of OAMP-LMMSE, GAMP-EP and traditional MP algorithms [13], [15], [18]. The computational cost is measured according to the total number of real-field multiplications,[1] exponential functions, and matrix inverses, respectively.

---

[1]Complex multiplication, inverse, and division account for three, four and six real-field multiplications, respectively.

TABLE I
COMPLEXITY COMPARISON OF DIFFERENT ALGORITHMS FOR EACH ITERATION

Algorithm	Real-field Multiplication	Exponential	Matrix Inverse
OAMP-LMMSE	$6M^2N(M+1)+10MNQ+34MN$	MNQ	$2N$
GAMP-EP	$4QD'+30D'+2MNQ+3MN$	QD'	-
R-OAMP-LMMSE	$3M^2N(M+3)+10MNQ+42MN$	MNQ	-
R-GAMP-EP	$2MNR(2Q+15)+2MNQ+3MN$	$MNRQ$	-
MP	$22QD'+2D'+MNQ$	QD'	-

We note that the OAMP-LMMSE detector complexity depends critically on matrix inverse. The GAMP-EP detector complexity is related to the number of non-zero channel terms (i.e., D') which represent channel matrix sparsity. Here, we write $\sum_{d=0}^{MN-1} |\mathcal{I}(d)| = \sum_{c=0}^{MN-1} |\mathcal{J}(c)| = D'$ for conciseness. However, the value D' can sometimes be relatively large due to many off-grid channel delays and Doppler shifts. To reduce receiver complexity, we propose the corresponding low-complexity alternatives for OAMP-LMMSE and GAMP-EP, respectively.

1) Reduced Complexity Algorithm for Stationary Users: The complexity of OAMP-LMMSE for stationary users mainly arise from the matrix inverse in (31) with order $\mathcal{O}(M^3N)$. As low-complexity approximations, similar to vector approximate message passing (VAMP) [47], we can use $\bar{\sigma}_{\mathcal{S}_k}\mathbf{I}$ and $\bar{\eta}_{\mathcal{S}_k}^{(\iota-1)}\mathbf{I}$ in place of $\boldsymbol{\Sigma}_{\mathcal{S}_k}$ and $\mathrm{diag}\{\boldsymbol{\eta}_{\mathcal{S}_k}^{(\iota-1)}\}$, respectively. The scalars $\bar{\sigma}_{\mathcal{S}_k}$ and $\bar{\eta}_{\mathcal{S}_k}^{(\iota-1)}$ are the sample average values of the diagonal elements of $\boldsymbol{\Sigma}_{\mathcal{S}_k}$ and the variance vector $\boldsymbol{\eta}_{\mathcal{S}_k}^{(\iota-1)}$, respectively. As a result, we can approximate (31) with

$$\mathbf{B}_{\mathcal{S}_k}^{(\iota)} \approx \mathbf{U}_{\mathcal{S}_k}^H \left((\bar{\sigma}_{\mathcal{S}_k})^{-1} \bar{\mathbf{H}}_{\mathcal{S}_k}^H \bar{\mathbf{H}}_{\mathcal{S}_k} + \left(\bar{\eta}_{\mathcal{S}_k}^{(\iota-1)} \right)^{-1} \mathbf{I} \right)^{-1} \mathbf{U}_{\mathcal{S}_k}, \quad (48)$$

where $\bar{\mathbf{H}}_{\mathcal{S}_k}$ and $\mathbf{U}_{\mathcal{S}_k}$ have the similar structures as $\bar{\mathbf{H}}_k$ in (14) and $\mathbf{U}_k$ in (16), respectively. Since $(\bar{\sigma}_{\mathcal{S}_k})^{-1}\bar{\mathbf{H}}_{\mathcal{S}_k}^H \bar{\mathbf{H}}_{\mathcal{S}_k} + \left(\bar{\eta}_{\mathcal{S}_k}^{(\iota-1)} \right)^{-1} \mathbf{I}$ is now diagonal, its inverse simply requires inverting the diagonal elements.

2) Reduced Complexity Algorithm for Mobile Users: As we can see, the channel factor graph in GAMP-EP has dense connections (edges), leading to relatively large value of D'. To reduce the resulting complexity, we adopt Gaussian approximation to trim part of these edges.

In particular, for each factor node $y_{\mathcal{M}_d}$, we would sort the corresponding $|\mathcal{I}(d)|$ channel coefficients according to their magnitudes, and choose R largest terms to connect the corresponding edges in the factor graph while eliminating others. Towards this, we can rewrite the received signal $y_{\mathcal{M}_d}$ at d-th factor node in (40) as

$$y_{\mathcal{M}_d} = \sum_{e\in\Phi(d)} H_{\mathcal{M}_{d,e}} x_{\mathcal{M}_e} + \underbrace{\sum_{e\in\bar{\Phi}(d)} H_{\mathcal{M}_{d,e}} x_{\mathcal{M}_e} + z_{\mathcal{M}_d}}_{z'_{\mathcal{M}_d}}, \quad (49)$$

where $\Phi(d)$ and $\bar{\Phi}(d)$ denote the index sets of the R largest terms and the rest $(|\mathcal{I}(d)|-R)$ terms in $\mathcal{I}(d)$, respectively. The eliminated terms plus the noise can be approximately modeled as a Gaussian random variable $z'_{\mathcal{M}_d}$, where its mean

and variance, respectively, given by

$$\mu_{z'_{\mathcal{M}_d}} = \sum_{e\in\bar{\Phi}(d)} H_{\mathcal{M}_{d,e}} \mu_{\mathcal{M}_e}, \quad (50a)$$

$$\sigma_{z'_{\mathcal{M}_d}} = \sum_{e\in\bar{\Phi}(d)} \left| H_{\mathcal{M}_{d,e}} \right|^2 \eta_{\mathcal{M}_e} + \sigma_{\mathcal{M}_d}. \quad (50b)$$

Through this approach, the channel factor graph is simplified and only the dominant edges shall participate in message updates to approximate the true posterior in GAMP-EP detector.

To summarize, we include the complexity analyses of the proposed reduced complexity algorithms for OAMP-LMMSE and GAMP-EP (denoted as R-OAMP-LMMSE and R-GAMP-EP, respectively) in TABLE I in comparison with the complexity of the original algorithms.

VI. SIMULATION RESULTS

The performance of our proposed coded uplink OBNOMA scheme and iterative SIC turbo receiver are evaluated for different deployment scenarios in this section. In our simulation setups, we apply carrier frequency of 4 GHz and subcarrier spacing $\Delta f = 15$ kHz. Unless otherwise mentioned, we modulate the symbols by Gray-mapped QPSK and set the rolloff factor of the RRC filters as 0.4 for both the transmitter and receiver. We generate a $(3, 6)$-regular LDPC of length 2048 with rate $1/2$ based on the progressive-edge growth (PEG) algorithm [48] and apply the belief propagation [49] with a maximum number of 100 iterations as channel decoder.

In OBNOMA, the delay-Doppler plane consists of $N = 32$ and $M = 128$. These delay-Doppler resources are allocated equally to $U = 4$ stationary users and $V = 4$ mobile users. A typical urban channel model [46] is applied with exponential power delay profile for both the stationary and mobile users. The velocity of the mobile user is set to $\lambda_v = 300$ km/h, resulting in a maximum Doppler spread $\nu_{\mathcal{M}_v,\max} = 1111$ Hz, $\forall v \in \mathcal{V}$. For simplicity, we generate the Doppler shift for the i-th delay of the v-th mobile user by utilizing the Jakes formulation [10], [13], [30], i.e., $\nu_{\mathcal{M}_v,i} = \nu_{\mathcal{M}_v,\max} \cos(\rho_{\mathcal{M}_v,i})$, where $\rho_{\mathcal{M}_v,i}$ is uniformly distributed over $[-\pi, \pi]$.

We first assume that full CSI is available at the receiver and also study the impact of channel uncertainties on receiver performance. Without loss of generality, we set $\delta_{\mathcal{S}} = \delta_{\mathcal{M}} = 0.3$, $\varepsilon = 10^{-8}$, $\varrho = 0.1$, $n_{\mathcal{S}} = n_{\mathcal{M}} = 20$ and choose $n_t = 4$. These parameters were selected after extensive experimentations as a compromise between convergence speed and accuracy. All simulation results are averaged over 500 independent realizations.
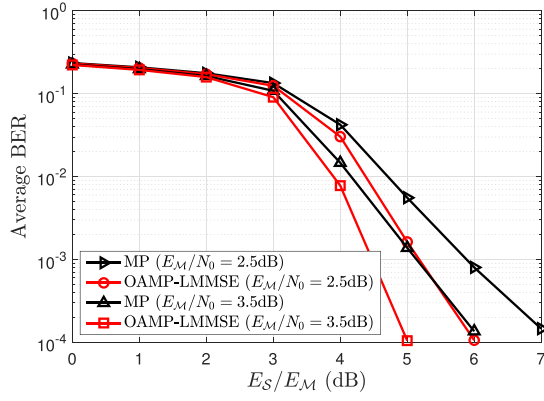
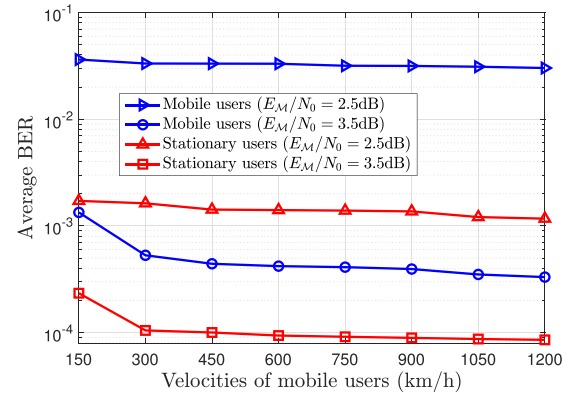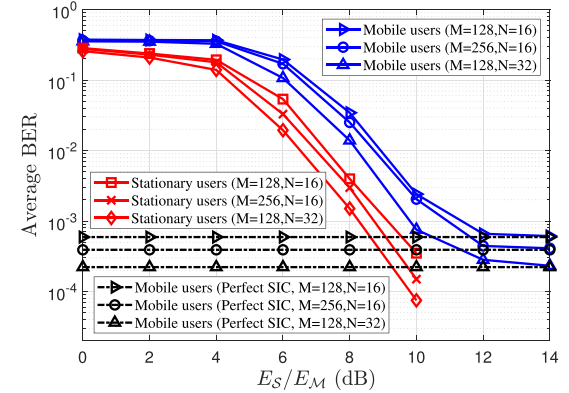(a) Average BER performance comparison for stationary users.
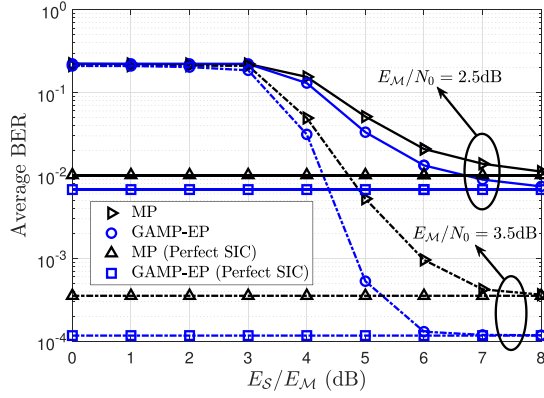
(b) Average BER performance comparison for mobile users.

Fig. 6. Average BER performance comparison of OBNOMA with different detector Algorithms.

Fig. 7. Average BER performance of OBNOMA users at different mobile velocities with $E_S/E_M = 5$ dB.

Fig. 8. Average BER performance of OBNOMA users utilizing 16-QAM symbols with $E_M/N_0 = 7.5$ dB.

Fig. 6 compares the average BER performance of the OBNOMA system with different detector algorithms for both the stationary users (Fig. 6(a)) and mobile users (Fig. 6(b)). To highlight the predominance of the proposed detecting algorithms, we also provide the performance of traditional MP algorithm [13], [15], [18] as baselines for both the stationary and mobile users in Fig. 6.

The results reveal that all the receivers deliver improved performance with higher E_S/E_M and E_M/N_0. However, our proposed OAMP-LMMSE and GAMP-EP detectors outperform the MP detectors for both the stationary and mobile users. We also observe that as E_S/E_M grows, the performance of both stationary and mobile users would improve. In particular, the mobile users' performance would asymptotically approach the performance of perfect SIC. Based on these analysis, we demonstrate that our proposed iterative SIC turbo receiver is practical and robustness against influence of imperfect SIC process.

Fig. 7 shows the average BER performance of OBNOMA system at different mobile users' velocities with $E_S/E_M = 5$ dB. As the velocities of the mobile users grow, we observe that the receiver performance improves modestly before saturation for velocities beyond 450 km/h. The underlying reason is that OTFS modulation can resolve high contrast paths in the Doppler dimension at higher mobile users' velocities. Consequently, performance advantage becomes conceivable at higher user velocities (i.e., high Doppler spread channels).
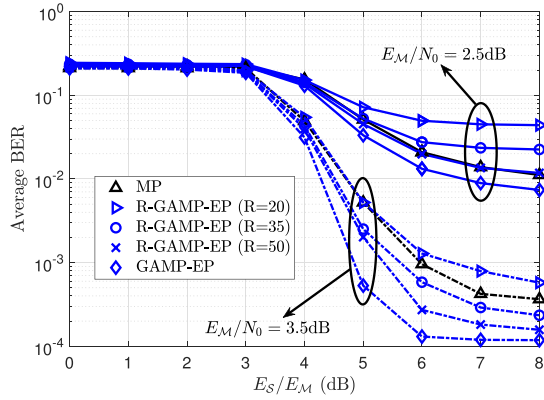
We also test the average BER performance of OBNOMA when user symbols are modulated as 16-QAM. Fig. 8 illustrates the average BER performance of OBNOMA users with 16-QAM under $E_M/N_0 = 7.5$ dB for different settings of M and N. We notice that the performance of both stationary and mobile users degrades for smaller M and N due to loss of delay-Doppler grid resolution. This results in the diversity loss as the receiver can only resolve a smaller number of signal paths. These tests and results strongly support the operational consistency of our proposed OBNOMA scheme and iterative SIC turbo receiver over different system parameters.

In terms of complexity reduction, Fig. 9 shows the average BER performance of OBNOMA system for receivers utilizing the proposed reduced complexity detectors. In Fig. 9(a), we observe that the R-OAMP-LMMSE detector achieves similar performance to that of OAMP-LMMSE detector for stationary users without costly matrix inverses, and achieves better performance than traditional MP detector. The results in Fig. 9(b) demonstrates a graceful performance degradation for mobile users as the algorithm complexity drops with smaller R. The results also reveal that as R increases, the performance of R-GAMP-EP detector would asymptotically approach to that of full GAMP-EP detector and even better than that of traditional MP detector for mobile users. We also notice that the value of R has slightly effect on the performance of stationary users. Therefore, our
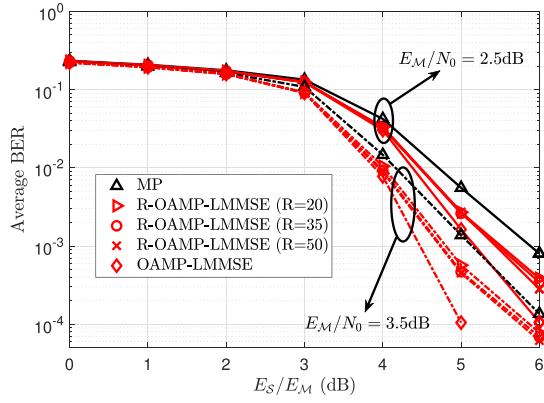
(a) Average BER performance for stationary users.

(b) Average BER performance for mobile users.

Fig. 9. Average BER performance of OBNOMA with reduced complexity detectors.

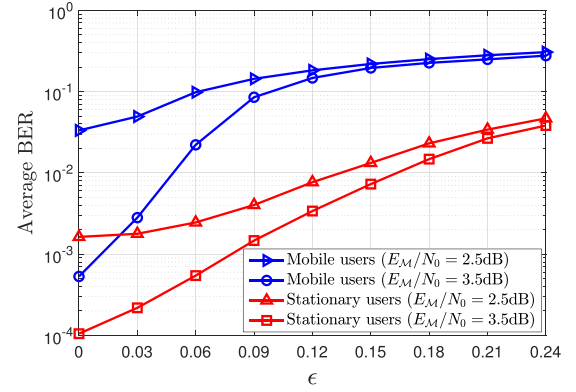Fig. 10. Average BER performance of OBNOMA users with imperfect CSI under $E_\mathcal{S}/E_\mathcal{M} = 5$ dB.

proposed R-OAMP-LMMSE and R-GAMP-EP detectors can yield attractive compromise between receiver performance and complexity.

Finally, the performance of the proposed iterative SIC turbo receiver are tested for imperfect CSI. Here, we characterize the channel uncertainties under norm-bounded CSI estimation errors, which can be modeled as

$$h_{x,i} = \hat{h}_{x,i} + \Delta h_{x,i}, \quad \|\Delta h_{x,i}\| \leq \epsilon_{h_{x,i}},$$
$$\tau_{x,i} = \hat{\tau}_{x,i} + \Delta \tau_{x,i}, \quad \|\Delta \tau_{x,i}\| \leq \epsilon_{\tau_{x,i}},$$
$$\nu_{\mathcal{M}_v,i} = \hat{\nu}_{\mathcal{M}_v,i} + \Delta \nu_{\mathcal{M}_v,i}, \quad \|\Delta \nu_{\mathcal{M}_v,i}\| \leq \epsilon_{\nu_{\mathcal{M}_v,i}},$$

where $\forall x \in \{\mathcal{S}_u, \mathcal{M}_v\}$, $\hat{h}_{x,i}$, $\hat{\tau}_{x,i}$ and $\hat{\nu}_{\mathcal{M}_v,i}$ denote the estimated values of $h_{x,i}$, $\tau_{x,i}$ and $\nu_{\mathcal{M}_v,i}$. The corresponding channel estimation errors $\Delta h_{x,i}$, $\Delta \tau_{x,i}$ and $\Delta \nu_{\mathcal{M}_v,i}$ are bounded in their norms. The model specifies the respective norm bounds of $\epsilon_{h_{x,i}}$, $\epsilon_{\tau_{x,i}}$ and $\epsilon_{\nu_{\mathcal{M}_v,i}}$, respectively. For brevity, we assume that $\epsilon_{h_{x,i}} = \epsilon \left\|\hat{h}_{x,i}\right\|$, $\epsilon_{\tau_{x,i}} = \epsilon \left\|\hat{\tau}_{x,i}\right\|$ and $\epsilon_{\nu_{\mathcal{M}_v,i}} = \epsilon \left\|\hat{\nu}_{\mathcal{M}_v,i}\right\|, \forall u, v, i$.

From the results in Fig. 10 under $E_\mathcal{S}/E_\mathcal{M} = 5$ dB, we observe that the performance loss of our proposed schemes is mild for the modest values of channel uncertainty ϵ. The performance of mobile users is more sensitive to the CSI uncertainty than the stationary users. The graceful degradation of receiver performance with increasing amount channel uncertainty demonstrate the robustness of our proposed OBNOMA framework and the iterative SIC turbo receiver against channel modeling errors.

VII. CONCLUSION

In this paper, we proposed a novel coded uplink multi-user system to achieve high spectrum efficiency through NOMA. Our proposed OBNOMA framework groups users with different mobility profiles for channel sharing. Based on the recently developed OTFS technology, we allocated the sub-vector resources of Doppler and delay dimensions respectively to the stationary and mobile users. We developed an iterative SIC turbo receiver for effective multi-user detection and decoding under strong CCI of OBNOMA system. We derived two detectors for different user mobility profiles within the turbo receiver. We also proposed reduced complexity variants for both the detector algorithms without significant performance drop. Our EXIT chart analysis further verified the rapid convergence of the proposed receivers. Our results demonstrated the feasibility of OBNOMA as well as strong performance and robustness of the proposed turbo receiver against channel uncertainty and errors from imperfect SIC.

APPENDIX

From (16), we have

$$
\begin{aligned}
H_k[\ell, m] &= \sum_{r=0}^{M-1} \Lambda_k^H[\ell, \ell] F_M^H[\ell, r] \bar{H}_k[r, r] F_M[r, m] \Lambda_k[m, m] \\
&= \frac{1}{M} \sum_{r=0}^{M-1} e^{j\frac{2\pi\ell k}{MN}} e^{j\frac{2\pi\ell r}{M}} H[k+rN] e^{-j\frac{2\pi rm}{M}} e^{-j\frac{2\pi km}{MN}} \\
&= \frac{1}{M} \sum_{r=0}^{M-1} H[k+rN] e^{j\frac{2\pi(\ell-m)k}{MN}} e^{j\frac{2\pi(\ell-m)r}{M}} \\
&= \frac{1}{M} \sum_{r=0}^{M-1} \sum_{p=0}^{P-1} h[p] e^{-j\frac{2\pi(k+rN)p}{MN}} e^{j\frac{2\pi(\ell-m)k}{MN}} e^{j\frac{2\pi(\ell-m)r}{M}} \\
&= \frac{1}{M} \sum_{p=0}^{P-1} h[p] e^{j\frac{2\pi(\ell-m-p)k}{MN}} \sum_{r=0}^{M-1} e^{j\frac{2\pi(\ell-m-p)r}{M}} \\
&= \sum_{p=0}^{P-1} h[p] e^{j\frac{2\pi(\ell-m-p)k}{MN}} \delta([\ell-m-p]_M). \quad (51)
\end{aligned}
$$

Thus, using the definition of $\bar{\gamma}(k,\ell,p)$ in (12), we have the following relationship from (15)

$$
\begin{aligned}
Y[\ell,k] &= \sum_{m=0}^{M-1} H_k[\ell,m]X[m,k] + \omega[\ell,k] \\
&= \sum_{m=0}^{M-1}\sum_{p=0}^{P-1} h[p]e^{j\frac{2\pi(\ell-m-p)k}{MN}}\delta([\ell-m-p]_M)X[m,k] \\
&\quad + \omega[\ell,k] \\
&= \sum_{p=0}^{P-1} h[p]e^{j\frac{2\pi(\ell-[\ell-p]_M-p)k}{MN}}X[[\ell-p]_M,k] + \omega[\ell,k] \\
&= \sum_{p=0}^{P-1} h[p]\bar{\gamma}(k,\ell,p)X[[\ell-p]_M,k] + \omega[\ell,k], \qquad (52)
\end{aligned}
$$

which completes the proof.

REFERENCES

[1] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[2] Z. Ding et al., "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[3] 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.

[4] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.

[5] S. Sharma, K. Deka, V. Bhatia, and A. Gupta, "Joint power-domain and SCMA-based NOMA system for downlink in 5G and beyond," *IEEE Commun. Lett.*, vol. 23, no. 6, pp. 971–974, Jun. 2019.

[6] R. Hadani et al., "Orthogonal time frequency space modulation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, San Francisco, CA, USA, Mar. 2017, pp. 1–6.

[7] W. Shen, L. Dai, J. An, P. Fan, and R. W. Heath, Jr., "Channel estimation for orthogonal time frequency space (OTFS) massive MIMO," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4204–4217, Aug. 2019.

[8] P. Raviteja, K. T. Phan, and Y. Hong, "Embedded pilot-aided channel estimation for OTFS in delay–Doppler channels," *IEEE Trans. Veh. Tech.*, vol. 68, no. 5, pp. 4906–4917, May 2019.

[9] Y. Liu, S. Zhang, F. Gao, J. Ma, and X. Wang, "Uplink-aided high mobility downlink channel estimation over massive MIMO-OTFS system," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 1994–2009, Sep. 2020.

[10] G. D. Surabhi and A. Chockalingam, "Low-complexity linear equalization for OTFS modulation," *IEEE Commun. Lett.*, vol. 24, no. 2, pp. 330–334, Feb. 2020.

[11] K. R. Murali and A. Chockalingam, "On OTFS modulation for high-Doppler fading channels," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2018, pp. 1–10.

[12] S. Tiwari, S. S. Das, and V. Rangamgari, "Low complexity LMMSE receiver for OTFS," *IEEE Commun. Lett.*, vol. 23, no. 12, pp. 2205–2209, Dec. 2019.

[13] P. Raviteja, K. T. Phan, Y. Hong, and E. Viterbo, "Interference cancellation and iterative detection for orthogonal time frequency space modulation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6501–6515, Oct. 2018.

[14] W. Yuan, Z. Wei, J. Yuan, and D. W. K. Ng, "A simple variational Bayes detector for orthogonal time frequency space (OTFS) modulation," *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 7976–7980, Jul. 2020.

[15] P. Raviteja, E. Viterbo, and Y. Hong, "OTFS performance on static multipath channels," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 745–748, Jun. 2019.

[16] G. D. Surabhi, R. M. Augustine, and A. Chockalingam, "On the diversity of uncoded OTFS modulation in doubly-dispersive channels," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3049–3063, Jun. 2019.

[17] P. Raviteja, Y. Hong, E. Viterbo, and E. Biglieri, "Effective diversity of OTFS modulation," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 249–253, Feb. 2020.

[18] M. K. Ramachandran and A. Chockalingam, "MIMO-OTFS in high-Doppler fading channels: Signal detection and channel estimation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 206–212.

[19] G. D. Surabhi, M. K. Ramachandran, and A. Chockalingam, "OTFS modulation with phase noise in mmWave communications," in *Proc. IEEE 89th Veh. Technol. Conf. (VTC-Spring)*, Apr. 2019, pp. 1–5.

[20] V. Khammammetti and S. K. Mohammed, "OTFS-based multiple-access in high Doppler and delay spread wireless channels," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 528–531, Apr. 2019.

[21] R. M. Augustine and A. Chockalingam, "Interleaved time-frequency multiple access using OTFS modulation," in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, Honolulu, HI, USA, Sep. 2019, pp. 1–5.

[22] G. D. Surabhi, R. Mary Augustine, and A. Chockalingam, "Multiple access in the delay-Doppler domain using OTFS modulation," 2019, *arXiv:1902.03415*. [Online]. Available: http://arxiv.org/abs/1902.03415

[23] M. Li, S. Zhang, F. Gao, P. Fan, and O. A. Dobre, "A new path division multiple access for the massive MIMO-OTFS networks," 2020, *arXiv:2003.08228*. [Online]. Available: http://arxiv.org/abs/2003.08228

[24] A. Chatterjee, V. Rangamgari, S. Tiwari, and S. S. Das, "Non orthogonal multiple access with orthogonal time frequency space signal transmission," 2020, *arXiv:2003.06387*. [Online]. Available: http://arxiv.org/abs/2003.06387

[25] K. Deka, A. Thomas, and S. Sharma, "OTFS-NOMA based on SCMA," 2020, *arXiv:2005.03216*. [Online]. Available: http://arxiv.org/abs/2005.03216

[26] G. Matz, H. Bolcskei, and F. Hlawatsch, "Time-frequency foundations of communications: Concepts and tools," *IEEE Signal Process. Mag.*, vol. 30, no. 6, pp. 87–96, Nov. 2013.

[27] Z. Ding, R. Schober, P. Fan, and H. V. Poor, "OTFS-NOMA: An efficient approach for exploiting heterogenous user mobility profiles," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7950–7965, Nov. 2019.

[28] Z. Ding, "Robust beamforming design for OTFS-NOMA," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 33–40, 2020.

[29] P. Raviteja, Y. Hong, E. Viterbo, and E. Biglieri, "Practical pulse-shaping waveforms for reduced-cyclic-prefix OTFS," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 957–961, Jan. 2019.

[30] Y. Ge, Q. Deng, P. C. Ching, and Z. Ding, "Receiver design for OTFS with fractionally spaced sampling approach," *IEEE Trans. Wireless Commun.*, pp. 1–15, doi: 10.1109/TWC.2021.3055585.

[31] Z. Jian, A. D. S. Jayalath, and Y. Chen, "Asymmetric OFDM systems based on layered FFT structure," *IEEE Signal Process. Lett.*, vol. 14, no. 11, pp. 812–815, Nov. 2007.

[32] X.-G. Xia, "Precoded and vector OFDM robust to channel spectral nulls and with reduced cyclic prefix length in single transmit antenna systems," *IEEE Trans. Commun.*, vol. 49, no. 8, pp. 1363–1374, Aug. 2001.

[33] Y. Li, I. Ngebani, X.-G. Xia, and A. Host-Madsen, "On performance of vector OFDM with linear receivers," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5268–5280, Oct. 2012.

[34] M. Tuchler and A. C. Singer, "Turbo equalization: An overview," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 920–952, Feb. 2011.

[35] C. Douillard et al., "Iterative correction of intersymbol interference: Turbo-equalization," *Eur. Trans. Telecommun.*, vol. 6, no. 5, pp. 507–511, Sep. 1995.

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

[37] K. Li and X. Wang, "EXIT chart analysis of turbo multiuser detection," *IEEE Trans. Wireless Commun.*, vol. 4, no. 1, pp. 300–311, Jan. 2005.

[38] J. Ma, L. Liu, X. Yuan, and L. Ping, "On orthogonal AMP in coded linear vector systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 12, pp. 5658–5672, Dec. 2019.

[39] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[40] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proc. IEEE*, vol. 95, no. 6, pp. 1295–1322, Jun. 2007.

[41] T. Minka, "Divergence measures and message passing," Microsoft Res., Redmond, WA, USA, Tech. Rep. MSR-TR-2005-173, 2005.

[42] I. Santos, J. J. Murillo-Fuentes, E. Arias-de-Reyna, and P. M. Olmos, "Turbo EP-based equalization: A filter-type implementation," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 4259–4270, Sep. 2018.

[43] S. Şahin, A. M. Cipriano, C. Poulliat, and M.-L. Boucheret, "Iterative equalization with decision feedback based on expectation propagation," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4473–4487, Oct. 2018.

[44] M. El-Hajjar and L. Hanzo, "EXIT charts for system design and analysis," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 127–153, 1st Quart., 2014.

[45] H. Lou and C. Xiao, "Soft-decision feedback turbo equalization for multilevel modulations," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 186–195, Jan. 2011.

[46] M. Failli, *Digital Land Mobile Radio Communications: COST 207.* Luxembourg City, Luxembourg: European Communities, 1989.

[47] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6664–6684, Oct. 2019.

[48] X.-Y. Hu, E. Eleftheriou, and D. M. Arnold, "Regular and irregular progressive edge-growth tanner graphs," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 386–398, Jan. 2005.

[49] T. Richardson and R. Urbanke, *Modern Coding Theory.* Cambridge, U.K.: Cambridge Univ. Press, 2008.

Yao Ge (Graduate Student Member, IEEE) received the B.Eng. degree in electronics and information engineering and the M.Eng. degree (research) in communication and information system from Northwestern Polytechnical University (NPU), Xi'an, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree in electronic engineering with The Chinese University of Hong Kong (CUHK), Hong Kong. From October 2015 to March 2016, he was a Visiting Student with the Department of Electrical and Computer Systems Engineering, Monash University, Melbourne, VIC, Australia. From April 2016 to August 2016, he was a Visiting Student with the Department of Computer, Electrical and Mathematical Science and Engineering, King Abdullah University of Science and Technology (KAUST), Saudi Arabia. From May 2019 to December 2019, he was a Visiting Student with the Department of Electrical and Computer Engineering, University of California at Davis (UC Davis), Davis, CA, USA. His current research interests include wireless communications, information theory, and statistical signal processing.

Qinwen Deng (Graduate Student Member, IEEE) received the B.Eng. degree in electronic information engineering from the University of Science and Technology of China (USTC), Hefei, China, in 2018. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of California at Davis (UC Davis), Davis, CA, USA. His current research interests include wireless communications and hypergraph signal processing.

P. C. Ching (Life Fellow, IEEE) received the B.Eng. (Hons.) and Ph.D. degrees from the University of Liverpool, U.K., in 1977 and 1981, respectively. From 1981 to 1982, he was a Research Officer with the University of Bath, U.K. In 1982, he returned to Hong Kong and then joined the Hong Kong Polytechnic as a Lecturer. Since 1984, he has been with the Department of Electronic Engineering, The Chinese University of Hong Kong (CUHK), where he is currently a Choh-Ming Li Professor of Electronic Engineering. He was the Department Chairman from 1995 to 1997, the Dean of Engineering from 1997 to 2003, and the Head of Shaw College from 2004 to 2008. He became the Director of the Shun Hing Institute of Advanced Engineering in 2004. From 2006 to 2014, he was appointed as a Pro-Vice-Chancellor/Vice-President with CUHK. From 2013 to 2014, he also took up the Directorship of the CUHK Shenzhen Research Institute. He is very active in promoting professional activities, both in Hong Kong and overseas. He was a Council Member of the Institution of Electrical Engineers (IEE) and a Past Chairman of the IEEE Hong Kong Section. He was also a member of the Technical Committee of the IEEE Signal Processing Society from 1996 to 2004. He has been an Honorary Member of the Editorial Committee for the *Journal of Data Acquisition and Processing* since 2000. He has been instrumental in organizing many international conferences in Hong Kong, including the 1997 IEEE International Symposium on Circuits and Systems, where he was the Vice-Chairman. He also served as a Technical Program Co-Chair for the 2003 and 2016 IEEE International Conference on Acoustics, Speech and Signal Processing. He received the IEEE Third Millennium Award in 2000 and the HKIE Hall of Fame in 2010. He is also plays an active role in community services. He received the Bronze Bauhinia Star (BBS) from the HKSAR Government and the Silver Bauhinia Star (SBS) in 2010 and 2017, respectively, in recognition of his long and distinguished public and community services. He is currently the Chairman of the Board of Directors of the Nano and Advanced Materials Institute Ltd., a Council Member of the Shaw Prize Foundation, and a member of the Museum Advisory Committee (MAC) and the Chairperson of its Science Sub-committee. He is elected as the President of the Hong Kong Academy of Engineering Sciences (HKAES) in 2018. He was appointed as the Editor-in-Chief of the *HKIE Transactions* from 2001 to 2004 and an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1997 to 2000 and IEEE SIGNAL PROCESSING LETTERS from 2001 to 2003.
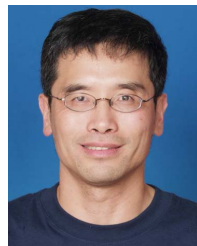
Zhi Ding (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Cornell University in 1990.

From 1990 to 2000, he was a Faculty Member with Auburn University and later, University of Iowa. He has held visiting positions with Australian National University, the Hong Kong University of Science and Technology, the NASA Lewis Research Center, and the USAF Wright Laboratory. He is with the Department of Electrical and Computer Engineering, University of California, Davis, where he currently holds the position of a Distinguished Professor. He has active collaboration with researchers from various universities in Australia, Canada, China, Finland, Hong Kong, Japan, South Korea, Singapore, Taiwan, and USA. He has been an Active Member of IEEE, serving on technical programs for several workshops and conferences. He was a member of Technical Committee on Statistical Signal and Array Processing and a member of Technical Committee on Signal Processing for Communications from 1994 to 2003. He served as a Steering Committee Member for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2007 to 2009 and its Chair from 2009 to 2010. He is a coauthor of the text: *Modern Digital and Analog Communication Systems* (5th edition, Oxford University Press, 2019). He received the IEEE Communication Society's WTC Award in 2012 and the IEEE Communication Society's Education Award in 2020. He was the General Chair of the 2016 IEEE International Conference on Acoustics, Speech, and Signal Processing and the Technical Program Chair of the 2006 IEEE Globecom. He was also an IEEE Distinguished Lecturer Circuits and Systems Society from 2004 to 2006, the Communications Society from 2008 to 2009. He was an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1994 to 1997 and 2001 to 2004, and an Associate Editor of IEEE SIGNAL PROCESSING LETTERS from 2002 to 2005.