Iterative NOMA Detection for Multiple Access in Satellite High-Mobility Communications

Jiawei Wang, *Graduate Student Member, IEEE*, Chunxiao Jiang, *Senior Member, IEEE*, and Linling Kuang, *Member, IEEE*

***Abstract*— Non-orthogonal multiple access (NOMA) is a promising technology for next generation multiple access (NGMA). However, traditional NOMA detection methods cannot cope with challenges of NGMA in satellite high-mobility communications. On the one hand, due to the high mobility and heterogeneity of terminals in terms of the velocity and acceleration, time-varying Doppler shifts caused by high-mobility terminals are relatively higher and different for each user, which degrades the demodulation performance seriously and makes multi-user interference cancellation become the bottleneck. On the other hand, owning to wide distribution of high-mobility terminals, the arriving time of users' signals is different at the receiver, which further incurs more difficulties for NOMA detection. To solve such a problem in satellite high-mobility communications, we propose a novel multi-user detection method based on the new three-dimensional (3D) factor graph for high-mobility environments, named the turbo iterative detection (TID) algorithm. Specially, the proposed algorithm consists of the interference cancellation loop, the Doppler elimination loop and the decoding loop. By means of message passing along edges in the proposed 3D factor graph, these three iterative loops can interact with each other to effectively eliminate time-varying Doppler shifts of heterogeneous high-mobility terminals and interference among multiple users. Simulation results show that the proposed algorithm improves the bit error ratio (BER) performance more than 0.9 dB with less computational complexity compared with traditional algorithms, which demonstrates the superiority of this algorithm in terms of the BER performance and computational complexity.**

***Index Terms*— NOMA, multi-user interference cancellation, satellite communications, factor graph, message passing.**

I. Introduction

SATELLITE communications can provide various services in global coverage with its wide coverage ability [1], which has become imperative to move forward for future integrated satellite-terrestrial networks [2], [3]. With the increasing communication demands of high speed terminals [4], next generation multiple access (NGMA) of terrestrial high-speed mobile terminals has drawn much attention from both the academia and industry [5], [6], such as high speed vehicles [7], high speed railways [8] and etc. As an indispensable part in next generation wireless networks, in addition to aforementioned terrestrial high-speed terminals, satellites can also provide various communication services for high-mobility terminals such as high speed flying objects [9], [10], aircrafts, low earth orbit (LEO) satellites, unmanned aerial vehicle (Uavs) [11] and etc. However, there are relatively few studies on NGMA schemes in satellite high-mobility communications.

Specially, non-orthogonal multiple access (NOMA) has been regarded as a key technique to provide massive connectivity by supporting multiple transmissions using the same wireless resources [12]–[14], which is a promising technology to help achieve targets of NGMA in satellite high-mobility communications [15]. However, traditional NOMA detection methods developed for terrestrial mobile communications and satellite-ground communications are not appropriate for satellite high-mobility communication systems [16], since there are still some important problems urgently to be solved.

For one thing, high mobility often means high speed, high acceleration and high jerk [9], [17], so high-mobility terminals usually cause high Doppler frequency offsets and high Doppler rates, which degrades the bit error ratio (BER) performance seriously. Moreover, due to the heterogeneity of different terminals in terms of the velocity and acceleration, time-varying Doppler shifts of each user are different, which further makes multi-user interference cancellation become more difficult. For another thing, because of the heterogeneity of different high-mobility terminals, locations of high-mobility terminals are usually geographically distributed widely and changed rapidly, which inevitably results in different arriving time offsets at the receiver [18]. Specially, inter-user interference caused by multi-user asynchronous simultaneous transmissions in high-mobility environments is difficult to eliminate with low computational complexity effectively. Therefore, it is significant to develop novel NOMA detection methods to cope with challenges of NGMA in satellite high-mobility communications.

In the literature, lots of NOMA detection methods have been introduced. Particularly, existing dominant NOMA schemes can be divided into two categories: code-domain (CD) NOMA

Manuscript received October 12, 2021; revised November 14, 2021; accepted December 17, 2021. Date of publication January 14, 2022; date of current version March 17, 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1804800 and in part by the National Natural Science Foundation of China under Grant 61922050. *(Corresponding author: Chunxiao Jiang.)*

Jiawei Wang is with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: wang-jw19@mails.tsinghua.edu.cn).

Chunxiao Jiang and Linling Kuang are with the Beijing National Research Center for Information Science and Technology, and the Tsinghua Space Center, Tsinghua University, Beijing 100084, China (e-mail: jchx@tsinghua.edu.cn; kll@tsinghua.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSAC.2022.3143254.

Digital Object Identifier 10.1109/JSAC.2022.3143254

and power-domain (PD) NOMA. For the former, CD-NOMA is capable of obtaining gain due to the use of spreading sequences of codewords [19]. In addition, the transmitted bandwidth can be spread by spreading sequences of codewords to improve the anti-interference ability of the transmitted signal. While for the later, it utilizes superposition coding at the transmitter and successive interference cancellation at the receiver to support multiple access [20]. However, most traditional NOMA detection algorithms cannot cope with NGMA challenges in satellite high-mobility communications. In [21] and [22], precoding and power allocation strategies were presented to improve PD-NOMA performance without considering mobility and asynchronous characteristics of multi-user signals. Then, the authors proposed user dissociation and re-association control methods to deal with mobility issues of NOMA [23]. Later, in [24], considering heterogenous mobility profiles of different users, the authors proposed the orthogonal time frequency space (OTFS) based PD-NOMA system, where receiving strategies were also designed to remove interference and harvest both multi-path and multi-user diversity. Whereas, this approach could not support asynchronous multiple access and the receiver in the satellite was difficult to obtain diversity in the light of sight (LoS) channel. In addition, most of these works assumed perfect synchronization at the receiver [25]–[27], i.e., the arriving time of each high-mobility terminals' signal was guaranteed to coincide. These methods reduced data transmission efficiency seriously and required accurate cooperation of each high-mobility terminal, which could not be achieved in satellite high-mobility communications. In the asynchronous case, the authors proposed a pre-coding and whitening-and-decomposing (WD) detection scheme [28]. However, this scheme assumed that the whole information of each terminal was known to the proposed precoding operator. Considering wide geographical distribution of high-mobility terminals, this assumption was impractical in high-mobility communication environments. Later, in [29], the authors proposed an asynchronous transmission system for multiple access channels, using the trellis-based detection methods to improve the performance. Nevertheless, the high Doppler frequency offset and the Doppler rate were not considered in this algorithm, therefore this scheme could not be applied in satellite high-mobility communication systems. Similarly, in [30], the authors proposed a novel triangular successive interference cancellation (T-SIC) that exploits a triangular pattern to perform the interference cancellation from all interfering users for the desired user without considering time-varying Doppler shifts in high-mobility environments. Different from aforementioned methods, in [31], jointly with asynchronous analog network coding (AANC), time offsets were added in the transmitter artificially to improve the performance of the cooperative communication system. Although these algorithms could exploit diversity gains to improve the system performance, high-mobility terminals would degrade diversity gains seriously, which inevitably leads to attenuation of the BER performance. Overall, these aforementioned multiple access approaches designed mainly for motionless or low speed ground terminals are not appropriate for NGMA in satellite high-mobility communications.

From above analysis, traditional multiple access detection methods have not taken both asynchronous arriving signals and time-varying Doppler shifts into consideration, which makes these methods difficult to be applied in satellite high-mobility communication systems for NGMA. Therefore, to cope with NGMA challenges mentioned above in satellite high-mobility communications, it is rather significant to develop the new asynchronous NOMA interference cancellation algorithm designed especially for satellite high-mobility communication systems, which is capable of achieving the better BER performance in high-mobility environments. Motivated by this, in this paper, we design the new factor graph considering the influence of high-mobility nature to cope with challenges of multiple access in high-mobility environments. Then a novel turbo iterative asynchronous NOMA detection algorithm is proposed to improve the interference cancellation performance for NGMA in satellite high-mobility communications. Concretely, the contributions of this paper are summarized as follows:

- We design a three-dimensional (3D) factor graph framework for asynchronous NOMA detection in satellite high-mobility communications, where different high Doppler frequency offsets and Doppler rates of heterogeneous high-mobility terminals are considered in the received signal. Then, based on the proposed factor graph model, we formulate passing messages of function nodes and variable nodes along edges in the proposed factor graph according to the sum-product rule.
- We propose the novel turbo iterative NOMA detection algorithm that consists of the Doppler elimination loop, the interference cancellation loop and the decoding loop. With the aid of the information interaction of these designed iterative loops in the proposed factor graph, the negative influence of time-varying Doppler shifts and multi-user interference can be eliminated effectively. In addition, the corresponding signal to interference plus noise ratio (SINR) analysis in high-mobility environments is also given.
- We evaluate the performance of the proposed algorithm by numerical simulations, which demonstrates that the turbo iterative NOMA detection algorithm can achieve the better BER performance with less computational complexity compared with traditional detection algorithms.

The rest of the paper is organized as follows. We first introduce the system model in Section II. Then, in Section III, we present the novel 3D factor graph in high-mobility environments. Later, in Section IV, passing messages of function nodes and variable nodes along edges in the proposed 3D factor graph are formulated. In Section V, based on the presented 3D factor graph, we propose the novel turbo iterative detection method of asynchronous NOMA for NGMA and give the SINR analysis of the proposed algorithm in high-mobility environments. Simulation results are then shown in Section VI. Finally, Section VII gives the conclusion.

II. System Model

Fig. 1 presents the system model of asynchronous NOMA in satellite high-mobility communications. In the transmitter,

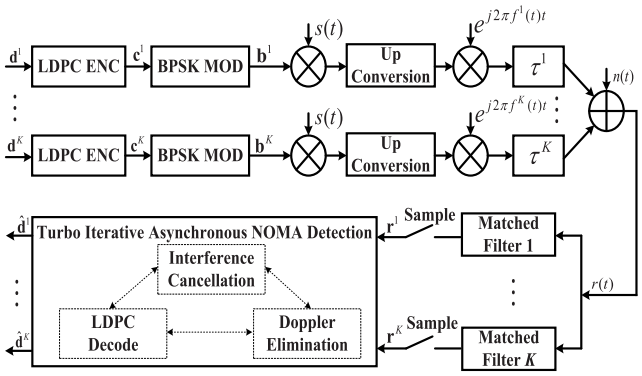

Fig. 1. System model of asynchronous NOMA in satellite high-mobility communications (ENC: Encoder, MOD: Modulation).

the transmitted bit sequence of the user-k is denoted as $\mathbf{d}^k = (d_0^k, d_1^k, \ldots, d_{L-1}^k)$ with the bit rate R_b, where $d_l^k \in \{0,1\}$, $k = 1, 2, \ldots, K$. In order to improve the reliability of data transmission, the low density parity check (LDPC) code [32] is widely used in satellite communications. The data sequence of user-k $\mathbf{d}^k$ is encoded by the LDPC encoder-k with the code rate $R = \frac{L}{N}$ $(L < N)$, to generate the coding sequence $\mathbf{c}^k = (c_0^k, c_1^k, \ldots, c_{N-1}^k)$, where $c_n^k \in \{0,1\}$. Then, the obtained coding sequence of user-k $\mathbf{c}^k$ is modulated to $\mathbf{b}^k = (b_0^k, b_1^k, \ldots, b_{N-1}^k)$ through a binary phase shift keying (BPSK) modulator, where $b_n^k \in \{-1, 1\}$. Then the user-k transmits the modulated symbol by sending $b_n^k s(t - nT)$, where $s(t)$ represents the transmitted signal waveform with $\int_0^T |s(t)|^2 dt = P$. Specially, the signal waveform vanishes outside interval $[0, T]$, where T represents the symbol period and satisfies $T = \frac{R}{R_b}$.

In the asynchronous case, as shown in Fig. 2, the arriving time of each user is different at the receiver. So let $\delta^k \in [0, T)$ denote the arriving time offset of user-k. In the practical satellite communication systems, due to the small time scale, achieving symbol synchronization is more difficult than achieving frame synchronization. Therefore, we assume that the acquisition of multi-user signals has been achieved in the receiver, i.e., time offsets $\forall k : \delta_k$ are known to the receiver. Without loss of generality, we assume $0 = \delta^1 \leq \cdots \leq \delta^K < T$. When $\delta_K \geq T$, it becomes a frame synchronization problem that is not considered in this paper, since frame synchronization is much easier to be achieved compared with symbol synchronization.

Generally, signal propagation between the UAV and the satellite is unshaded due to high operation altitude of UAVs and satellites. In this case, considering that K high-mobility terminals are in the LoS of the satellite, the received signal can be expressed as

$$r(t) = \sum_{k=1}^{K} \sum_{n=0}^{N-1} b_n^k s\left(t - nT - \delta^k\right) e^{j2\pi\left[f_c + f^k(t)\right]t} + w(t), \tag{1}$$

where

$$f^k(t) = \frac{v_k(t) f_c}{c} \cos\theta_k, \tag{2}$$

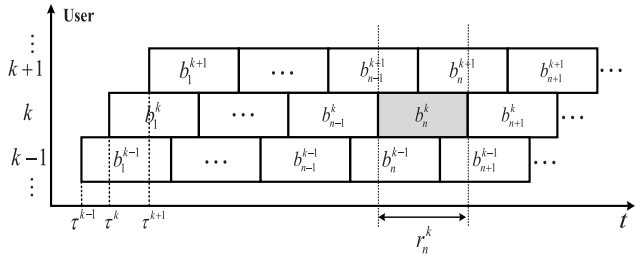

Fig. 2. Different arriving times of users' signals in asynchronous NOMA.

where $f_d^k(t)$ represents the Doppler shift of the high-mobility terminal-k, $v_k(t)$ represents the relative velocity between the high-mobility terminal-k and the satellite, θ_k denotes the angle of moving direction with respect to signal propagation direction of the high-mobility terminal-k, $w(t)$ represents the complex additive white Gaussian noise (AWGN) with zero mean and variance σ^2, f_c represents the carrier frequency and c is the speed of light. Specially, the jerk of high-mobility terminal can be equivalent to an additional acceleration by linearisation. Thus, the Doppler shifts of user-k can be further written as

$$f^k(t) = f_k + \frac{a_k t}{2}, \tag{3}$$

where f_k refers to the Doppler frequency offset caused by the high-mobility terminal-k in the k-signal and a_k refers to the Doppler rate caused by the high-mobility terminal-k in the k-signal.

Assuming that perfect down-conversion has been achieved in the receiver, the continuous-time signal $r(t)$ is passed through K matched filters to obtain M discrete-time outputs. Then samples at the output of the received filter can be written as

$$\begin{aligned} r_n^k &= \int_{nT+\delta^k}^{(n+1)T+\delta^k} r(t) s^*\left(t - nT - \delta^k\right) dt \\ &= b_n^k \cdot \eta_n^k + \sum_{i<k} \left[\rho_n^i(\delta_{k,i}) b_n^i + \rho_{n+1}^i(\delta_{k,i} - T) b_{n+1}^i\right] \\ &\quad + \sum_{i>k} \left[\rho_{n-1}^i(T + \delta_{k,i}) b_{n-1}^i + \rho_n^i(\delta_{k,i}) b_n^i\right] + w_n^k, \end{aligned} \tag{4}$$

where

$$\eta_n^k = \int_{nT}^{(n+1)T} s(t) \cdot s^*(t)\, e^{j2\pi(f_k + \frac{a_k t}{2})t} dt, \tag{5}$$

$$\rho_n^i(\delta) = \int_{nT}^{(n+1)T} s(t)\, s^*(t - \delta)\, e^{j2\pi(f_i + \frac{a_i t}{2})t} dt, \tag{6}$$

$$\delta_{k,i} = \tau^k - \delta^i, \tag{7}$$

and $s^*(t)$ represents the conjugate of $s(t)$. Basically, $s(t)$ refers to signal waveform, which be expressed as

$$s(t) = \sum_{m=0}^{M-1} z_m g_{T_s}(t - mT_s), \tag{8}$$

where

$$g_{T_s}(t) = \begin{cases} 1 & 0 \leq t < T_s \\ 0 & others, \end{cases} \tag{9}$$

and $T_s = \frac{T}{M}$ is the chip period, z_m represents the mth chip, $\forall m : |z_m| = \sqrt{\frac{P}{M}}$.

Then, sample sequences $\forall k, n : \ r_n^k$ are used to perform the turbo iterative detection algorithm to extract the data sequence $\forall k : \hat{\mathbf{d}}^k$. By means of message passing among the Doppler elimination loop, the interference cancellation and the decoding loop, time-varying Doppler shifts and multi-user interference can be eliminated effectively, which can improve the BER performance significantly.

III. Factor Graph Representation

In this section, based on the factor graph theory [33], the novel factor graph representation framework and the computational tool for asynchronous NOMA in satellite high-mobility communications are proposed.

Basically, the goal of asynchronous NOMA detection is to recover the transmitted data in the superimposed multi-user signal. Therefore, the maximum a posteriori (MAP) criterion can be used to minimize the BER of each user, i.e.,

$$d_n^k = \arg\max_{d_n^k} \ p\left(d_n^k \mid \mathbf{r}^k\right). \tag{10}$$

Then, a posteriori probability $p\left(d_n^k \mid \mathbf{r}^k\right)$ can be obtained by solving the marginal probability of the joint posterior probability density function $p\left(\mathbf{d}^k, f^k, a^k \mid \mathbf{r}^k\right)$, i.e.,

$$p\left(d_n^k \mid \mathbf{r}^k\right) = \sum_{\mathbf{d}^k/d_n^k} \int_{f_k, a_k} p\left(\mathbf{d}^k, f_k, a_k \mid \mathbf{r}^k\right). \tag{11}$$

Due to the high computational complexity of integrals of continuous variables, Doppler frequency offsets f_k and Doppler rates a_k are both discretized, i.e.,

$$f^k \in S_f = \left\{ f^{(q)} = -f_{\max} + \frac{2(q-1)\cdot f_{\max}}{D_f - 1} \mid q = 1, 2, \ldots, D_f \right\}, \tag{12}$$

$$a^k \in S_a = \left\{ a^{(q)} = -a_{\max} + \frac{2(q-1)\cdot a_{\max}}{D_a - 1} \mid q = 1, 2, \ldots, D_a \right\}, \tag{13}$$

where $f_{\max}$ is the maximum of Doppler frequency offsets and $a_{\max}$ is the maximum of Doppler rates. Then, a posteriori probability $p\left(d_n^k \mid \mathbf{r}^k\right)$ can be rewritten as

$$p\left(d_n^k \mid \mathbf{r}^k\right) = \sum_{f^k, a^k, \mathbf{d}^k / d_n^k} p(\mathbf{d}^k, f^k, a^k \mid \mathbf{r}^k). \tag{14}$$

By applying Bayes theorem, $p(\mathbf{d}^k, f^k, a^k \mid \mathbf{r}^k)$ can be further expressed as

$$\begin{aligned} p(\mathbf{d}^k, f^k, a^k \mid \mathbf{r}^k) &\propto p(\mathbf{r}^k \mid \mathbf{d}^k, f^k, a^k) \cdot p(\mathbf{d}^k, f^k, a^k) \\ &= \Upsilon[\mathbf{b}^k = \kappa(\mathbf{d}^k)] p(\mathbf{r}^k \mid \mathbf{b}^k, f^k, a^k) \\ &\quad \cdot p(\mathbf{b}^k) p(f^k) p(a^k), \end{aligned} \tag{15}$$

where $\mathbf{b}^k = \kappa(\mathbf{d}^k)$ refers to that $\mathbf{b}^k$ obeys the encoding constraints of $\mathbf{d}^k$, $p(\mathbf{b}^k)$ is the probability mass function (PMF) of $\mathbf{b}^k$, $p(f^k)$ represents the probability density function (PDF) of f^k, $p(a^k)$ is the PDF of a^k, and $\Upsilon[\cdot]$ denotes the indictor function, i.e.,

$$\Upsilon[p] \triangleq \begin{cases} 1 & p \text{ is true} \\ 0 & p \text{ is flase.} \end{cases} \tag{16}$$

Since $\mathbf{r}^k$ is a sequence of independent random variables with interference cancellation, the joint PDF of $p(\mathbf{r}^k \mid \mathbf{b}^k, f^k, a^k)$ can be expressed as

$$p(\mathbf{r}^k \mid \mathbf{b}^k, f^k, a^k) = \prod_{n=0}^{N-1} p\left(r_n^k \mid b_n^k, a^k, f^k\right), \tag{17}$$

where

$$\begin{aligned} &p\left(r_n^k \mid b_n^k, a^k, f^k\right) \\ &\quad = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-\left|r_n^k - I_n^k - b_n^k \cdot \eta_n^k\right|^2}{2\sigma^2} \right) \end{aligned} \tag{18}$$

where I_n^k represents the interference caused by other users at r_n^k.

As a matter of the fact, Doppler frequency offsets and Doppler rates usually obey the uniform distribution, respectively. Therefore, the PDF of f^k can be expressed as

$$p(f^k) = \frac{1}{D_f}, \tag{19}$$

and the PDF of a^k can be expressed as

$$p(a^k) = \frac{1}{D_a}. \tag{20}$$

From above analysis, the probabilistic structure defined by the factorization (15), (17), (19), (20) can be represented by the factor graph depicted in Fig. 3. Specially, the factor graph can deal with complicated global functions of many variables in the manner of factorization [33]. Based on the presented structure of the factor graph, various marginal functions derived from the global function can be easily computed by means of message passing.

Different from the exiting graph representation, both high-mobility characteristics of received signals and multi-user interference are considered in the proposed factor graph. On the one hand, Doppler nodes are established in the proposed factor graph to analyze Doppler frequency offsets and Doppler rates caused by high-mobility terminals. On the other hand, the proposed factor graph is built into a three-dimensional structure. Based on the proposed 3D factor graph, interference caused by signal overlap among multiple users can be estimated accurately and cancelled effectively.

IV. Formulation of Massage Passing

In this section, based on the proposed structure of the 3D factor graph, passing messages along each edge are formulated in detail. By means of the sum-product rule, variable nodes and function nodes can exchange messages along edges to achieve interference cancellation and eliminate time-varying Doppler shifts effectively. Specially, the sum-product rule is that the message passed from a node o on an edge e is the product of all messages sent from neighbouring edges

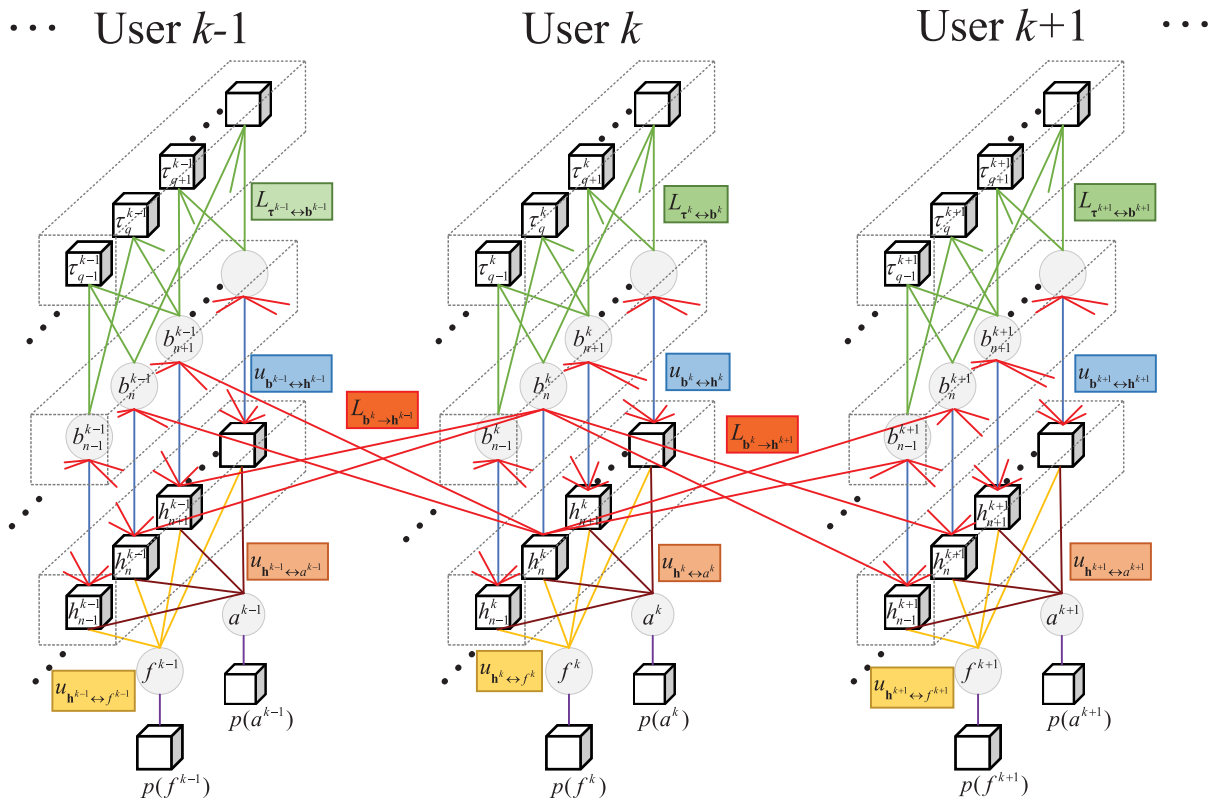

Fig. 3. The 3D factor graph model of asynchronous NOMA in satellite high-mobility communications. (1. Note that red lines indicate unidirectional message transmission and lines in other colors indicate dual-direction message transmission. 2. Note that b_n^k also has edges to channel function nodes h of users who are not adjacent to user k. Specially, b_n^k have edges to h_n^i and h_{n+1}^i for $i = 1, 2, \ldots, k-1$, and have edges to h_n^i and h_{n-1}^i for $i = k+1, \ldots, K$. This figure only draws part of these edges to keep neat.)

(except for the edge e) to the node o, summarized for the variable associated with e [33].

For the message passing schedule, passing messages are formulated from top to bottom of the factor graph. Let $L^l_{\tau_q^k \to b_n^k}$ denote the log-likelihood ratio (LLR) message passed from the LDPC check node (function node) of user-k τ_q^k to the chip node (variable node) of user-k b_n^k in the l iteration. According to the LDPC theory [34], $L^l_{\tau_q^k \to b_n^k}$ can be expressed as

$$L^l_{\tau_q^k \to b_n^k} = 2\tanh^{-1}\left[\prod_{b_v^k \in N(\tau_q^k),\, v \neq n} L^l_{b_v^k \to \tau_q^k}\right], \tag{21}$$

where $N(\cdot)$ is the set of neighbors of a given node in the factor graph, and $b_v^k \in N(\tau_q^k)$, $v \neq n$ represents the neighbors of the LDPC check node τ_q^k except for the chip node b_n^k. Specially, $L^l_{b_n^k \to \tau_q^k}$ represents the LLR message sent from the chip node b_n^k to the LDPC check node τ_q^k in the iteration l, which can be expressed as

$$L^l_{b_n^k \to \tau_q^k} = L_{h_n^k \to b_n^k} + \sum_{\tau_v^k \in N(b_n^k),\, v \neq n} L^l_{\tau_v^k \to b_n^k}, \tag{22}$$

where $\tau_v^k \in N(b_n^k)$, $v \neq q$ represents the neighbors of the chip node b_n^k except for the LDPC check node τ_q^k, and $L_{h_n^k \to b_n^k}$ refers to the LLR message passed from the channel function node h_n^k to the chip node b_n^k in the iteration l, which is calculated in Eq. (41).

Based on the sum-product update rule, the LLR message sent from the chip node b_n^k to the channel function node h_n^k in the l iteration is the sum of LLR messages passed from the channel function node h_n^k and LLR messages passed from the LDPC check nodes τ_q^k that are neighbour to the chip node b_n^k, say,

$$L^l_{b_n^k \to h_n^k} = L_{h_n^k \to b_n^k} + \sum_{\tau_v^k \in N(b_n^k)} L^l_{\tau_v^k \to b_n^k}. \tag{23}$$

Since $L^l_{b_n^k \to h_n^k}$ satisfies

$$L^l_{b_n^k \to h_n^k} \triangleq \ln\left[\frac{u^l_{b_n^k \to h_n^k}\left(b_n^k = 1\right)}{u^l_{b_n^k \to h_n^k}\left(b_n^k = -1\right)}\right], \tag{24}$$

where $u^l_{b_n^k \to h_n^k}\left(b_n^k = 1\right)$ represents the reliability message of $b_n^k = 1$ passed form the chip node b_n^k to the channel function node h_n^k, and $u^l_{b_n^k \to h_n^k}\left(b_n^k = -1\right)$ represents the reliability message of $b_n^k = -1$ passed form the chip node b_n^k to the channel function node h_n^k in the iteration l. Thus, $u^l_{b_n^k \to h_n^k}\left(b_n^k = 1\right)$ and $u^l_{b_n^k \to h_n^k}\left(b_n^k = -1\right)$ can be further written as

$$u^l_{b_n^k \to h_n^k}\left(b_n^k = 1\right) = \frac{e^{L^l_{b_n^k \to h_n^k}}}{1 + e^{L^l_{b_n^k \to h_n^k}}}, \tag{25}$$

$$u^l_{b_n^k \to h_n^k}\left(b_n^k = -1\right) = \frac{1}{1 + e^{L^l_{b_n^k \to h_n^k}}}. \tag{26}$$

Then, let $u^l_{h^k_n \to f^k}$ represent the message sent from the channel function node h^k_n to the Doppler frequency offset node (variable node) f^k in the iteration l. According to the sum-product rule, $u^l_{h^k_n \to f^k}$ involves nontrivial function multiplications, followed by an application of the summary operator, which can be written as

$$u^l_{h^k_n \to f^k} = \sum_{b^k_n, a^k} h\left(b^k_n, a^k, f^k\right) \cdot u^l_{b^k_n \to h^k_n} \cdot u^l_{a^k_n \to h^k_n}. \tag{27}$$

Similarly, messages passed from the channel function node h^k_n to the Doppler rate node (variable node) a^k can be expressed as follows

$$u^l_{h^k_n \to a^k} = \sum_{b^k_n, f^k} h\left(b^k_n, a^k, f^k\right) \cdot u^l_{b^k_n \to h^k_n} \cdot u^l_{f^k \to h^k_n}. \tag{28}$$

where $h\left(b^k_n, a^k, f^k\right) = p\left(r^k_n \mid b^k_n, a^k, f^k\right)$.

Considering the multi-user interference, I^k_n in Eq. (18) can be written as

$$\begin{aligned} I^k_n = &\sum_{i<k} \left[\rho^i_n(\delta_{k,i}) b^i_n + \rho^i_{n+1}(\delta_{k,i} - T) b^i_{n+1}\right] \\ &+ \sum_{i>k} \left[\rho^i_{n-1}(T + \delta_{k,i}) b^i_{n-1} + \rho^i_n(\delta_{k,i}) b^i_n\right]. \end{aligned} \tag{29}$$

Then, the mean $E\left(I^k_n\right)$ and variance $V\left(I^k_n\right)$ of I^k_n can be expressed as

$$\begin{aligned} &E\left(I^k_n\right) \\ &= \sum_{i<k} \left[\rho^i_n(\delta_{k,i}) E\left(b^i_n\right) + \rho^i_{n+1}(\delta_{k,i} - T) E\left(b^i_{n+1}\right)\right] \\ &\quad + \sum_{i>k} \left[\rho^i_{n-1}(T + \delta_{k,i}) E\left(b^i_{n-1}\right) + \rho^i_n(\delta_{k,i}) E\left(b^i_n\right)\right], \end{aligned} \tag{30}$$

$$\begin{aligned} &V\left(I^k_n\right) \\ &= \sum_{i<k} \left[\left|\rho^i_n(\delta_{k,i})\right|^2 V\left(b^i_n\right) + \left|\rho^i_{n+1}(\delta_{k,i} - T)\right|^2 V\left(b^i_{n+1}\right)\right] \\ &\quad + \sum_{i>k} \left[\left|\rho^i_{n-1}(T + \delta_{k,i})\right|^2 V\left(b^i_{n-1}\right) + \left|\rho^i_n(\delta_{k,i})\right|^2 V\left(b^i_n\right)\right], \end{aligned} \tag{31}$$

where $E\left(b^i_n\right)$ is the mean of b^i_n and $V\left(b^i_n\right)$ is the variance of b^i_n. Specially, according to the relation between the variable's LLR and its mean, $E\left(b^i_n\right)$ and $V\left(b^i_n\right)$ can be calculated by the incoming messages from other users' chip nodes b^i_k $(i \neq k)$, i.e.,

$$E\left(b^i_n\right) = \tanh\left[\frac{L_{b^i_n \to h^k_n}}{2}\right], \tag{32}$$

$$V\left(b^i_n\right) = 1 - E^2\left(b^i_n\right), \tag{33}$$

where

$$L^l_{b^i_n \to h^k_n} = L^l_{h^i_n \to b^i_n} + \sum_{\tau^i_v \in N(b^i_n)} L^l_{\tau^i_v \to b^i_n}. \tag{34}$$

Specially, messages passed from other users' chip nodes $b^i_v \in N(h^k_n), i \neq k$ to the channel function node h^k_n is unidirectionally. In addition, considering that the actual Doppler frequency offset and the actual Doppler rate of user-i are unknown in Eq. (6), i.e., f^k_i and a^k_i are both unknown in the correlation coefficient $\rho^i_n(\delta)$. Therefore, the Doppler frequency offset $\tilde{f}^k_i$ and the Doppler rate $\tilde{a}^k_i$ with the largest confidence in the passing message are used to replace f^k_i and a^k_i to calculate $\rho^i_n(\delta)$. Then, $h\left(b^k_n, a^k, f^k\right)$ can be further expressed as

$$\begin{aligned} h\left(b^k_n, a^k, f^k\right) = &\frac{1}{\sqrt{2\pi(\sigma^2 + V^k_n)}} \exp \\ &\times \left(\frac{-\left|r^k_n - E\left(I^k_n\right) - b^k_n \cdot \eta^k_n\right|^2}{2(\sigma^2 + V\left(I^k_n\right))}\right). \end{aligned} \tag{35}$$

Using Eq. (27), the local belief of the discrete Doppler frequency offset is computed as the product of all messages directed toward the user-k's Doppler frequency offset node f^k, say,

$$\Gamma^l\left(f^k\right) = \prod_n u^l_{h^k_n \to f^k} \cdot p\left(f^k\right). \tag{36}$$

Then, let $u^l_{f^k \to h^k_n}$ denote the message passed from the Doppler frequency offset node (variable node) f^k to the channel function node h^k_n. Specially, $u^l_{f^k \to h^k_n}$ is equal to the product of these messages except for $u^l_{h^k_n \to f^k}$, i.e.,

$$u^l_{f^k \to h^k_n} = \frac{\Gamma^l\left(f^k\right)}{u^l_{h^k_n \to f^k}}. \tag{37}$$

Basically, $u^l_{f^k \to h^k_n}$ refers to soft-decision information (metrics) on all valid local configurations of Doppler frequency offset, which is sent from the Doppler frequency offset node to the channel function node. The value of $u^l_{f^k \to h^k_n}$ represents the soft-decision information of discrete Doppler frequency offset. The larger value of $u^l_{f^k \to h^k_n}$, the higher reliability of configurations of Doppler frequency offset f^k.

Similarly, by Eq. (28), the local belief of the discrete Doppler rate is computed as the product of all messages directed toward the user-k' Doppler frequency offset node a^k, i.e.,

$$\Gamma^l\left(a^k\right) = \prod_n u^l_{h^k_n \to a^k} \cdot p\left(a^k\right). \tag{38}$$

Then, the message passed from the Doppler rate node (variable node) a^k to the channel function node h^k_n can be computed as the product of all but one of these messages, say,

$$u^l_{a^k \to h^k_n} = \frac{\Gamma^l\left(a^k\right)}{u^l_{h^k_n \to a^k}}. \tag{39}$$

After finishing the message passing about time-varying Doppler shifts, according to messages of $u^l_{f^k \to h^k_n}$ and $u^l_{a^k \to h^k_n}$, messages sent from the channel function node h^k_n to the chip node b^k_n can be expressed as

$$u^l_{h^k_n \to b^k_n} = \sum_{a^k, f^k} h\left(b^k_n, a^k, f^k\right) \cdot u^l_{f^k \to h^k_n} \cdot u^l_{a^k \to h^k_n}. \tag{40}$$

Similar to Eq. (24), the LLR message of $u^l_{h^k_n \to b^k_n}$ in the iteration l can be expressed as

$$L^l_{h^k_n \to b^k_n} \triangleq \ln\left[\frac{u^l_{h^k_n \to b^k_n}\left(b^k_n = 1\right)}{u^l_{h^k_n \to b^k_n}\left(b^k_n = -1\right)}\right]. \tag{41}$$

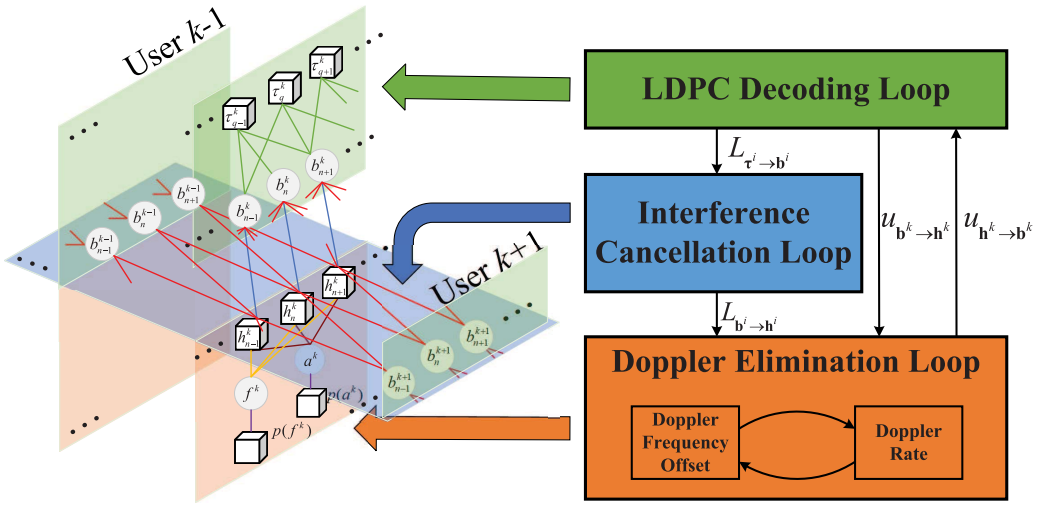

Fig. 4. The framework of turbo iterative asynchronous NOMA detection algorithm. (Note that b_n^k also has edges to channel function nodes h of users who are not adjacent to user k. Specially, b_n^k have edges to h_n^i and h_{n+1}^i for $i = 1, 2, \ldots, k-1$, and have edges to h_n^i and h_{n-1}^i for $i = k+1, \ldots, K$.)

From above analysis, passing messages along each edge in the proposed 3D factor graph have been formulated. In the next section, based on passing messages and the factor graph, the turbo iterative detection algorithm for asynchronous NOMA is proposed to improve the BER performance of NGMA in satellite high-mobility communications. In addition, the SINR analysis of the proposed algorithm is also given in the next section.

V. Novel Iterative NOMA Detection

To improve the BER performance of asynchronous NOMA detection, the turbo iterative detection algorithm based on the proposed factor graph framework and message passing formulation is proposed. Specially, the proposed algorithm is comprised of three cycles, i.e., the Doppler elimination loop, the interference cancellation loop and the LDPC decoding loop. By message interaction among these loops, time-varying Doppler shifts and multi-user interference can be eliminated effectively. In this section, the turbo iterative detection algorithm and the SINR analysis are introduced in detail.

A. Turbo Iterative Detection Algorithm

The framework of the turbo iterative detection algorithm for asynchronous NOMA in satellite high-mobility communications is depicted in Fig. 4. In the proposed algorithm, the process of message passing is divided into three cycles, the Doppler elimination loop, the interference cancellation loop and the LDPC decoding loop. For the Doppler elimination loop, based on the passing messages $L_{\mathbf{b}^i \to \mathbf{h}^k}$ $(i \neq k)$, $u_{\mathbf{b}^k \to \mathbf{h}^k}$ sent by the interference cancellation loop and the LDPC decoding loop, it is used to estimate the influence of time-varying Doppler shifts caused by high-mobility terminals by means of message interactions among Doppler frequency offset nodes $\forall k : f^k$, Doppler rate nodes $\forall k : a^k$ and channel function nodes $\forall n, k : h_n^k$. When it comes to the interference cancellation loop, it is mainly used to eliminate the multi-user interference with aid of message interactions among chip nodes $\forall n, i : b_n^i$ $(i \neq k)$ and the channel function nodes $\forall n, k : h_n^k$. Specially, this loop can utilize messages of other users' chip nodes to cancel interference at the user-k, which helps improve the reliability of the user sequence $\mathbf{b}^k$. Then, more reliable decision results of $\mathbf{b}^k$ can help other users cancel interference in turn. While for the LDPC decoding loop, its main function is to calculate LLR messages of chip nodes $\forall n, k : b_n^k$ according to messages $u_{\mathbf{h}^k \to \mathbf{b}^k}$ sent from the Doppler elimination loop. In addition, it also can generate transmitted data sequences $\hat{\mathbf{d}}^k$ of each user after finishing total iterations.

Similar to the turbo principle [35], information interactions can improve the system performance effectively. By means of soft information sent by the LDPC decoding loop, the interference cancellation loop is capable of canceling multi-user interference effectively, which helps other loops generate more reliable soft information. In addition, the Doppler elimination loop can provide the reliability messages of time-varying Doppler shifts for the interference cancellation loop and the LDPC decoding loop, which makes other iterative loops generate more reliable soft information. Meanwhile, more reliable soft information can help the LDPC decoding loop generate more reliable soft information, which helps the interference cancellation loop and the Doppler elimination loop in turn. With the aid of information interaction among three loops, passing messages (soft information) become more reliable, which consequently results in the better BER performance for NGMA in satellite high-mobility communications.

Specially, it is worth noting that the proposed algorithm can be also viewed as an approximation to the maximum likelihood (ML) detection, much in the same way that a turbo decoder is an approximation of the ML decoder for a concatenated code. Therefore, the proposed algorithm inevitably suffers the performance degradation compared with the ML detection. However, considering complex compositions

and constraints in the received signal, it is impossible to perform the exhaustive search to eliminate time-varying Doppler shifts and multi-user interference, because the complexity of ML detection is so high that ML detection cannot be applied in satellite high-mobility communication systems.

In the proposed algorithm, iterations of the Doppler elimination loop and the LDPC decoding loop are L_d and L_c, respectively. In addition, let $L_{\max}$ represent total iterations of the proposed algorithm. Specially, the proposed algorithm starts with the interference cancellation loop, then the Doppler elimination loop, and finally the LDPC decoding loop.

At last, the proposed algorithm is summarized as follows. Specially, rows 6-9 represent the interference cancellation loop, rows 10-15 represent the Doppler elimination loop and rows 20-23 represent the LDPC decoding loop in the **Algorithm 1**.

Algorithm 1 Turbo Iterative Asynchronous NOMA Detection

Input:

$\mathbf{r}$, N, K, σ^2, L_{max}, L_d, L_c, S_f, S_a, $s(t)$.

Initialize:

$\forall n,k : L^0_{b^k_n \to h^k_n} = 0$, $\forall n,k : u^0_{a^k_n \to h^k_n} = 1$, $\forall k : \tilde{f}^k = 0$, $\forall k : \tilde{a}^k = 0$.

1: **for** l to L_{max} **do**
2: **if** $l == 2$
3: Compensate Doppler shifts in $\forall n,k : r^k_n$ with $\forall k : \tilde{a}^k, \tilde{f}^k$;
4: **end if**
5: Update $\forall n,k : u^l_{b^k_n \to h^k_n}$ by Eq. (25) and Eq. (26);
6: Update $\forall n,k : L^l_{b^i_v \to h^k_n}(b^i_v \in N(h^k_n), i \neq k)$ by Eq. (34);
7: Calculate $\forall n,k : E\left(I^k_n\right)$ by Eq. (30);
8: Calculate $\forall n,k : V\left(I^k_n\right)$ by Eq. (31);
9: Update $\forall n,k : h\left(b^k_n, a^k, f^k\right)$ by Eq. (35);
10: **for** l_d to L_d **do**
11: Calculate $\forall n,k : u^{(l-1)L_d+l_d}_{h^k_n \to f^k}$ by Eq. (27);
12: Calculate $\forall n,k : u^{(l-1)\cdot L_d+l_d}_{f^k \to h^k_n}$ by Eq. (36) and Eq. (37);
13: Calculate $\forall n,k : u^{(l-1)\cdot L_d+l_d}_{h^k_n \to a^k_n}$ by Eq. (28);
14: Calculate $\forall n,k : u^{(l-1)\cdot L_d+l_d}_{a^k \to h^k_n}$ by Eq. (38) and Eq. (39);
15: **end for**
16: Update $\forall k : \tilde{a}^k = \arg\max_{a^k}\left(u_{a^k \to h^k_n}\right)$;
17: Update $\forall k : \tilde{f}^k = \arg\max_{f^k}\left(u_{f^k \to h^k_n}\right)$;
18: Update $u^l_{h^k_n \to b^k_n}$ by Eq. (40);
19: Update $\forall n,k : L^l_{h^k_n \to b^k_n}$ by Eq. (41);
20: **for** l_c to L_c **do**
21: Calculate $\forall q,k : L^{(l-1)L_c+l_c}_{b^k_n \to \tau^k_q}$ by Eq. (22);
22: Calculate $\forall n,k : L^{(l-1)I_c+i_c}_{\tau^k_q \to b^k_n}$ by Eq. (21);
23: **end for**
24: Update $\forall n,k : L^l_{b^k_n \to h^k_n}$ by Eq. (23);
25: **end for**
26: $\forall k : \hat{\mathbf{d}}^k \leftarrow \forall n,k : L^l_{b^k_n \to h^k_n}\ (LDPC\ \ Decode)$
27: **Output:** $\forall k : \hat{\mathbf{d}}^k$

B. SINR Analysis

According to Eq. (35), the belief of b^k_n after interference cancellation can be expressed as

$$\begin{aligned}&\Gamma\left(b^k_n\right)\\&= \frac{2}{V\left(I^k_n\right)+\sigma^2} R\left[\left(r^k_n - E\left(I^k_n\right)\right)\eta^{*k}_n\right]\\&= \frac{2}{V\left(I^k_n\right)+\sigma^2} R\left[b^k_n \cdot \left|\eta^k_n\right|^2 + \left(I^k_n - E\left(I^k_n\right) + w^k_n\right)\eta^{*k}_n\right],\end{aligned} \tag{42}$$

where $R[\cdot]$ represents the real part of a given value. It is important to note that b^k_n is the useful signal, $\left(I^k_n - E\left(I^k_n\right) + w^k_n\right)$ contains residual interference from the other users and noise, and $\left|\eta^k_n\right|^2$ represents the signal attenuation caused by time-varying Doppler shifts. Then, the average SINR of the channel function node is calculated as

$$\begin{aligned}SINR^k_n &= \frac{E\left[\left|b^k_n\right|^2\left|\eta^k_n\right|^4\right]}{E\left[\left|R\left[I^k_n\eta^{*k}_n\right] - R\left[E\left(I^k_n\right)\eta^{*k}_n\right] + R\left[w^k_n\eta^{*k}_n\right]\right|^2\right]}\\&= \frac{P\cdot\left|\eta^k_n\right|^2}{V_R\left(I^k_n\eta^{*k}_n\right) + \frac{N_0}{2}},\end{aligned} \tag{43}$$

where $\left|b^k_n\right|^2 = 1$, N_0 represents power spectral density of AWGN, and $V_R\left(I^k_n\eta^{*k}_n\right)$ represents the variance of the real part of $I^k_n\eta^{*k}_n$. As shown at the bottom of the page, the calculation of $V_R\left(I^k_n\eta^{*k}_n\right)$ is in Eq. (44), where $I[\cdot]$ represents the imaginary part of a given value.

Remark 1: The derived results in Eq. (44) also demonstrate that the SINR usually decreases with the increase of the number of users. This is due to more multi-user interference brought by more users, i.e., $V_R\left(I^k_n\eta^{*k}_n\right)$ becomes larger with the increase of the number of users, which degrades the BER performance indirectly.

In order to analysis the influence of Doppler shifts for simplicity, we assume the Doppler shift is a constant,

$$\begin{aligned}V_R\left(I^k_n\eta^{*k}_n\right) = &\sum_{i<k}\Big[\left(R\left[\rho^i_n(\delta_{k,i})\right]R\left[\eta^{*k}_n\right] - I\left[\rho^i_n(\delta_{k,i})\right]I\left[\eta^{*k}_n\right]\right)^2 V\left(b^i_n\right)\\&+\left(R\left[\rho^i_{n+1}(\delta_{k,i}-T)\right]R\left[\eta^{*k}_n\right] - I\left[\rho^i_{n+1}(\delta_{k,i}-T)\right]I\left[\eta^{*k}_n\right]\right)^2 V\left(b^i_{n+1}\right)\Big]\\&+\sum_{i>k}\Big[\left(R\left[\rho^i_{n-1}(T+\delta_{k,i})\right]R\left[\eta^{*k}_n\right] - I\left[\rho^i_{n-1}(T+\delta_{k,i})\right]I\left[\eta^{*k}_n\right]\right)^2 V\left(b^i_n\right)\\&+\left(R\left[\rho^i_n(\delta_{k,i})\right]R\left[\eta^{*k}_n\right] - I\left[\rho^i_n(\delta_{k,i})\right]I\left[\eta^{*k}_n\right]\right)^2 V\left(b^i_{n+1}\right)\Big]\end{aligned} \tag{44}$$

TABLE I
SYSTEM PARAMETERS

Variable	Parameters
Channel model	AWGN
Modulation	BPSK
Encode	LDPC (N,3,6)
Code rate	$R = \frac{1}{2}$
Data length	$L = 50$
Chip sequence length	$M = 100$
Doppler frequency offset	$f_{\max} = 50kHz$
Doppler rate	$a_{\max} = 4kHz/s$
Bit rate	$R_b = 100$kbps
Discrete samples of f	$D_d = 100$
Discrete samples of a	$D_a = 10$
Iterations of loops L_c	$L_c = 25$
Iterations of loops L_d	$L_d = 3$
Total Iterations L_{max}	$L_{max} = 5$

i.e., $f^k(t) = f_d$, then η_n^k can be calculated as

$$\eta_n^k = \frac{P}{M} \cdot \sum_{m=nM}^{m=(n+1)M-1} e^{j2\pi f_d T_s \cdot m} = \frac{P}{M} \cdot e^{j2\pi f_d T_s \cdot nM} \left| \frac{\sin(\pi f_d T_s M)}{\sin(\pi f_d T_s)} \right|. \quad (45)$$

Remark 2: The derived results in Eq. (45) demonstrate that the attention factor η_n^k decreases with the increase of the Doppler shift, when $|f_d| \leq \frac{1}{MT_s}$. Thus, the useful signal power is reduced with the increase of the Doppler shift, which decreases the SINR of the desired symbol and degrades the BER performance indirectly.

In summary, it is clear from Eq. (43) that the magnitude of the SINR is a function of the accuracy of the most recent detected symbols and time-varying Doppler shifts. According to Eq. (45), it is easy to demonstrate that higher Doppler shifts can cause more attenuation of useful signal power, which increases the probability of error detection indirectly. In addition, error detected symbols usually increases residual interference in turn, which inevitably increases the SINR of the desired symbol.

VI. SIMULATION RESULTS

In this section, simulation results based on MATLAB are provided to evaluate the performance of the proposed algorithm in terms of the BER performance, computational complexity and comparisons with other classical algorithms.

Specially, considering a practical scenario that the speed of high-mobility terminals can be up to 20 mach (about $6800m/s$) in satellite high-mobility communications, the maximum Doppler frequency offset $f_{\max}$ can be up to $50kHz$ and the maximum Doppler rate $a_{\max}$ can be up to $4kHz/s$. In addition, some essential parameters are shown in Tab. I.

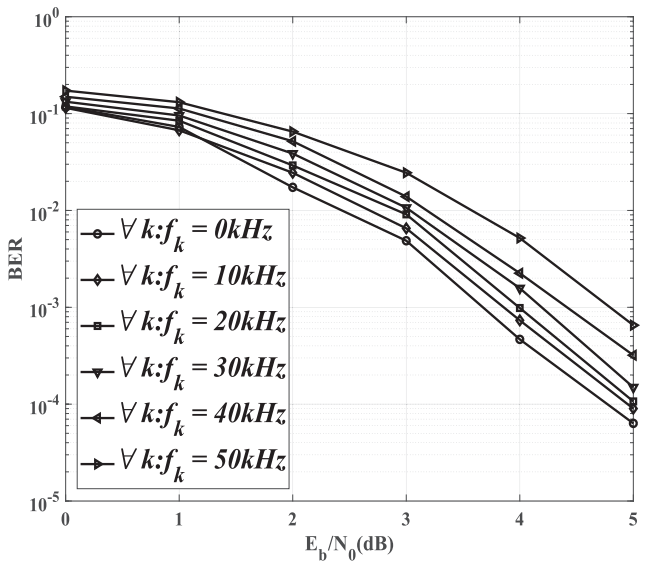

Fig. 5. BER performance of the proposed algorithm versus Doppler frequency offsets.

A. BER Performance

Fig. 5-7 show the BER performance of the proposed algorithm for different Doppler offsets, Doppler rates and user numbers. Specially, time offsets of users are varied to investigate the performance.

Fig. 5 shows the BER performance of the proposed asynchronous NOMA detection algorithm for different Doppler frequency offsets, when the user number $K = 3$ and the Doppler rate $\forall k : a^k = 3kHz/s$. Specially, each user's Doppler frequency offset is same in this BER result. It can be seen from the figure that the BER performance degrades gradually with the increase of the Doppler frequency offset. It is because Doppler frequency offsets degrade the correlation accumulation gain in the matched filter. It also verifies our obtained insights in **Remark 2**. However, by means of the message passing among three proposed loops, the proposed algorithm can eliminate the Doppler frequency offset effectively, where the loss of the correction gain can be reduced. It is also obvious in the figure that the performance degradation is about 1dB at BER=10^{-3} when the Doppler frequency offset is $50kHz$.

Fig. 6 shows the BER performance of the proposed asynchronous NOMA detection algorithm for different Doppler rates, when the user number $K = 3$ and the Doppler frequency offset $\forall k : f^k = 40kHz$. Specially, each user's Doppler rate is same in this BER result. It can be seen from the figure that the BER performance degrades with the increase of the Doppler rate, because Doppler shifts much faster with the increase of Doppler rates, which makes Doppler shifts become more difficult to be tracked and compensated in the Doppler elimination loop effectively. Overall, with the aid of the proposed algorithm, the influence of Doppler rates is relatively small. Specially, the degradation of the BER performance caused by Doppler rates is less than 0.15dB at BER=10^{-3} when the Doppler rate is up to $3kHz/s$.

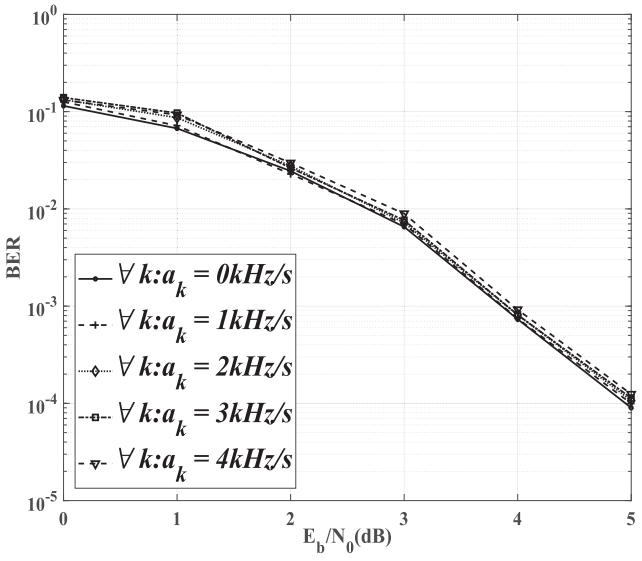

Fig. 6. BER performance of the proposed algorithm versus Doppler rates.

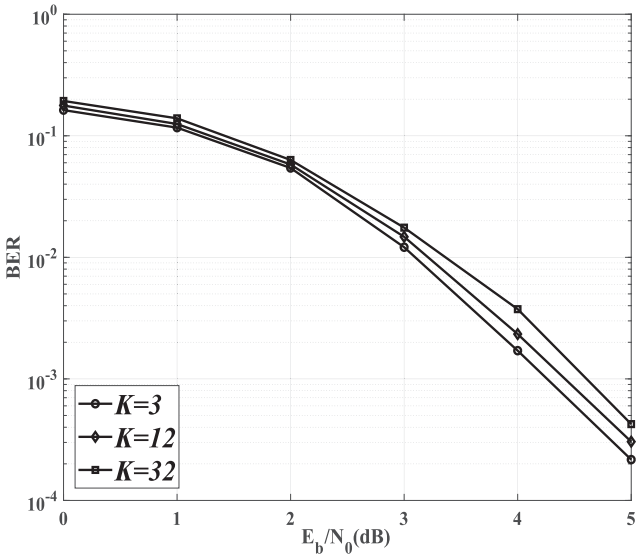

Fig. 7. BER performance of the proposed algorithm versus users.

Fig. 7 shows the BER performance of the proposed asynchronous NOMA detection algorithm for different user numbers, when the Doppler frequency offset $\forall k : f^k = 40kHz$ and the Doppler rate $\forall k : a^k = 2kHz/s$. Specially, users' Doppler shifts are same in this BER result. Obviously, it can be seen from the figure that the BER degrades gradually with the increase of the number of users, which is consistent with **Remark 1**. This is because more users increase multi-user interference in the received signal, which inevitably increases residual interference in the channel function nodes $\forall n, k : h_n^k$. Therefore, user-k would meet relatively more residual interference from adjacent symbols, which inevitably degrades the BER performance of asynchronous NOMA in high-mobility environments. Overall, with the aid of the proposed algorithm, the degradation of the BER performance can be reduced effectively. Compared with the BER performance with $K = 3$,

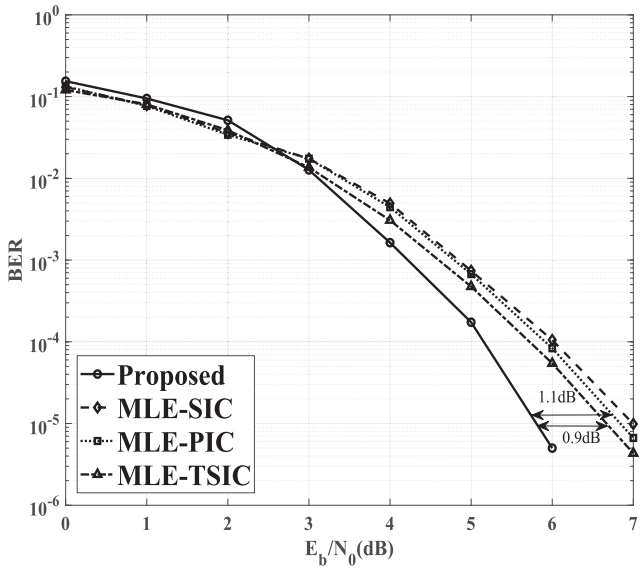

Fig. 8. BER performance versus different algorithms.

the degradation of the proposed algorithm with $K = 12$ is about 0.2dB and the degradation of $K = 32$ is about 0.4dB at BER=10^{-3}, respectively.

B. Comparisons With Classical Algorithms

In order to demonstrate the superiority of the proposed algorithm in terms of the BER performance, three classical interference cancellation methods, i.e., the conventional successive interference cancellation (Conv-SIC) algorithm [36], the parallel interference cancellation (PIC) algorithm [37], and the T-SIC algorithm [30] are selected to compare with the proposed algorithm. Owning to that traditional algorithms cannot work well in high-mobility environments, these three interference cancellation algorithms are improved by combining the maximum likelihood estimation (MLE) of Doppler shifts with Conv-SIC, PIC and T-SIC, named MLE-SIC, MLE-PIC and MLE-TSIC, respectively. For the former, MLE-SIC first estimates Doppler shifts by the MLE principle [38], and then performs Conv-SIC. While for the latter, MLE-PIC also first eliminates Doppler shifts by MLE, then performs PIC. For the last one, MLE-TSIC estimates Doppler shifts by MLE and performs T-SIC in the iterative manner. Moreover, for fair comparisons, the LDPC code is also applied in the MLE-SIC algorithm, MLE-PIC algorithm and MLE-TSIC algorithm. In addition, LPDC decoding iterations of MLE-SIC, MLE-PIC, MLE-TSIC and the proposed algorithm are same.

Fig. 8 shows the BER performance of the proposed asynchronous NOMA detection algorithm, the MLE-SIC algorithm, the MLE-PIC algorithm and the MLE-TSIC algorithm, when the user number $K = 3$. Specially, the Doppler frequency offset of each user is randomly and uniformly distributed on $[-50kHz, 50kHz]$ and the Doppler rate of each user is randomly and uniformly distributed on $[-4kHz/s, 4kHz/s]$. Obviously, it can be seen from the figure that the proposed algorithm can achieve the better BER performance gain of about 0.9dB over the MLE-TSIC algorithm, about 1.1dB over the MLE-PIC algorithm and about 1.2dB over the MLE-SIC

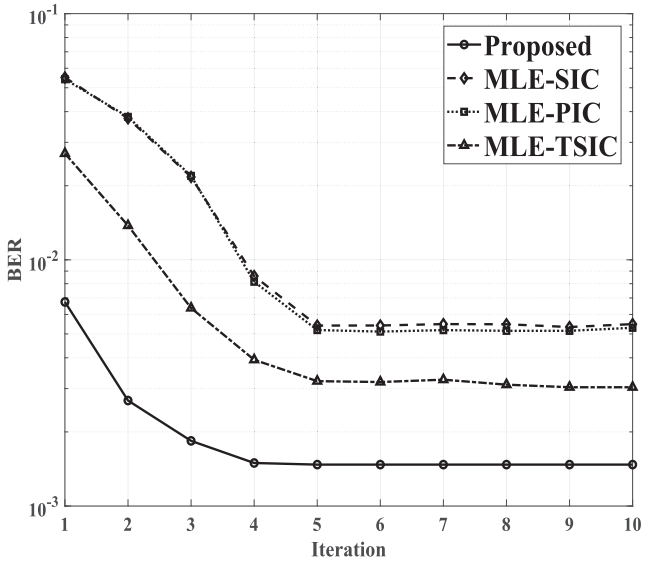

Fig. 9. BER performance of different algorithms versus iterations.

algorithm at BER=10^{-5}, respectively. Compared with traditional algorithms, the proposed algorithm is capable of achieving the BER performance of more than 0.9dB gain, which demonstrates the superiority of this algorithm in terms of the BER performance.

Fig. 9 shows the BER performance of the proposed asynchronous NOMA detection algorithm, the MLE-PIC algorithm, the MLE-SIC algorithm and the MLE-TSIC algorithm, when $E_b/N_0 = 4dB$. Similarly, the Doppler frequency offset of each user is also randomly and uniformly distributed on $[-50kHz, 50kHz]$ and the Doppler rate of each user is randomly and uniformly distributed on $[-4kHz/s, 4kHz/s]$. It is observed that BER performances of different algorithms gradually converges to constants with the increase of iterations. Moreover, the MLE-PIC algorithm, the MLE-SIC algorithm and the MLE-TSIC algorithm cannot converge to the BER performance of the proposed algorithm. Specially, BER convergence of the proposed algorithm is better than other traditional algorithms. This is because the proposed algorithm can better eliminate multi-user interference.

In addition, the proposed scheme also has potential advantages in spectrum efficiency and energy efficiency compared with other algorithms. It is shown in Fig. 9 that the proposed scheme can reach the same demodulation threshold with less E_b/N_0. In satellite communication systems, the demodulation threshold mainly refers to the value of E_b/N_0 at the specific BER threshold. In most cases, the specific BER threshold is set at BER=10^{-5}. Thus, the proposed scheme increases link margin in the satellite-terminal link. Particularly, link budget is the key of multi-user access in satellite communications. Generally, based on the actual system, bandwidth and transmitting power are given in advance. Then, the bit rate is determined according to the link margin. Given the same bandwidth transmitting power, the satellite receiver with the proposed scheme has more link margin, which is expected to realize higher spectrum efficiency and energy efficiency.

In summary, although MLE of Doppler shifts has been employed in the Conv-SIC algorithm, the PIC algorithm and the T-SIC algorithm to help them improve the BER performance in high-mobility environments, the proposed algorithm also outperforms them. There are two main reasons. On the one hand, based on the proposed 3D factor graph, multi-user interference can be eliminated more effectively in high-mobility environments. This is mainly due to more accurate detection results of multi-user sequences obtained by means of the message passing along edges of the proposed factor graph. Less multi-user interference from adjacent symbols inevitably results in the better BER performance. On the other hand, the proposed algorithm obtains gains from joint decoding, Doppler shift elimination and interference cancellation to approximate the MAP receiver. Basically, the proposed algorithm can achieve MAP estimation of data sequences approximately in high-mobility environments, which improves the BER performance effectively.

C. Complexity Analysis

In this paper, computational complexity of the proposed algorithm is evaluated by the amount of addition operations, multiplication operations and operations of looking up table. Note that looking up table operations mainly calculate the exponentiation $\exp(\cdot)$ and the logarithm $\ln(\cdot)$. For the proposed algorithm, in the Doppler elimination loop, it needs about $4NKD_dD_a$ multiplication operations and $4NKD_dD_a$ addition operations. In the interference cancellation loop, it needs about $2NK^2$ multiplication operations, $2NK^2$ addition operations and $2NK$ operations of looking up table. While for the LDPC decoding loop, it needs about $6NKW$ multiplication operations and $\frac{3NKW}{2}$ addition operations, where W represents the column weigh of LDPC check matrix. In this case, $W = 3$. In addition, in the process of message passing among these loops, it needs about $2NK(2D_dD_a + M)$ multiplication operations, $NK(2D_dD_a+M)$ addition operations and $3NK$ operations of looking up table. From above analysis, the proposed algorithm needs about $NKL_{\max}(4D_dD_aL_d + 6L_cW + 4D_dD_a)$ multiplication operations, $NKL_{\max}(4D_dD_aL_d + \frac{3L_cW}{2} + 2D_dD_a)$ addition operations and $5NKL_{\max}$ operations of looking up table totally. For comparison, the computational complexity of traditional algorithms is also listed in Tab. II. Note that $K!$ represents the factorial of K.

In order to compare the proposed algorithm with traditional algorithms, the computational complexity of different algorithms is compared by the number of floating-point operations (FLOPs), which is similar to [39] that the complexity of addition, subtraction, multiplication, and division is not distinguished for simplicity. In this case, the complexity of the proposed algorithm is about $\frac{1}{7}$ of the MLE-TSIC algorithm and about $\frac{4}{5}$ of the MLE-SIC algorithm and the MLE-PIC algorithm. Thus, the computational complexity of the proposed algorithm is less than traditional algorithms. Compared with tradition algorithm, the proposed algorithm reduces the computational complexity by about 20% at least, which demonstrates the superiority of this algorithm in terms of the

TABLE II
COMPLEXITY OF COMPARED ALGORITHMS

Algorithm	Addition Operations	Multiplication Operations
MLE-SIC	$NK(D_aD_fM + M + 6WL_cL_{\max})$	$NK(D_aD_fM + 2M + \frac{3WL_cL_{\max}}{2})$
MLE-PIC	$NK(D_aD_fM + KM + M + 6WL_cL_{\max})$	$NK(D_aD_fM + KM + \frac{3WL_cL_{\max}}{2})$
MLE-TSIC	$D_aD_fMNK! + NK(M + 6WL_cL_{\max})$	$D_aD_fMNK! + NK(2M + \frac{3WL_cL_{\max}}{2})$
Proposed	$NKL_{\max}(4D_dD_aL_d + \frac{3L_cW}{2} + 2D_dD_a)$	$NKL_{\max}(4D_dD_aL_d + 6L_cW + 4D_dD_a)$

computational complexity. Specially, the BER performance of the proposed algorithm is better than both of them. Overall, the proposed algorithm is capable of achieving the better performance with less computational complexity.

VII. CONCLUSION

NOMA is a promising technique for NGMA in satellite high-mobility communications. However, high-mobility terminals impose more challenges on traditional NOMA detection methods. Specially, time-varying high Doppler shifts caused by heterogeneous high-mobility terminals and different arriving times of multi-user signals due to wide geographical distribution of different terminals degrade the BER performance seriously and incur more difficulties for multi-user interference cancellation. To solve such a problem, a novel 3D factor graph and the turbo iterative asynchronous NOMA detection algorithm are proposed in this paper. By means of message passing among these loops designed specially for asynchronous NOMA in high-mobility environments, multi-user interference and time-varying Doppler shifts can be eliminated effectively, which improves the BER performance significantly. Finally, simulations results show that the BER performance of the proposed algorithm is at least 0.9dB better than traditional algorithms with less computational complexity. Overall, the proposed algorithm is expected to be well applied in NGMA of satellite high-mobility communications for its good performance in terms of the BER performance and computational complexity.

REFERENCES

[1] C. Jiang and X. Zhu, "Reinforcement learning based capacity management in multi-layer satellite networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4685–4699, Jul. 2020.

[2] C. Jiang, N. Ge, and L. Kuang, "AI-enabled next-generation communication networks: Intelligent agent and AI router," *IEEE Wireless Commun.*, vol. 27, no. 6, pp. 129–133, Dec. 2020.

[3] X. Yue *et al.*, "Outage behaviors of NOMA-based satellite network over Shadowed-Rician fading channels," *IEEE Trans. Veh. Technol.*, vol. 69, no. 6, pp. 6818–6821, Jun. 2020.

[4] J. Wu and P. Fan, "A survey on high mobility wireless communications: Challenges, opportunities and solutions," *IEEE Access*, vol. 4, pp. 450–476, 2016.

[5] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

[6] Y. Liu, Z. Qin, M. Elkashlan, A. Nallanathan, and J. A. McCann, "Non-orthogonal multiple access in large-scale heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2667–2680, Dec. 2017.

[7] Y. Liao, G. Sun, Z. Cai, X. Shen, and Z. Huang, "Nonlinear Kalman filter-based robust channel estimation for high mobility OFDM systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 11, pp. 7219–7231, Nov. 2021.

[8] J. Liu, X. Wang, G. Yue, and S. Shen, "Data sharing in VANETs based on evolutionary fuzzy game," *Future Gener. Comput. Syst.*, vol. 81, no. 1, pp. 141–155, 2018.

[9] X. Xia, J. Zhao, H. Long, G. Yang, J. Sun, and W. Yang, "Fractional Fourier transform-based unassisted tracking method for global navigation satellite system signal carrier with high dynamics," *IET Radar, Sonar Navigat.*, vol. 10, no. 3, pp. 506–515, 2016.

[10] J. Wang, C. Jiang, L. Kuang, and B. Yang, "Iterative Doppler frequency offset estimation in satellite high-mobility communications," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 12, pp. 2875–2888, Dec. 2020.

[11] R. Duan, J. Wang, C. Jiang, H. Yao, Y. Ren, and Y. Qian, "Resource allocation for multi-UAV aided IoT NOMA uplink transmission systems," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 7025–7037, Aug. 2019.

[12] J. Du, C. Jiang, J. Wang, Y. Ren, and M. Debbah, "Machine learning for 6G wireless networks: Carrying forward enhanced bandwidth, massive access, and ultrareliable/low-latency service," *IEEE Veh. Technol. Mag.*, vol. 15, no. 4, pp. 122–134, Dec. 2020.

[13] Y. Liu, Z. Qin, M. Elkashlan, Y. Gao, and L. Hanzo, "Enhancing the physical layer security of non-orthogonal multiple access in large-scale networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1656–1672, Mar. 2017.

[14] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, Apr. 2016.

[15] Y. Chen *et al.*, "Toward the standardization of non-orthogonal multiple access for next generation wireless networks," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 19–27, Mar. 2018.

[16] Z. Fang, J. Hu, Y. Lu, and W. Ni, "Three-user cooperative NOMA transmission," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 465–469, Apr. 2020.

[17] J. Wang, C. Jiang, and L. Kuang, "Turbo iterative DSSS acquisition in satellite high-mobility communications," *IEEE Trans. Veh. Technol.*, vol. 70, no. 12, pp. 12998–13009, Dec. 2021.

[18] X. Lin, L. Kuang, Z. Ni, C. Jiang, and S. Wu, "Approximate message passing-based detection for asynchronous NOMA," *IEEE Commun. Lett.*, vol. 24, no. 3, pp. 534–538, Mar. 2020.

[19] 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.

[20] 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.

[21] Z. Ding, L. Dai, and H. V. Poor, "MIMO-NOMA design for small packet transmission in the Internet of Things," *IEEE Access*, vol. 4, pp. 1393–1405, 2016.

[22] Z. Ding *et al.*, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[23] M. K. Naeem, R. Abozariba, M. Asaduzzaman, and M. Patwary, "Towards the mobility issues of 5G-NOMA through user dissociation and re-association control," in *Proc. IEEE 21st Int. Symp. 'World Wireless, Mobile Multimedia Netw.' (WoWMoM)*, Aug. 2020, pp. 427–432.

[24] 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.

[25] X. Dai, Z. Zhang, B. Bai, S. Chen, and S. Sun, "Pattern division multiple access: A new multiple access technology for 5G," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 54–60, Apr. 2018.

[26] B. Wang, L. Dai, T. Mir, and Z. Wang, "Joint user activity and data detection based on structured compressive sensing for NOMA," *IEEE Commun. Lett.*, vol. 20, no. 7, pp. 1473–1476, Jul. 2016.
[27] Y. Yin, Y. Peng, M. Liu, J. Yang, and G. Gui, "Dynamic user grouping-based NOMA over Rayleigh fading channels," *IEEE Access*, vol. 7, pp. 110964–110971, 2019.
[28] J. Cui, G. Dong, S. Zhang, H. Li, and G. Feng, "Asynchronous NOMA for downlink transmissions," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 402–405, Feb. 2017.
[29] M. Ganji, X. Zou, and H. Jafarkhani, "Asynchronous transmission for multiple access channels: Rate-region analysis and system design for uplink NOMA," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4364–4378, Jul. 2021.
[30] H. Haci, H. Zhu, and J. Wang, "Performance of non-orthogonal multiple access with a novel asynchronous interference cancellation technique," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1319–1335, Mar. 2017.
[31] X. Zhang and H. Jafarkhani, "Asynchronous network coding for multi-user cooperative communications," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8250–8260, Dec. 2017.
[32] A. Dehghan and A. H. Banihashemi, "On the Tanner graph cycle distribution of random LDPC, random protograph-based LDPC, and random quasi-cyclic LDPC code ensembles," *IEEE Trans. Inf. Theory*, vol. 64, no. 6, pp. 4438–4451, Jun. 2018.
[33] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
[34] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inf. Theory*, vol. 8, no. 1, pp. 21–28, Jan. 1962.
[35] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.
[36] Z. Ding, R. Schober, and H. V. Poor, "Unveiling the importance of SIC in NOMA systems—Part II: New results and future directions," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2378–2382, Nov. 2020.
[37] Y. Zouine, Z. Madini, K. S. Aloui, and J. Foshi, "Performance analysis of parallel interference cancellation receiver applied to hybrid OCDMA/WDM System for high-speed optical access," in *Proc. Asia–Pacific Microw. Conf. (APMC)*, vol. 2, Dec. 2015, pp. 1–3.
[38] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
[39] S. Wu, L. Kuang, Z. Ni, D. Huang, Q. Guo, and J. Lu, "Message-passing receiver for joint channel estimation and decoding in 3D massive MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8122–8138, Dec. 2016.

Jiawei Wang (Graduate Student Member, IEEE) received the B.S. degree from Xidian University, Xi'an, China, in 2019. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering, Tsinghua University, Beijing, China. His research interests include signal processing and satellite communications.

Chunxiao Jiang (Senior Member, IEEE) received the B.S. degree (Hons.) in information engineering from Beihang University, Beijing, in 2008, and the Ph.D. degree (Hons.) in electronic engineering from Tsinghua University, Beijing, in 2013. He is currently an Associate Professor with the School of Information Science and Technology, Tsinghua University. From 2011 to 2012 (as a Joint Ph.D. Student) and from 2013 to 2016 (as a Post-Doctoral Research), he was with the Department of Electrical and Computer Engineering, University of Maryland, College Park, under the supervision of Prof. K. J. Ray Liu. His research interests include application of game theory, optimization, and statistical theories to communication, networking, and resource allocation problems, in particular space networks and heterogeneous networks. He is a fellow of IET. He was a recipient of the Best Paper Award from IEEE GLOBECOM in 2013, the IEEE Communications Society Young Author Best Paper Award in 2017, the Best Paper Award from ICC 2019, the IEEE VTS Early Career Award 2020, the IEEE ComSoc Asia-Pacific Best Young Researcher Award 2020, the IEEE VTS Distinguished Lecturer 2021, and the IEEE ComSoc Best Young Professional Award in Academia 2021. He received the Chinese National Second Prize in Technical Inventions Award in 2018 and the Natural Science Foundation of China Excellent Young Scientists Fund Award in 2019. He has served as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE INTERNET OF THINGS JOURNAL, IEEE WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, *IEEE Network*, and IEEE COMMUNICATIONS LETTERS; and a Guest Editor for *IEEE Communications Magazine*, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, and IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING. He has also served as a member for the technical program committee and the symposium chair for a number of international conferences.

Linling Kuang (Member, IEEE) received the B.S. and M.S. degrees from the National University of Defense Technology, Changsha, China, in 1995 and 1998, respectively, and the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in 2004. Since 2007, she has been with the Tsinghua Space Center, Tsinghua University. Her research interests include wireless broadband communications, signal processing, and satellite communications. She is a member of the IEEE Communications Society.