IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 19, 2018, accepted November 16, 2018, date of publication November 26, 2018,
date of current version December 31, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2883399

A Framework of Non-Orthogonal Slotted Aloha (NOSA) Protocol for TDMA-Based Random Multiple Access in IoT-Oriented Satellite Networks

QIWEI WANG[1], GUANGLIANG REN[1], (Member, IEEE), STEVEN GAO[2], (Senior Member, IEEE), AND KUN WU[3]

[1]State Key Laboratory of ISN, Xidian University, Xi'an 710071, China
[2]School of Electronics and Digital Arts, University of Kent, Canterbury CT2 7NT, U.K.
[3]Wireless research center of Huawei Co., Ltd., Shanghai 200135, China

Corresponding author: Guangliang Ren (glren@mail.xidian.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 91538105 and Grant 61801352, in part by the Fundamental Research Funds for Central Universities under Grant JB170106, in part by the China Postdoctoral Science Foundation under Grant 2016M590924, in part by EPSRC under Project EP/N032497/1, and in part by the National Key Basic Research Program of China (973) under Grant 2014CB340206.

ABSTRACT There is an urgent demand for massive machine-type terminals to have access into time-division multiple access (TDMA)-based satellite networks by means of random multiple access (RMA). Several RMA protocols have been proposed by exploiting packet repetitions and interference cancellation to achieve high throughput. In this paper, a framework of non-orthogonal slotted aloha (NOSA) protocol is reported to achieve even higher throughput. With a specifically designed tile-based frame structure, it introduces the intra-tile sparse mapping as a special kind of pre-coded packet repetitions and exploits the joint multi-packet detection to blindly detect superimposed packets. By further employing inter-tile packet repetitions and interference cancellation, the NOSA protocol is able to achieve high throughput with affordable complexity while keeping the same transmission efficiency as and comparable power consumptions to available protocols. Simulation results show that the NOSA prototype has the potential in providing RMA for massive machine-type terminals in practical TDMA-based satellite networks.

INDEX TERMS Non-orthogonal slotted aloha (NOSA), random multiple access (RMA), sparse code multiple access (SCMA), time division multiple access (TDMA).

I. INTRODUCTION

With the development of Internet-of-Things (IoT) as well as wireless sensor networks, random multiple access (RMA) has represented a promising solution for a massive amount of machine-type communication (MTC) terminals to have access into satellite networks through a shared contention channel [1], [2]. As the transmission of massive MTC terminals has the features of power conservation, intermittency, and short packet transmissions, the traditional demand assignment multiple access (DAMA) protocol [3] appears not suitable for MTC terminals due to several reasons. For instance, there is a lack of global coordination among MTC terminals, while frequent applications and assignments would lead to massive signaling overheads as well as insufficient resources for a contention-free assignment, resulting in an unacceptable transmission latency via satellite channels. A great counterpart to the DAMA protocol, RMA protocols enable massive MTC terminals to have simultaneous access into the network through a shared contention channel in a dynamical and opportunistic way, and it has received increasingly attention in literature.

Several RMA protocols have been proposed for and widely utilized in satellite and terrestrial networks during the last three decades [4]–[10]. Slotted Aloha (SA) [4] and Diversity Slotted Aloha (DSA) [5] protocols are today widely used in Time Division Multiple Access (TDMA) based satellite networks for short packet transmissions and initial network access. The throughput of SA has increased over the primal pure Aloha, yet still poor in an absolute term. The DSA protocol introduces packet repetitions into the SA to achieve

a diversity gain in detection and thus a slight performance improvement is achieved in low load regions. By extending the concept of the DSA, significant throughput improvements have been brought by the Contention Resolution Diversity Slotted Aloha (CRDSA) protocol [6], [7], which introduces the intra-frame successive interference cancellation (SIC) technique into the DSA. Once a packet replica is successfully detected, its other replicas would be canceled from their respective slots, thus enabling further detection in these slots. Two extensions of CRDSA have been presented in the literature. Irregular Repetition Slotted Aloha (IRSA) [8] presented an optimized irregular packet repetitions on a bipartite graph, by which the number of packet replicas follows a distribution mass function instead a fixed number of repetitions in CRDSA. Coded Slotted Aloha (CSA) [9], [10] further generalized IRSA and CRDSA by introducing erasure correcting codes. In this case, the packet replicas are encoded rather than simply being repeated. Compared to traditional SA and DSA which simply discard the packets experiencing collisions as a waste, CRDSA and its extensions enable a favorable exploitation of repeated replicas of packets, and share features of cancelling interference from successfully detected packets and resolving replicas that were originally collided, thus achieving dramatic performance improvements. More specifically, the basic idea behind these protocols is that if a 'clean' replica in a collision channel model (i.e., a packet is assumed to be always successfully detected once there is no interfering ones in the current slot) is successfully detected, its inter-slot replicas could be recovered and canceled from other slots so as to make it possible to detect more replicas that were originally collided. Unfortunately, as the throughput of any MAC protocol under a collision channel is proved to have an asymptotical limit of 1 packet/slot, the reliance on the detection of 'clean' replicas prevent the original versions of these protocols from working in high load regions.

As the collision channel model is rather simple and not accurate, several studies [11], [14] have been presented to evaluate the performance of CRDSA, IRSA and CSA protocols under more practical models and prove that the performance of these protocols could be further boosted by taking advantages of forward error correction (FEC) coding and the capture effect [11]. In addition to the line of researches originated from TDMA-based systems, different works have applied spectrum spreading techniques into the RMA such as Enhanced Spread-Spectrum Access (E-SSA) [12], which implements symbol-level spreading and interference cancellation to accommodate a large number of simultaneous terminals and achieves high performance. A generalized technique [13] is further performed to combine E-SSA and Asynchronous Contention Resolution Diversity Aloha (ACRDA, which is the asynchronous version of CRDSA [7]) into an integrated framework. Most recently, a work in [15] has been reported to compare the performances of these protocols, namely CRDSA, IRSA, CSA, and E-SSA under practical physical (PHY) channel models. It demonstrates that CRDSA outperforms both original IRSA and CSA since the latter two schemes have not been optimized according to the practical system model. It also proves that E-SSA may achieve better performance and energy efficiency than existing TDMA based RMA schemes while suffering lower peak power and less synchronization overheads.

In this paper, in line with researches originating from CRDSA, we would like to provide an alternative way to realizing the RMA for massive MTC terminals in TDMA-based satellite networks that are widely employed in current standards. Toward this target, it is natural to resort to the state-of-the-art physical techniques. A family of advanced spreading techniques named low-dimensional signatures CDMA (LDS-CDMA) [16], [17] and sparse code multiple access (SCMA) [18], [19] has been proposed as one of the promising non-orthogonal multiple access techniques for 5G mobile communication systems. As opposed to traditional CDMA techniques, the utilization of low-dimensional sparse codes makes it possible for the system to work in an non-orthogonal overloading status, enabling an opportunity for massive terminals to have access into the network through a shared contention channel. A few works attempted to apply LDS-CDMA/SCMA techniques into the RMA transmission in LTE systems [20], [21] by spreading the data of a fixed number of terminals onto subcarriers with randomly selected sparse codes, and a message passing detector (MPD) is utilized to detect superimposed signals. However, since they rely highly on the LTE system with a fixed number of terminals, there is still a lack of an effective protocol by exploiting LDS-CDMA/SCMA techniques to realizing full RMA in wireless networks, especially in TDMA based satellite networks.

By exploiting advantages of both LDS-CDMA/SCMA and CRDSA techniques, we would like to present a novel framework of Non-Orthogonal Slotted Aloha (NOSA) for TDMA based satellite networks. With a specifically defined term *tile* that indicates a group of slots within a frame, the NOSA protocol is composed of two aspects. One is *intra-tile* packet-level sparse mapping and the other is *inter-tile* packet repetitions. The former one actually adopts the advantage of LDS-CDMA/SCMA techniques, by means of which original MAC packets would be mapped in packet-level by *randomly* selected low-density sparse codes on *randomly* selected tiles. This is essentially equivalent to a special form of pre-coded repetitions to form a tile-based packet replica. An MPD is then utilized at the demodulator to *blindly* resolve superimposed packets based on a specifically built bipartite graph in each tile. The latter one in fact exploits the advantage of CRDSA in generating twin replicas of a tile-based packet replica. In this case, should any of the tile-based packet replicas be resolved in one tile its twin-tile replica would be interference canceled from another tile by iterations, thus enabling the successful detection of remaining packets. Due to the intra-tile mapping and inter-tile repetitions, one MAC packet is also formed into several PHY replicas such that the NOSA protocol could therefore be regarded as a special extension of CRDSA.

The major contribution of this paper lies in that the NOSA framework provides an alternative approach to realizing IoT-oriented RMA for TDMA-based satellite networks by introducing multi-packet detection. With a specifically designed tile structure, we propose the packet-level sparse code mapping at the modulator and a tile-based joint detector using dynamically built bipartite graphs at the demodulator. The implementation of the NOSA and its connections with existing protocols are also depicted in detail. Note that the full RMA fashion of NOSA is determined by the aforementioned *random* transmission at modulators and *blind* detection at the demodulator, and that the term *Non-Orthogonal* comes from the intrinsic feature of low-density sparse codes and from possibly overloaded amounts of terminals in a unit resource.

This paper is structured as follows. Section II introduces the NOSA protocol in detail in terms of basic definitions, the NOSA modulator, demodulator and practical considerations. Simulation results are shown in Section III along with discussions, and the conclusion is presented in Section IV. The following notations are utilized throughout this paper. Matrices and vectors are denoted by symbols in boldface, and variables in italic. $\mathbf{1}_{M\times 1}$ and $\mathbf{0}_{M\times 1}$ stand for $M\times 1$ all-one and all-zeros vectors, and $\mathbf{I}_M$ for an $M\times M$ identity matrix. Notations $(\cdot)'$ and $(\cdot)^\dagger$ are used to represent transpose and conjugate transpose, respectively.

II. NON-ORTHOGONAL SLOTTED ALOHA FRAMEWORK

In this section, some basic definitions are provided for the proposed NOSA protocol, and the modulator and demodulator are presented in detail. Some practical considerations and discussion are illustrated as well.

A. PRELIMINARY DEFINITIONS

The purpose of this section is to briefly describe the TDMA-based satellite networks and the contention-based RMA channel, and provides necessary definitions in advance. One of the typical application scenarios of NOSA, the return link of satellite networks is mainly considered, by means of which signals of massive terminals are able to be transferred through the satellite to any gateway providing ground network access.

We will consider a TDMA frame of duration $\mathbb{T}_F$, which is composed of N TDMA slots with the duration of each being $\mathbb{T}_S = \mathbb{T}_F/N$. As for TDMA systems, we assume MTC terminals connecting with a gateway through the satellite have established coarse synchronization with the system such that only timing offsets between TDMA signals and the local timing reference are taken into account. To perform a fair comparison with existing protocols, the average normalized MAC load G expressed in bits/symbol is defined as

$$G = \frac{\lambda_f}{N \cdot G_p}, \tag{1}$$

where λ_f is defined as the mean MAC packet arrival rate per frame, and

$$G_p = \frac{R_c}{R_b} = \frac{1}{r_c \cdot log_2 Q} \tag{2}$$

with R_c the symbol rate, R_b the bit rate, r_c the FEC code rate, and Q the modulation order. Due to the introduction of packet repetitions in CRDSA-like schemes in a frame, the number of PHY replicas may be several times the number of MAC packets.

In a certain frame, among total N slots, we begin with defining $T = N/M$ *tiles*, where M refers to the total number of slots in one tile such that the duration of each tile is given as $\mathbb{T}_T = M\mathbb{T}_S$. Similar to the CRDSA, any MTC terminal should randomly select two distinct tiles to transmit its MAC packet, which is first intra-tile mapped and then inter-tile repeated such that each MAC packet forms into several PHY replicas. The intra-tile PHY replicas are used to resolve superimposed packets of different terminals, and inter-tile PHY replicas are utilized to cancel the interference on the other selected tile if successfully detected in one tile.

Since it is aforementioned that the NOSA protocol employs LSD-CDMA/SCMA techniques, a sparse indication matrix and sparse codebooks should be defined to map information bits onto corresponding slots of their selected tiles. The sparse indication matrix is expressed as $\mathbf{F} = [\mathbf{f}_1, ..., \mathbf{f}_v, ..., \mathbf{f}_V]$, where $\mathbf{f}_v \in \mathbb{B}^{M\times 1}$, V is the cardinality of $\mathbf{F}$ and $\mathbb{B}$ stands for the space of boolean values. Suppose $F(m, v)$ as the element on the m^{th} row and v^{th} column of $\mathbf{F}$. If $F(m, v) = 1$, it indicates there is a connection between the m^{th} slot and the v^{th} element of a code, and vice versa. Suppose $M = 4$ and $V = 6$, its indication matrix is given as

$$\mathbf{F}_{4,6} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

Associated with the indication matrix, a set of low-density sparse spreading codebooks $\mathcal{C} = [\mathbf{C}_1, ..., \mathbf{C}_v, ..., \mathbf{C}_V] \in \mathbb{C}^{M\times VQ}$ is defined in order to perform packet-level sparse mapping onto slots, where $\mathbb{C}$ stands for the space of complex values. $\mathbf{C}_v = \left[\mathbf{c}_{v,1}, ..., \mathbf{c}_{v,q}, \ldots, \mathbf{c}_{v,Q}\right] \in \mathbb{C}^{M\times Q}$ is defined as the v^{th} codebook with the dimension of each code being M, where $\mathbf{c}_{v,q}$ stands for the q^{th} code in the codebook $\mathbf{C}_v$. The cardinality of the codebook corresponds to the modulation order Q. The higher the modulation order, the more codes in each codebook. Every codebook $\mathbf{C}_v$ is a one-to-one mapping on the indication vector $\mathbf{f}_v$, meaning that positions of non-zero elements in $\mathbf{f}_v$ are identical to that in columns of $\mathbf{C}_v$. The number of non-zero elements in $\mathbf{f}_v$ and $\mathbf{c}_{v,q}$ is defined as K_v, which determines the sparsity of codebooks. If $K_1 = ... = K_v = ... = K_V$, it is called a regular NOSA system; otherwise, it is called an irregular NOSA system [22]. For example, the indication matrix $\mathbf{F}_{4,6}$ given above expresses a regular indication matrix with $K_{v,v=1,...,6} = 2$, while an irregular indication matrix is given as $\mathbf{F}_{4,10} = \left[\mathbf{F}_{4,6}\ \mathbf{I}_4\right]$ with $K_{v,v=1,...,6} = 2$ and $K_{v,v=7,...,10} = 1$. Moreover, in order for the ground demodulator to identify which specific codebook that a terminal employs and estimate corresponding channel parameters, we define a set of preambles $\mathcal{P} = \left[\mathbf{P}_1^{Pre}, ..., \mathbf{P}_v^{Pre}, \ldots, \mathbf{P}_V^{Pre}\right]$, where $\mathbf{P}_v^{Pre} \in \mathbb{C}^{1\times L_{Pre}}$ denotes

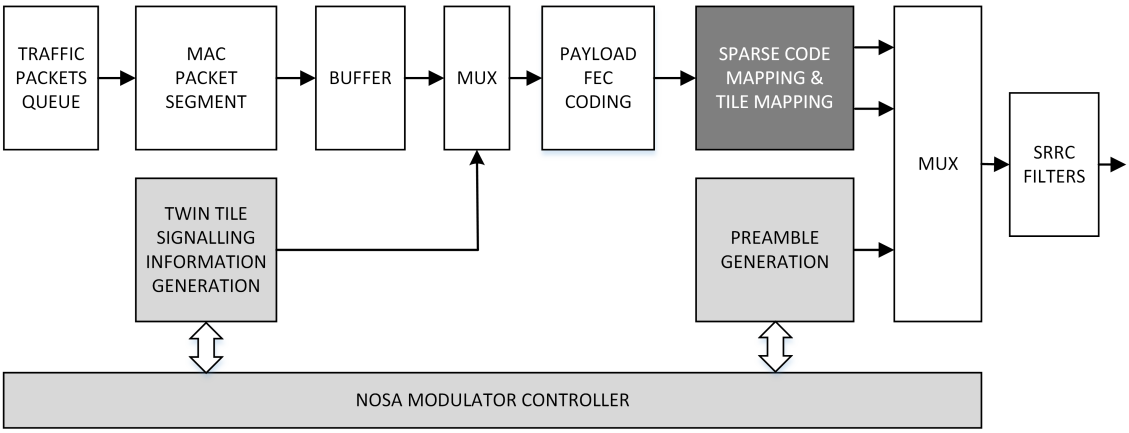


FIGURE 1. Baseband block diagram of a NOSA modulator. In dark grey the key blocks, in light grey modified ones compared to a CRDSA modulator.

one preamble with length L_{Pre} and each preamble $\mathbf{P}_v^{Pre}$ is in association with each codebook $\mathbf{C}_v$. Once a codebook $\mathbf{C}_v$ has been randomly selected by a terminal, an indication vector $\mathbf{f}_v$ and preamble $\mathbf{P}_v^{Pre}$ are naturally selected as well.

B. NOSA MODULATOR

A baseband functional block diagram of the NOSA modulator is depicted in detail in Fig. 1, in which blocks in dark gray indicates the key ones different from a CRDSA modulator, while blocks in light gray modified ones. The NOSA modulator is interpreted as follows. The traffic packet queue is first segmented into MAC packets and buffered for transmission. In order to realize the RMA feature, the NOSA modulator controller indicates any terminal to perform random selections of one codebook from the codebook set and of two distinct tile indexes of a frame. The twin-tile signaling information that indicates the tile index of its twin-tile replica should be generated according to the selected tile indexes, and it is then multiplexed with the MAC packet to form a FEC coded payload. This payload would then be mapped in packet-level by the SCMA/LDS-CDMA sparse codebook onto corresponding slots of its selected tiles. By multiplexing with its corresponding preamble, the baseband signal is finally filtered through a square root raised cosine (SRRC) filter for transmission. As the block of sparse code mapping and tile mapping is the key one that distinguishes with a CRDSA modulator, we will focus on its interpretation in what follows.

1) SPARSE CODE MAPPING

According to the mean MAC packet arrival rate λ_f, we define by U_{pkt}^{MAC} the number of MAC packets in the current frame. Define the FEC coded bits of a terminal u as $\mathbf{b}_u = [b_u(1), ..., b_u(log_2\ Q \cdot L_{Pay})]' \in \mathbb{B}^{log_2 Q \cdot L_{pay} \times 1}$, where L_{pay} is the length of the payload of a PHY replica. The coded bits $\mathbf{b}_u$ should then be mapped onto its selected twin tiles with indexes of t and t' ($t, t' \in [1, \ldots, T]$, $t \neq t'$), respectively. As the mapping operations for payloads on tile t and t' are exactly the same, we will then mainly focus on the t^{th} tile without loss of generality. Note that the FEC coded bits on tile t and t' are different due to different signaling bits. On the current tile t, we define $U_t \leq U$ as the number of terminals that have randomly selected this tile such that the overloading factor on this tile is expressed as $\eta_t = U_t/M$, which actually indicates the overloading condition in one tile.

By using the randomly selected codebook $\mathbf{C}_{v_u}$, where $v_u \in [1, ..., V]$ stands for a codebook index that has been chosen by a terminal $u \in [1, ..., U_t]$, every group of $log_2 Q$ coded payload bits should be mapped by one of the sparse codes in the codebook, and the mapping operation is represented as

$$\mathbf{b}_{u,l} \triangleq [b_u((l-1)\cdot log_2 Q+1), ..., b_u(l \cdot log_2 Q)]' \rightarrow \mathbf{c}_{v_u,q_l}, \tag{3}$$

where $\mathbf{b}_{u,l}$ is defined as the l^{th} group of the coded bits $\mathbf{b}_u$ with $l = 1, ..., L_{Pay}$, and $q_l \in [1, \ldots, Q]$ is the index of codes associated with the combination of bits in $\mathbf{b}_{u,l}$. This mapping operation is repeated until all coded bits of terminal u have been mapped to corresponding codes of its codebook. Note that eqn. (3) represents both the sparse code mapping of LDS-CDMA and SCMA techniques, which is stated as follows.

- For SCMA, the codebook $\mathbf{C}_v$ should be well designed to optimize the performance and the mapping is implemented directly from bits to codes as eqn. (3). A couple of researches have been conducted to optimize the design of codebooks [23], [24].
- For LDS-CDMA, every group of coded bits $\mathbf{b}_{u,l}$ is first modulated onto constellations $\mathbf{\Omega} = [\Omega_1, ..., \Omega_Q]$ as $\mathbf{b}_{u,l} \rightarrow \Omega_{q_l}$. A set of low-density spreading codes is then defined as $\mathcal{D} = [\mathbf{D}_1, \ldots, \mathbf{D}_v, ..., \mathbf{D}_V] \in \mathbb{Z}^{M \times V}$, where $\mathbb{Z}$ is defined as the space of $\{+1, -1\}$ and locations of non-zero elements in $\mathbf{D}_v$ are in accordance with that of $\mathbf{f}_v$. The modulation and spreading by LDS could be deemed as a special case of SCMA by expressing $\mathbf{c}_{v,q} = \Omega_q \mathbf{D}_v$.

2) TILE MAPPING

Focusing on the m^{th} slot of the current tile t, the payload on this slot is composed of the m^{th} element of all the mapped

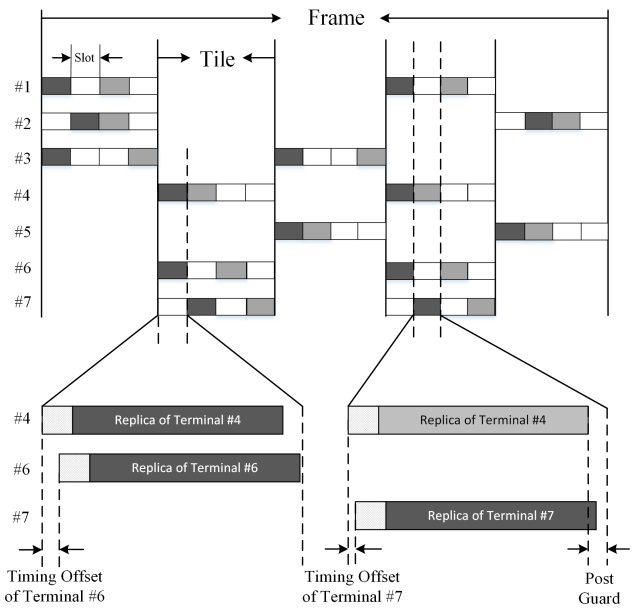


FIGURE 2. A schematic diagram of the NOSA modulator. We assume there are $U=7$ active terminals in $T=5$ demonstrative tiles, each equipped with $M=4$ slots. In dark gray represents the first non-zero packet, in light gray the second one, and in white vacant packets.

codes $\mathbf{c}_{v_u,q_l}$,and the PHY payload could be represented as

$$\mathbf{S}_{t,u}^{Pay}[m]=\left[\mathbf{c}_{v_u,q_1}(m)\,,\,...,\,\mathbf{c}_{v_u,q_{L_{Pay}}}(m)\right]\in\mathbb{C}^{1\times L_{Pay}},$$

where $m=1,\ldots,M$, $\mathbf{c}_{v_u,q_l}(m)$ stands for the m^{th} element of the q_l^{th} code in the codebook $\mathbf{C}_{v_u}$.

By adding the corresponding preamble ahead and a post guard at the end, the PHY replica on the m^{th} slot is given as

$$\mathbf{S}_{t,u}^{Slot}[m]=\left[\mathbf{P}_{v_u}^{Pre},\,\mathbf{S}_{t,u}^{Pay}[m]\,,\,\mathbf{S}^{Grd}\right]\in\mathbb{C}^{1\times L_{Slot}},$$

where $\mathbf{S}^{Grd}=\mathbf{0}_{1\times L_{Grd}}$ is defined as the post guard of length L_{Grd}, and L_{Slot} is the length of a slot. The length of a post guard should be large enough to accommodate the maximum timing offset among terminals.

By aggregating all mapped PHY replicas of M slots in a tile, a tile-based replica of terminal u is finally given as

$$\mathbf{S}_{t,u}^{Tile}=\left[\mathbf{S}_{t,u}^{Slot}[1],\ldots,\mathbf{S}_{t,u}^{Slot}[m],\ldots,\mathbf{S}_{t,u}^{Slot}[M]\right].$$

The tile-based replica above is sparse in accordance with the sparse indication vector $\mathbf{f}_{v_u}$, i.e., among total M slots in a tile there are only K_{v_u} ones that are non-zero. If a replica on the slot m does not actually exist, its payload should be vacant, i.e., $\mathbf{S}_{t,u}^{Pay}[m]=\mathbf{0}_{1\times N_{L_{Pay}}}$. Note that the PHY replicas are only mapped by sparse codes in packet-level, meaning that the information bits are mapped to several PHY replicas with each being the same as a traditional TDMA packet.

Fig. 2 offers a schematic diagram about the NOSA modulator, in which we assume $U=7$ active terminals in total and $T=5$ tiles for demonstration with each one composed of $M=4$ slots. The sparsity of codebooks is $K=2$, indicating a regular NOSA system, and there are 2 vacant slots for any terminal in its tile. The $1st$ non-zero slot of a terminal in a tile is demonstrated in dark gray, the $2nd$ in light gray, and the preamble in grid. By numbering these tiles as $t=1,..,5$, we may have $U_1=3$, $U_2=3$, $U_3=2$, $U_4=4$, and $U_5=2$ active terminals in corresponding tiles, respectively. Every terminal occupies two tiles to transmit its twin-tile replicas, indicating that each terminal has transmitted 4 PHY replicas in total. Taking terminal #1 as an example, it has randomly selected the $1st$ and $4th$ tiles, and since slots $m=1,3$ are randomly selected its codebook should be $\mathbf{C}_2$ with the indication vector being the second column of $\mathbf{F}_{4,6}$ as $\mathbf{f}_2=[1\ 0\ 1\ 0]'$. The lower half of the diagram demonstrates slots $m=1$ of the $2nd$ tile and $m=2$ of the $4th$ tile, where the PHY replicas of terminals #4 and #6 are active in the $1st$ slot of the $2nd$ tile and terminals #4 and #7 in the $2nd$ slot of the $4th$ tile. As for terminal #4, there exists no timing offset; yet for terminal #6 and #7, the post guard is utilized to avoid inter-slot interference due to timing offsets.

As for the inter-tile repetitions, it is worth mentioning that in the twin tile t' that is selected by a terminal u its payload bits should be operated exactly in the same way to form the tile-based replica. Due to the signaling information that points the tile index of the other twin-tile replica, the payload bits in its twin tiles are supposed to be different after FEC coding. However, as they both bear the original MAC packet, if one tile-based packet is successfully decoded, the other one could be simply regenerated for further interference cancellation.

C. NOSA DEMODULATOR

A baseband block diagram of a typical NOSA demodulator is depicted in Fig. 3, in which blocks in dark gray indicates the key ones and blocks in light gray modified ones compared to a CRDSA demodulator. The demodulator is controlled by a NOSA demodulator controller while first reserving the received signal into the frame sample memory. The controller will indicate qualified tiles according to the searching results of preambles, and then a tile-based joint detector is utilized to jointly detect replicas in qualified tiles. After that, detected tile-based replicas should be processed through the FEC decoder and circular redundant checking (CRC). The tile-based Replicas that have been successfully checked will be utilized for further interference cancellation, and signaling bits embedded in the packet are then utilized to indicate the location of its twin-tile replica. At the current or subsequent iterations, the tile-based replicas would be interference canceled from their own tiles. The key models of the NOSA demodulator will be elaborated in detail in this subsection with the received signal model presented at the very beginning.

The complex baseband signals could be represented after the matched filter at 1 sample/symbol with integer timing offsets. By assuming there are total U_t terminals in the tile t, the received signal at the demodulator could be represented in baseband as

$$\mathbf{r}_{t,m}=\sum_{u=1}^{U_t}A_{t,m,u}\left(e^{j(\varpi_{t,u}\cdot\varsigma+\phi_{t,m,u})}\odot\mathbf{S}_{t,u}^{Slot}[m]\right)\mathcal{Z}^{-\tau_{t,u}}+\mathbf{w}_{t,m},\tag{4}$$

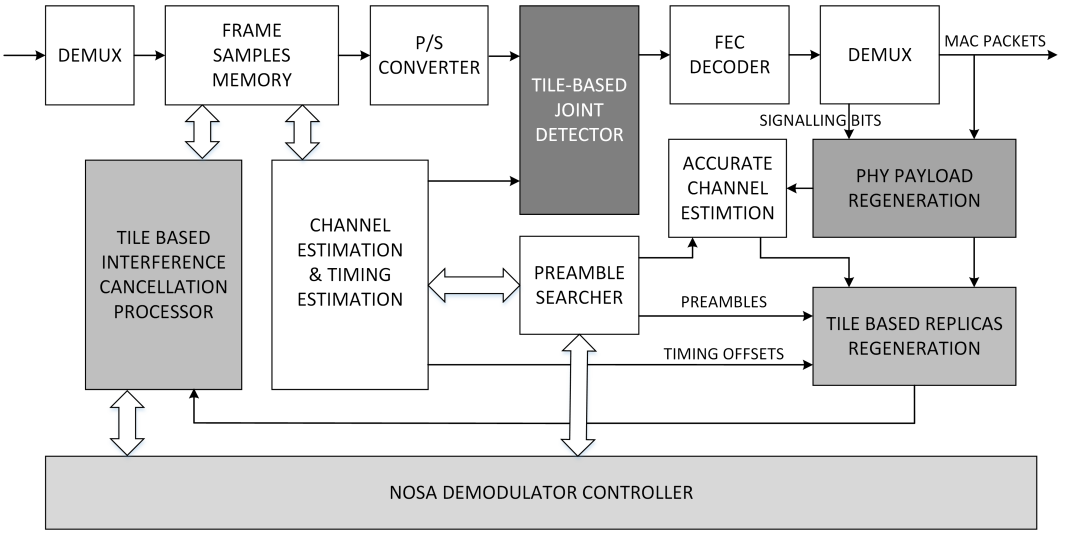


FIGURE 3. **Baseband block diagram of a NOSA demodulator. In dark grey the key blocks, and in light grey the modified ones compared to a CRDSA demodulator.**

where

- $\mathbf{R}_t \triangleq \left[\mathbf{r}_{t,1}..., \mathbf{r}_{t,m}, \ldots, \mathbf{r}_{t,M}\right]$ is defined as the received signal in the current tile, $\odot$ stands for the element-wise multiplication of vectors with the same length, $\mathbf{W}_t \triangleq \left[\mathbf{w}_{t,1}..., \mathbf{w}_{t,m}, \ldots, \mathbf{w}_{t,M}\right]$ is defined as the additive Gaussian white noise (AWGN) with zero mean and variance of σ_n^2, and $\mathbf{w}_{t,m}$ is the AWGN in each slot;
- $A_{t,m,u} = \sqrt{p_{t,m,u}}$ represents the amplitude of a PHY replica of terminal u in the m^{th} slot of the t^{th} tile, where $p_{t,m,u}$ is the received power that follows log-normal distribution with a mean power $\overline{p}_u$ and standard deviation in dB σ_{p_u}.
- $\varpi_{t,u}$ is defined as the angular frequency offset normalized by the symbol rate, ς is a vector standing for the symbol indexes corresponding to the start of the replica, and $\phi_{t,m,u}$ that follows uniform distribution in $[0, 2\pi)$ is the phase offset for terminal u in each slot.
- $\tau_{t,u}$ is the timing offset of terminal u in the tile t, and $\mathcal{Z}^{-\tau_{t,u}}$ is defined as a delay operator that shifts rightward the received signal by $\tau_{t,u}$ integer samples, which operates by adding at the head and by deleting at the tail $\tau_{t,u}$ zeros of $\mathbf{S}_{t,u}^{Slot}[m]$.
- The frequency and timing offsets are assumed to keep constant in a tile since we can regard a tile-based replica as an integration, while the amplitude and phase offset are assumed to be constant only in one slot but vary among slots due to possible fluctuation and drifting.

1) SIGNAL MODEL FOR TILE-BASED JOINT DETECTION

Based on the received signal model above, a signal model for the joint detection of tile-based replicas. By offsetting $L_{Pre}+\iota$ symbols from the beginning of each slot, we may sample one single symbol from $\mathbf{r}_{t,m}$ in each slot to form a signal with M symbols in total for the joint detection, i.e.,

$$\mathbf{y}_{t,\iota} = \sum_{u=1}^{U_t} \mathbf{h}_{t,u,\iota} \odot \mathbf{x}_{t,u,\iota} + \mathbf{w}_{t,\iota}, \tag{5}$$

where $\iota = 1, \ldots, L_{Slot} - L_{Pre}$ is the index of samples for the joint detection, $\mathbf{y}_{t,\iota} = \left[y_{t,1}(\iota), ..., y_{t,m}(\iota), ..., y_{t,M}(\iota)\right]'$ with each sample being

$$y_{t,m}(\iota) = r_{t,m}(L_{Pre}+\iota) = \sum_{u=1}^{U_t} h_{t,m,u}(\iota)\, S_{t,m,u}^{Pay}\left(\iota - \tau_{t,u}\right) + w_{t,m}(L_{Pre}+\iota).$$

$\mathbf{w}_t = \left[w_{t,1}(L_{Pre}+\iota), \ldots, w_{t,M}(L_{Pre}+\iota)\right]$ are AWGN samples. $S_{t,m,u}^{Pay}\left(\iota - \tau_{t,u}\right)$ and $r_{t,m}(L_{Pre}+\iota)$ represent elements of $\mathbf{S}_{t,u}^{Pay}[m]$ and $\mathbf{r}_{t,m}$, respectively. $S_{t,m,u}^{Pay}\left(\iota - \tau_{t,u}\right) \triangleq 0$ in case of $\iota - \tau_{t,u} \leq 0$ or vacancy. $\mathbf{x}_{t,u,\iota}$ contains M samples as

$$\begin{aligned}\mathbf{x}_{t,u,\iota} &= \left[x_{t,1,u}(\iota), ..., x_{t,m,u}(\iota), \ldots, x_{t,M,u}(\iota)\right]' \\ &= \left[S_{t,1,u}^{Pay}\left(\iota - \tau_{t,u}\right), ..., S_{t,M,u}^{Pay}\left(\iota - \tau_{t,u}\right)\right]',\end{aligned}$$

and the equivalent channel coefficient vector is given as

$$\mathbf{h}_{t,u,\iota} = \left[h_{t,1,u}(\iota), \ldots, h_{t,m,u}(\iota), \ldots, h_{t,M,u}(\iota)\right]$$

with $h_{t,m,u}(\iota) \triangleq A_{t,m,u} e^{j(\varpi_{t,u}(L_{Pre}+\iota-\tau_{t,u})+\phi_{t,m,u})}$.

It is worth noting that $\mathbf{x}_{t,u,\iota}$ contains K_{v_u} non-zero samples (by supposing its codebook is known as $\mathbf{C}_{v_u}$ yet in practical systems it is not aware of) or all zero samples (in case of $S_{t,m,u}^{Pay}\left(\iota - \tau_{t,u}\right) = 0$). Therefore, $\mathbf{x}_{t,u,\iota}$ is actually an all zero vector or one of the codes in $\mathbf{C}_{v_u}$, i.e., $\mathbf{x}_{t,u,\iota} \in \mathbf{C}_{v_u} \bigcup \mathbf{0}_{M\times 1}$. The signal $\mathbf{y}_{t,\iota}$ actually forms an SCMA or LDS-CDMA system. A bipartite factor graph $\mathcal{G}(\mathcal{F}, \mathcal{V})$ could be used to describe the system with $\mathcal{F}$ standing for the function nodes (FN) and $\mathcal{V}$ variable nodes (VN). Elements of $\mathbf{y}_{t,\iota}$ could be deemed as FNs, and elements of the payload $\mathbf{x}_{t,u,\iota}$ as VNs.

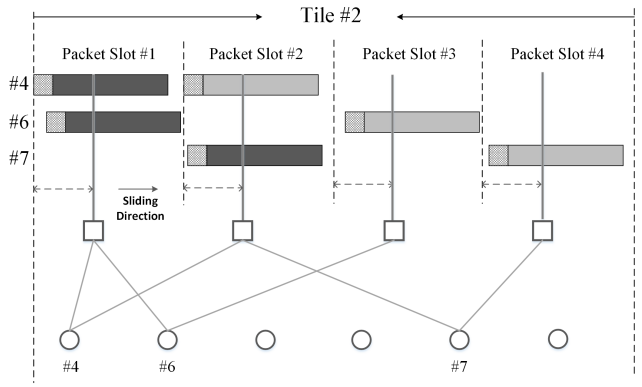


FIGURE 4. A schematic diagram about the formation of signal $\mathbf{y}_{t,\iota}$ for the joint detection with the corresponding bipartite graph representation, where squares stands for FNs, circles for VNs.

Therefore, there are M FNs, U_t VNs (if correctly detected), and $\sum_{u=1}^{U_t} K_{v_u}$ connections between FNs and VNs in total.

Fig. 4 gives an example about the formation of $\mathbf{y}_{t,\iota}$ for the joint detection with the bipartite graph presented. The tile represented in Fig. 4 is actually the 2*nd* tile in Fig. 2 with $U_2 = 3$ terminals and the terminal indexes are #4, #6, and #7. The number of codebooks is defined as $V = 6$, it can be seen that terminal #4 uses the 1*st* codebook, that terminal #6 uses the 2*nd* codebook, and that terminal #7 uses the 5*th* codebook. The sampling locations slide rightward to obtain the signal $\mathbf{y}_{t,\iota}$ that would be detected each time. By using the signal model above, it is possible to detect PHY replicas in a tile with different timing offsets, and this process will be repeated until all payload samples have been detected. The FEC decoder is employed to decode the payload symbols with the knowledge of estimated timing offsets. The joint detection for $\mathbf{y}_{t,\iota}$ would be elaborated in the next sub-section.

2) MESSAGE PASSING DETECTOR FOR NOSA

The received signal above shall be adopted to resolve packets of terminals tile by tile. With the assistance of preambles, the detector at the gateway is feasible to obtain information of superimposed replicas about adopted codebook indexes, amplitudes, and offsets of frequency, phase and timing. As for the local preamble searcher, the correlation of a local preamble $\mathbf{P}_v^{Pre}$ with $\mathbf{r}_{t,m}$ is given as

$$I_{t,m}(v,\tau) = \mathbf{P}_v^{Pre} \cdot \mathbf{r}_{t,m}^{\dagger}[\tau : \tau + L_{Pre}], \tag{6}$$

where $\tau = 1, ..., \tau_{\max} + 1$ stands for the index of possible timing offsets, $\tau_{\max}$ is the maximum possible timing offset, and $\mathbf{r}_{t,m}[\tau : \tau + L_{Pre} - 1]$ accounts for a vector of length L_{Pre} extracted from $\mathbf{r}_{t,m}$. The existence of adopted codebooks and corresponding parameters could be estimated according to the correlation results. Due to constant frequency and timing offsets in one tile, we may further exploit the coherency of K_v preambles of a codebook $\mathbf{C}_v$ to improve the performance. Note that if more than one terminals have reused the same codebook and preamble, they are expected to be distinguished thanks to their possible different frequency and timing offsets.

The signal model in eqn. (5) is only provided by supposing perfect knowledge of terminals. However, in practical systems, the NOSA demodulator is aware of nothing but the estimation results from the preamble searcher. Therefore, we should define by $\widehat{U}_t$ the number of detected terminals (including both successfully and falsely detected ones) and by $\widehat{v}_{t,u}$, $\widehat{h}_{t,m,u}(\iota)$, and $\widehat{\tau}_{t,u}$ the estimated codebook index, channel coefficient, and timing offset of terminal u with $u = 1, \ldots, \widehat{U}_t$. The signal model based on the estimation results is re-expressed as

$$\mathbf{y}_{t,\iota} = \sum_{u=1}^{\widehat{U}_t} \widehat{\mathbf{h}}_{t,u,\iota} \mathbf{x}_{t,u,\iota} + \overline{\mathbf{w}}_{t,\iota} \tag{7}$$

with $\mathbf{x}_{t,u,\iota} = [S_{t,1,u}^{Pay}(\iota - \widehat{\tau}_{t,u}), ..., S_{t,M,u}^{Pay}(\iota - \widehat{\tau}_{t,u})]'$, $\widehat{\mathbf{h}}_{t,u,\iota}$ the estimation of $\mathbf{h}_{t,u,\iota}$, and $\overline{\mathbf{w}}_{t,\iota}$ AWGN plus interference due to estimation errors. The objective of the detector is to determine which specific code $\mathbf{x}_{t,u,\iota}$ is supposed to be in $\mathbf{C}_{\widehat{v}_{t,u}}$. In case of $\iota - \widehat{\tau}_{t,u} \leq 0$ or $\iota - \widehat{\tau}_{t,u} > L_{Pay}$, $\mathbf{x}_{t,u,\iota}$ is set to $\mathbf{0}_{M\times 1}$.

The MPD would be used repeatedly to detect the payload symbols with detected codebooks. Defining $\Theta_{\mathcal{F}\to\mathcal{V}}$ and $\Theta_{\mathcal{V}\to\mathcal{F}}$ as the messages sent along edges from an FN to a VN and vice versa, and bearing in mind that $\mathbf{x}_{t,u,\iota}$ is deemed as one of the codes in codebook $\mathbf{C}_{\widehat{v}_{t,u}}$, the MPD could be expressed as [26]

$$\begin{aligned}\Theta_{\mathcal{F}_m\to\mathcal{V}_u}(x_{t,m,u}(\iota)) &= \sum\nolimits_{\sim x_{t,m,u}(\iota)} \frac{1}{\sqrt{2\pi}\sigma_n} \\ &\times \exp\left(-\frac{1}{2\sigma_n^2}\| y_{t,m}(\iota) - \sum_{\mu\in\xi_m} \widehat{h}_{t,m,\mu}(\iota)\, x_{t,m,\mu}(\iota)\|^2\right) \\ &\times \prod_{\mu\in\xi_m/\{u\}} \Theta_{\mathcal{V}_\mu\to\mathcal{F}_m}(x_{t,m,\mu}(\iota)),\end{aligned} \tag{8}$$

and

$$\Theta_{\mathcal{V}_u\to\mathcal{F}_m}(x_{t,m,u}(\iota)) = \prod_{n\in\zeta_u/\{m\}} \Theta_{\mathcal{F}_n\to\mathcal{V}_u}(x_{t,n,u}(\iota)), \tag{9}$$

where ξ_m and ζ_u stand for the sets of VNs and FNs that connect to the FN $y_{t,m}(\iota)$ and VN $\mathbf{x}_{t,u,\iota}$, $x_{t,m,u}(\iota)$ is actually a variable that represents the m^{th} element connecting with the FN $y_{t,m}(\iota)$ and has Q possible values (i.e., $\mathbf{c}_{\widehat{v}_{t,u},q}(m)$ with $q = 1, \ldots, Q$), μ and n are defined as indexes of VNs and FNs that are distinct with u and m, respectively. The probability that the current samples belong to the q^{th} code is given as

$$\mathcal{P}(\mathbf{b}_{u,l} \to \mathbf{c}_{v_u,q_l}) = \prod_{m\in\xi_m} \Theta_{\mathcal{F}_m\to\mathcal{V}_u}(x_{t,m,u}(l)), \tag{10}$$

based on which the Log-Likelihood Ratio (LLR) of every bit could be easily derived with $\Theta_{\mathcal{F}_m\to\mathcal{V}_u}(x_{t,m,u}(l)) = \Theta_{\mathcal{F}_m\to\mathcal{V}_u}(\mathbf{c}_{\widehat{v}_{t,u},q}(m))$, $q = 1, \ldots, Q$. Note that even though the total number of samples of each terminal is $L_{Pkt} - L_{Pay}$, only LLRs of L_{Pay} symbols should be calculated as the NOSA demodulator is aware of the estimated timing offset. That is

to say, the l^{th} payload symbol is calculated by utilizing the sample with index of $\iota = L_{Pre} + \widehat{\tau}_{t,u} + l$.

3) PRACTICAL CONSIDERATIONS

In tradional LDS-CDMA/SCMA systems, since codebooks are specifically allocated by the based station to terminals, each FN has a fixed number of connections with VNs. For example, if the code length is M with the number of non-zero elements being K, the total number of codebooks is given as $V = \binom{M}{K}$. Suppose that all the codebooks are exclusively allocated to V terminals, the number of colliding elements on each FN is given as $d = \binom{M-1}{K-1}$. Unfortunately, the NOSA protocol actually works in a complete RMA way with an indefinite number of terminals as well as randomly selected codebooks in each tile, leading to a codebook collision problem. By focusing on the tile t occupied by at least one terminal, the probability that given G there exist $U_t \geq 1$ terminals (i.e., an existing terminal plus $U_t - 1$ coexisting ones) is given as

$$\mathcal{P}_t(U_t - 1|G) = \binom{NGG_p - 1}{U_t - 1} \cdot p_t^{U_t - 1} (1 - p_t)^{NGG_p - U_t},$$

where $p_t = \frac{2}{T}$ is the probability that a terminal has randomly selected the current tile by taking into consideration twin-tile repetitions. As there are total V codebooks, the probability that a codebook collision occurs (i.e., one or more terminals have selected the same codebook with the existing one) in the t^{th} tile is given as

$$\mathcal{P}_{Coll}(G) = \sum_{U_t=1}^{NGG_p} \mathcal{P}_t(U_t - 1|G) \cdot \left[1 - \frac{(V-1)^{U_t - 1}}{V^{U_t - 1}}\right]. \quad (11)$$

The codebook collision is an increasingly frequent occurring event with the increase in the normalized load and it happens quite often in the NOSA scheme. Intuitively speaking, the NOSA demodulator will successfully detect replicas not experiencing codebook collisions with a higher probability while those reusing codebooks with a smaller probability. However, due to the employment of interference cancellation, once replicas have been detected, the interference cancellation of them enables other remaining ones to be detected later. The NOSA demodulator is destructive when there are too many tile-based replicas superimposed in each tile such that few replicas could be detected and interference canceled.

Besides the codebook collision problem, the RMA way of NOSA actually leads to probabilistic connections on each FN. As the calculation of $\Theta_{\mathcal{F}_m \to \mathcal{V}_u}\left(x_{t,m,u}(\iota)\right)$ necessitates a maximum likelihood detection of all elements that connect to the FN $y_{t,m}(\iota)$ by excluding the current value of $x_{t,m,u}(\iota)$, the condition of connections determines both the complexity and performance of the NOSA demodulator. In this case, we need to make some proper modifications for the MPD. Based on the estimation results, an indication matrix $\widehat{\mathbf{F}}_t$ for the current tile could be generated by stacking up all indication vectors of detected codebook indexes, i.e.,

$$\widehat{\mathbf{F}}_t = \left[\mathbf{f}_{\widehat{v}_{t,1}}, ..., \mathbf{f}_{\widehat{v}_{t,u}}, \ldots, \mathbf{f}_{\widehat{v}_{t,\widehat{U}_t}}\right].$$

A vector that contains numbers of connections on each FN is then defined as $\mathbf{d}_t$ to indicate possible collisions on each slot, i.e., $\mathbf{d}_t = [d_{t,1}, \ldots, d_{t,m}, \ldots, d_{t,M}]' = \widehat{\mathbf{F}}_t \cdot \mathbf{1}_{\widehat{U}_t \times 1}$, where $d_{t,m}$ means the number of estimated colliding symbols on the m^{th} slot in the t^{th} tile. As the calculation of $\Theta_{\mathcal{F}_m \to \mathcal{V}_u}\left(x_{t,m,u}(\iota)\right)$ necessitates a maximum likelihood detection (MLD) of all elements that connect to $\mathcal{F}_m$, the computational complexity increases exponentially with the number of connections $d_{t,m}$. By further defining the maximum number of connections on a FN as $d_{\max}$, only $d_{\max}$ most dominant VNs on $\mathcal{F}_m$ instead of all $d_{t,m}$ ($> d_{\max}$) ones would be taken into account, while all K_{v_u} connections of an abandoned codebook v_u should be neglected during the detection. The replicas of abandoned codebooks are regarded as interference in the current detection, and they are probably detected in further detections after interference cancellation. Similarly to an E-SSA demodulator, this operation indicates that the joint detection also exploits power unbalance starting with the most dominant ones. In addition, due to the difference of timing offsets among replicas, within the first $\tau_{\max}$ samples, the influence of preambles should be canceled as well. Note that the optimization of the joint detection, codebook design and power distribution are expected yet beyond the scope of this paper.

Similar to existing RMA protocols, interference cancellation is of significance to the NOSA scheme as well, and it requires more accurate estimation of parameters than the joint detection. In this case, the NOSA demodulation employs exactly the same operation as that in existing RMA schemes to estimate the parameter set $\{A_{t,m,u}, \varpi_{t,u}, \phi_{t,m,u}\}$ by using both the preamble and successfully detected replica payload, the detail of which can be found in the Appendix of [8]. After interference cancellation in the current tile, the preamble searcher needs to further go through the preambles and timing offsets of interference canceled replicas such that codebooks that may probably be under collisions before could be identified along with estimated parameters. The implementation of the NOSA demodulator is presented in the Algorithm I in detail.

D. FURTHER DISCUSSIONS

In this subsection, we will further discuss some mostly concerned issues about the NOSA scheme.

1) CONNECTIONS WITH EXISTING RMA SCHEMES

In the NOSA scheme, every MAC packet should be first LDS-CDMA/SCMA mapped and then tile repeated to form several PHY replicas, each one being totally the same as a traditional TDMA packet in the CRDSA family. By using the same frame structure, the only difference is that the PHY replicas in NOSA should be placed according to the tile structure with each symbol belonging to the sparse code constellation, more specifically to one of the rotated QPSK constellation in this paper. As there is no symbol-level spreading on the replica, the transmission efficiency, bandwidth and Baud rate of a NOSA replica are exactly the same as that of a traditional TDMA packet with their preambles and

Algorithm 1 NOSA Demodulator

Require: $\mathbf{y}_{t,\iota}$, $d_{\max}$, and $\Theta_{\mathcal{V}_u\to\mathcal{F}_m}\left(x_{t,m,u}(\iota)\right)=\frac{1}{Q}$.
repeat
for Each $t \in [1,\ldots,T]$ **do**
repeat
Obtain by the preamble searcher estimates of $\widehat{U}_t$, $\widehat{v}_{t,u}$, $\widehat{h}_{t,m,u}$, and $\widehat{\tau}_{t,u}$ with $u=1,\ldots,\widehat{U}_t$, and calculate $\mathbf{d}_t$. If $d_{t,m} > d_{\max}$, keep only $d_{\max}$ dominant replicas on $\mathcal{F}_m$ by ignoring weak ones. Update $\widehat{U}_t$ and cancel interference of preambles at the start.
1) Message-Passing Detection
if $\widehat{U}_t > 0$ and any replica has been detected or canceled previously **then**
for $\iota \in \left[1,\ldots,L_{pkt}-L_{Pre}\right]$ **do**
Based on $\widehat{\tau}_{t,u}$, obtain $\widehat{U}_{t,\iota} \le \widehat{U}_t$ replicas among the remaining ones that are active in the current sample ι, and obtain codebook indexes of replicas connecting $\mathcal{F}_m$ as $u_{\alpha,\alpha=1,\ldots,d_{t,m}(\iota)}$.
repeat
for $m \in [1,\ldots,M]$ **do**
Calculate $\Theta_{\mathcal{F}_m\to\mathcal{V}_u}\left(x_{t,m,u_\alpha}(\iota)\right)$ as (8).
end for
Normalize $\Theta_{\mathcal{F}_m\to\mathcal{V}_u}\left(x_{t,m,u}(\iota)\right)$ over total Q possible values.
for $u \in [1,\ldots,\widehat{U}_t]$ **do**
Update $\Theta_{\mathcal{V}_u\to\mathcal{F}_m}\left(x_{t,m,u}(\iota)\right)$ as (9).
end for
until Designated MPD iterations done.
end for
end if
OUTPUT: Decode the LLRs of detected replicas to check if they are successfully detected, and obtain twin-tile indexes of successfully detected replicas.
2) Interference Cancellation
For each successfully detected MAC packet, obtain the parameter set $\left\{A_{t,m,u}, \varpi_{t,u}, \phi_{t,m,u}\right\}$ on both tiles, and cancel its interference on both tiles by reconstructing both tile-based replicas.
OUTPUT: Interference canceled signals $\mathbf{y}_{t,\iota}$.
until None of PHY replicas could be detected.
end for
until None of PHY replica is newly detected or canceled or the maximum number of iterations is reached.
Output: Payloads of successfully detected MAC packets.

payloads of equal size, and there is no extra sparse coding overhead in NOSA replicas but instead mapping one MAC packet into one or two PHY replicas. In this regard, we may deem the NOSA scheme as a special case of the CRDSA by introducing a specific tile structure and mapping onto sparse code constellations. It is worth noting that the TDMA based RMA schemes require extra synchronization overheads and higher peak power than the E-SSA scheme. However, as the LDS-CDMA/SCMA technique normally employs pretty short sparse codes, it is not in nature a spreading technique as CDMA. Therefore, the NOSA scheme relates closely to TDMA based RMA schemes rather than the E-SSA. The main purpose of the NOSA is to introduce multi-packet detection into TDMA systems that are widely employed in satellite standards such as DVB-RCS2.

2) COMPUTATIONAL COMPLEXITY

Compared to CRDSA, the complexity of NOSA mainly resides in two aspects, one is the preamble detection, and the other is the joint detection of replicas. As for the former one, it is necessary for the NOSA scheme to identify different codebooks that have been randomly selected by different terminals. Therefore, the number of parallel correlators that the NOSA scheme requires is V instead of 1 as in CRDSA. The complexity of preamble detection of NOSA should also take into account correlations of those successfully detected replicas in identifying originally collided replicas. As for the latter one, the complexity mainly comes from the MLD calculations at FNs with the maximum number of connections, which can be expressed as $\mathcal{O}(L_{Pay}Q\cdot 2^{\overline{d}_{\max}})$, where $\overline{d}_{\max}$ is defined as the maximum number of connections overall with $\overline{d}_{\max} \le d_{\max}$, and L_{Pay} is used instead of $L_{Pkt}-L_{Pay}$ since there are at most L_{Pay} samples whose FN has a maximum number of connections $\overline{d}_{\max}$. Note that there are many researches demonstrating low-complexity joint detection techniques which could be exploited in the NOSA scheme [26]. In addition, the complexity of interference cancellation for each PHY replica of NOSA is exactly the same as that of CRDSA, and it grows with the achievable throughput and average number of PHY replicas of each MAC packet.

3) ENERGY EFFICIENCY

As most MTC terminals are powered by batteries or solar energy, the energy efficiency of satellite based RMA schemes is of great significance, which is expressed as [15]

$$\psi_e = \frac{r_c log_2 Q}{\overline{N}_{rep} N}, \tag{12}$$

where $\overline{N}_{rep}$ represents the average number of PHY replicas, and as aforementioned $\overline{N}_{rep}=4$ for a regular NOSA system and $\overline{N}_{rep}=3.2$ for an irregular system. This indicates a regular NOSA system has a half energy efficiency of the CRDSA-2, and that an irregular NOSA system has comparable energy efficiency to the CRDSA-3. In line with [15], the energy efficiency of a regular and irregular NOSA system with $N=128$ slots is 1.3×10^{-3} and 1.6×10^{-3}, respectively.

III. SIMULATION RESULTS

A. SIMULATION PARAMETERS

In this paper, the simulation of NOSA is carried out by utilizing a physical simulator to verify the practical performance. In order to make a fair comparison with existing RMA schemes, we follow the simulation parameters as mentioned

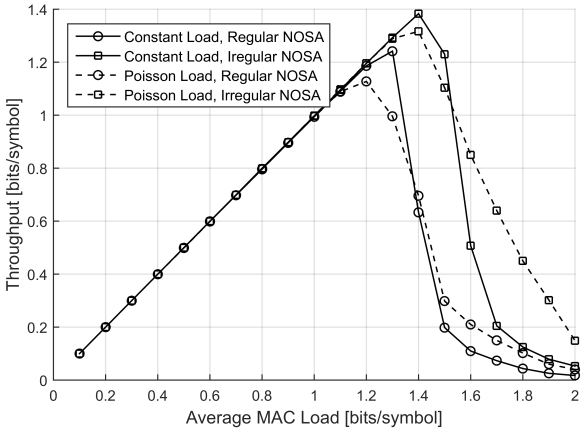


FIGURE 5. Comparison of throughput between regular and irregular NOSA systems under constant and Poisson MAC loads.

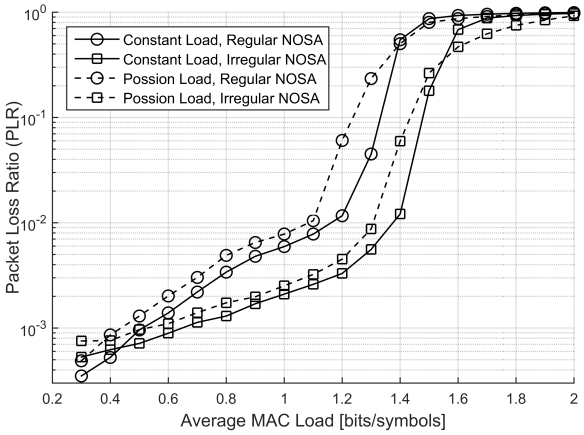


FIGURE 6. Comparison of PLR between regular and irregular NOSA systems under constant and Poisson MAC loads.

in [15]. A TDMA based satellite system is considered in which a frame is composed of $N = 128$ slots. We will investigate both the regular and irregular NOSA prototypes to verify the performance of throughput measured in bits/symbol. The TDMA frame will be divided into $T = 32$ tiles, each consisting of $M = 4$ slots. The normalized MAC load G will vary from 0.1 to 2 bits/symbol to verity the performance in both low and high load regions. Without loss of generality, codebooks that are generated by simply rotating QPSK constellations are utilized throughout the simulation [27], indicating a QPSK transmission of the NOSA system with $Q = 4$, and thus the number of codes in each codebook is given as 4 as well.

The NOSA replicas are totally the same as a traditional TDMA packet except that the information bits are mapped to sparse code elements that correspond to a rotated QPSK constellation. The signal-to-noise ratio (SNR) is defined as E_s/N_0 ranging uniformly in dB from a minimum value $2dB$ to a maximum value of $\{6, 9, 12\}dB$. The carrier phase shift follows uniform distribution in $[0, 2\pi)$, and the maximum timing offset is set as 5 samples after coarse synchronization. The parameter estimation for interference cancellation is assumed to be perfect as proved in [8]. The payload of every terminal is first FEC coded by 3GPP Turbo coding with the code rate of 1/3. The number of inner iterations of an MPD detector is set as 3, and the number of maximum number of connections on a FN is fixed as $d_{\max} = 5$ standing for the maximum calculation complexity the demodulator can afford typically. We will take CRDSA-3 as a reference via a semi-analytical simulator in [14].

B. PERFORMANCE EVALUATION

Fig. 5-6 demonstrate the throughput and packet loss ratio (PLR) performances of both regular and irregular NOSA systems with the SNR uniformly distributed in [2,9]dB. One one hand, it can be seen that under constant MAC loads the throughput that an irregular NOSA system could achieve is 1.384 bits/symbols at $PLR \simeq 1 \times 10^{-2}$. While the throughput of a regular NOSA

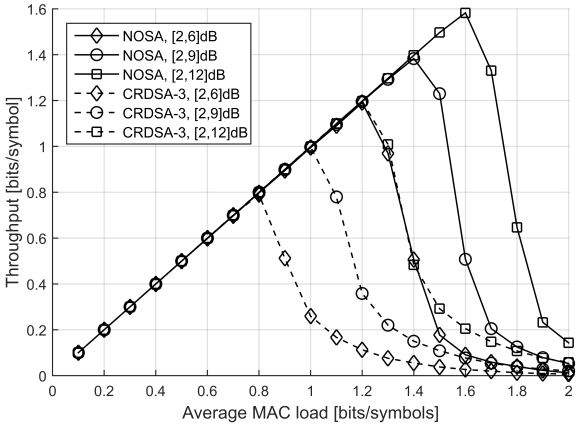


FIGURE 7. Throughput of NOSA and CRDSA-3 with the SNR of packets uniformly distributed in [2,6]dB, [2,9]dB, and [2,12]dB.

system is 1.242 bits/symbols at $PLR \simeq 4.5 \times 10^{-2}$, and 1.186 bits/symbols at $PLR \simeq 1 \times 10^{-2}$. This indicates that the irregular NOSA system outperforms the regular counterpart in most load regions rather than those under 0.4 bits/symbols. In this case, the irregular NOSA system is recommended and utilized in the rest of this section. On the other hand, the performances of NOSA systems under constant and Poisson distributed MAC loads are evaluated as well. It can be found that the irregular NOSA system under Poisson distributed MAC loads could achieve a maximum throughput of 1.316 bits/symbols at $PLR \simeq 6 \times 10^{-2}$ and 1.289 bits/symbols at $PLR \simeq 1 \times 10^{-2}$, respectively. A better performance could be achieved under constant MAC loads rather than Poisson distributed MAC loads due to the reason that the Poisson traffic model is more realistic than the constant model with the presence of varying load peaks. The performance loss of the irregular NOSA system is approximately 6.7% at $PLR \simeq 1 \times 10^{-2}$.

Fig. 7-8 present the comparisons of throughput and PLR performances between an irregular NOSA and CRDSA-3 schemes under different SNR conditions, i.e., the SNR uniformly distributed within [2, 6]dB, [2, 9]dB, and [2, 12]dB, respectively. The constant traffic model is utilized

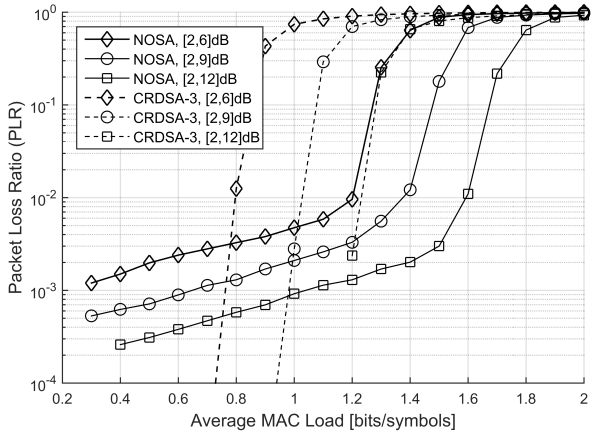


FIGURE 8. PLR of NOSA and CRDSA-3 with the SNR of packets uniformly distributed in [2,6]dB, [2,9]dB, and [2,12]dB.

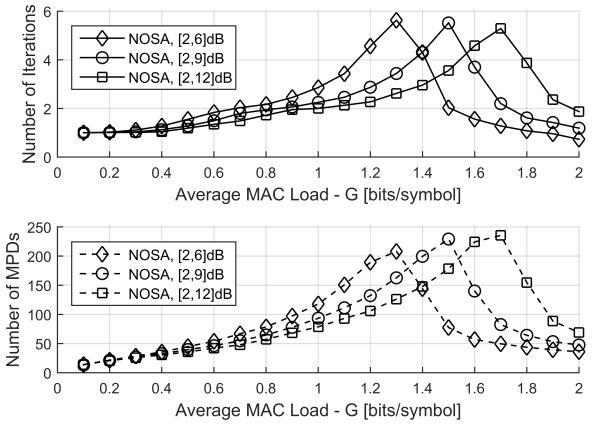


FIGURE 9. The number of iterations and MPDs initiated of an irregular NOSA system with the SNR uniformly distributed in [2,6]dB, [2,9]dB, and [2,12]dB.

in the simulation for both schemes without loss of generality. As for the irregular NOSA scheme, it can be seen that the maximum throughput is 1.2, 1.4, and 1.6 bits/symbols at $PLR \simeq 10^{-2}$ with SNR distributed in [2,6]dB, [2,9]dB, and [2,12]dB, respectively. While the maximum throughput of CRDSA-3 is 0.8, 1.0, and 1.2 bits/symbols at $PLR \simeq 10^{-2}$, respectively. The higher the maximum SNR, the higher throughput both schemes could achieve. Simulation results show that the prototype of the NOSA scheme significantly outperforms the CRDSA scheme at $PLR \simeq 10^{-2}$ in high load regions, providing an efficient alternative to the existing schemes. Unfortunately, the major drawback of the NOSA prototype lies on the PLR performance under 1×10^{-2}, which drops slowly with the decrease in the average MAC load, indicating that the NOSA scheme would have more retransmissions than CRDSA-3 in the low load regions. This may be resulted by the loop phenomenon in the NOSA scheme, i.e., several terminals have randomly selected the same codebook and tile indexes, and PHY replicas of these terminals have pretty low receive power at the demodulator. In this case, there is a chance that all these remaining replicas cannot be successfully detected. This prompts us to further improve the performance of NOSA in both low and high load regions. Alternatively, as the architecture and packet structure of the NOSA are similar to that of the CRDSA, it is also possible to switch NOSA to CRDSA in low load regions to avoid this problem. It is worth emphasizing that the NOSA scheme could work also work well with $PLR < 1 \times 10^{-2}$ yet suffering a little bit higher rate of retransmissions than existing RMA schemes. In addition, the E-SSA scheme is reported in [15] to have comparable achievable throughput to the NOSA scheme in high load regions and better PLR performance in low load regions. However, as the main focus of this paper is to introduce the multi-packet detection into TDMA based satellite systems, the NOSA scheme is proposed as an alternative to existing TDMA based RMA schemes in achieving higher performances in high load regions. Further improvements of the NOSA scheme are expected to fully exploit its potential in achieving even higher performances.

Fig. 9 represents the number of iterations and MPD detections that an irregular NOSA demodulator initiates verses the average MAC load with the SNR uniformly distributed within [2,6]dB, [2,9]dB, and [2,12]dB, respectively. From the upper figure, it can be found that the maximum number of iterations is smaller than 6 at the average MAC load $G = 1.3, 1, 5$ and 1.7 with the SNR distributed within [2,6]dB, [2,9]dB, and [2,12]dB, respectively. The indicates that the number of iterations required arrives its peak value around the average MAC loads corresponding to the peak throughput. The lower figure demonstrates the total number of MPD attempts that a NOSA demodulator initiates during the packet detection. As the complexity of the NOSA demodulator is dominated by the MLD contained in each MPD shown in Section II.D, and the peak values of MPD attempts range from $N_{MPD} = 200$ to 240 times, it means that the peak complexity of the NOSA demodulator is approximately $\mathcal{O}\left(N_{MPD} L_{Pay} Q \cdot 2^{d_{\max}}\right)$. It is seen that the curves of the number of MPD attempts match that of the number of iterations as they both depend on the average MAC load as well as the number of successfully detected packets. It is worth noticing that many low-complexity detectors instead of the MPD for LDS-CDMA/SCMA systems could be simply employed to lower the complexity of the NOSA demodulator such as [26].

IV. CONCLUSION

A framework of NOSA protocol is proposed in this paper to provide an alternative way in realizing the RMA for TDMA based satellite networks. The proposed NOSA protocol is also expected to be properly modified as one possible attempt to enable the LDS-CDMA/SCMA techniques to work in a complete RMA fashion for terrestrial IoT-oriented networks. Further studies are expected to exploit the potential of the NOSA system in terms of the power control, codebooks design, parameters estimation, and joint detection, etc.

REFERENCES

[1] M. de Sanctis, E. Cianca, G. Araniti, I. Bisio, and R. Prasad, "Satellite communications supporting Internet of remote things," *IEEE Internet Things J.*, vol. 3, no. 1, pp. 113–123, Feb. 2016.

[2] M. Hasan, E. Hossain, and D. Niyato, "Random access for machine-to-machine communication in LTE-advanced networks: Issues and approaches," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 86–93, Jun. 2013.
[3] T. Le-Ngoc and J. I. Mohammed, "Combined free/demand assignment multiple access (CFDAMA) protocols for packet satellite communications," in *Proc. 2nd Int. Conf. Pers. Commun.*, vol. 2, Oct. 1993, pp. 824–828.
[4] R. De Gaudenzi, O. Del Rio Herrero, G. Gallinaro, S. Cioni, and P.-D. Arapoglou, "Random access schemes for satellite networks, from VSAT to M2M: A survey," *Int. J. Satell. Commun. Netw.*, vol. 36, no. 1, pp. 66–107, 2018.
[5] G. Choudhury and S. Rappaport, "Diversity ALOHA—A random access scheme for satellite communications," *IEEE Trans. Commun.*, vol. COMM-31, no. 3, pp. 450–457, Mar. 1983.
[6] E. Casini, R. De Gaudenzi, and O. del Rio Herrero, "Contention resolution diversity slotted ALOHA (CRDSA): An enhanced random access scheme for satellite access packet networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1408–1419, Apr. 2007.
[7] R. De Gaudenzi, O. del Rio Herrero, G. Acar, and E. G. Barrabes, "Asynchronous contention resolution diversity ALOHA: Making CRDSA truly asynchronous," *IEEE Trans. Wireless Commun.*, vol. 13, no. 11, pp. 6193–6206, Nov. 2014.
[8] G. Liva, "Graph-based analysis and optimization of contention resolution diversity slotted ALOHA," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 477–487, Feb. 2011.
[9] E. Paolini, G. Liva, and M. Chiani, "High throughput random access via codes on graphs: Coded slotted ALOHA," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2011, pp. 1–6.
[10] E. Paolini, G. Liva, and M. Chiani, "Coded slotted ALOHA: A graph-based method for uncoordinated multiple access," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6815–6832, Dec. 2015.
[11] F. Clazzer, E. Paolini, I. Mambelli, and Č. Stefanović, "Irregular repetition slotted ALOHA over the Rayleigh block fading channel with capture," in *Proc. IEEE Int. Conf. Commun.*, May 2017, pp. 1–6.
[12] O. del Río Herrero and R. De Gaudenzi, "Generalized analytical framework for the performance assessment of slotted random access protocols," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 809–821, Feb. 2014.
[13] O. Del Rio Herrero and R. De Gaudenzi, "High efficiency satellite multiple access scheme for machine-to-machine communications," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 4, pp. 2961–2989, Oct. 2012.
[14] A. Mengali, R. De Gaudenzi, and P.-D. Arapoglou, "Enhancing the physical layer of contention resolution diversity slotted ALOHA," *IEEE Trans. Commun.*, vol. 65, no. 10, pp. 4295–4308, Oct. 2017.
[15] A. Mengali, R. De Gaudenzi, and Č. Stefanović, "On the modeling and performance assessment of random access with SIC," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 2, pp. 292–303, Feb. 2018.
[16] R. Hoshyar, F. P. Wathan, and R. Tafazolli, "Novel low-density signature for synchronous CDMA systems over AWGN channel," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1616–1626, Apr. 2008.
[17] R. Razavi, R. Hoshyar, M. A. Imraxn, and Y. Wang, "Information theoretic analysis of LDS scheme," *IEEE Commun. Lett.*, vol. 15, no. 8, pp. 798–800, Aug. 2011.
[18] S. Zhang, X. Xu, L. Lu, Y. Wu, G. He, and Y. Chen, "Sparse code multiple access: An energy efficient uplink approach for 5G wireless systems," in *Proc. IEEE Globecom Conf.*, Dec. 2014, pp. 4782–4787.
[19] F. Wei and W. Chen, "Low complexity iterative receiver design for sparse code multiple access," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 621–634, Feb. 2017.
[20] L. Lei, C. Yan, G. Wenting, Y. Huilian, W. Yiqun, and X. Shuangshuang, "Prototype for 5G new air interface technology SCMA and performance evaluation," *China commun.*, vol. 12, no. 9, pp. 38–48, Sep. 2015.
[21] J. Zhang *et al.*, "PoC of SCMA-based uplink grant-free transmission in UCNC for 5G," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1353–1362, Jun. 2017.
[22] L. Yu, P. Fan, Z. Ma, X. Lei, and D. Chen, "An optimized design of irregular SCMA codebook based on rotated angles and EXIT chart," in *Proc. IEEE 84th Veh. Technol. Conf.*, Sep. 2016, pp. 1–5.
[23] M. Alam and Q. Zhang, "Designing optimum mother constellation and codebooks for SCMA," in *Proc. IEEE Int. Conf. Commun.*, May 2017, pp. 1–6.
[24] J. Peng, W. Chen, B. Bai, X. Guo, and C. Sun, "Joint optimization of constellation with mapping matrix for SCMA codebook design," *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 264–268, Mar. 2017.
[25] 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.
[26] J. Lian, S. Zhou, X. Zhang, and Y. Wang, "Low complexity decoding method For SCMA in uplink random access," in *Proc. IEEE Globecom Conf.*, Dec. 2016, pp. 1–6.
[27] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "SCMA codebook design," in *Proc. 80th Veh. Technol. Conf.*, Sep. 2014, pp. 1–5.

QIWEI WANG was born in Zhengzhou, Henan, China, in 1987. He received the B.S. degree in communication engineering from the Harbin University of Science and Technology, Harbin, China, in 2009, and the Ph.D. degree in communication and information systems from Xidian University, Xi'an, China, in 2015.

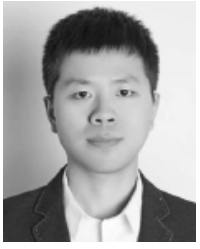
He is currently a Lecturer with Xidian University. His research interests include wireless communication and signal processing, especially the random multiple access process of massive terminals and multi-user detection for M2M networks.

GUANGLIANG REN (M'06) was born in Xuzhou, Jiangsu, China, in 1971. He received the B.S. degree in communication engineering from Xidian University, Xi'an, China, in 1993, the M.S. degree in signal processing from the Academy of China Ordnance, Beijing, China, in 1996, and the Ph.D. degree in communication and information systems from Xidian University in 2006.

He is currently a Professor with the School of Telecommunications Engineering, Xidian University. He has authored more than 40 research papers in journals and conference proceedings, such as the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and authored or co-authored three books. His research interests include wireless communication and digital signal processing, particularly multiple-input-multiple-output systems, WiMax, and LTE.

STEVEN GAO (M'01–SM'16) is currently a Professor and the Chair of RF and microwave engineering with the University of Kent, U.K. He has authored two books, including *Space Antenna Handbook* (Wiley, 2012) and *Circularly Polarized Antennas* (IEEE and Wiley, 2014), over 200 papers. He holds several patents. His research covers smart antennas, phased arrays, MIMO, satellite antennas, RF/microwave/mm-wave circuits, satellite communications, UWB radars, synthetic-aperture radars, and mobile communications.

Dr. Gao was the General Chair of LAPC 2013 and a Keynote Speaker or an Invited Speaker at some international conferences, such as AES'2014 in China, IWAT'2014 in Sydney, SOMIRES'2013 in Japan, and APCAP'2014 in China. He is an IEEE AP-S Distinguished Lecturer, an Associate Editor of the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, an Associate Editor of *Radio Science*, and the Editor-in-Chief for Wiley Book Series on Microwave and Wireless Technologies.



KUN WU was born in Xuzhou, Jiangsu, China, in 1989. He received the B.S. degree in communication engineering and the Ph.D. degree in communications and information systems from Xidian University, Xi'an, China, in 2011 and 2017, respectively.

He is currently an Engineer with Huawei Co., Ltd. His current research interests include key techniques for uplink multiuser data transmission in single-carrier frequency-division multiple access (SC-FDMA)-based communications systems, including channel estimation in LTE uplink, PAPR reduction of SC-FDMA signals, and multiuser detection.