Sparse Code Multiple Access with Enhanced K-Repetition Scheme: Analysis and Design

Ke Lai, Zilong Liu *Senior Member, IEEE*, Jing Lei, Gaojie Chen *Senior Member, IEEE*, and Pei Xiao *Senior Member, IEEE*, Lei Wen

***Abstract*—This work presents a novel K-Repetition based Hybrid Automatic Repeat reQuest (HARQ) scheme for uplink sparse code multiple access (SCMA) systems. Our core idea is to apply network coding (NC) principle to re-encode different packets (after channel coding and interleaving) or their fragments, where K-Repetition is an emerging HARQ technique (recommended in 3GPP Release 15) for enhanced reception in future massive machine-type communications. Such a proposed scheme is referred to as the NC aided K-repetition SCMA (NCK-SCMA) in this paper. We aim to understand the optimal NCK-SCMA design criteria for maximizing the channel diversity as well as the efficient receiver processing for superior error rate performances. It is found that NC can enable a larger diversity order for NCK-SCMA with fewer resources (i.e., higher spectrum efficiency). Toward this objective, some novel design criteria are developed for the efficient configuration of NCK-SCMA. Moreover, we propose an iterative network decoding and SCMA detection (INDSD) algorithm for robust and low-complexity recovery of the transmit data from a low-density parity-check (LDPC) coded uplink NCK-SCMA system. Simulation results demonstrate that the proposed NCK-SCMA lead to higher throughput and improved reliability over the conventional K-SCMA.**

***Index Terms*—Sparse Code Multiple Access (SCMA), K-Repetition, Network Coding, LDPC, Pairwise Error Probability (PEP), Diversity Order, Iterative Detection.**

I. Introduction

A. Background

Wireless communication systems are evolving towards providing efficient and reliable data services over a massive number of user equipment (UEs), which is referred to as machine-type communications (MTC) [1]. An emerging paradigm for MTC is non-orthogonal multiple access (NOMA) which enables higher spectral efficiency by overloading multiple users upon finite resource nodes. In this work, we focus on sparse code multiple access (SCMA), which is a promising code-domain NOMA (CD-NOMA) scheme for the next generation MTC systems [2], [3]. Moreover, in order to correct as many data errors as possible, whilst at the same time, attaining faster communication speed, Hybrid Automatic Repeat reQuest (HARQ) is widely deployed in modern wireless networks [4]. However, conventional HARQ allows for retransmissions only upon reception of a Negative ACKnowledgement (NACK). This could lead to large handshaking latency as it requires the base station (BS) to first receive the packet for detection, then issue the feedback. As such, it is desirable to have an enhanced HARQ that can provide reduced latency and improved reliability for future MTC networks. Such an enhanced HARQ scheme is expected to enable grant-free (GF) transmission (or semi-GF transmission) in view of the increasingly stringent latency requirements in the next generation wireless networks.

K. Lai, J. Lei, L. Wen are with Department of Communication Engineering, College of Electronic Science and Engineering, National University of Defence technology. (e-mail: laike12@nudt.edu.cn; leijing@nudt.edu.cn; newton1108@126.com).

Zilong Liu is with School of Computer Science and Electronics Engineering, University of Essex, UK (e-mail: zilong.liu@essex.ac.uk).

G. Chen is School of Flexible Electronics (SoFE) State Key Laboratory of Optoelectronic Materials and Technologies, Sun Yat-sen University, Guangdong, China. (e-mail: gaojie.chen@ieee.org).

P. Xiao is with State Key Laboratory of ISN, Xidian University, Xi'an, 710071, China. (email: pxiao2022@163.com).

B. Related works

Recently, K-Repetition has been introduced in 3GPP Release 15 as a novel HARQ scheme towards GF random access [5]. In K-Repetition, a few consecutive replicas of the same packet are transmitted without waiting for the feedback, and with the additional benefit of enlarged diversity order (DO). System-level simulations capturing the main performance influencing factors of K-repetition are reported in [6]. An analytical expression of the success probability of K-Repetition is derived in [7] from a stochastic geometry perspective. Moreover, as reported in [8], random network coding (NC) has been employed in the K-Repetition scheme to generate distinguished linear combinations of preambles to enhance transmission reliability. This idea was soon extended to sliding NC in [9] with a reduced decoding latency and a high reliability.

To improve throughput and reduce the number of retransmission rounds, network coded (NCed) HARQ has proven to be an efficient means. A theoretical analysis of NCed HARQ schemes for multicast and multiple unicast has been reported in [10] and [11], respectively. For practical implementation of NCed HARQ, an NC based retransmission strategy that combines two incorrectly detected packets is proposed for both HARQ with chase combining (HARQ-CC) and HARQ with incremental redundancy (HARQ-IR) [12], [13]. Moreover, the authors in [14] considered an NC-Turbo joint coded HARQ protocol in a wireless broadcasting system. The extension of NCed HARQ to satellite Internet of things to achieve higher energy efficiency is presented in [15].

From the SCMA standpoint, although numerous research attempts have been made for enhancing its error performance in recent years [16]–[18], very few works are known on SCMA with retransmissions. In [17], a blanking based HARQ scheme for SCMA is proposed, while in [19], the performance of multi-packet HARQ with blankings is analyzed for both power-domain NOMA (PD-NOMA) and SCMA[1]. However, these two works are tailored for the conventional HARQ, such as HARQ-CC and HARQ-IR based NOMA.

C. Motivations and Contributions

Based on the above introduction, one can see that there are several drawbacks in the current state-of-the-art: 1) The conventional NCed HARQ mainly focuses on the scenario of point-to-point or downlink transmissions but with no straightforward application to multi-user uplink channel; 2) The employment of NC in the K-Repetition scheme of [8] intends to improve error rate performance through resolving preamble collision, however, the NC on data transmission is missing; and 3) Albeit numerous works have investigated PD-NOMA with HARQ [20], [21], there is a clear paucity on CD-NOMA with the conventional HARQ, let alone the combination with advanced HARQ like K-Repetition.

This work aims to address the above drawbacks to deliver *reliable*, *rapid*, and *massive connectivity* in future MTC networks. The main contributions of this article are summarized as follows:

- We propose a novel NC assisted K-repetition scheme for efficient integration with SCMA. The resultant system is thus referred to as NCK-SCMA throughout this paper. The core idea of NCK-SCMA is to re-encode multiple packets (after channel coding and interleaving) or their fragments with NC before SCMA mapping is applied. We show that such an inter-packet coding scheme can bring in extra diversity gain in comparison to the conventional K-SCMA.
- The DO and asymptotic average symbol error probability (ASEP) of both uncoded NCK-SCMA and K-SCMA are analyzed with the aid of pairwise error probability (PEP). Based on this analysis, we show that the utilization of NCed packets can enable NCK-SCMA to achieve a larger DO than the conventional K-SCMA with the same or even fewer resource nodes. To guide the system development, some important design criteria and properties of NCK-SCMA are also presented.
- Finally, we study an LDPC coded NCK-SCMA and develop an iterative network decoding and SCMA detection (INDSD) algorithm with a novel soft combining approach building upon the message passing principle. Our key innovation is that information in the failed transmission rounds can be fully utilized and as a result, the throughput of NCK-SCMA gets improved with fewer HARQ retransmissions and resources, especially in the moderate to high SNR regions.

Notations: The operation $[\cdot]^T$ denotes transpose of a matrix. $\oplus$ represents the XOR operation. $\mathbb{C}$ and $\Re\{\cdot\}$ denote the complex field and the real part of a complex number, respectively. $\lceil\cdot\rceil$ and $|\cdot|$ are the ceil function and 1-norm respectively. $A\backslash a$ indicates all the elements in set A except a. $\mathrm{LLR}(\cdot)$ represents the logorithm likelihood ratio (LLR).

II. System model

In this section, we first introduce the traditional K-Repetition based SCMA system model followed by the basic principles of our proposed NCK-SCMA.

A. K-Repetition SCMA

Let us consider a single-cell network where each uplink UE transmits data to their serving BS using SCMA. Assume that J layers/users transmit over K allocated resource elements ($J > K$). An SCMA codeword for the jth layer/user can be represented as a K-dimensional complex vector $\mathbf{x}_j = [x_{j,1}, \cdots, x_{j,K}]^T$, where $\mathbf{x}_j \in \mathbb{C}^{K\times 1}$, $j \in \{1, 2, \cdots, J\}$ is selected from a pre-defined codebook $\mathcal{X}_j$ with cardinality of $|\mathcal{X}_j| = M = 2^b$, meaning that b bits are mapped to an SCMA codeword. As such, the encoder of SCMA can be defined as a mapping function from $b = \log_2(M)$ bits to a K-dimensional codeword $\mathbf{x}_j$. Note that $\mathbf{x}_j$ has a sparse structure, where the number of non-zero elements d_μ, i.e., the effective number of function nodes (FNs) occupied by each user in a codeword, are determined by a $K \times J$ sparse indicator matrix[2] that characterizes an SCMA system [3].

In K-SCMA, K_{rep} consecutive replicas of the same packet are transmitted without waiting for feedback. Fig. 1 presents an example of K-SCMA with $K_{rep} = 2$. It can be seen from the figure that the instantaneous uncoded packet message of the jth UE is $\mathbf{Pa}_j$. After channel coding and bit-level interleaving, each coded packet $\mathbf{P}_j$ consists of N bits. $K_{rep} = 2$ replicas are to be sent out in consecutive transmission time intervals for SCMA encoding[3]. After SCMA mapping, each packet contains $L = \lceil \frac{N}{b} \rceil$ SCMA codewords[4], and thus one sends LK_{rep} codewords to the wireless channel. Therefore, for each K-SCMA codeword, the received column vector $\mathbf{y}$ which has the dimension of $U = KK_{rep}$ can be expressed as:

$$\begin{aligned}\mathbf{y} &= \sum_{j=1}^{J} \mathrm{diag}(\mathbf{h}_j) \underbrace{\left[\mathbf{x}_j^T, \mathbf{x}_j^T, \cdots, \mathbf{x}_j^T\right]^T}_{K_{rep}} + \mathbf{n} \\ &= \sum_{j=1}^{J} \mathrm{diag}(\mathbf{h}_j) \mathbf{X}_j + \mathbf{n},\end{aligned} \tag{1}$$

where $\mathbf{X}_j = [X_{j,1}, X_{j,2}, \cdots, X_{j,U}]^T \in \mathbb{C}^{U\times 1}$, $\mathbf{h}_j = [h_{j,1}, h_{j,2}, \cdots, h_{j,U}]^T \in \mathbb{C}^{U\times 1}$ and $\mathbf{n} = [n_1, n_2, \cdots, n_U]^T \in \mathbb{C}^{U\times 1}$ are the concatenation of K_{rep} repetitions of $\mathbf{x}_j$, the

[1]HARQ with blanking enforces the UEs without detection errors to be silent, i.e., transmitting no information in the retransmissions, leading to reduced spectral efficiency.

[2]We use "$K \times J$" to depict the indicator matrix used in both NCK-SCMA and K-SCMA in the sequel of this paper.

[3]As depicted in Fig. 1, it should be noted that every repetition of uncoded packet $\mathbf{Pa}_j$ undergoes an independent transmission line of channel coding, interleaving, and SCMA encoding, before combining together for radio propagation in the wireless channel. Furthermore, in our system model, it is assumed that the same interleaver is used for these different transmission lines.

[4]For simplicity, it is assumed that L is an integer, and N is an even number in this paper. Even if N may be an odd number in practical implementation, we can make it an even number by adding zeros.

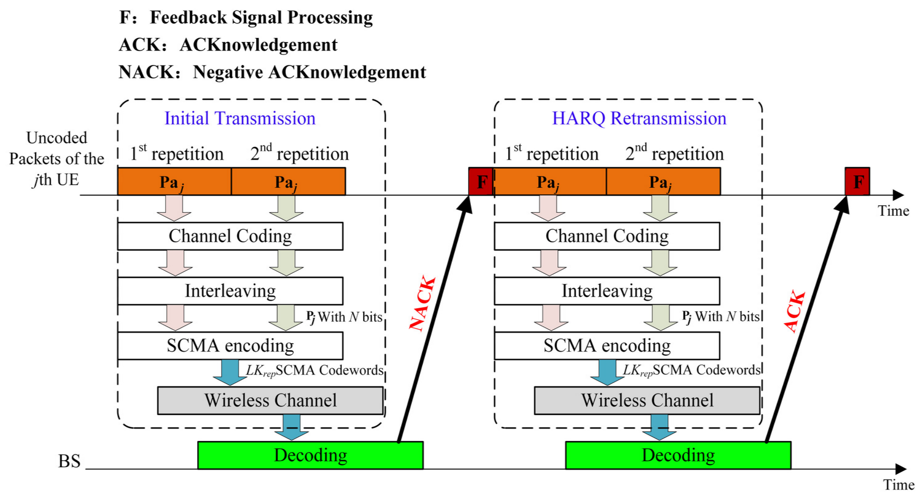

Fig. 1: Illustration of K-SCMA with $K_{rep} = 2$.

channel fading coefficient vector associated to the jth UE, and the additive white Gaussian noise vector, respectively. Note that $h_{j,u} \sim \mathcal{CN}(0,1)$, and $n_u \sim \mathcal{CN}(0, N_0)$, where $u \in \{1, 2, \cdots, U\}$. By defining $\mathbf{h} \in \mathbb{C}^{1\times JU}$:

$$\mathbf{h} = \left[\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \cdots, \tilde{\mathbf{h}}_U\right] = [h_{1,1}, \cdots, h_{J,1}, \cdots, h_{1,U}, \cdots, h_{J,U}], \tag{2}$$

where $\tilde{\mathbf{h}}_u = [h_{1,u}, h_{2,u}, \cdots, h_{J,u}]$, and

$$\mathbf{X}_U = \begin{bmatrix} \mathbf{C}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{C}_U \end{bmatrix}_{JU\times U}, \tag{3}$$

where $\mathbf{C}_u = [X_{1,u}, X_{2,u}, \cdots, X_{J,u}]^T \in \mathbb{C}^{J\times 1}$, (2) can be rewritten in a compact form as:

$$\mathbf{y}^T = \mathbf{h}\mathbf{X}_U + \mathbf{n}^T. \tag{4}$$

At the receiver, decoding is performed by using the message passing algorithm (MPA) and soft combining after receiving K_{rep} repetitions. Note that if the detected packet $\hat{\mathbf{P}}_j$ fails to pass the cyclic redundancy check (CRC), a retransmission should be conducted as NACK is received from HARQ feedback.

B. Proposed NC based K-Repetition SCMA (NCK-SCMA)

Different from K-SCMA which simply transmits K_{rep} replicas, NCK-SCMA leverages NC to combine T_p transmitted packets (after channel coding and interleaving) based on XOR operation to enhance the error rate performance and to better utilize physical resources. To proceed, let us define R_{in} and R_{nc} as the numbers of initial repetitions and NCed packets repetitions, respectively. It is worth noting that NCK-SCMA is determined by R_{in}, R_{nc}, and T_p simultaneously; hence, the design of NCK-SCMA can be very flexible. To ease understanding, three different types of NCK-SCMA to achieve the same equivalent value of repetitions, denoted by K_{eq} are introduced in Fig. 2. For brevity, $K_{eq} = 3$ is considered in the example.

As for Type-A NCK-SCMA in Fig. 2(a), after transmiting $T_p = 2$ packets $\mathbf{P}_{j,1}$ and $\mathbf{P}_{j,2}$ with $R_{in} = 1$ in the first round-trip time (RTT), $R_{nc} = 2$ replicas of NCed packets, i.e., $\mathbf{P}_{j,1} \oplus \mathbf{P}_{j,2}$ along with the feedback signal are transmitted in the next RTT. In this case, it is equivalent to send both $\mathbf{P}_{j,1}$ and $\mathbf{P}_{j,2}$ three times, and thus $K_{eq} = 3$. Fig. 2(b) demonstrates the Type-B NCK-SCMA that obtains the same K_{eq} through changing R_{nc} based on Type-A NCK-SCMA. It is shown that the NCed packet $\mathbf{P}_{j,1} \oplus \mathbf{P}_{j,2}$ packet is sent with $R_{in} = 2$ initial repetitions. To elaborate further, Fig. 2(c) plots Type-C NCK-SCMA. As can be observed from the figure, $T_p = 3$ packets are to be transmitted, which indicates Type-C NCK-SCMA is obtained by using different T_p. Let us denote W as the total number of NCed combinations of the T_p packets. Since XOR operations are performed in pairs, we have $W = 3$ in this example. As such, $K_{eq} = 3$ can be achieved with $R_{in} = R_{nc} = 1$. Let N_R be the number of resources for a complete NCK-SCMA or K-SCMA transmission with the same number of transmitted packets T_p, neglecting the retransmissions. From above discussions, one can see that NCK-SCMA with the same K_{eq} consumes various number of resources, i.e., N_R, which may lead to diverse energy efficiency and error rate performance. Because of this, NCK-SCMA should be carefully designed. Note that the examples above mainly focus on the circumstances when $K_{eq} = 3$, once K_{eq} varies, different types of NCK-SCMA can be obtained.

In the following, we consider a generalized version of NCK-SCMA, and give a general representation of NCK-SCMA. For an NCK-SCMA that intends to transmit T_p packets, after generating the same initial repetitions R_{in} of each packet $\mathbf{P}_{j,t}$, where $t \in \{1, 2, \cdots, T_p\}$, NC is performed to combine T_p packets in pairs. Assuming further that $t_\alpha, t_\beta \in \{1, 2, \cdots, T_p\}$ are the indices of two different packets before performing NC, and let Ξ be the set of all the possible $\{\alpha, \beta\}, (\beta > \alpha)$, such that all the NCed packets can be uniformly represented as $\mathbf{P}_{j,t_\alpha} \oplus \mathbf{P}_{j,t_\beta}$. Note that the cardinality of Ξ is $|\Xi| = W$, where

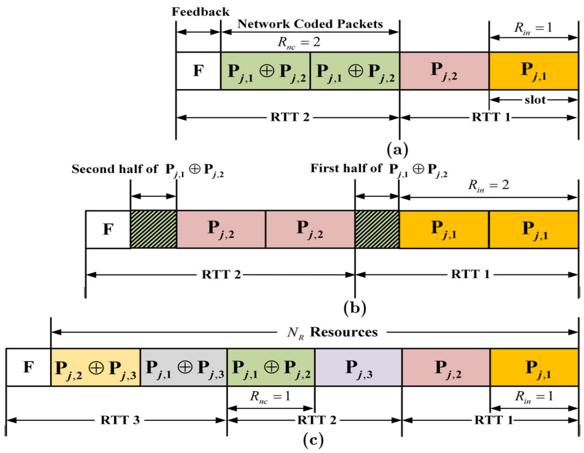

Fig. 2: Illustration of NCK-SCMA with $K_{eq}=3$. (a) Type-A with $T_p=2, R_{in}=1$; (b) Type-B with $T_p=2, R_{in}=2$; (c) Type-C with $T_p=3, R_{in}=1$.

$W=C(T_p,2)$, and $C(\cdot,\cdot)$ denotes the binomial coefficient [5]. Therefore, the equivalent repetitions K_{eq} in NCK-SCMA can be expressed as:

$$K_{eq}=R_{in}+\left(T_p-1\right)R_{nc}=R_{in}+\frac{2\left(N_R-T_pR_{in}\right)}{T_p}. \quad (5)$$

According to the above definitions, it can be inferred that K-SCMA is a special case of NCK-SCMA by defining $R_{in}=K_{rep}, T_p=1$ and $R_{nc}=0$. Therefore, to transmit T_p different packets in NCK-SCMA manner without retransmissions, N_R times transmission is needed:

$$N_R=T_pR_{in}+WR_{nc}, \quad (6)$$

From the above discussions, it is clear that we can use R_{in}, R_{nc}, and T_p to uniquely characterize an NCK-SCMA system. From hereon, we will use (R_{in},R_{nc},T_p) as a parameter group which is associated to an NCK-SCMA. The advantage of NCK-SCMA can be intuitively interpreted from the above discussions, i.e., a higher K_{eq} can be achieved compared to K-SCMA under the condition that approximately the same N_R are utilized. After mapping messages of the jth UE to codewords selected from codebook $\mathcal{X}_j$, the NCK-SCMA signals are superimposed at the receiver, and the receiver combines each retransmission with previous packets until the maximum N_{re} retransmission rounds are met on which packet failure is declared. Different from the works in [17], [19], NCK-SCMA allows to transmit new packets and retransmit unsuccessfully decoded data simultaneously.

III. Analysis of NCK-SCMA

To reveal the performance gain of NCK-SCMA over K-SCMA, we investigate their DOs and error rate properties with an emphasis on uplink Rayleigh fading channels.

[5]Without loss of generality, we consider the generic case when all transmitted packets are coded to exploit the performance gain of the NCK-SCMA. However, not all NCed packets are necessarily used in practical implementation due to resource constraints or different performance requirements. On the other hand, T_p does not need to be high in practical systems as higher DO can be achieved by NCK-SCMA when T_p takes a small value, and higher T_p may lead to more resource consumptions.

A. Analysis of Diversity Order (DO)

For simplicity, let us assume that ideal channel state information (CSI) is available at the receiver. Moreover, as the CSI is considered as identical Rayleigh distribution, the PEP events within different SCMA codewords are identical. As a result, it is sufficient to investigate the PEP events by assuming $N=b, L=1$ and $N=bT_p, L=T_p$ for K-SCMA and NCK-SCMA, respectively, in the following analysis.

As reported in [22], the asymptotic unconditional PEP (UPEP) of SCMA under maximum likelihood detection (MLD) can be expressed as:

$$\Pr\left\{\mathbf{x}\to\hat{\mathbf{x}}\right\}\approx\frac{1}{12}\prod_{k=1}^{K}\frac{1}{1+\frac{\Delta_k^2}{4N_0}}+\frac{1}{6}\prod_{k=1}^{K}\frac{1}{1+\frac{\Delta_k^2}{3N_0}}, \quad (7)$$

where $\Delta_k\triangleq\sqrt{\sum_{j=1}^{J}|x_{j,k}-\hat{x}_{j,k}|^2}$, and $|x_{j,k}-\hat{x}_{j,k}|$ is the element-wise distance between the jth SCMA codeword $\mathbf{x}_j=\left[x_{j,1},x_{j,2},\cdots,x_{j,K}\right]^T$ and the corresponding erroneously detected codeword $\hat{\mathbf{x}}_j=\left[\hat{x}_{j,1},\hat{x}_{j,2},\cdots,\hat{x}_{j,K}\right]^T$, $k\in\{1,2,\cdots,K\}$.

It is noted that in K-SCMA, the received signals from multiple replicas should be properly combined. Among several combining schemes which are proposed in [23], [24], distance level combining (DLC) is deemed to be optimal. To analyze K-SCMA and NCK-SCMA with UPEP given in (7), we first consider the conditional PEP (CPEP) of K-SCMA under DLC, which makes a decision by summing the Euclidean distances for all possible transmit signal estimation $\hat{\mathbf{X}}$ over all repetitions [24]. According to (4), we have where $\mathbf{y}_r$, $\mathbf{n}_r\in\mathbb{C}^{1\times K}$, $\mathbf{h}_r\in\mathbb{C}^{1\times JK}$, and $\mathbf{x}\in\mathbb{C}^{JK\times K}$ are the received signal, channel gain coefficient vector, noise vector and transmit signal matrix of the rth repetition. $[\cdot]^\dagger$ denotes the conjugate transpose of a vector. Moreover, $\mathbf{y}_r^T$, $\mathbf{h}_r$, and $\mathbf{n}_r^T$ are the sub-vector of $\mathbf{y}$, $\mathbf{h}$, and $\mathbf{n}$, respectively, defined in Subsection II-A by taking elements from $u=(r-1)K+1$ to $u=rK$. $\mathbf{X}_{U,e}=\mathbf{X}_U-\hat{\mathbf{X}}_U$ $(\mathbf{X}_U\neq\hat{\mathbf{X}}_U)$, and $\hat{\mathbf{X}}_U$ is the erroneously detected codeword corresponding to $\mathbf{X}_U$. Recalling that we have defined $U=KK_{rep}$ in Subsection II-A. It can be seen that the last step of (8) is the CPEP of K-SCMA with MLD. Since the error rate is determined by CPEP, we can conclude that DLC and MLD have the same decoding performance.

Eq. (8) indicates that the K-SCMA is equivalent to mapping b bits symbol to a U-dimensional codeword space by repeating the original K dimensional SCMA codeword K_{rep} times. Therefore, we assert that the DO[6] of K-SCMA is $d_\mu K_{rep}$ by noting that

$$\mathrm{DO}=\min_{\mathbf{X}_U\neq\hat{\mathbf{X}}_U}\operatorname{rank}\left(\mathbf{X}_{U,e}\right), \quad (9)$$

and there are d_μ different non-zero elements between every pair of SCMA codewords [28]. As for NCK-SCMA, the

[6]It is noted that DO is an important concept in communication theory. For example, it has been widely used in the study of space-time coding where DO arises due to the use of multiple antennas. Formally, we have DO = $\lim_{\mathrm{SNR}\to\infty}-\frac{\log\mathrm{BER}}{\log\mathrm{SNR}}$ [25]. That is, DO measures the effective number of independent paths/channels over which the data is received. In the context of K-SCMA and the proposed NCK-SCMA, the asymptotic DO refers to the minimum rank of difference matrix between the transmitted and detected codeword matrices [16], [26], [27].

$$\begin{aligned}\Pr\left(\mathbf{X}_U \to \hat{\mathbf{X}}_U \mid \mathbf{h}\right) &= \Pr\left(\sum_{r=1}^{K_{rep}} \left\|\mathbf{y}_r - \mathbf{h}_r\mathbf{x}\right\|^2 > \sum_{r=1}^{K_{rep}} \left\|\mathbf{y}_r - \mathbf{h}_r\hat{\mathbf{x}}\right\|^2\right) \\ &= \Pr\left(2\sum_{r=1}^{K_{rep}} \Re\left\{\mathbf{h}_r\left(\mathbf{x}-\hat{\mathbf{x}}\right)\mathbf{n}_r^{\dagger}\right\} > \sum_{r=1}^{K_{rep}} \left\|\mathbf{h}_r\left(\mathbf{x}-\hat{\mathbf{x}}\right)\right\|^2\right) \\ &= \Pr\left(2\Re\left\{\mathbf{n}^{\dagger}\left(\mathbf{h}\mathbf{X}_{U,e}\right)^T\right\} > \left\|\mathbf{h}\mathbf{X}_{U,e}\right\|^2\right),\end{aligned} \tag{8}$$

T_p different packets are encoded with NC, and thus the PEP analysis of NCK-SCMA is more complicated. To ease understanding, in the sequel, we provide an example of NCK-SCMA with $R_{in} = 1$ and $R_{nc} = 1$ as follows, because all NCK-SCMA schemes with fixed T_p can be derived from this case.

Example 1: Let us consider the NCK-SCMA with $(R_{in}, R_{nc}, T_p) = (1, 1, 2)$ as an instance, i.e., trasnmitting $T_p = 2$ packets with $R_{in} = 1, R_{nc} = 1$, and the same settings for K-SCMA that $N = b, L = 1$ in each packet are made for simplicity. Note that if the transmitted bits of $\mathbf{P}_{j,1}$ and $\mathbf{P}_{j,2}$ are determined, then the NCed packets can be directly obtained through $\mathbf{P}_{j,1} \oplus \mathbf{P}_{j,2}$, indicating that the NCed packets do not carry any new information but help to detect $\mathbf{P}_{j,1}$ and $\mathbf{P}_{j,2}$. In this case, the DO is $2d_\mu$, which can be obtained when $\mathbf{P}_{j,1} \oplus \mathbf{P}_{j,2} = \hat{\mathbf{P}}_{j,1} \oplus \hat{\mathbf{P}}_{j,2}$, where $\hat{\mathbf{P}}_{j,1}$ and $\hat{\mathbf{P}}_{j,2}$ are the incorrectly detected packets[7].

It can be inferred from this example that NCK-SCMA with $(R_{in}, R_{nc}, T_p) = (1, 1, 2)$ can achieve the same DO as the K-SCMA with $K_{rep} = 2$, whilst reaching the same repetitions with fewer resource consumptions. As such, NCK-SCMA is more spectral efficient compared to K-SCMA. In general, NCK-SCMA with (R_{in}, R_{nc}, T_p) is equivalent to map $\left[\mathcal{P}_1\mathcal{P}_2 \cdots \mathcal{P}_{T_p}\right]$ with parameters R_{in} and R_{nc} to a codeword matrix $\mathbf{X}_V$, where $\mathcal{P}_t = [\mathbf{P}_{1,t}, \cdots, \mathbf{P}_{J,t}]$. Therefore, $\mathbf{X}_V \in \mathbb{C}^{JV \times V}$ can be represented in a form similar to (3), where $V = N_R K$, and $\mathbf{C}_v = [X_{1,v}, X_{2,v}, \cdots, X_{J,v}]^T \in \mathbb{C}^{J\times 1}$, $v = \{1, 2, \cdots, V\}$. Note that $\mathbf{C}_v$ is the concatenation of T_p different packets and W NCed packets with R_{in} and R_{nc} repetitions, respectively. Likewise, to obtain the DO of NCK-SCMA, the difference matrix between $\mathbf{X}_V$ and the corresponding erroneously detected codeword matrix $\hat{\mathbf{X}}_V$, i.e., $\mathbf{X}_{V,e}$, should be considered. To compare the DO of NCK-SCMA and K-SCMA for a generic setting, we provide the following proposition:

Proposition: 1: For a NCK-SCMA with (R_{in}, R_{nc}, T_p), the UPEP can be expressed as

$$\Pr\left\{\mathbf{X}_V \to \hat{\mathbf{X}}_V\right\} \simeq \Gamma(G_d, N_0) \prod_{v=1}^{G_d} \Delta_{j,v}^{-2}, \tag{10}$$

where $\Gamma(x, y) = (1/y)^{-x} \cdot (4^{-x}/12 + 3^{-x}/6)$, $\Delta_{j,v} = \sqrt{|X_{j,v} - \hat{X}_{j,v}|^2}$, and

$$G_d = d_\mu \cdot \min\left(R_{in}T_p, K_{eq}\right). \tag{11}$$

Proof: See Appendix A. ■

Remark 1: To achieve higher DO in NCK-SCMA, the goal is to maximize $\min\left(R_{in}T_p, K_{eq}\right)$. According to (5), K_{eq} can be further written as $K_{eq} = 2N_R/T_p - R_{in}$. Let us assume $G_d = d_\mu K_{eq}$ under a certain parameter configuration. To increase G_d in this case, it is desirable to enlarge N_R or decrease $R_{in}T_p$. However, larger N_R implies more resource consumption, while the DO may get reduced since G_d will be equal to $d_\mu R_{in}T_p$ as the decrease of $R_{in}T_p$. Therefore, a judicious design between R_{in}, R_{nc}, T_p is needed for a larger DO of NCK-SCMA.

B. Key properties of NCK-SCMA

From the analysis in *Proposition 1*, the natural questions are 1) how NCK-SCMA can achieve a larger DO and 2) what is its largest DO? In this subsection, we discuss these two questions under different N_R configurations. To proceed, the following proposition is first presented:

Proposition 2: Let us assume that NCK-SCMA and K-SCMA communicate over the same number of resources, i.e., $N_R = K_{rep}T_p$. The DO of NCK-SCMA can be higher than K-SCMA under the condition that

$$\frac{R_{nc}}{2} < R_{in} \le R_{nc}, \tag{12}$$

when $G_d = d_\mu R_{in}T_p$, or

$$R_{nc} \le R_{in}, \tag{13}$$

when $G_d = d_\mu K_{eq}$, where the maximum value of G_d in this case is:

$$G_{d,max} = d_\mu\left(2K_{rep} - \left\lceil\frac{2K_{rep}}{T_p + 1}\right\rceil\right). \tag{14}$$

The equality in (12) and (13) are both achieved if and only if $R_{in}T_p = K_{eq}$.

Proof: We consider the following three cases:

(i) $G_d = d_\mu R_{in}T_p$. In this case, $R_{in}T_p < R_{in} + (T_p - 1)R_{nc}$, such that $R_{in} < R_{nc}$. To achieve a higher DO,

[7]This conclusion can be obtained following the facts that: 1) the minimum rank of difference matrix between transmitted and estimated codeword matrices can be only obtained when there is only one user in error; 2) provided that one of the detected packets contains the same bits as the transmitted packet, the minimum rank can be obtained; 3) each SCMA codeword contains d_μ non-zero elements. As the symmetry of XOR operation, when $\mathbf{P}_{j,1} \oplus (\mathbf{P}_{j,1} \oplus \mathbf{P}_{j,2}) = \hat{\mathbf{P}}_{j,1} \oplus \left(\hat{\mathbf{P}}_{j,1} \oplus \hat{\mathbf{P}}_{j,2}\right)$, and $\mathbf{P}_{j,2} \oplus (\mathbf{P}_{j,1} \oplus \mathbf{P}_{j,2}) = \hat{\mathbf{P}}_{j,2} \oplus \left(\hat{\mathbf{P}}_{j,1} \oplus \hat{\mathbf{P}}_{j,2}\right)$, the minimum rank of $\mathbf{X}_{V,e}$ can be also obtained.

$R_{in}T_p > K_{rep}$ should be satisfied. As the same number of resources are utilized, $R_{in}T_p > \frac{N_R}{T_p}$ holds. After substituting (6) into the inequality, and with some algebra manipulations, $R_{in} > \frac{R_{nc}}{2}$ can be obtained.

(ii) $G_d = d_\mu K_{eq}$. This case happens when $R_{in}T_p > R_{in} + (T_p - 1)R_{nc}$, we have

$$R_{in} + (T_p - 1)R_{nc} > \frac{N_R}{T_p}. \tag{15}$$

By using the same steps in the proof of (12), the inequality in (15) holds under the condition that $R_{nc} < R_{in}$.

(iii) $G_d = d_\mu K_{eq} = d_\mu R_{in}T_p$. It can be inferred that $R_{in}T_p = R_{in} + (T_p - 1)R_{nc}$ in this case, as $T_p \neq 1$, $R_{in} = R_{nc}$ holds. By combining the results in cases (i) and (ii), (12) and (13) can be proved.

Next, we calculate $G_{d,max}$ under the condition that $G_d = d_\mu K_{eq}$. As NCK-SCMA and K-SCMA share the same N_R in the proposition, by substituting $N_R = K_{rep}T_p$ into (5), K_{eq} can be expressed as:

$$K_{eq} = 2K_{rep} - R_{in}. \tag{16}$$

As $G_d = d_\mu K_{eq}$ in this case, $2K_{rep} - R_{in} \leq R_{in}T_p$ holds so that $R_{in} \geq \frac{2K_{rep}}{T_p+1}$. According to (16), $G_{d,max}$ can be obtained when R_{in} takes the minimum value, i.e., $R_{in} = \frac{2K_{rep}}{T_p+1}$, considering the fact that $R_{in} \in \mathbb{N}^+$, $G_{d,max}$ can be written as (14). After combining the above results, this proposition is proved. ■

Remark 2: Proposition 2 shows that NCK-SCMA can enjoy a larger DO with a proper configuration. It should be noted that the condition to achieve higher DO in NCK-SCMA mainly relies on the selection of R_{in} and R_{nc}. Furthermore, by using the same number of utilized resources for K-SCMA, it is clear that an NCK-SCMA with the same or larger DO exists, and the maximum DO can be calculated by (14). More importantly, another interesting property is that with the increase of K_{rep}, the DO discrepancy between NCK-SCMA and K-SCMA becomes more significant according to (14). This property is useful in the practical design of NCK-SCMA.

It can be observed that the above propositions mainly consider the case when $N_R = K_{rep}T_p$, i.e., NCK-SCMA and K-SCMA consume the same number of resources. The following corollaries provide conditions of NCK-SCMA to achieve larger DO without the constraint of N_R based on the above analysis.

Corollary 1: Under the constraint that $G_d = d_\mu K_{eq}$, the DO of NCK-SCMA exceeds that of K-SCMA if and only if $K_{rep} < K_{eq}$ is satisfied.

Proof: (*If*)- For $K_{rep} < K_{eq}$, we can directly obtain $d_\mu K_{rep} < G_d = d_\mu K_{eq}$, it can be easily seen that the DO of NCK-SCMA is higher than that of K-SCMA.

(*Only if*)- For $K_{rep} \geq K_{eq}$, when $G_d = d_\mu K_{eq}$, then $G_d = d_\mu K_{eq} \leq d_\mu K_{rep}$ is obtained. However, this contradicts with the premise of this corollary. By combining the above results, this corollary is proved. ■

Corollary 2: If NCK-SCMA can achieve higher DO than K-SCMA, then $K_{rep} < K_{eq}$.

Proof: The case when $G_d = d_\mu K_{eq}$ has been proved in *Corollary 1*. Hence, we only consider the case if $G_d = d_\mu R_{in}T_p$. In this case, $G_d = d_\mu R_{in}T_p > K_{rep}$, considering $R_{in}T_p < K_{eq}$, and thus we have $K_{eq} > R_{in}T_p > K_{rep}$. This completes the proof. ■

Furthermore, it is interesting to investigate whether a larger DO of NCK-SCMA is possible under the condition that $N_R < K_{rep}T_p$. In this case, NCK-SCMA can attain both higher reliability and energy-efficiency. To illustrate this issue, we present the following design criteria based on previous results.

Corollary 3: In NCK-SCMA, to achieve larger DO with fewer resources than K-SCMA with $K_{rep} \geq 3$, the following inequality must be satisfied:

$$R_{in} \geq R_{nc} > \frac{2}{T_p - 1}, \tag{17}$$

when $G_d = d_\mu K_{eq}$, and

$$2R_{in} - \frac{2}{T_p - 1} > R_{nc} \geq R_{in}, \tag{18}$$

when $G_d = d_\mu R_{in}T_p$. The equality in (17) and (18) are both achieved if and only if $R_{in}T_p = K_{eq}$.

Proof: See Appendix B. ■

Remark 3: Corollary 3 indicates that NCK-SCMA has the potential of achieving larger DO with fewer resources. It is also found that when $G_d = d_\mu K_{eq}$, only one repetition should be conducted on the NCed packets when $T_p > 3$. This implies that one can achieve a larger DO by increasing T_p when K_{rep} is low. Therefore, for a given K-SCMA with K_{rep}, an NCK-SCMA with fewer resources and larger DO can be constructed. In addition, higher R_{in} and R_{nc} should be used in the case that $G_d = d_\mu R_{in}T_p$ compared to $G_d = d_\mu K_{eq}$ even if T_p is large. In this regard, the condition specified in (17) can be applied to more scenarios.

C. ASEP analysis

To further understand the error performance of NCK-SCMA, we also consider the ASEP performance, where a multi-user union bound [16], [22] can be approximated by

$$P_s \leq \frac{1}{M^{T_pJ}} \cdot \sum_{\mathbf{X}} \sum_{\hat{\mathbf{X}} \neq \mathbf{X}} \Pr(\mathbf{X} \to \hat{\mathbf{X}}). \tag{19}$$

In the high SNR region, the probability that two or more users have erroneous transmissions are sufficiently small and hence can be negligible, provided that the number of users is significantly smaller than 1/ASEP. Hence, for a tractable analysis, we assume that there is at most one user among the remaining $J-1$ users suffering from error transmission. Such an approximation in the ASEP analysis is termed as single-user error pattern (SUEP) condition [27], [29]. For clarification, we provide the following example that shares the same parameters of *Example 1*.

Example 2: Based on the previous analysis, the DO of NCK-SCMA with $(R_{in}, R_{nc}, T_p) = (1, 1, 2)$ is $G_d = 2d_\mu$. As the UPEP is mainly determined by the SUEP, we first consider the number of PEPs with $D(\mathbf{X}_{V,e}) = G_d = 2d_\mu$. To elaborate further, since $G_d = d_\mu K_{eq} = d_\mu R_{in}T_p$ in this example, not only the case that the error occurs in one packet, such as $\mathbf{P}_j = [\mathbf{P}_{j,1}, \mathbf{P}_{j,2}, \mathbf{P}_{j,1} \oplus \mathbf{P}_{j,2}] = [000000]$ is erroneously detected as $\hat{\mathbf{P}}_j = \left[\hat{\mathbf{P}}_{j,1}, \mathbf{P}_{j,2}, \hat{\mathbf{P}}_{j,1} \oplus \mathbf{P}_{j,2}\right] = [010001]$, but also $\hat{\mathbf{P}}_j = \left[\hat{\mathbf{P}}_{j,1}, \hat{\mathbf{P}}_{j,2}, \hat{\mathbf{P}}_{j,1} \oplus \hat{\mathbf{P}}_{j,2}\right] = [010100]$ can make $\mathbf{X}_{V,e}$ attain the minimum rank. After enumerating all the possible cases, 9×16 different PEP events with $G_d = 2d_\mu$ should be calculated for each candidate codeword in each user, whereas the remaining 6×16 PEP events are with $D(\mathbf{X}_{V,e}) = 3d_\mu$.

Let us extend the above example to a more general case. First, let us consider a set $\mathcal{A}$ of codeword pairs $\left(\mathbf{X}_{j,V_\alpha}, \hat{\mathbf{X}}_{j,V_\alpha}\right)$ that makes $\mathbf{X}_{V,e}$ to have the minimum rank under SUEP condition, then the ASEP of NCK-SCMA can be rewritten as:

$$P_s \lesssim \underbrace{\frac{1}{M^{T_p}} \sum_{\mathbf{X}_{j,V_\alpha}} \sum_{\hat{\mathbf{X}}_{j,V_\alpha} \neq \mathbf{X}_{j,V_\alpha}} \Gamma(G_d, N_0) \prod_{v_\alpha=1}^{G_d} \Delta_{j,v_\alpha}^{-2}}_{\text{Part I}} + \underbrace{\frac{1}{M^{T_p}} \sum_{\mathbf{X}_{j,V_\beta}} \sum_{\hat{\mathbf{X}}_{j,V_\beta} \neq \mathbf{X}_{j,V_\beta}} \Gamma\left(D(\mathbf{X}_{V,e}), N_0\right) \prod_{v_\beta=1}^{D(\mathbf{X}_{V,e})} \Delta_{j,v_\beta}^{-2}}_{\text{Part II}}, \tag{20}$$

where codeword pairs $\left(\mathbf{X}_{j,V_\beta}, \hat{\mathbf{X}}_{j,V_\beta}\right) \notin \mathcal{A}$. Note that Part I in (20) is the summation of all PEPs with $D(\mathbf{X}_{V,e}) = G_d$, whereas Part II sums the PEPs that satisfy $D(\mathbf{X}_{V,e}) > G_d$. In addition, P_s is mainly determined by Part I in the high SNR region, and thus the cardinality of $\mathcal{A}$ should be evaluated. As for NCK-SCMA with (R_{in}, R_{nc}, T_p), we can infer that

$$|\mathcal{A}| = T_p M^{T_p} (M-1), \quad \text{if } G_d = d_\mu K_{eq}. \tag{21a}$$

$$|\mathcal{A}| = M^{T_p}(M-1), \quad \text{if } G_d = d_\mu R_{in} T_p. \tag{21b}$$

$$|\mathcal{A}| = M^{T_p}(M-1)(T_p+1), \quad \text{if } G_d = d_\mu K_{eq} = d_\mu R_{in} T_p. \tag{21c}$$

From the above analysis, (21a) can be obtained by considering all the cases that error occurs in only one packet; (21b) holds following from the fact that when $\mathbf{P}_{j,t_\alpha} \oplus \hat{\mathbf{P}}_{j,t_\beta}$ takes the same value for all $t_\alpha, t_\beta \in \{1, 2, \cdots, T_p\}$, $\beta > \alpha$, then the minimum rank of $\mathbf{X}_{V,e}$ can be attained, and (21c) results from the fact that the situations in both (21a) and (21b) should be taken into consideration when $G_d = d_\mu K_{eq} = d_\mu R_{in} T_p$. Based on the above discussions, we provide the following proposition to compare K-SCMA and NCK-SCMA:

Proposition 3: When the unimodular codebook of $M = 4$ is utilized by both NCK-SCMA and K-SCMA that share the same DO, and $K_{rep} = K_{eq} = R_{in} T_p$, then the ASEP performance of K-SCMA is superior to that of NCK-SCMA with more resource consumption.

Proof: See Appendix C. ■

The above proposition seems to imply a limitation of NCK-SCMA, however, the following two observations showing a trade-off between the error rate performance and consumed resources in this case should be noted:

1) The reason for unimodular codebooks to be considered lies in their excellent error rate performance [16]. When NCK-SCMA and K-SCMA share the same DO, we can always find an NCK-SCMA that consumes fewer resources. This is because $K_{rep} T_p - N_R = W R_{nc}$ under the condition that $K_{eq} = K_{rep} = R_{in} T_p$; hence, the transmitted energy can be reduced in this case.

2) Based on *Proposition 3*, we can further estimate the performance gain of K-SCMA compared to NCK-SCMA when they share the same DO according to the same method in [30]. Assuming the ASEP performances of K-SCMA and NCK-SCMA are the same, then the SNR discrepancy between K-SCMA and NCK-SCMA can be written as:

$$\Delta\text{SNR} \simeq \frac{10}{K_{rep} d_\mu} \log_{10} \left(\frac{M^{T_p}(M-1)}{|\mathcal{A}|} \right). \tag{22}$$

As can be observed from (22), this approximation can be used to estimate the performance gain of K-SCMA over NCK-SCMA. With an increasing K_{rep}, the performance gain of K-SCMA decreases. This indicates that although the ASEP of K-SCMA has superiority over that of NCK-SCMA when $K_{eq} = K_{rep} = R_{in} T_p$, the performance loss is acceptable. The accuracy of (22) will be examined in Section V.

Corollary 4: Based upon the assumptions in *Proposition 3*, the ASEP of NCK-SCMA with $M = 4$ cannot be improved by increasing R_{nc} in the high SNR region when $K_{eq} > R_{in} T_p$.

Proof: Note that $G_d = d_\mu R_{in} T_p$ in this case, which is irrelavent with R_{nc}. Hence, this corollary can be simply proved by integrating the results into (21b) and (46), leading to the fact that $\sum_{i=1}^{|\mathcal{A}|-M(M-1)} \delta_i = (M^{T_p} - M)(M-1) P_{s,k}$ when unimodular codebooks are utilized. This indicates that for any R_{nc}, provided that $K_{eq} > R_{in} T_p$, the Part I of ASEP in (20) are the same, and thus the results can be proved. ■

An important observation from the above analysis is that the simple repetition of NCed packets is inefficient for the ASEP improvement as no extra information can be extracted, even though the improvement of DO in NCK-SCMA stems from the transmission of NCed packets. That being said, increasing R_{in} under the constraint of N_R to obtain higher DO, whereas selecting reasonable R_{nc} to ensure $K_{eq} \leq R_{in} T_p$ is a preliminary guideline to design NCK-SCMA.

IV. Proposed Detection for coded NCK-SCMA

To examine the error performance of NCK-SCMA with advanced channel coding, we consider LDPC coded NCK-SCMA and propose an improved iterative algorithm called INDSD in this section.

A. INDSD for LDPC coded NCK-SCMA

In LDPC coded NCK-SCMA, to recover the transmitted data, one needs to carry out SCMA detection, network decoding, and LDPC decoding at the receiver side. By employing bit-level combining (BLC) [24] for soft combining, we propose a novel receiver to iteratively update the belief messages among the LDPC, NC, and SCMA decoders. By doing so, we will show that the proposed scheme can fully utilize the advantages of NC and HARQ retransmission.

For clarity, we define SCMA variable nodes (SVNs) and LDPC variable nodes (LVNs) as $sv_{l,j}$ and $lv_{\tau,n}$, where $l \in \{1, 2, \cdots, N_R L\}$, $\tau \in \{1, 2, \cdots, JT_p\}$, and $n \in$

$\{1,2,\cdots,N\}$ denote the indices of SCMA codewords, LDPC decoders, and LVNs, respectively. Furthermore, a virtual node for network decoding and message conversion is denoted as NC check node (NCN) $nc_{\eta}^{t,j}$, where $\eta \in \{1,2,\cdots,N\}$ is the index of NCN for the tth transmitted packet of the jth UE. In addition, the function nodes (FN) of SCMA and parity check nodes (PN) of LDPC update should also be taken into consideration. Before elaborating on the update rules of these nodes, let us define the following messages:

- $I_{l,k\to j}$, $G_{l,j\to k}$: Message passed from FN $f_{l,k}$ to SVN $sv_{l,j}$, and SVN $sv_{l,j}$ to FN $f_{l,k}$;
- $Q_{\tau,c\to n}$, $S_{\tau,n\to c}$: Message passed from PN $p_{\tau,c}$ to LVN $lv_{\tau,n}$, and LVN $lv_{\tau,n}$ to PN $p_{\tau,c}$;
- $\Lambda_{\eta\to\gamma}^{t,j}$, $\Lambda_{\eta\to n}^{t,j}$: Message passed from NCN $nc_{\eta}^{t,j}$ to SVN $sv_{l,j}$, and NCN $nc_{\eta}^{t,j}$ to LVN $lv_{j,n}$,

where $c \in \{1,2,\cdots,C\}$, C is the number of PNs for the LDPC code, and $\gamma \in \{1,2,\cdots,\cdots,JN_RL\}$ is the index of SVN for each packet.

1) FN Update and PN Update: After receiving the superimposed signals of NCK-SCMA at FNs, the detection will start from FN and PN simultaneously. In our proposed algorithm, FN and PN only exchange information with SVN and LVN, respectively; hence, the update of FN and PN is given as:

$$I_{l,k\to j}(\mathbf{X}_j) = \sum_{\mathbf{X}_j : j\in\xi_k\setminus j} \frac{1}{\pi N_0} e^{\left(-d_l^{j,k}/N_0\right)} \cdot \prod_{j\in\xi_k\setminus j} G_{l,j\to k}(\mathbf{X}_j), \tag{23}$$

where $d_l^{j,k} = \frac{1}{N_0/2}|Y_{l,k} - \sum_{k\in\xi_k} h_l^{j,k} X_{j,k}|^2$, and $h_l^{j,k}$ denotes the channel coefficient between the jth UE and BS during the transmission of the kth resource element of the lth codeword, $Y_{l,k}$ the received signal at the kth FN of the lth codeword, N_0 the noise variance, and ξ_k the node set consisting of SVNs connect to the kth FN. Meanwhile, the PN update is given by:

$$Q_{\tau,c\to i} = 2\times \tanh^{-1}\left(\prod_{n\in\phi_c\setminus i} \tanh(S_{\tau,n\to c})\right), \tag{24}$$

where ϕ_c is the set that contains all the LVNs connecting to the cth PN. Note that the messages in both (23) and (24) should be normalized in each iteration for the purpose of numerical stability.

2) SVN Update: In the update of SVN, messages that come from FN and NCN are involved. For clarity, we first consider SVN update for the non-NCed packets, and thus

$$G_{l_1+\lceil\frac{\eta}{b}\rceil,j\to\kappa}(\mathbf{X}_j) = \prod_{k\in\zeta_j\setminus\kappa} I_{l_1+\lceil\frac{\eta}{b}\rceil,k\to j}(\mathbf{X}_j)\cdot\mathcal{M}^{-1}\left(\Lambda_{\eta\to\tilde{j}}^{t,j}\right), \tag{25}$$

where $l_1 = (t-1)R_{in}L + (r_{in}-1)L, r_{in} \in \{1,2,\cdots,R_{in}\}$, $\tilde{j} = J(l_1+\lceil\frac{\eta}{b}\rceil-1)+j$, ζ_j is the FN set of nodes that connect to the jth SVN, and $\mathcal{M}^{-1}$ is the inverse marginalization which transform bit-wise soft information to symbol level[8] in this paper. Specifically, it can be represented as:

$$\mathcal{M}^{-1}(\boldsymbol{a}) = \prod_{i_b=1}^{b} \Pr\left(\boldsymbol{a}^{(i_b)}\right), \tag{26}$$

and

$$\Pr\left(\boldsymbol{a}^{(i_b)}\right) = \begin{cases} \dfrac{\exp\left(\mathrm{LLR}\left(\boldsymbol{a}^{(i_b)}\right)\right)}{1+\exp\left(\mathrm{LLR}\left(\boldsymbol{a}^{(i_b)}\right)\right)}, & \text{if } \boldsymbol{a}^{(i_b)} = 0; \\ 1-\dfrac{\exp\left(\mathrm{LLR}\left(\boldsymbol{a}^{(i)}\right)\right)}{1+\exp\left(\mathrm{LLR}\left(\boldsymbol{a}^{(i_b)}\right)\right)}, & \text{if } \boldsymbol{a}^{(i_b)} = 1. \end{cases}$$

where $\boldsymbol{a} \in \{0,1\}^b$ represents a symbol with b bits, and $\boldsymbol{a}^{(i_b)}$ is its i_bth element. However, since NC is employed in NCK-SCMA, to improve the performance of iterative detection, the soft information that comes from NCN should be re-encoded to generate messages for the NCed packets. Hence,

$$G_{l_2+\lceil\frac{\eta}{b}\rceil,j\to\kappa}(\mathbf{X}_j) = \prod_{k\in\zeta_j\setminus\kappa} I_{l_2+\lceil\frac{\eta}{b}\rceil,k\to j}(\mathbf{X}_j) \\ \mathcal{M}^{-1}\left(\Lambda_{\eta\to\tilde{j}_\alpha}^{t_\alpha,j} \boxplus \Lambda_{\eta\to\tilde{j}_\beta}^{t_\beta,j}\right), \tag{27}$$

where $l_2 = T_pLR_{in} + (w-1)R_{nc}L + (r_{nc}-1)L, r_{nc} \in \{1,\cdots,R_{nc}\}$, and $\tilde{j}_\Xi = J((t_\Xi - 1)LR_{in} + \lceil\frac{\eta}{b}\rceil - 1) + j$. Moreover, $\boxplus$ is the soft XOR operator that is defined as [31]:

$$L_1 \boxplus L_2 = 2\tanh^{-1}\left(\tanh(L_1/2)\cdot\tanh(L_2/2)\right) \\ \approx \mathrm{sign}(L_1)\cdot\mathrm{sign}(L_2)\cdot\min(|L_1|,|L_2|). \tag{28}$$

3) LVN Update: Likewise, the update of LVN involves the information coming from PN and NCN. As such, for a certain packet t, the update rule of LVN can be represented as

$$S_{t+(j-1)T_p,n\to q} = \sum_{c\in\psi_n\setminus q} Q_{t+(j-1)T_p,c\to n} + \pi_{t,j}^{-1}\left(\Lambda_{\eta\to n}^{t,j}\right), \tag{29}$$

where ψ_n is the node set whose elements connect to the nth LVN, and $\pi_{t,j}^{-1}$ denotes the bit level de-interleaver operation.

4) NCN Update: It is clear that NCN bridges the SCMA detection and LDPC decoding, and thus the update of NCN should be carefully designed to fully utilize the NC packets. A partial view of the NCN update is presented in Fig. 3. As can be seen from the left side of the figure, to update $\Lambda_{\eta\to n}^{t,j}$, we should first consider the messages at SVN after soft combining of the R_{in} initial repetitions[9], which can be expressed as:

$$I_{\eta}^{t,j} = \sum_{r_{in}=1}^{R_{in}}\left[\mathcal{M}\left(\prod_{k\in\zeta_j} I_{\lceil\frac{\eta}{b}\rceil+l_1,k\to j}(\mathbf{X}_j)\right)\right], \tag{30}$$

where $\mathcal{M}$ is the inverse operation of $\mathcal{M}^{-1}$ that converts M-dimensional message vector to b messages. It should be noted that the combining in (30) follows from BLC principle [23], which is a suboptimal combining scheme with low computational complexity. To process the NCN update, the nth message of NCed packets after R_{nc} times soft combining is also necessary, which can be represented as:

$$I_{\eta}^{w,j} = \sum_{r_{nc}=1}^{R_{nc}}\left[\mathcal{M}\left(\prod_{k\in\zeta_j} I_{\lceil\frac{\eta}{b}\rceil+l_2,k\to j}(\mathbf{X}_j)\right)\right]. \tag{31}$$

It is also shown in Fig. 3 that messages $I_{\eta_1}^{t_\alpha,j}$ to $I_{\eta_b}^{t_\alpha,j}$ pass to b differernt NCNs after marginalization and combining R_{in} soft information of packet t_α. Furthermore, once NCK-SCMA

[8]For the ceil function, note that there exist b different elements that satify $\lceil\frac{n_1}{b}\rceil = \cdots = \lceil\frac{n_b}{b}\rceil$ for a given $b \in \mathbb{N}^+$. This indicates that b different messages can be generated from the same node indexed by $\lceil\frac{n}{b}\rceil$ in this paper.

[9]It is worth noting that the subscript of SVN in Fig. 3 is obtained by substituting $t_\alpha = 1$ into l_1.

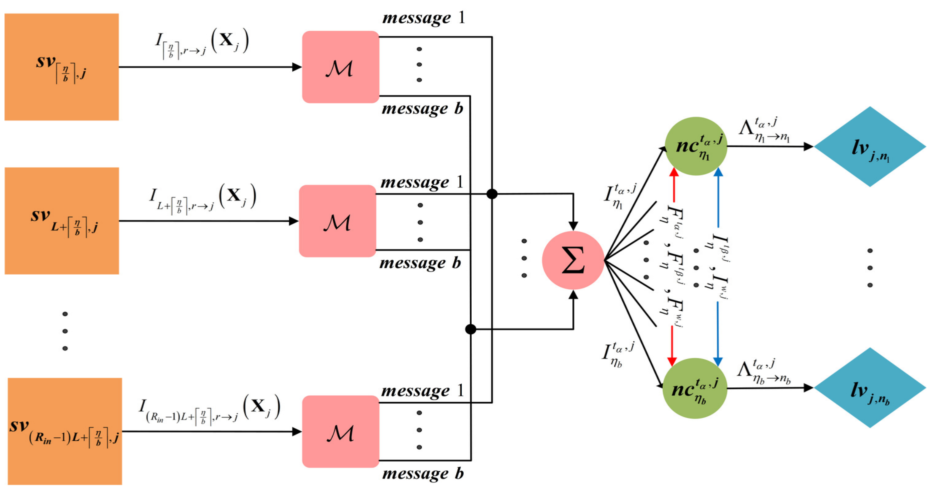

Fig. 3: Partial view of NCN update for a non-NCed packet indexed by $t_\alpha = 1$.

retransmission is requested, the soft information generated from previous unsuccessful detections can be utilized. It is assumed that $F_\eta^{w,j}$, $F_\eta^{t_\alpha,j}$, and $F_\eta^{t_\beta,j}$ are the previous soft information of NCed packet w and its corresponding packets t_α and t_β. After receiving all the soft information at NCNs, the receiver can output $\Lambda_{\eta\to n}^{t_\alpha,j}$. In order to properly use these messages, the following three cases should be taken into consideration. For simplicity, we take the message $\Lambda_{\eta\to n}^{t_\alpha,j}$ as an example to demonstrate the NCN update, and the updating rules of $\Lambda_{\eta\to n}^{t_\beta,j}$ can be derived similarly.

Case 1: *The previous T_p packets all fail.* In this case, as long as the maximum transmission round N_{re} has not been reached, retransmission is necessary. To ensure the accuracy of soft information, we update NCN by using

$$\Lambda_{\eta\to n}^{t_\alpha,j} = \left[(I_\eta^{t_\beta,j} + F_\eta^{t_\beta,j}) \boxplus (I_\eta^{w,j} + F_\eta^{w,j})\right] + I_\eta^{t_\alpha,j} + F_\eta^{t_\alpha,j}. \tag{32}$$

The updating rule in Case 1 follows from the BLC principle, and thus the messages generated in previous and new transmissions are combined in the first step to ensure the accuracy of soft information. By doing so, the initial transmission and retransmissions can be fully utilized.

Case 2: *The previous T_p packets all successful.* In this instance, T_p new packets are sent at the current RTT, which indicates that $F_\eta^{t_\beta,j} = F_\eta^{w,j} = F_\eta^{t_\alpha,j} = 0$, for $\forall t, w, j, \eta$ in (32). Therefore, the NCN update can be written as

$$\Lambda_{\eta\to n}^{t_\alpha,j} = (I_\eta^{t_\beta,j} \boxplus I_\eta^{w,j}) + I_\eta^{t_\alpha,j}. \tag{33}$$

Case 3: *The previous T_p packets partially fail.* In this case, the messages corresponding to newly transmitted packets update follow the rule in Case 2, while the NCN update for those packets that are unsuccessfully detected is:

$$\Lambda_{\eta\to n}^{t_\alpha,j} = \underbrace{I_\eta^{t_\alpha,j} + F_\eta^{t_\alpha,j}}_{\text{Part I}} + \underbrace{\left[I_\eta^{t_\beta,j} \boxplus I_\eta^{w,j}\right] + \left[F_\eta^{t_\beta,j} \boxplus F_\eta^{w,j}\right]}_{\text{Part II}}. \tag{34}$$

As can be observed from Part I of (34), to avoid distortion of soft information caused by soft XOR in (28), we combine the messages of packet t_α that suffer from failure in previous and current transmissions at first. After that, as can be seen from Part II, the information that comes from NCed packets in previous and current transmissions involved in the NCN update is combined to further enhance the reliability of this case.

On the other hand, the messages passing from NCN to SVN can be expressed as

$$\Lambda_{\eta\to\gamma}^{t,j} = \pi_{t,j}\left(Q_{t+(j-1)T_p,n}\right), \tag{35}$$

where $Q_{\tau,n} = \sum_{c\in\psi_n} Q_{\tau,c\to n}$, $\pi_{t,j}$ is the interleaver of the tth packet of UE j.

5) Decision and Output Messages: In the proposed joint detection scheme, the final decision is made at LVN, and the accumulated LLR is:

$$Q_{t+(j-1)T_p,n} = \sum_{c\in\psi_n} Q_{t+(j-1)T_p,c\to n} + \Lambda_{\eta\to n}^{t,j}, \tag{36}$$

whereas the LLRs for the possible HARQ retransmission output at SVN, are given by (30) and (31). The iteration stops once the syndromes are zeros or reaching the maximum number of iterations. Finally, the proposed detector is summarized at a glance in **Algorithm 1**, which can be found in the last page of this manuscript.

V. Numerical Evaluation

In this section, the error rate and throughput performances of NCK-SCMA and K-SCMA are first evaluated via numerical simulations. Afterwards, we analyze and compare the implementation complexities of these two systems.

A. Comparisons between uncoded K-SCMA and NCK-SCMA

We first consider the uncoded K-SCMA and NCK-SCMA with indicator matrix "4 × 6" [3] to verify the analysis in Section III [10]. Without any special announcement, we adopted the optimal unimodular codebooks (i.e., sparse codebooks with unimodular nonzero elements) [16] for $M = 4$. In addition,

[10]Although this paper focuses on the enhancement of SCMA with HARQ, the uncoded performance comparison between NCK-SCMA and K-SCMA is presented to demonstrate the advantages of NCK-SCMA.

Algorithm 1: Iterative detection of LDPC coded NCK-SCMA

1 Initialization: $I_{l,r\to j}=1/M$, $Q_{k,c\to n}=0$, $\Lambda^{t,j}_{\eta\to\gamma}=1/M$, $\Lambda^{t,j}_{\eta\to n}=0$, $F^{t,j}_{\eta}=0$, and $F^{w,j}_{\eta}=0$;
2 **for** $i=0:N_{re}$ **do**
3   **while** $iter<N_{iter}$ **do**
4   Execute FN update by using (23), return $I_{l,k\to j}$;
5   Execute PN update by using (24), return $Q_{\tau,c\to n}$;
6   Execute NCN update by using (35), return $\Lambda^{t,j}_{\eta\to\gamma}$;
7   Find different types of SVN by calculating l_1 and l_2, then judging whether $l\le l_1$ or $l_1<l\le l_2$;
8   **if** $l\le l_1$ **then**
9   Execute SVN update by using (25) and (26), return $I_{l,k\to j}$;
10   **else**
11   Execute SVN update by using (26), (27), and (28), return $I_{l,k\to j}$;
12   Execute LVN update by using (29), return $S_{\tau,n\to c}$;
13   Combining the messages of two types SVN by using (30) and (31);
14   **if** *Case 1* **then**
15   Execute NCN update by using (32);
16   **if** *Case 2* **then**
17   Execute NCN update by using (33);
18   **if** *Case 3* **then**
19   Execute NCN update by using (34);
20   Calculate the accumulated LLR by using (36), then $\hat{\mathbf{P}}_{j,t}$ can be recoverd;
21   **if** *$iter<N_{iter}$ and the syndromes are not zeros* **then**
22   Proceed to Step **4**;
23   **else**
24   break;
25   **if** *T_p packets are not correctly detected* **then**
26   Generate $F^{w,j}_{\eta}$, $F^{t_\alpha,j}_{\eta}$, and $F^{t_\beta,j}_{\eta}$ according to $F^{w,j}_{\eta}=F^{w,j}_{\eta}+I^{w,j}_{\eta}$ and $F^{t,j}_{\eta}=F^{t,j}_{\eta}+I^{t,j}_{\eta}$; Proceed to Step **3**;
27   **else**
28   break;
29 Return $\hat{\mathbf{P}}_{j,t}$.

to make fair comparisons with K-SCMA, we also define the average utilized resources over the same number of packets as $\bar{N}_R=N_R/T_p$. Specifically, for K-SCMA, we have $K_{rep}=N_R=\bar{N}_R$. Note that "Ana." and "Sim." denote "Analytical" and "Simulation", respectively.

In Fig. 4, the ASEP performances of uncoded K-SCMA and NCK-SCMA are considered. The "Sim." curves are obtained following the DLC principle. As can be observed from Fig. 4, the analytical upper bound in (19) and simulated results of K-SCMA almost coincide at high SNRs, and the upper bound becomes tighter as the increase of K_{rep}. Since the simulated results are obtained by using DLC, the assertation that DLC and MLD have the same decoding performance in (8) can be verified. On the other hand, we also plot the ASEP performance of NCK-SCMA with $(R_{in},R_{nc},T_p)=(2,1,2)$ and $(3,2,2)$, which have the same DO as K-SCMA with $K_{rep}=3$, and $K_{rep}=5$, respectively. It is shown in the figure that although NCK-SCMA suffers from slight degradation of ASEP in the low-to-medium SNR region, it consumes fewer resources. For instance, to transmit $T_p=2$ different packets with $K_{rep}=5$ in K-SCMA, at least 10 transmissions should be conducted, whereas only 8 transmissions are necessary for NCK-SCMA with $(3,2,2)$. In this context, NCK-SCMA leads to throughput improvement.

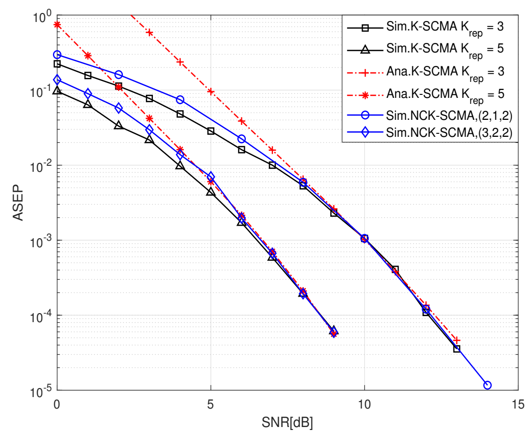

Fig. 4: ASEP comparison for uncoded "4×6" K-SCMA and NCK-SCMA with the same K_{eq}.

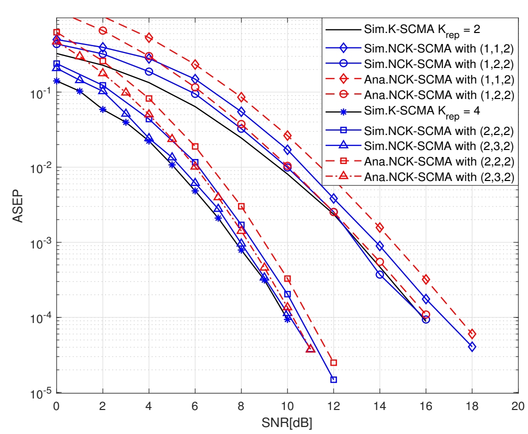

Fig. 5: ASEP comparison for uncoded "4×6" K-SCMA and NCK-SCMA under the same DO.

Fig. 5 presents the ASEP performance of NCK-SCMA and K-SCMA with the same DO, and the theoretical ASEP by combining the PEP of NCK-SCMA under the SUEP condition in (39) and union bound in (19). It is observed that the analytical results of NCK-SCMA verify the feasibility of the SUEP condition in approximating the asymptotic ASEP of

NCK-SCMA. We also observe that the NCK-SCMA with $K_{rep} = R_{in}T_p = K_{eq}$ exhibits worse ASEP performance than K-SCMA, which is consistent with the analysis of *Proposition 3*. Furthermore, we can calculate the approximated performance gains of K-SCMA according to (22) for $K_{rep} = 2$ and $K_{rep} = 4$, which are 1.193 dB and 0.596 dB, respectively. This is well aligned with the actual performance gains based on the simulation results which are 1.1 dB and 0.5 dB, respectively. Together with the observations from Fig. 4, our analysis about DO and ASEP of both K-SCMA and NCK-SCMA can be verified.

Fig. 6 compares the ASEP performances between K-SCMA and NCK-SCMA under the same resources consumption, i.e, NCK-SCMA and K-SCMA have the same $\bar{N}_R$. It is shown in the figure that when $R_{in} \geq R_{nc}$, NCK-SCMA with $(2, 2, 2)$ and $(3, 2, 2)$ outperform their corresponding K-SCMA with approximately 1.8 dB and 1.6 dB gain at ASEP = 10^{-4}, respectively. This follows from the assertion in *Proposition 2* that NCK-SCMA can achieve a larger DO in this case. In addition, with the increase of available resources, a higher DO can be achieved by selecting reasonable R_{in} and R_{nc} in NCK-SCMA. As for $\bar{N}_R = 2$, since it is impossible to design an NCK-SCMA with a higher DO than K-SCMA in this case and according to *Proposition 2*, NCK-SCMA and K-SCMA have nearly the same performance, despite a slight ASEP loss in low SNR regions.

Fig. 7 investigates the impact of R_{nc} on the ASEP performance of NCK-SCMA. Although the improvement of DO arises from NC, as can be observed from the figure, when $R_{in} = 1$, NCK-SCMA with $R_{nc} = 2$ and $R_{nc} = 3$ exhibits almost the same ASEP performances, especially in the region of medium to high SNRs. As for NCK-SCMA with $R_{in} = 2$, a similar observation can be witnessed. This indicates that the increase of R_{nc} is unable to improve ASEP when $K_{eq} > R_{in}T_p$, which is consistent with the analysis in *Corollary 4*. However, before K_{eq} reaching $R_{in}T_p$ via adding more repetitions of NCed packets, the ASEP performance can be slightly improved. This is attributed to the fact that the repetitions of NCed packets help to increase K_{eq} when $K_{eq} < R_{in}T_p$.

B. Throughput and PER of LDPC coded NCK-SCMA

In this subsection, the throughput and PER performances of LDPC coded K-SCMA and NCK-SCMA are evaluated to demonstrate the effectiveness of our proposed NCK-SCMA scheme and the INDSD algorithm. Two different SCMA signature matrices and codebooks having orders of "4×6" and "5×10" [16] with overloading factor 150% and 200%, respectively, are considered. Moreover, two 5G new radio LDPC codes, as specified in [32], with rates of 1/2 and 5/6, respectively, are used for simulation. Note that each block consists of 270 and 260 bits for codes with rates of 1/2 and 5/6, respectively.[11] We consider the Rayleigh fading channel and the maximum LDPC iteration number of 50. As indicated

[11] As K-repetition scheme is proposed to reduce latency to meet the stringent latency requirement in the future wireless network, we mainly consider the transmitted packets with short blocklength in this paper.

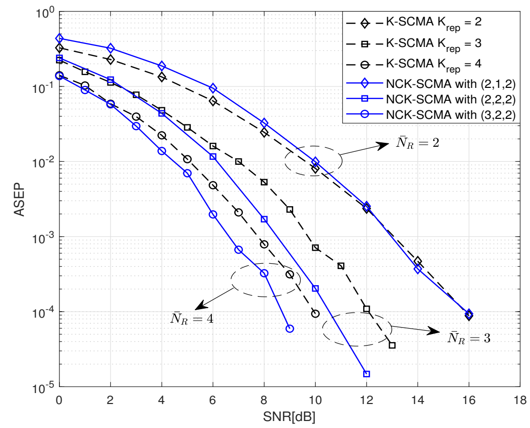

Fig. 6: ASEP comparison for uncoded "4×6" K-SCMA and NCK-SCMA with the same resources consumption.

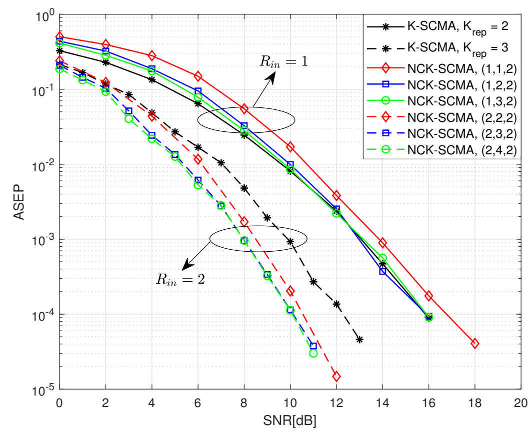

Fig. 7: ASEP performance comparison of uncoded "4×6" NCK-SCMA under different R_{nc}.

and defined in [17], [33], the average throughput is treated as an important criterion to evaluate the performance of the proposed NCK-SCMA. Therefore, we have

$$\theta = \frac{T_{correct}}{T_{total}}, \tag{37}$$

where $T_{correct}$ and T_{total} are the number of correctly decoded and the total number of transmitted packets, respectively.

Fig. 8 compares the throughputs of "4×6" NCK-SCMA and K-SCMA with a code rate of 5/6 under different N_{re}. As can be observed from the figure, NCK-SCMA with $(2, 2, 2)$ and $\bar{N}_R = 3$ can achieve higher throughput than K-SCMA with $K_{rep} = 3$ without any retransmissions when SNR ≥ 5 dB, thanks to the fact that higher DO can be obtained by NCK-SCMA with $(2, 2, 2)$. By performing retransmission using NCK-SCMA with $(2, 2, 2)$, the throughput can be significantly improved at low and medium SNRs. Nevertheless, K-SCMA with $K_{rep} = 3$ and $N_{re} = 2$ still outperforms NCK-SCMA with $(2, 2, 2)$ in the relatively low SNR region, which matches the uncoded ASEP performance in the previous subsection. It is worth noting that the increase of N_{re} leads to marginal improvement of throughput in NCK-SCMA. This can be attributed to the fact that the soft XOR operation may have

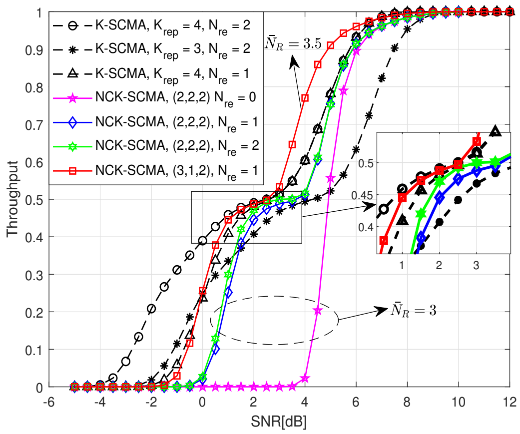

Fig. 8: Throughput comparison for "4×6" K-SCMA and NCK-SCMA with code rate of 5/6 , and $N = 260$.

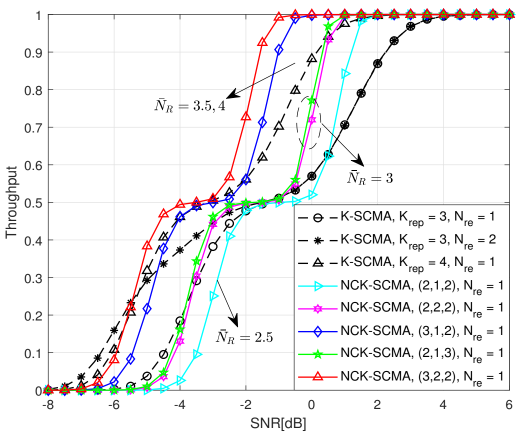

Fig. 9: Throughput comparison for "5×10" K-SCMA and NCK-SCMA with code rate of 1/2, and $N = 270$.

negative impacts on the effectiveness of soft information, especially at low SNRs. As a result, it may lead to error propagation in the proposed INDSD algorithm when reusing the soft information in previous transmissions. Note that NCK-SCMA with $K_{eq} = R_{in}T_p = 4$ can approach the throughput of K-SCMA with $K_{rep} = 4$ in the medium-to-high SNR region for diverse N_{re}, where the resource consumption of NCK-SCMA with $K_{eq} = 4$ is lower than K-SCMA with $K_{rep} = 4$. These observations are consistent with the findings in *Proposition 3* . Moreover, NCK-SCMA with $(3, 1, 2)$ and $\bar{N}_R = 3.5$ is also simulated, in which a slightly more resources are utilized compared to the cases when $\bar{N}_R = 3$. In this case, higher throughput can be achieved compared to $K_{rep} = 4$ and $K_{rep} = 3$ K-SCMA with fewer retransmission rounds.

Fig. 9 investigates the throughput performances of "5×10" K-SCMA and NCK-SCMA with a code rate of 1/2 under different parameter configurations. We can observe that the NCK-SCMA with $(2, 1, 3)$ outperforms NCK-SCMA with $(2, 2, 2)$ when $N_{re} = 1$ and $\bar{N}_R = 3$. This is because a potentially higher performance gain can be achieved by NCK-SCMA as more NCed packets are transmitted, which can be obtained by using (22). Moreover, both of them show superiority over the K-SCMA with $K_{rep} = 3$ and $N_{re} = 1$. When SNR ≥ -3 dB, NCK-SCMA with $(2, 2, 2)$ also shows superiority albeit with a slight degradation in the region of SNR < -3 dB. For NCK-SCMA with $(2, 1, 3)$ and $(3, 2, 2)$ that possesses $\bar{N}_R = 3.5$ and $\bar{N}_R = 4$, respectively. In spite of slight throughput loss in the low SNR regime, both NCK-SCMA schemes exhibit far higher throughput compared to K-SCMA with $K_{rep} = 4$ at medium-to-high SNRs, and the advantages of NCK-SCMA in throughput become more significant with the increase of SNR.

As for PER performances, if a single bit of a packet cannot be correctly decoded after N_{re} HARQ retransmissions, the packet is deemed to be in error. Based on this definition, we evaluate the PER performances of K-SCMA and NCK-SCMA in this subsection. Fig. 10 shows the PER performances of K-

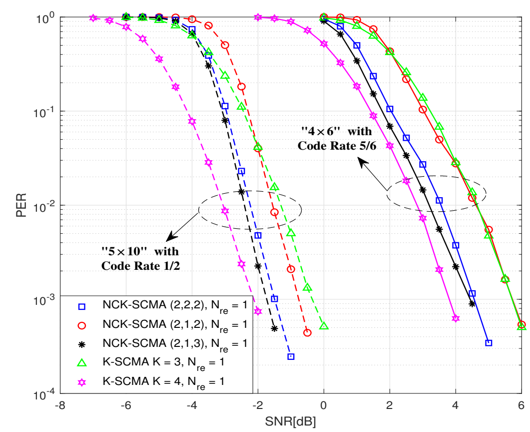

Fig. 10: PER comparison for K-SCMA and NCK-SCMA.

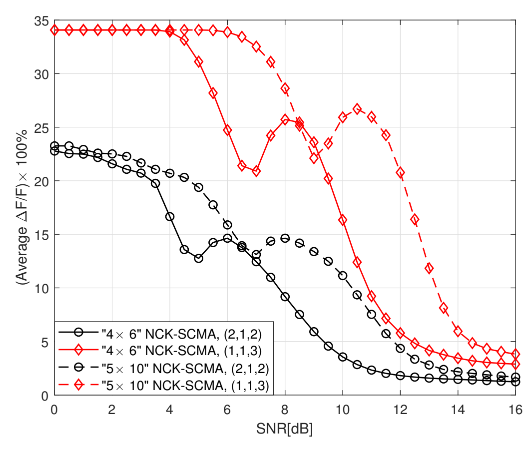

Fig. 11: Complexity comparison for NCK-SCMA and K-SCMA with code rate of 5/6 , and $N = 260$.

SCMA and NCK-SCMA with two different LDPC codes and overloading factors. It can be seen from the figure that for both "4×6" and "5×10" NCK-SCMA with $(2, 1, 3)$, improved PER performance is achieved across all SNRs compared to that with $(2, 2, 2)$, which is consistent with the throughput results discussed above. Furthermore, we can observe that "4×6" NCK-SCMA with $(2, 1, 2)$ has approximately the same PER performance as K-SCMA with $K_{rep} = 3$ at a code rate of $5/6$, while "5×10" NCK-SCMA with $(2, 1, 2)$ can achieve 0.3 dB gain at PER =10^{-3}. We can also observe that the gap between K-SCMA with $K_{rep}=4$ and NCK-SCMA with $(2,1,3)$ is approximately 0.4 dB at PER = 10^{-3} for both "4×6" and "5×10" NCK-SCMA, while 25% resources can be saved by the proposed NCK-SCMA.

C. Complexity analysis

To further illustrate the feasibility of the proposed detection algorithm for coded NCK-SCMA, we compare the computational complexity of INDSD with the detection and decoding of K-SCMA. To proceed, we utilize the average number of floating-point operations (FLOPs) to measure the complexity, where the calculation principle is based on the assumptions in [34]. As for LDPC coded K-SCMA, the computational complexity is mainly determined by MPA in SCMA detection and belief propagation (BP) in LDPC decoding, while INDSD for NCK-SCMA needs additional operations, such as network decoding, to recover the transmitted data. To make a fair comparison, it is assumed that the computational complexity is calculated under the condition that T_p packets are transmitted with the same DO. As such, without loss of generality, we mainly focus on the extra FLOPs ΔF that are necessary for the INDSD, which can be expressed as:

$$\begin{aligned}\Delta F = N_{iter}\sum_{j=1}^{J}\Delta F_j = \underbrace{6N_{iter}WJN}_{\text{Re-encode in (27)}}\\ +\sum_{j=1}^{J}\underbrace{9N_{iter}T_p\left(T_p-1\right)N\cdot\mathbb{1}\left(\forall t,\mathbf{P}_{j,t}\neq\hat{\mathbf{P}}_{j,t}\right)}_{\text{Case 1 in (32)}}\\ +\underbrace{6N_{iter}T_p\left(T_p-1\right)N\cdot\mathbb{1}\left(\forall t,\mathbf{P}_{j,t}=\hat{\mathbf{P}}_{j,t}\right)}_{\text{Case 2 in (33)}}\\ +\sum_{T_{err}=1}^{T_p-1}\underbrace{13N_{iter}\left(T_pT_{err}-T_{err}^2\right)N\cdot\mathbb{1}\left(\exists t,\mathbf{P}_{j,t}\neq\hat{\mathbf{P}}_{j,t}\right)}_{\text{Case 3 in (34)}},\end{aligned} \tag{38}$$

where T_{err} is the number of erroneously detected packets. Note that the extra complexity in (27) is fixed once the parameters of NCK-SCMA are determined, whereas the complexity derived from different combining strategies with previous packets in NCN update varies with SNR. Therefore, we compute the average FLOPs $\Delta\bar{F}$ through Monte-Carlo simulations to evaluate the increase of computational complexity.

Fig. 11 characterizes the percentage of $\Delta\bar{F}/F$ for both "4×6" and "5×10" LDPC coded NCK-SCMA at different SNRs, where F is the FLOPs of K-SCMA[12]. It is obvious that NCK-SCMA with the INDSD algorithm needs more FLOPs to recover data compared to K-SCMA due to $\Delta F>0$, and approximately 35% additional FLOPs is necessary at low SNR region for NCK-SCMA with (1,1,3). The increase of complexity reduces to about 23% as fewer resources are utilized in NCK-SCMA with (2,1,2). Nevertheless, the growth of computational complexity overall tends to decrease with increasing SNR, because a smaller value of N_{iter} is needed for the convergence of INDSD. Moreover, an interesting observation is that there exists an inflection point as the $\Delta\bar{F}$ decreases with SNR. This is mainly because although fewer number of iterations N_{iter} are needed as the SNR increases, the FLOPs of INDSD when dealing with the retransmission of T_{err} packets are higher than the cases when T_p packets are all successfully or unsuccessfully transmitted. Due to the probability of T_{err} out of T_p packets being incorrectly detected is high in the region of medium SNRs, and thus an inflection point can be witnessed.

[12]As illustrated in [35], the FLOPs of MPA detection per iteration in SCMA is $Kd_f\left[\left(7d_f-1\right)M^{d_f}+\left(d_\mu-1\right)M\right]$.

VI. Conclusion

In this paper, we have introduced a novel NC aided K-repetition scheme for SCMA called NCK-SCMA. We have conducted the PEP and ASEP analysis on uncoded NCK-SCMA, and made a comparison with K-SCMA. It is found that NCK-SCMA can obtain higher DO than K-SCMA with fewer resource consumption when it is properly configured. Under the same DO, it can also save transmission resources with negligible ASEP loss. These properties bring in higher throughput for NCK-SCMA. For efficient decoding at the receiver, we have studied the LDPC coded NCK-SCMA system by developing a joint SCMA-NC-LDPC detector as detailed in the proposed INDSD algorithm. Our results have demonstrated that the proposed NCK-SCMA achieves higher throughput at medium-to-high SNR regime with fewer resources and retransmissions, while attaining an enhanced PER under the same $\bar{N}_R$ and N_{re}.

Future Directions: One of the important directions for future study of NCK-SCMA is the design of more efficient inter-packets coding schemes in order to strike a proper balance between throughput and resource consumption. Besides, slight throughput loss can be witnessed for NCK-SCMA in the low SNR region compared to K-SCMA. In this case, it is meaningful to seek a coding based K-SCMA scheme with a higher DO.

Acknowledgement

The authors are deeply indebted to the Associate Editor Dr. Maxime Guillaud and the anonymous Reviewers for many of their insightful comments which have greatly helped improve the quality of this work.

APPENDIX A
PROOF OF PROPOSITION 1

According to (7), the UPEP of NCK-SCMA can be expressed as:

$$\begin{aligned}&\Pr\left\{\mathbf{X}_V \to \hat{\mathbf{X}}_V\right\}\\ &\overset{(a)}{\approx} \frac{1}{12}\prod_{v=1}^{V}\left(4N_0/\Delta_v^2\right)+\frac{1}{6}\prod_{v=1}^{V}\left(3N_0/\Delta_v^2\right)\\ &\overset{(b)}{=}\left(\frac{1}{N_0}\right)^{-D(\mathbf{X}_{V,e})}\left(\frac{4^{-D(\mathbf{X}_{V,e})}}{12}+\frac{3^{-D(\mathbf{X}_{V,e})}}{6}\right)\prod_{v\in d_E(\mathbf{X}_{V,e})}\Delta_v^{-2}\\ &=\Gamma\left(D(\mathbf{X}_{V,e}),N_0\right)\prod_{v\in d_E(\mathbf{X}_{V,e})}\Delta_v^{-2}\overset{(c)}{\le}\Gamma\left(G_d,N_0\right)\prod_{v=1}^{G_d}\Delta_{j,v}^{-2}\end{aligned} \tag{39}$$

where $G_d = \min\left(D(\mathbf{X}_{V,e})\right)$; (a) follows from the approximation of $1+\Delta_v^2/3N_0 \approx \Delta_v^2/3N_0$ and $1+\Delta_v^2/4N_0 \approx \Delta_v^2/4N_0$ in the high SNR region; (b) can be obtained by defining:

$$\begin{aligned}d_E(\mathbf{X}_{V,e}) &\triangleq \{v:\Delta_v \neq 0, 1\le v\le V\}\\ D(\mathbf{X}_{V,e}) &\triangleq \sum_{v=1}^{V}\mathbb{1}\left(\Delta_v\neq 0\right),\end{aligned} \tag{40}$$

$\mathbb{1}(x)$ takes the value of one if condition x is satisfied and zero otherwise; (c) holds after defining $\Delta_{j,v}=\sqrt{|X_{j,v}-\hat{X}_{j,v}|^2}$ for the jth user, and the fact that the PEP performance of SCMA is mainly determined by SUEP. That being said, there is at most one user j among the J interfering users that may be in error. After obtaining (39), the main problem of this proposition is to obtain the DO, i.e., G_d. In order to prove this proposition, two cases should be taken into account.

(i) $R_{in} \ge R_{nc}$. As the calculation of DO is equivalent to obtain $\min\operatorname{rank}\left(\mathbf{X}_{V,e}\right)$ [22], [27]. To achieve this goal, let us assume that $\hat{\mathbf{X}}_V$ contains only one erroneous codeword under SUEP. In other words, there exists a unique $t\in\{1,2,\cdots,T_p\}$ that makes $\mathbf{P}_{j,t}\neq\hat{\mathbf{P}}_{j,t}$ for jth UE. Since

$$\hat{\mathbf{P}}_{j,t_\alpha}\oplus\mathbf{P}_{j,t_\beta}\neq\mathbf{P}_{j,t_\alpha}\oplus\mathbf{P}_{j,t_\beta},\quad \forall\hat{\mathbf{P}}_{j,t_\alpha}\neq\mathbf{P}_{j,t_\alpha} \tag{41}$$

then $\min\operatorname{rank}\left(\mathbf{X}_{V,e}\right)=R_{in}+(T_p-1)R_{nc}=K_{eq}$ in this case.

(ii) $R_{in} < R_{nc}$. In this case, the minimum rank of $\mathbf{X}_{V,e}$, i.e., $d_\mu R_{in} T_p$ can be obtained by taking the case that $\mathbf{P}_{j,t_\alpha}\oplus\mathbf{P}_{j,t_\beta}=\hat{\mathbf{P}}_{j,t_\alpha}\oplus\hat{\mathbf{P}}_{j,t_\beta}$ for any $t_\alpha,t_\beta\in\{1,2,\cdots,T_p\}$ into account. By combining these two cases, the conclusion can be proved.

APPENDIX B
PROOF OF COROLLARY 3

Likewise, we should consider two different cases:

(i) $G_d = d_\mu K_{eq}$. In this case, we have $N_R < K_{rep}T_p$, and $K_{eq} > K_{rep}$, which can be rewritten as:

$$R_{in}+\frac{(T_p-1)R_{nc}}{2} < K_{rep}, \tag{42}$$

and

$$R_{in}+(T_p-1)R_{nc} > K_{rep}, \tag{43}$$

respectively. Therefore, the existence of higher DO is equivalent to finding out a $K_{rep}\in\mathbb{N}^+$ which satisfies $K_{rep}\in\left(R_{in}+\frac{(T_p-1)R_{nc}}{2}, R_{in}+(T_p-1)R_{nc}\right)$. As $R_{in}+(T_p-1)R_{nc}$ is a positive integer, to make sure such a K_{rep} can be found, we have $(T_p-1)R_{nc} > 2$. Note that $R_{nc}\le R_{in}$ according to *Proposition 2*, and thus $R_{in}\ge R_{nc} > \frac{2}{T_p-1}$

(ii) $G_d = d_\mu R_{in} T_p$. In this case, (43) should be modified as $R_{in}T_p > K_{rep}$, then we have

$$R_{in}\left(T_p-1\right)-\frac{(T_p-1)R_{nc}}{2} > 1, \tag{44}$$

which can be also attributed to the reason that K_{rep} is a positive integer. Thus, we can directly obtain $R_{nc} > 2R_{in}+\frac{2}{T_p-1}$. According to (12) in *Proposition 2*, $R_{in}\le R_{nc}$ must be satisfied as $G_d = d_\mu R_{in} T_p$.

Next, we intend to show the restriction of K_{rep} in this corollary. As for case (i), it can be easily spotted that $K_{rep} > R_{in}+1$, and thus $K_{rep}\ge 3$ must be satisifed. On the other hand, in case (ii), it is assumed that when $K_{rep}=2$, then we can simply verify that (42) and $R_{in}T_p > K_{rep}$ cannot be satisfied simultaneously, such that $K_{rep}\ge 3$. After combining the results in (i) and (ii), the corollary can be proved.

APPENDIX C
PROOF OF PROPOSITION 3

The exact ASEP of both NCK-SCMA and K-SCMA are not tractable, the upper bound under SUEP is used for comparison. Without loss of generality, let us first consider the case when the DO of NCK-SCMA can be written as $G_d = d_\mu K_{eq}$. The ASEP of K-SCMA can be obtained immediately by substituting (7) into (19), which can be written as:

$$P_{s,k}\lesssim\tilde{P}_{s,k}=\frac{1}{M}\Gamma\left(K_{rep}d_\mu,N_0\right)\sum_{i=1}^{M(M-1)}\rho_i^{K_{rep}}, \tag{45}$$

where $\tilde{P}_{s,k}$ is the upper bound of $P_{s,k}$ under SUEP, $\rho_i=\prod_{k=1}^{K}\Delta_{j,k}^{-2}$. On the other hand, note that the ASEP of NCK-SCMA is mainly determined by part I in (20). Thereby, as for the ASEP of NCK-SCMA, let us first define $\delta_i=\prod_{v_\alpha=1}^{V}\Delta_{j,v_\alpha}^{-2}$, then we can have:

$$\begin{aligned}&P_{s,nck}\lesssim\tilde{P}_{s,nck}\\ &\overset{(a)}{=}\frac{1}{M^{T_p}}\left(M\tilde{P}_{s,k}+\Gamma\left(G_d,N_0\right)\sum_{i=1}^{|\mathcal{A}|-M(M-1)}\delta_i\right)\\ &\overset{(b)}{>}\frac{1}{M^{T_p}}\left(M\tilde{P}_{s,k}+\frac{1}{3}\Gamma\left(G_d,N_0\right)\left(|\mathcal{A}|-M(M-1)\right)\left(\delta+\delta^*\right)\right)\\ &\overset{(c)}{\ge}M^{1-T_p}\tilde{P}_{s,k}+2M\Gamma\left(R_{in}T_p,N_0\right)\left(M^{T_p-1}-1\right)\delta_{min},\end{aligned} \tag{46}$$

where (a) follows from the fact that the DOs of NCK-SCMA and K-SCMA are the same, and thus $\sum_{i=1}^{M(M-1)}\rho_i^{K_{rep}}$ can be a part of the summations in $\sum_{\mathbf{X}_{j,V_\alpha}}\sum_{\hat{\mathbf{X}}_{j,V_\alpha}}\prod_{v_\alpha=1}^{G_d}\Delta_{j,v_\alpha}^{-2}$; (b) results from the ASEP performance of NCK-SCMA is mainly determined by the terms that $\delta_i > 1$. This can be attributed to the fact that for those $\delta_i < 1$, $\delta_i \ll 1$ holds, especially when the DO is large; (c) is obtained by noticing that $|\mathcal{A}|-M(M-1)=(M^{T_p}(T_p+1)-M)(M-1)$ is obtained when

$K_{eq} = R_{in}T_p$, and defining $\delta_{min} = \min(\delta, \delta^*)$, where δ and δ^* are the $\delta_i > 1$ under the condition that unimolar codebook is utilized. A lower bound of $\tilde{P}_{s,nck}$ is hereto obtained, and thus we only have to prove that the lower bound of $\tilde{P}_{s,nck}$ is higher than $\tilde{P}_{s,k}$. Let us assume that $\tilde{P}_{s,nck} > \tilde{P}_{s,k}$, by substituting (46) into the inequality, and after some algebra manipulations, we have:

$$2M\left(\frac{T_p}{1-M^{1-T_p}}+1\right)\delta_{min} > \sum_{i=1}^{M(M-1)} \rho_i^{K_{rep}} \tag{47}$$

It can be easily spotted that the inequality holds. This completes the proof.

References

[1] F. Guo, F. R. Yu, H. Zhang, X. Li, H. Ji, and V. C. M. Leung, "Enabling massive IoT toward 6G: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 11 891–11 915, Aug. 2021.

[2] L. Yu, Z. Liu, M. Wen, D. Cai, S. Dang, Y. Wang, and P. Xiao, "Sparse code multiple access for 6G wireless communication networks: Recent advances and future directions," *IEEE Communications Standards Magazine*, vol. 5, no. 2, pp. 92–99, Jun. 2021.

[3] H. Nikopour and H. Baligh, "Sparse code multiple access," in *in Proc. IEEE 24th Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2013, pp. 332–336.

[4] A. Ahmed, A. Al-Dweik, Y. Iraqi, H. Mukhtar, M. Naeem, and E. Hossain, "Hybrid automatic repeat request (HARQ) in wireless communications systems and standards: A contemporary survey," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 4, pp. 2711–2752, 4th quarter. 2021.

[5] "5G; NR; Physical layer procedures for data," 3GPP, Tech. Rep. TS 38.214 v15.9.0, Mar. 2020.

[6] T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, I. Z. Kovacs, and P. Mogensen, "System level analysis of K-repetition for uplink grant-free URLLC in 5G NR," in *European Wireless 2019; 25th European Wireless Conference*, Aarhus, Denmark, May. 2019, pp. 1–5.

[7] Y. Liu, Y. Deng, M. Elkashlan, A. Nallanathan, and G. K. Karagiannidis, "Analyzing grant-free access for URLLC service," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 3, pp. 741–755, Mar. 2021.

[8] J. Choi and J. Ding, "Network coding for K-repetition in grant-free random access," *IEEE Wireless Communications Letters*, vol. 10, no. 11, pp. 2557–2561, Nov. 2021.

[9] J. Choi, "Sliding network coding for URLLC," *IEEE Transactions on Wireless Communications*, vol. 21, no. 6, pp. 4424–4433, Dec. 2022.

[10] P. Larsson, B. Smida, T. Koike-Akino, and V. Tarokh, "Analysis of network coded HARQ for multiple unicast flows," *IEEE Transactions on Communications*, vol. 61, no. 2, pp. 722–732, Jan. 2013.

[11] J. Manssour, A. Osseiran, and S. B. Slimane, "A unicast retransmission scheme based on network coding," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 2, pp. 871–876, Dec. 2012.

[12] Y. Lang, D. Wübben, A. Dekorsy, V. Braun, and U. Doetsch, "Improved HARQ based on network coding and its application in LTE," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Paris, France, Jun. 2012, pp. 1958–1963.

[13] M. Woltering, D. Wubben, A. Dekorsy, V. Braun, and U. Doetsch, "Performance of HARQ with reduced size retransmissions using network coding principles," in *IEEE 77th Vehicular Technology Conference (VTC Spring)*, Dresden, Germany, Jan. 2014, pp. 1–6.

[14] K. Xu, W. Ma, L. Zhu, Y. Xu, Y. Gao, D. Zhang, and W. Xie, "NTC-HARQ: Network–turbo-coding based HARQ protocol for wireless broadcasting system," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 10, pp. 4633–4644, Nov. 2015.

[15] J. Jiao, Z. Ni, S. Wu, Y. Wang, and Q. Zhang, "Energy efficient network coding HARQ transmission scheme for S-IoT," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 1, pp. 308–321, Oct. 2021.

[16] Z. Liu and L.-L. Yang, "Sparse or dense: A comparative study of code-domain NOMA systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 4768–4780, Aug. 2021.

[17] Y. Long, Z. Chen, Z. Guo, and J. Fang, "A novel HARQ scheme for SCMA systems," *IEEE Wireless Communications Letters*, vol. 5, no. 5, pp. 452–455, Oct. 2016.

[18] X. Li, Z. Gao, Y. Gui, Z. Liu, P. Xiao, and L. Yu, "Design of power-imbalanced scma codebook," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 2140–2145, Dec. 2022.

[19] Z. Mheich, W. Yu, P. Xiao, A. U. Quddus, and A. Maaref, "On the performance of HARQ protocols with blanking in NOMA systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7423–7438, Nov. 2020.

[20] Z. Shi, C. Zhang, Y. Fu, H. Wang, G. Yang, and S. Ma, "Achievable diversity order of HARQ-aided downlink NOMA systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 471–487, Oct. 2020.

[21] Y. Xu, D. Cai, F. Fang, Z. Ding, C. Shen, and G. Zhu, "Outage constrained power efficient design for downlink NOMA systems with partial harq," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 5188–5201, May. 2020.

[22] J. Bao, Z. Ma, G. K. Karagiannidis, M. Xiao, and Z. Zhu, "Joint multiuser detection of multidimensional constellations over fading channels," *IEEE Transactions on Communications*, vol. 65, no. 1, pp. 161–172, Jan. 2017.

[23] S. Park and S. Choi, "Performance of symbol-level combining and bit-level combining in MIMO multiple ARQ systems," *IEEE Transactions on Communications*, vol. 64, no. 4, pp. 1517–1528, Feb. 2016.

[24] E. W. Jang, J. Lee, H.-L. Lou, and J. M. Cioffi, "On the combining schemes for MIMO systems with hybrid ARQ," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 836–842, Feb. 2009.

[25] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York, NY, USA: Cambridge University Press.

[26] S. Sharma, A. Thomas, and K. Deka, "OTFS-SCMA: A code-domain NOMA approach for orthogonal time frequency space modulation," *IEEE Transactions on Communications*, vol. 69, no. 9, pp. 5043–5058, Aug. 2021.

[27] S.-C. Lim, N. Kim, and H. Park, "Uplink SCMA system with multiple antennas," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 6982–6992, Aug. 2017.

[28] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "SCMA codebook design," in *in Proc. IEEE Veh. Technol. Conf. (VTC-Fall)*, Sep. 2014, pp. 1–5.

[29] Y. Liu, L.-L. Yang, and L. Hanzo, "Spatial modulation aided sparse code-division multiple access," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1474–1487, Mar. 2018.

[30] S. Li, J. Yuan, W. Yuan, Z. Wei, B. Bai, and D. W. K. Ng, "Performance analysis of coded OTFS systems over high-mobility channels," *IEEE Transactions on Wireless Communications*, vol. 20, no. 9, pp. 6033–6048, Sep. 2021.

[31] S. Salamat Ullah, S. C. Liew, G. Liva, and T. Wang, "Short-packet physical-layer network coding," *IEEE Transactions on Communications*, vol. 68, no. 2, pp. 737–751, Feb. 2020.

[32] "5G; NR; Multiplexing and channel coding," 3GPP, Tech. Rep. TS 38.212, Rel. 15, Jul. 2018.

[33] F. Nadeem, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "Non-orthogonal HARQ for URLLC: Design and analysis," *IEEE Internet of Things Journal*, vol. 8, no. 24, pp. 17 596–17 610, Dec. 2021.

[34] L. Xiang, Y. Liu, L.-L. Yang, and L. Hanzo, "Low complexity detection for spatial modulation aided sparse code division multiple access," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 12, pp. 12 858–12 871, Oct. 2021.

[35] Z. Tang, J. Wang, J. Wang, and J. Song, "A low-complexity detection algorithm for uplink NOMA system based on gaussian approximation," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, San Francisco, CA, USA, May. 2017, pp. 1–6.