SIC Aided K-Repetition for Mission-Critical MTC in Cell-Free Massive MIMO

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

25-11-2021 / 06-12-2021

CITATION

Ding, Jie; Choi, Jinho (2021): SIC Aided K-Repetition for Mission-Critical MTC in Cell-Free Massive MIMO. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.17078402.v1>

DOI

[10.36227/techrxiv.17078402.v1](https://doi.org/10.36227/techrxiv.17078402.v1)

SIC Aided K -Repetition for Mission-Critical MTC in Cell-Free Massive MIMO

Jie Ding and Jinho Choi

School of Information Technology, Deakin University, Geelong, VIC 3220, Australia

Email: yxdj2010@gmail.com; jinho.choi@deakin.edu.au

Abstract—In this paper, a successive interference cancellation (SIC) aided K -repetition scheme is proposed to support contention-based mission-critical machine-type communication (MTC) in cell-free (CF) massive multiple-input and multiple-output (MIMO) systems. With the assistance of a tailored deep neural network (DNN) based preamble multiplicity estimator, the proposed SIC in K -repetition is capable of fully cancelling the interference signals, which leads to the reliability improvement in CF massive MIMO. Simulation results show the accuracy of preamble multiplicity estimation by the proposed DNN, and demonstrate that, compared to the existing schemes, the proposed SIC scheme can achieve an improvement of two orders of magnitude in terms of block error rate (BLER) under a given latency constraint. Moreover, when the number of access points (APs) is sufficiently large, employing the proposed SIC scheme provides a great potential to meet ultra-reliable and low-latency requirements, e.g., 10^{-5} BLER and 1 ms access latency, for crowd mission-critical applications, which is far beyond the capabilities of the existing schemes.

I. INTRODUCTION

Machine-type communication (MTC) has been recognized as a key enabler of the Internet of Things (IoT) in the 5th generation (5G) and beyond [1]. Unlike conventional human-type communication, MTC is mainly dominated by uplink traffic driven by a number of devices which usually have low duty cycles while transmitting short bursts of data. In MTC, one of the major categories is referred to as mission-critical MTC, which aims to support services of high reliability under given latency constraints. To meet the ambitious requirements for mission-critical MTC, employing effective transmission schemes based on advanced technologies becomes essential.

MTC devices are usually triggered to perform random access when they have new data packets to transmit. In 5G new radio (NR) standards, grant-free (GF) or two-step random access scheme is introduced for MTC [2], which can reduce access delay and control-signaling overhead compared to traditional grant-based random access. In GF random access, an active device does not wait for a scheduling grant from a base station (BS). That is, once a device becomes active, it is to directly transmit a randomly selected preamble along with data (payload) in a time division multiplexing (TDM) manner and waits for a feedback for acknowledgement from the BS. To enhance the reliability in contention-based random access,

various re-transmission and back-off strategies based on feedback are developed [3], [4]. However, when the delay due to feedback is overwhelming or unacceptable, the K -repetition scheme can be considered, which has been part of the recent 3rd generation partnership project (3GPP) agreements for NR [5]. In GF random access with K -repetition, active devices are allowed to autonomously transmit preamble and data K times before receiving a feedback from the BS. In [6] and [7], the performance of K -repetition was evaluated for mission-critical MTC in GF random access. Nevertheless, these works are simply based on the assumption that each active device has a unique preamble, which overlooks the key impact of preamble collisions¹ on the data detection performance when a number of devices contend for the shared channel resources in GF random access. In practice, preamble collisions do happen in contention-based random access, which result in contaminated channel estimation and thus lead to a potential transmission failure for the collided devices [8]. Therefore, if the preamble collision issue can be addressed, the reliability in K -repetition can be further improved.

Massive multiple-input multiple-output (MIMO) has been considered a promising solution that enables reliable and efficient GF random access with a high bandwidth-utilization efficiency [9]. In co-located massive MIMO, where a large number of antennas are co-located at a BS, two prominent features such as channel hardening and favorable propagation could be asymptotically achieved, which can be exploited to resolve the preamble collisions in K -repetition for reliability enhancement. For instance, in [10], a successive interference cancellations (SIC) based data detection scheme was proposed on a multiple-slot transmission basis in co-located massive MIMO. By utilizing the two prominent features under perfect power control, the scheme applies SIC to resolve the preamble collisions by removing the signals of the decoded devices with their estimated channel powers in all the transmission slots.

Cell-free (CF) massive MIMO, i.e., distributed massive MIMO without the concept of cell boundaries [11], has emerged as an alternative deployment to co-located massive MIMO for the 6th generation (6G) networks [12], where a large number of distributed access points (APs) are employed to serve a number of devices jointly. Recently, various studies

This research was supported by the Australian Government through the Australian Research Council's Discovery Projects funding scheme (DP200100391)

¹A preamble collision happens when multiple active devices select the same preamble.

on CF massive MIMO have been carried out. However, how to exploit CF massive MIMO to support mission-critical MTC in GF random access with K -repetition is still unexplored. In particular, although CF massive MIMO inherits the features of channel hardening and favorable propagation from massive MIMO, their gains are less than those of co-located massive MIMO [13], [14]. Therefore, the existing SIC scheme in [10] may become ineffective when it is directly applied in a CF massive MIMO scenario.

Motivated by the above, this paper proposes an effective SIC based scheme which can work well in the context of CF massive MIMO to enable reliable GF transmissions in K -repetition under a given latency constraint. Both the conventional scheme without SIC and the existing SIC scheme in [10] are considered for performance comparisons. Specifically, we first review their principles and identify their performance bottlenecks when applied in CF massive MIMO. To address the bottlenecks and improve the reliability, we then propose a more powerful SIC scheme in K -repetition that can fully cancel the interference signals from decoded devices and resolve the associated preamble collisions. To make the proposed SIC feasible, enabling accurate preamble multiplicity estimation, i.e., determining how many active devices select the same preamble in a transmission slot, is crucial. To this end, we further propose a simple yet effective deep neural network (DNN) based estimator in CF massive MIMO, which predicts the preamble multiplicity based on the received preamble signals over all the APs.

In simulations, we consider a factory automation scenario with a realistic channel model and evaluate the reliability of the proposed SIC scheme under a stringent access latency constraint, i.e., 1 milliseconds (ms). Numerical results demonstrate that accurate and reliable preamble multiplicity estimation can be achieved by the proposed DNN and the proposed SIC scheme can significantly outperform the other two existing schemes. Moreover, it is shown that when the number of APs is large in CF massive MIMO, employing the proposed SIC scheme is able to simultaneously support a number of mission-critical devices while meeting their stringent reliability requirement, e.g., 10^{-5} , in term of block error rate (BLER) under the 1 ms latency constraint.

Notation: $\mathbf{I}_n$ is the $n \times n$ identity matrix. The conjugate, transpose, and complex conjugate transpose operators are denoted by $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$. $\|\cdot\|$ denotes the Euclidean norm.

II. SYSTEM MODEL

We consider a CF massive MIMO network, where M simple and low-cost APs are spatially distributed to jointly serve a number of sporadically active MTC devices over the shared channel resources for uplink. Each AP is equipped with N antennas and is connected to a BS central processing unit (CPU) via an error-free fronthaul. In GF random access, each active device first transmits a randomly selected preamble from an orthogonal pool of size L in the preamble-transmission phase and then a data payload in the data-transmission phase.

Each data payload consists of a data packet and device's identification number (ID) for device identification [15].

Fig. 1. Illustration of K -repetition in GF random access.

As shown in Fig. 1, the K -repetition scheme is employed to improve the reliability for mission-critical MTC in GF random access, where each active device transmits preamble and the same data K times before receiving a feedback from the BS. The selection of preamble in each transmission slot can be independent. The value of K is set in advance based on a given access latency constraint and frame structure adopted in mission-critical scenarios.

Note that as the K -repetition scheme is used for the purpose of reliability enhancement in this work, we treat the K transmission slots as a time slot block (SB) in GF random access, which are within a coherence interval and can be used only by the mission-critical devices that perform access at the beginning of the SB. In addition, due to the sporadic traffic pattern, we assume that the devices become active randomly and independently and the number of active devices in each SB follows a Poisson distribution with an average of λ_D .

We consider both small-scale fading and independent large-scale fading in the channel model. In particular, the channel vector between AP m and device u is given by

$$\mathbf{h}_{u,m} = \sqrt{\beta_{u,m}} \mathbf{g}_{u,m} \in \mathbb{C}^N, \quad (1)$$

where $\beta_{u,m}$ denotes the large-scale fading coefficient including path loss and shadowing and $\mathbf{g}_{u,m} \sim \mathcal{CN}(0, \mathbf{I}_N)$ stands for the i.i.d Rayleigh small-scale fading vector.

III. CONVENTIONAL SCHEME WITHOUT SIC

In this section, we explain the conventional data detection scheme in K -repetition without using SIC. The conventional scheme detects data based on the received signals in each transmission slot independently and can be considered as a baseline for performance comparison.

A. Channel Estimation

In GF random access, the APs have to estimate devices' channel state information (CSI) by detecting the preambles before attempting to decode any data using coherent decoding. Denote by $\mathcal{A}$ the set of active devices and $\mathbf{p}_l$ the l th orthogonal preamble vector of length L , where $\|\mathbf{p}_l\|^2 = L$ and $\mathbf{p}_l^T \mathbf{p}_{l'}^* = 0$, for $l \neq l'$, $l, l' \in \{1, 2, \dots, L\}$. The received preamble signal, $\mathbf{Y}_k \in \mathbb{C}^{MN \times L}$, over all the APs in the k th transmission slot ($k = \{1, 2, \dots, K\}$) can be given by

$$\mathbf{Y}_k = \sum_{l=1}^L \sum_{u \in \mathcal{Q}_k^l} \sqrt{P_T} \mathbf{h}_u \mathbf{p}_l^T + \mathbf{N}_k, \quad (2)$$

where P_T is the transmit power of each active device, $\mathbf{h}_u = [\mathbf{h}_{u,1}^T, \mathbf{h}_{u,2}^T, \dots, \mathbf{h}_{u,M}^T]^T \in \mathbb{C}^{MN}$ is the channel response vector between active device u and the APs, $\mathcal{Q}_k^l$ denotes the set of active devices selecting preamble l , i.e., $\mathbf{p}_l$, in the k th transmission slot, $\bigcup_{l=1}^L \mathcal{Q}_k^l = \mathcal{A}$, and $\mathbf{N}_k$ is the noise matrix with i.i.d. elements distributed as $\mathcal{CN}(0, \sigma^2)$ in the k th transmission slot.

With (2), the least-squares (LS) based channel estimation can be used for active devices in $\mathcal{Q}_k^l$ and the estimated channel vector over all the APs is given by

$$\phi_{k,l} = \frac{\mathbf{Y}_k \mathbf{p}_l^*}{\sqrt{P_T L}} = \sum_{u \in \mathcal{Q}_k^l} \mathbf{h}_u + \frac{1}{\sqrt{\rho_T L}} \mathbf{n}_{k,l}, \quad (3)$$

where $\rho_T = P_T/\sigma^2$ is defined as the uplink transmit signal-to-noise ratio (SNR) corresponding to each device, and $\mathbf{n}_{k,l} \sim \mathcal{CN}(0, \mathbf{I}_{MN})$. From (3), we see that the estimated channel is a noisy superposition of all the channels of devices selecting the same preamble.

B. Beamforming

Following the preamble, each active device transmits its data in each transmission. The received data symbol vector $\mathbf{r}_k \in \mathbb{C}^{MN}$ over all the APs in the k th transmission slot is given by

$$\mathbf{r}_k = \sum_{u \in \mathcal{A}} \sqrt{P_T} \mathbf{h}_u s_u + \bar{\mathbf{n}}_k, \quad (4)$$

where $\bar{\mathbf{n}}_k$ the background noise vector distributed as $\mathcal{CN}(0, \sigma^2 \mathbf{I}_{MN})$ in the k th transmission slot and s_u is a data symbol transmitted by active device u and $\mathbb{E}[|s_u|^2] = 1$.

Then, all the APs forward their channel estimates and received data signals to the BS CPU for data detection slot by slot. In particular, based on estimated channel $\phi_{k,l}$, $\forall l$, the BS CPU detects data by using receive beamforming on the received data signal $\mathbf{r}_k$, $\forall k$.

For example, with $\phi_{k,l}$, the received data signal after conjugate beamforming [10] [11] corresponding to the devices in $\mathcal{Q}_k^l$ in the k th transmission slot is given by

$$\begin{aligned} z_{k,l} &= \frac{\phi_{k,l}^H \mathbf{r}_k}{\sqrt{P_T}} \\ &= \sum_{u \in \mathcal{Q}_k^l} \left(\|\mathbf{h}_u\|^2 + \sum_{u' \in \mathcal{Q}_k^l \setminus u} \mathbf{h}_{u'}^H \mathbf{h}_u \right) s_u \\ &\quad + \sum_{v \in \mathcal{A} \setminus \mathcal{Q}_k^l} \left(\sum_{u \in \mathcal{Q}_k^l} \mathbf{h}_u^H \mathbf{h}_v \right) s_v + w_k, \end{aligned} \quad (5)$$

where $w_k = \sum_{u \in \mathcal{A}} \frac{\mathbf{n}_{k,l}^H \mathbf{h}_u s_u}{\sqrt{\rho_T L}} + \frac{\phi_{k,l}^H \bar{\mathbf{n}}_k}{\sqrt{P_T}}$ is the filtered noise. In a preamble-collision-free case that only one active device, e.g., device u , selects preamble l in the k th transmission slot, i.e., $|\mathcal{Q}_k^l| = 1$, (5) can be simplified as

$$z_{k,l} = \|\mathbf{h}_u\|^2 s_u + \sum_{v \in \mathcal{A} \setminus u} \mathbf{h}_u^H \mathbf{h}_v s_v + w_k. \quad (6)$$

In theory, when $M \rightarrow \infty$ and based on Chebyshev's theorem, we have [16]

$$\lim_{M \rightarrow \infty} \frac{z_{k,l}}{M} = \sum_{u \in \mathcal{Q}_k^l} \frac{\|\mathbf{h}_u\|^2}{M} s_u, \quad (7)$$

which asymptotically becomes a linear combination of data symbols of all the active devices selecting preamble l . In CF massive MIMO, since devices experience different levels of macro diversity depending on their relative locations to the APs, the channel power of each device, i.e., $\|\mathbf{h}_u\|^2$, can be significantly different even for large M . Thus, in a preamble-collision case of $|\mathcal{Q}_k^l| > 1$, i.e., multiple active devices select preamble l in the k th transmission slot, if one of these collided devices has a dominant channel power, its data is still likely to be successfully decoded. In addition, (7) also indicates that, in a preamble-collision-free case of $|\mathcal{Q}_k^l| = 1$, the received data signal of the unique device selecting preamble l is noise- and interference-free thanks to the favorable propagation when $M \rightarrow \infty$. However, unlike co-located massive MIMO, as revealed in [13], the levels of favorable propagation and channel hardening in a realistic CF massive MIMO network are limited even when M is very large. Therefore, the multi-user interference² in (6), i.e., $\sum_{v \in \mathcal{A} \setminus u} \mathbf{h}_u^H \mathbf{h}_v s_v$, may have a non-negligible impact on the data detection in practice, especially when the channel power of the unique device is small compared to those of other active devices. As a result, even though there is no preamble collision for an active device in a transmission slot, its data may still not be successfully decoded. Therefore, $|\mathcal{Q}_k^l| = 1$ is neither necessary nor sufficient condition for the statement that $z_{k,l}$ is decodable in CF massive MIMO.

IV. PROPOSED SCHEME WITH SIC

From the above analysis, we see that the preamble collision is the major issue that contaminates channel estimation and leads to the transmission failure for the collided devices. Moreover, the devices that have no preamble collision but small channel powers could suffer from strong multi-user interference imposed by other devices and also experience transmission failure due to the weak favorable propagation and channel hardening effects in CF massive MIMO. To address the issues and resolve the preamble collisions for reliability enhancement, a viable solution is to apply SIC by exploiting the repetition time diversity over the K transmission slots.

A. Existing SIC Scheme

The existing SIC scheme that can be directly applied to K -repetition was proposed in [10], which was originally designed in the context of co-located massive MIMO under the condition of perfect power control. In this scheme, as in the conventional scheme, the BS CPU first obtains all the immediately decodable data and identifies decoded devices over the K transmission slots. Then, if $z_{k,l}$ is decodable, it

²Employing other beamforming techniques may help suppress the multi-user interference, the study of which is left for future work.

deems $|\mathcal{Q}_k^l| = 1$ and treats $\|\phi_{k,l}\|^2$ as the estimated channel power of the decoded device. As the decoded data embeds the information of the preamble choices over the K transmission slots, the BS CPU is able to locate all the replicas of the same data sent by the decoded devices and carry out SIC based on the decoded data and estimated channel powers.

For instance, suppose that the data of device u is successfully decoded in $z_{k,l}$. Then, with known s_u and $\|\hat{\mathbf{h}}_u\|^2 = \|\phi_{k,l}\|^2$, the key step of the SIC procedure can be presented as follows for all the pairs of k and l corresponding to $u \in \mathcal{Q}_k^l$:

$$\begin{aligned} z_{k,l} &\leftarrow z_{k,l} - \|\hat{\mathbf{h}}_u\|^2 s_u \\ \|\phi_{k,l}\|^2 &\leftarrow \|\phi_{k,l}\|^2 - \|\hat{\mathbf{h}}_u\|^2. \end{aligned} \quad (8)$$

After SIC to all the decoded data, the BS CPU further obtains new immediately decodable data in the k th transmission slot with updated $z_{k,l}$ and repeats the process until it cannot find any slot with new decoded data.

Since the SIC procedure in (8) was tailored by exploiting the strong favorable propagation and channel hardening effects in co-located massive MIMO under the condition of perfect power control, its viability relies on the two following features: 1) the multi-user interference is negligible, thanks to the favorable propagation; 2) $|\mathcal{Q}_k^l| = 1 \iff z_{k,l}$ is decodable, thanks to the channel hardening with perfect power control.

However, as discussed in Section III-B, CF massive MIMO does not inherit the above two features. Thus, the existing SIC scheme cannot be effective when applied in the context of CF massive MIMO. In particular, as shown in (8), by subtracting $\|\hat{\mathbf{h}}_u\|^2 s_u$ from $z_{k,l}$, the BS CPU could only cancel the coherent interference caused by the replicas from device u due to the preamble collisions. Although this might lead to new cases that new data can be decoded, the signal from device u is not fully cancelled in $z_{k,l}$ and $\mathbf{r}_k$, which still causes multi-user interference to other devices and error propagation that affects the cancellation and decoding performance in consecutive SIC iterations.

B. Proposed SIC Scheme

To fully remove the interference imposed by those decoded devices, a more sensible SIC procedure is to reconstruct the signals of the decoded data with the estimated CSI and remove them from the received data signals and estimated channels, i.e., $\mathbf{r}_k$ and $\phi_{k,l}$, $\forall k, l$. As a result, the preamble collision and multi-user interference issues in CF massive MIMO can be effectively solved.

Specifically, the proposed SIC scheme proceeds as follows. The BS CPU initially obtains all the immediately decodable data and identifies decoded devices over the K transmission slots. For an arbitrary decoded device, e.g., device u , we denote by $\bar{\mathcal{S}}_u$ and $\mathcal{I}_u = \{l_{u,1}, l_{u,2}, \dots, l_{u,K}\}$ the set of slots in which its data is successfully decoded and the set of its selected preamble indices, respectively. Based on the acquired $\mathcal{I}_u$ from the decoded data, the BS CPU performs the preamble multiplicity estimation for preamble $l_{u,k}$ for $k \in \bar{\mathcal{S}}_u$, which is used to estimate the number of active devices selecting

preamble $l_{u,k}$, i.e., $|\mathcal{Q}_k^{l_{u,k}}|$. If $|\mathcal{Q}_k^{l_{u,k}}|$ is estimated as 1, it means that device u has a preamble-collision-free transmission in the k th transmission slot ($k \in \bar{\mathcal{S}}_u$), which results in an interference-free channel estimation. Denote by $\mathcal{S}_u \in \bar{\mathcal{S}}_u$ the set of slots in which device u has preamble-collision-free transmissions. The estimated interference-free channel for device u can be given by

$$\hat{\mathbf{h}}_u = \frac{1}{|\mathcal{S}_u|} \sum_{k \in \mathcal{S}_u} \phi_{k,l_{u,k}} = \mathbf{h}_u + \frac{\sum_{k \in \mathcal{S}_u} \mathbf{n}_{k,l_{u,k}}}{|\mathcal{S}_u| \sqrt{\rho_T L}}, \quad (9)$$

which is distorted by noise only. Thanks to the macro diversity in CF massive MIMO, the impact of noise on the channel estimation at the neighbouring APs of a device can be negligible and the channel estimation is thus reliable in a interference-free case. Then, with the obtained s_u , $\mathcal{I}_u$, and estimated interference-free channel $\hat{\mathbf{h}}_u$, the key steps of the proposed SIC procedure can be presented as follows for all k :

$$\begin{aligned} \phi_{k,l_{u,k}} &\leftarrow \phi_{k,l_{u,k}} - \hat{\mathbf{h}}_u \\ \mathbf{r}_k &\leftarrow \mathbf{r}_k - \sqrt{P_T} \hat{\mathbf{h}}_u s_u. \end{aligned} \quad (10)$$

It is important to remark that locating the preamble-collision-free transmission slots for each decoded device is essential in the proposed SIC scheme. As discussed in Section III-B, in a preamble-collision case of $|\mathcal{Q}_k^l| > 1$, the data of the collided device that has a dominant channel power is still likely to be decoded in CF massive MIMO. If the contaminated channel is used in (10), the coherent and multi-user interference caused by the decoded device cannot be fully removed in all the slots. The caused error propagation inevitably makes the other collided devices undecodable and affects the whole SIC procedure. For this reason, among all the decoded devices, only the ones with $\mathcal{S} \neq \emptyset$ are allowed to be removed in the current SIC iteration with their estimated interference-free channels $\hat{\mathbf{h}}$. For the others with $\mathcal{S} = \emptyset$, they can be removed only when their $\mathcal{S} \neq \emptyset$ in the following SIC iterations. In summary, the proposed SIC procedure is described in Algorithm 1.

C. Preamble Multiplicity Estimation

Since locating the preamble-collision-free transmission slots for each decoded device relies on the preamble multiplicity estimation, a reliable and accurate preamble multiplicity estimator based on the received preamble signals needs to be designed in the context of CF massive MIMO for the proposed SIC scheme. In this paper, for an arbitrary preamble l in an arbitrary transmission slot k , the information of $\phi_{k,l}$ is exploited to obtain the estimated preamble multiplicity, denoted by $|\widehat{\mathcal{Q}}_k^l|$, by using a feed-forward DNN which is well known for powerful prediction capability [17].

In the proposed DNN, as illustrated in Fig. 2, there are $J + 2$ layers, including one input layer (layer 0), J hidden layers, and one output layer (layer $J + 1$). Let x_j denote the number of neurons at layer j , $j = 0, 1, \dots, J + 1$. Layer 0 consists of $x_0 = M$ neurons, which forwards the instantaneous information of $\phi_{k,l}$ to the following layers. In

Algorithm 1 Proposed SIC Scheme

Input: $\mathbf{r}_k$ and $\phi_{k,l}, \forall k, l$.
Initialization: Create empty sets $\mathcal{V}$, $\mathcal{R}$, and $\mathcal{W}$.
 1: **for** $k = 1$ to K **do**
 2: **for** $l = 1$ to L **do**
 3: Obtain $z_{k,l}$ based on (5);
 4: **if** $z_{k,l}$ is decodable **then**
 5: Store the decoded device's data;
 6: **end if**
 7: **end for**
 8: **end for**
 9: Add all the decoded devices' IDs to set $\mathcal{V}$.
 10: For each device $u \in \mathcal{V}$, with $\bar{\mathcal{S}}_u$ and $\mathcal{I}_u$, perform preamble multiplicity estimation to determine its $\mathcal{S}_u$. If $\mathcal{S}_u \neq \emptyset$, add it to set $\mathcal{R}$ and obtain its $\hat{\mathbf{h}}_u$ based on (9); Otherwise, add it to set $\mathcal{W}$;
 11: **while** $\mathcal{R}$ is not empty **do**
 12: Apply SIC to remove the signals of all the decoded devices in $\mathcal{R}$, as presented in (10);
 13: Repeat Steps 1 to 8 to find and store new decoded data.
 14: Empty $\mathcal{V}$ and add all the new decoded devices' IDs to set $\mathcal{V}$;
 15: Update $\mathcal{V} = \mathcal{V} \cup \mathcal{W}$;
 16: Repeat Step 10 to obtain new $\mathcal{R}$ and $\mathcal{W}$.
 17: **end while**

this paper, we utilize the estimated channel power over the M distributed APs as the information. Specifically, from (3), we have $\phi_{k,l} = [\phi_{k,l,1}^T, \phi_{k,l,2}^T, \dots, \phi_{k,l,M}^T]^T$, where $\phi_{k,l,m} \in \mathbb{C}^N$, $m = 1, 2, \dots, M$, represents the estimated channel vector at AP m for preamble l in the k th transmission slot. Based on $\phi_{k,l,m}$, the corresponding estimated channel power is given by

$$\xi_{k,l,m} = \frac{\|\phi_{k,l,m}\|^2}{N}. \quad (11)$$

Let $\xi_{k,l} = [\xi_{k,l,1}, \xi_{k,l,2}, \dots, \xi_{k,l,M}]^T \in \mathbb{R}^M$ denote the estimated channel power vector over all the APs for preamble l in the k th transmission slot. The input vector of layer 0, denoted by $\mathbf{I}_0 = [I_{0,1}, I_{0,2}, \dots, I_{0,M}]^T \in \mathbb{R}^M$, becomes

$$\mathbf{I}_0 = \text{Sort}_D(\xi_{k,l}), \quad (12)$$

where $\text{Sort}_D(\cdot)$ is a function sorting the values of the elements in descending order.

Fig. 2. A diagram of the proposed DNN modelled as a preamble multiplicity estimator.

With $\mathbf{I}_0$, the J hidden layers of the DNN are built by the

following J iterative processing steps:

$$\mathbf{I}_j = \psi(\mathbf{W}_{j-1}\mathbf{I}_{j-1} + \mathbf{b}_{j-1}), j = 1, 2, \dots, J, \quad (13)$$

where $\mathbf{I}_j \in \mathbb{R}^{x_j}$ is the output of layer j , $\psi(v) = \frac{1}{1+e^{-v}}$ represents a non-linear activation function, and $\mathbf{W}_{j-1} \in \mathbb{R}^{x_j \times x_{j-1}}$ and $\mathbf{b}_{j-1} \in \mathbb{R}^{x_j}$ respectively stand for the weighting matrix and bias vector at layer $j-1$, which are used to encode the output of layer $j-1$.

At the end, the output of the proposed DNN in layer $J+1$ is given by

$$\mathbf{I}_{J+1} = \text{Softmax}(\mathbf{W}_J\mathbf{I}_J + \mathbf{b}_J), \quad (14)$$

where $\mathbf{I}_{J+1} = [I_{J+1,1}, I_{J+1,2}, \dots, I_{J+1,x_{J+1}}]^T \in \mathbb{R}^{x_{J+1}}$ and $\text{Softmax}(\cdot)$ is the softmax function. In this paper, x_{J+1} is set to $Q_{\max} + 1$, where $Q_{\max}$ is the maximum number of active devices selecting the same preamble in the network that we are interested in estimating. As $\mathbf{I}_{J+1}$ represents a predicted probability distribution over the $Q_{\max} + 1$ different possible preamble multiplicities, the value of the entry $I_{J+1,q}$ indicates the predicted probability that the preamble multiplicity equals $q-1$, $q = 1, 2, \dots, Q_{\max} + 1$. Therefore, the estimated preamble multiplicity for preamble l in the k th transmission slot is obtained as

$$|\widehat{\mathcal{Q}}_k^l| = \arg \max(\mathbf{I}_{J+1}) - 1, \quad (15)$$

where $\arg \max(\hat{\mathbf{I}}_{J+1})$ returns the index of the largest entry of $\mathbf{I}_{J+1}$.

In this paper, with a sufficient number of training samples, the proposed DNN is trained offline to minimize the cross-entropy loss between the actual and predicted outputs by using back-propagation algorithms. As a result, the parameter sets of all layers including the weighting matrices and bias vectors can be optimized and the DNN model can be constructed for implementation.

Note that the DNN based preamble multiplicity estimation is achieved by exploiting the received preamble signals. Due to the random nature of wireless channels over distributed APs and spatial randomness with respect to the locations of sporadically active devices, estimation error can occur. As mentioned, the information of the preamble choices over the K transmission slots for a decoded device $u \in \mathcal{V}$, i.e., $\mathcal{I}_u$, is made available after its data is successfully decoded at the BS CPU. This information can be utilized to potentially enhance the estimation accuracy to some extent. For instance, when $|\widehat{\mathcal{Q}}_k^l|$ is predicted as 1 for a decoded device, the BS checks if other decoded devices also select preamble l based on their preamble choices in the k th transmission slot. If there do exist other devices selecting preamble l , it means that $|\widehat{\mathcal{Q}}_k^l| > 1$ in practice and the estimation is incorrect. As a result, a false alarm estimation error can be avoided.

V. SIMULATION RESULTS

In this section, the simulation results are presented to evaluate the reliability performance of the proposed SIC scheme in GF random access with K -repetition. In GF random access,

a 7-symbol mini-slot with a subcarrier spacing of 60 KHz is adopted for transmissions. As a result, each slot occupies 0.125 ms. Suppose that a mission-critical application has a target access latency of 1 ms, a maximum number 8 of slots can be exploited in K -repetition. Like [6], we assume that the BS CPU spends 1 slot for processing and 1 slot for transmitting feedback. As a result, $K = 8 - 3 = 5$ slots are used for repetitions. In addition, the data payload length of each device is set to 200. The convolutional coding with rate 1/2 and constraint length 7 and the quadrature phase shift keying (QPSK) modulation scheme are used.

To depict the performance of a realistic network, we consider a factory automation scenario with a square area of 200 m $\times$ 200 m and capitalize on [18] to implement a 3GPP industrial channel model. Within this area, M distributed APs are deployed on the ceiling on a square grid. Each AP has a height of 6 m and is equipped with $N = 4$ antennas. The devices have a height of 2 m. For each realization of K -repetition, the active devices are randomly distributed in the considered area and their number follows the Poisson distribution with an average of λ_D . In the proposed DNN, $Q_{\max}$ is set to 3. 80000 training samples are generated evenly among different multiplicities, which are randomly divided into a training sample set (75% of total instances), a validation sample set (12.5% of total instances), and a test sample set (12.5% of total instances). Based on [18], the large-scale fading coefficient is modeled by $\beta_{u,m} [\text{dB}] = -32.45 - 20 \log_{10}(f_c) - 31.9 \log_{10}(d_{u,m}) - \zeta$, where f_c [GHz] is the carrier frequency, $d_{u,m}$ [m] is the distance between AP m and device u , and ζ represents an independent shadowing which follows a normal distribution, i.e., $\zeta \sim \mathcal{N}(0, 7.56^2)$. The additive thermal noise is assumed to have a power spectral density of -174 dBm/Hz while the front-end receiver at each AP is assumed to have a noise figure of 9 dB. The key parameter setups are summarized in Table I.

TABLE I
PARAMETER SETUP

Symbol	Definition	Value
K	Number of repetitions	5
L	Preamble pool size	32
N	Number of antennas per AP	4
$Q_{\max}$	Maximum multiplicity for estimation	3
f_c	Carrier frequency	3.5 GHz
B	Channel bandwidth	10 MHz
P_T	Transmit power	23 dBm

A. Performance of Preamble Multiplicity Estimation

In Tables II and III, we present the estimation performance of the proposed DNN with $M = 100$ and $M = 400$, respectively. Here, the proposed DNN is constructed by 5 hidden layers. In the case of $M = 100$, the numbers of neurons in the hidden layers are set to 128, 64, 32, 16, and 8, respectively. In the case of $M = 400$, they are set to 256, 128, 64, 32, and 16, respectively. From both tables, it is evident that accurate estimations for a wide range of preamble

multiplicities can be achieved by the proposed DNN and the estimation accuracy enhances as M increases. In particular, for estimating *Multiplicity 1*, i.e., determining whether a decoded device has a preamble-collision-free transmission or not, the accuracy can be increased from 98.7% to 99.7% when M increases from 100 to 400. Furthermore, there are two types of estimation errors for *Multiplicity 1*, i.e., false alarm error and miss detection error. As shown, the false alarm error results from that *Multiplicity 2* is wrongly estimated as *Multiplicity 1*, while the miss detection error results from that *Multiplicity 1* is wrongly estimated as *Multiplicity 2*. Nevertheless, we see that both the estimation error probabilities are very small. For instance, when $M = 400$, the false alarm and miss detection error probabilities are only 0.004 and 0.003, respectively.

TABLE II
CONFUSION MATRIX OF THE DNN FOR PREAMBLE MULTIPLICITY ESTIMATION WITH $M = 100$.

		Estimated			
		0	1	2	3
Target	0	1	0	0	0
	1	0	0.987	0.013	0
	2	0	0.017	0.933	0.050
	3	0	0	0.082	0.918

TABLE III
CONFUSION MATRIX OF THE DNN FOR PREAMBLE MULTIPLICITY ESTIMATION WITH $M = 400$.

		Estimated			
		0	1	2	3
Target	0	1	0	0	0
	1	0	0.997	0.003	0
	2	0	0.004	0.970	0.026
	3	0	0	0.043	0.957

B. Performance of BLER

To evaluate the reliability performance of the proposed SIC scheme within the given access latency, we plot the BLER of an arbitrary active device in Figs. 3 and 4. In Fig. 3, the BLER as a function of M with $\lambda_D = 10$ is presented. As shown, increasing M results in a lower BLER. Moreover, the existing SIC scheme in [10] only provides a small improvement in terms of BLER compared to the conventional scheme without SIC. This verifies our analysis in Section IV-A that the SIC proposed in [10] becomes ineffective due to the weak favorable propagation and channel hardening effects in CF massive MIMO. Compared to the two schemes, the proposed SIC scheme is effective in enhancing the reliability, i.e., improvements of two orders of magnitude in terms of BLER can be achieved over a wide range of M . Furthermore, as M increases to over 350, a BLER of 10^{-5} can be achieved by the proposed SIC scheme. This means that on average 10 mission-critical devices can be accommodated simultaneously over the

shared channel resources in GF random access, while their ultra-reliable and low-latency requirements, e.g., 10^{-5} BLER and 1 ms access latency, can be satisfied.

Fig. 3. BLER as a function of M with $\lambda_D = 10$.

Fig. 4. BLER as a function of λ_D with $M = 256$.

In Fig. 4, the BLER as a function of λ_D with $M = 256$ is depicted. Since a larger λ_D indicates that on average more active devices access the same channel resources for GF transmissions, the BLER increases with λ_D . Nevertheless, even when λ_D is large, e.g., $\lambda_D = 20$, the proposed SIC scheme is still able to offer a BLER as low as 10^{-4} compared to 10^{-2} by the existing SIC scheme in [10]. In addition, to achieve an ultra reliability, e.g., 10^{-5} BLER, we see that the proposed SIC scheme still can support on average $\lambda_D = 5$ mission-critical devices at $M = 256$, which is far beyond the capabilities of the other two schemes.

VI. CONCLUSIONS

In this paper, we proposed a SIC-aided K -repetition scheme in CF massive MIMO to support contention-based mission-critical MTC. Using the proposed DNN based preamble multiplicity estimator, the proposed SIC in K -repetition becomes

effective in resolving preamble collisions and removing the interference imposed by the decoded devices, which thus can significantly enhance the reliability under a given latency constraint compared to the existing schemes. Simulation results, obtained in a realistic mission-critical scenario with CF massive MIMO deployment, showed the accuracy of preamble multiplicity estimation by the proposed DNN and verified the suitability and effectiveness of the proposed SIC scheme. In particular, when the number of APs is sufficiently large, employing the proposed SIC scheme provides a great potential to meet ultra-reliable and low-latency requirements, e.g., 10^{-5} BLER and 1 ms access latency, for crowd mission-critical applications.

REFERENCES

- [1] J. Ding, M. Nemati, C. Ranaweera, and J. Choi, "IoT connectivity technologies and applications: A survey," *IEEE Access*, vol. 8, pp. 67646–67673, 2020.
- [2] 3rd Generation Partnership Project (3GPP), *Evolved Universal Terrestrial Radio Access (EUTRA) and Evolved Universal Terrestrial Radio Access Network (EUTRAN); Overall Description, TS 36.300 v.16.5.0*, March 2021.
- [3] Lenovo and Motorola Mobility, "HARQ design for uplink grant-free transmission," *3GPP R1-1717857*, Oct. 2017.
- [4] J. Choi, "On fast retrieval for two-step random access in MTC," *IEEE Internet of Things Journal*, vol. 8, no. 3, pp. 1428–1436, 2021.
- [5] 3GPP RP-181477, "SID on physical layer enhancements for NR URLLC," Jun. 2018.
- [6] T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, P. Mogensen, I. Z. Kovacs, and T. K. Madsen, "System level analysis of uplink grant-free transmission for URLLC," in *2017 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, 2017.
- [7] N. H. Mahmood, R. Abreu, R. Böhnke, M. Schubert, G. Berardinelli, and T. H. Jacobsen, "Uplink grant-free access solutions for URLLC services in 5G new radio," in *2019 16th International Symposium on Wireless Communication Systems (ISWCS)*, pp. 607–612, 2019.
- [8] J. Ding, M. Nemati, S. R. Pokhrel, O.-S. Park, J. Choi, and F. Adachi, "Enabling grant-free URLLC: An overview of principle and enhancements by massive MIMO," *IEEE Internet of Things Journal*, pp. 1–1, 2021.
- [9] J. Choi, J. Ding, N. P. Le, and Z. Ding, "Grant-free random access in machine-type communication: approaches and challenges," *IEEE Wireless Communications*, pp. 1–6, 2021 (accepted).
- [10] J. H. Sørensen, E. de Carvalho, C. Stefanovic, and P. Popovski, "Coded pilot random access for massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 8035–8046, 2018.
- [11] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834–1850, 2017.
- [12] I. F. Akyildiz, A. Kak, and S. Nie, "6G and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133995–134030, 2020.
- [13] R. Gholami, L. Cottatellucci, and D. Slock, "Favorable propagation and linear multiuser detection for distributed antenna systems," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5190–5194, 2020.
- [14] Z. Chen and E. Björnson, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry," *IEEE Trans. Communications*, vol. 66, pp. 5205–5219, Nov 2018.
- [15] J. Ding and J. Choi, "Comparison of preamble structures for grant-free random access in massive MIMO systems," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 166–170, 2020.
- [16] J. Ding, D. Qu, P. Liu, and J. Choi, "Machine learning enabled preamble collision resolution in distributed massive MIMO," *IEEE Transactions on Communications*, vol. 69, no. 4, pp. 2317–2330, 2021.
- [17] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.
- [18] 3GPP, "Scenarios, frequencies and new field measurement results from two operational factory halls at 3.5 GHz for various antenna configurations," *3GPP R1-1813177*, Nov. 2018.