IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 5 November 2023, accepted 1 December 2023, date of publication 5 December 2023, date of current version 15 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3339576

RESEARCH ARTICLE

Beam Selection for Two-Step Random Access in MTC With a Small Number of Antennas

JINHO CHOI, (Fellow, IEEE)
School of Information Technology, Deakin University, Geelong, VIC 3220, Australia
e-mail: jinho.choi@deakin.edu.au

This work was supported in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korean Government (MSIT) under Grant 2021-0-00794, and in by the Development of 3D Spatial Mobile Communication Technology.

ABSTRACT In 5th generation (5G) systems, two-step random access has been introduced for machine-type communication (MTC) to lower signaling overhead. It is also shown that when a base station (BS) is equipped with a large number of antennas, the notion of massive multiple-input multiple-output (MIMO) can be exploited to improve the performance in terms of throughput and spectral efficiency. In this paper, we consider the case that a BS is equipped with a small number of antennas, in which a sufficiently high spatial selectivity cannot be obtained, and propose an approach to two-step random access based on beam selection that can perform well with a small number of antennas. In the proposed approach, spreading sequences are also used in conjunction with beam selection to mitigate interference due to limited spatial selectivity. To analyze the performance of the proposed approach, the distribution of the signal-to-interference-plus-noise ratio (SINR) is derived as a closed-form expression and the throughput is found. We compare the throughput of the proposed approach with those of conventional two-step random access approaches through analysis and simulations. While the theoretical results agree with simulation results, we can see that the proposed approach outperforms conventional ones when the number of antennas is small for a wide range of traffic intensity.

INDEX TERMS Random access, machine type communication, beam selection.

I. INTRODUCTION

For the Internet of Things (IoT), it is expected to connect a large number of sensors and devices to the Internet. In order to provide devices' connectivity, various approaches are considered [1]. Machine-type communication (MTC) in 5th generation (5G) cellular systems has been actively studied [2], [3], [4] to provide connectivity for a large number of devices with a wide area coverage. Diverse use-cases of MTC in IoT applications are presented in [5]. In [6], various issues of MTC in cellular IoT networks including device- and network-level challenges as well as machine learning-aided solutions are extensively discussed. In MTC, although there are a large number of devices to be connected, in general, they have short data packets with sporadic activity. Thus, random access based transmission schemes are considered for MTC thanks to their low signaling overhead [7], [8].

In order to further reduce signal overhead for short packet transmissions in MTC, two-step random access has been proposed in 5G [9], [10]. Two-step random access is also called grant-free random access, because active devices transmit data packets without obtaining any reserved channel resources [11], [12], [13], which differs from a well-known 4-step random access [7]. As discussed in [14], two-step random access can be used with massive multiple input multiple output (MIMO) [15], [16] to improve the spectral efficiency by exploiting a high spatial selectivity, while its throughput is limited by the number of preambles [17], [18], [19], [20].

Although massive MIMO can provide a number of advantages to two-step random access, in practice, a base station (BS) may not be able to have a large number of antennas because of cost and space. In [21], through an

The associate editor coordinating the review of this manuscript and approving it for publication was Nurul I. Sarkar.

analysis of the signal-to-interference-plus-noise ratio (SINR), it is shown that ideal performance of massive MIMO cannot be expected with a relatively small number of antennas (e.g., few tens).

In this paper, we consider a two-step random access approach that can be used when a BS has only a few antennas. In order to exploit the spatial selectivity with a small number of antennas, beam selection is considered together with spread spectrum to reduce collisions. In the proposed approach, multiple beams are formed, and each active device is to choose one of beams. Each beam is associated with a subset of spreading sequences, which is called a sub-pool. Within a sub-pool, the spreading sequences are orthogonal, while those in different sub-pools are not, which may result in a strong interference. However, because of the spatial selectivity of multiple beams, the interference from other active devices in different beams becomes limited and the resulting SINR can be reasonably high. Note that spreading sequences had been widely used for code division multiple access (CDMA) systems in order to mitigate co-channel interference (CCI) [22]. As in [23] and [24], the effective use of beamforming and spreading can successfully mitigate the CCI in the space-code domain.

The main contributions of the paper are as follows:

i) a two-step random access approach that can effectively exploit the beam space with a small number of antennas is proposed;
ii) in order to understand the performance of the proposed approach, the distribution of SINR is analyzed with an upper-bound and approximations;
iii) the proposed approach is compared with conventional approaches in terms of throughput based on the analysis and simulations.

While the first one is the main one, the other two would be the consequences of the first one. Note that when grant-free random access schemes exploit the space domain, they usually require massive MIMO, e.g., the approaches in [11] and [12], to exploit channel hardening and favorable propagation [16]. However, when the number of antennas is relative small (which may be a more realistic case), such approaches cannot be employed. The proposed approach is to exploit the space domain with a small number of antennas using beam selection: to the best of our knowledge, there is no similar approach to the proposed approach.

The rest of the paper is organized as follows. In Section II, we review two-step random access and discuss its limitations. The proposed approach is presented with its system model in Section III. The SINR is analyzed in Section IV and the throughput is obtained and compared with those of conventional approaches in Section V. We present simulation results in Section VI and conclude the paper with remarks in Section VII.

NOTATION

Matrices and vectors are denoted by upper- and lower-case boldface letters, respectively. The superscripts T and H denote the transpose and complex conjugate, respectively. $\mathbb{E}[\cdot]$ and $\mathrm{V}ar(\cdot)$ denote the statistical expectation and variance, respectively. $\mathcal{CN}(\mathbf{a}, \mathbf{R})$ represents the distribution of circularly symmetric complex Gaussian (CSCG) random vectors with mean vector $\mathbf{a}$ and covariance matrix $\mathbf{R}$.

II. BACKGROUND

In this section, we briefly explain grant-free or two-step random access that is employed for MTC in 5G [9] and discuss its limitations.

A. TWO-STEP RANDOM ACCESS

In two-step random access, active devices transmit signals to a BS in the first step, and in the second step, the BS informs devices with the outcomes of transmissions (i.e., acknowlegement (ACK) for successful transmissions or negative ACK (NACK) for unsuccessful transmissions). For the first step that has preamble transmission phase (PTP) and data transmission phase (DTP) as shown in Fig. 1, one time slot can be used. Each active device chooses a preamble from a pool of L preambles, $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_L\}$, where $\mathbf{c}_l$ represents the lth preamble, uniformly at random, and transmits it in the PTP. In the DTP, a data packet is then transmitted. Since all the devices are synchronized, it is expected that the length of data packet is the same for all devices (or the length of data transmission phase is decided by the maximum length of data packet among all the devices).

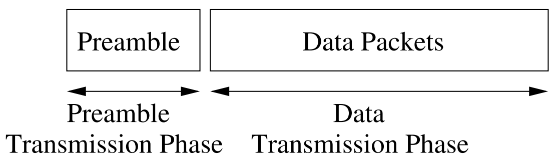


FIGURE 1. Two phases in a slot (i.e., the PTP and DTP) for two-step random access.

B. LIMITATIONS

In two-step random access, the size of resource block for DTP is proportional to the number of preambles [10]. In particular, we can assume that there are L mini-slots for DTP, and there is a one-to-one correspondence between L preambles and L mini-slots. For example, if an active device chooses the lth preamble, it transmits its data packet through the lth mini-slot during DTP. The resulting two-step random access approach is referred to as type-1 approach (i.e., the approach in 5G standards [10]) for convenience. In type-1 approach, if an active device experiences a preamble collision, its packet will be collided with the packets transmitted by the active devices that choose the same preamble. This implies that the performance is limited by the number of preambles, L, while L cannot be arbitrarily large as more resources are required. To see this, consider time division multiplexing (TDM) for

L mini-slots and assume that the preambles are orthogonal. In this case, the length of preambles is L (in unit time[1]) and the length of slot (in unit time) becomes

$$T = L + LB = L(1 + B), \tag{1}$$

where B is the length of mini-slot or data packet, which shows that T increases with L. In addition, T is limited to be less than the coherence time.

If the BS is equipped with multiple antennas, it is possible to support multiple active devices in the same resource block by exploiting the spatial selectivity [14]. That is, there is no need to have L mini-slots, and the spectral efficiency can be significantly improved. The resulting approach is referred to as type-2 approach (i.e., the approach in [14]), in which the length of slot becomes

$$T = L + B. \tag{2}$$

Compared to type-1 approach, type-2 approach can have a shorter slot length or a longer data packet (with the same length of slot), which can lead to a higher throughput. Unfortunately, if the number of antennas at the BS is not sufficiently large to exploit the spatial selectivity, type-2 approach may not be efficient unless the number of active devices is small.

In general, type-2 approach is preferable to type-1 when the BS is equipped with a large number of antennas as the spatial selectivity can be well exploited. On the other hand, if the BS has one antenna, there is no choice but to use the type 1 approach. However, as mentioned above, if the number of antennas is small,[2] type-2 approach may fail to provide a good performance, while type-1 approach is unable to exploit multiple antennas. In this paper, we propose a two-step random access approach different from the conventional ones that can provide better performance when the conventional approaches do not provide good performance.

III. SYSTEM MODEL

In this section, we present the system model for the proposed approach with one BS equipped with M antennas, where $M \geq 1$, and a number of devices. Each device is equipped with a single antenna. Throughout the paper, we assume that time division duplexing (TDD) is assumed so that the channel reciprocity can be exploited as in massive MIMO [15], [16], [25], while M is not sufficiently large so as to exploit the key properties of massive MIMO such as channel hardening effect.

A. THE PROPOSED APPROACH

As mentioned earlier, the proposed approach is based on the channel reciprocity in order to effectively use beam-space. A slot is divided into two sub-slots. In the first sub-slot, the BS sends orthogonal pilot signals through D beams, where each beam is characterized by beamforming vector d, $\mathbf{w}_d \in \mathbb{C}^M$, in order to allow active devices to choose one of the D beams as illustrated in Fig. 2. That is, the BS transmits the following signal through M antennas: $\sum_{d=1}^{D} \mathbf{b}_d \mathbf{w}_d^{\mathrm{H}} \in \mathbb{C}^{D \times M}$, where the $\mathbf{b}_d$'s are orthogonal beacon signals of length D. For convenience, this phase is referred to as the beam-space probing phase (BSPP).

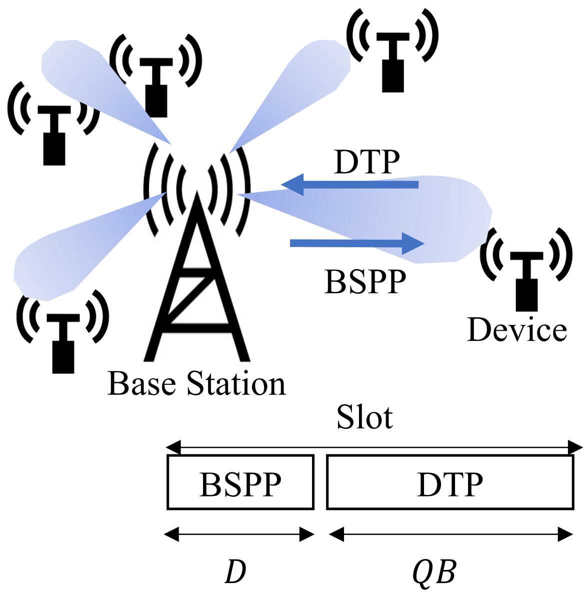


FIGURE 2. An illustration of the proposed approach based on beam selection with a slot consisting of BSPP and DTP.

The second sub-slot is used for DTP where active devices transmit their data packets. During DTP, the BS also uses the same set of D beams to receive the signals from active devices thanks to the channel reciprocity. As in CDMA [22], in order to differentiate transmitted packets from different active devices, spreading sequences are used, which will be explained in the next subsection. The length of slot, T, in the proposed approach becomes

$$T = D + QB, \tag{3}$$

where Q is the length of the spreading sequences or spreading gain. However, unlike conventional CDMA, each active device does not have own spreading sequence, but chooses a spreading sequence that is associated with its selected beam.

In summary, the proposed approach has the following key steps:

- **S1**: (Downlink) The BS broadcasts orthogonal pilots through D beams, which is the BSSP as shown in Fig. 1.
- **S2**: (Uplink) Each active device chooses a beam among D and transmit its packet with spreading, which is the DTP (see also Fig. 1). We will explain how to select a beam at each device in Subsection III-C.

[1]The unit time is equivalent to the symbol duration, which is inversely proportional to the system bandwidth.

[2]For type-2 approach, it is expected to exploit channel hardening and favorable propagation from massive MIMO [16], which is based on the law of large numbers. Thus, if the number of antennas is not large enough, these properties cannot be exploited. Thus, we focus on the case of a small number of antennas, say up to 10.

B. BEAM SELECTION ASSOCIATED WITH SPREADING SEQUENCES

In the proposed approach, we aim to associate multiple beam spaces or beams with subsets of spreading sequences. To this end, we assume that the set of spreading sequences is divided into D subsets as follows:

$$\mathcal{C} = \cup_{d=1}^{D} \mathcal{C}_d, \; \mathcal{C}_d \cap \mathcal{C}_{d'} = \emptyset, \; d \neq d', \tag{4}$$

where $\mathcal{C}_d$ represents the dth subset of spreading sequences, which is referred as the dth sub-pool of spreading sequences. For convenience, let $\mathbf{c}_{d,q}$ denote the qth spreading sequence in sub-pool d. We assume that $|\mathcal{C}_d| = Q$, i.e., the number of spreading sequences per beam is Q.

Throughout the paper, we assume that the spreading sequences in a sub-pool are orthogonal, while those in different sub-pools are not. Thus, the length of the spreading sequences is Q. For a specific design, we can consider Alltop sequences [26], [27] (Zadoff-Chu sequences [28] can also be used). Let $\boldsymbol{\phi}_l$ denote the lth Alltop sequence of length Q with $||\boldsymbol{\phi}_l|| = 1$ for all l, where $Q \geq 5$ is a prime. Then, the dth sub-pool becomes $\mathcal{C}_d = \{\boldsymbol{\phi}_{Q(d-1)+1}, \ldots, \boldsymbol{\phi}_{Qd}\}$, $d \in \{1, \ldots, D\}$, i.e., $\mathbf{c}_{d,q} = \boldsymbol{\phi}_{Q(d-1)+q}$. Since Alltop sequences are used, $D \leq Q$ and the correlation between two spreading sequences in different sub-pools is $\frac{1}{\sqrt{Q}}$.

In Fig. 3, we show D sub-pools of spreading sequences associated with D beams. If active devices in the same beam choose different spreading sequences, they do not interfere with each other. On the other hand, active devices in different beams interfere with each other as the inter-correlation of sub-pools of spreading sequences is $\frac{1}{\sqrt{Q}}$. However, since they are in different beams, the their channels are less correlated in the spatial domain and interference is mitigated.

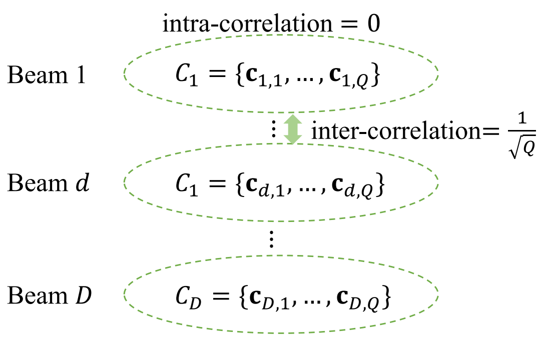


FIGURE 3. Sub-pools of spreading sequences associated with *D* beams.

C. BEAM SELECTION AT DEVICES

Suppose that device k is active and receives the signal in the BSPP from the BS. Then, the received signal is given by

$$\mathbf{y}_k = \left(\sum_{d=1}^{D} \mathbf{b}_d \mathbf{w}_d^{\mathrm{H}} \right) \mathbf{h}_k + \mathbf{n}_k, \tag{5}$$

where $\mathbf{h}_k \in \mathbb{C}^{M \times 1}$ is the channel vector from the BS to device k and $\mathbf{n}_k \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$ is the background noise at device k. Here, σ^2 is the noise variance and $[\mathbf{h}_k]_m$ represents the channel coefficient from the mth antenna at the BS to device k. The output of the dth correlator with $\mathbf{b}_d$ at device k is given by

$$\begin{aligned} v_{k,d} &= \mathbf{b}_d^{\mathrm{H}} \mathbf{y}_k \\ &= \mathbf{w}_d^{\mathrm{H}} \mathbf{h}_k + \mathbf{b}_d^{\mathrm{H}} \mathbf{n}_k, \end{aligned} \tag{6}$$

which can be seen as an estimate of the *composite-channel* coefficient, $\mathbf{w}_d^{\mathrm{H}} \mathbf{h}_k$. Then, device k can use $v_{k,d}$ to choose the spreading sequence sub-pool corresponding to the largest composite-channel gain as follows:

$$d(k) = \underset{d \in \{1, \ldots, D\}}{\operatorname{argmax}} \; |v_{k,d}|^2, \tag{7}$$

where $d(k)$ is the index of the selected spreading sequence sub-pool at device k.

D. RECEIVED SIGNALS AT BS

As mentioned earlier, spreading sequences are used for active devices to transmit their packets. Suppose that an active device, say device k, chooses spreading sequence $\mathbf{c}_l = [c_{l,0} \; \ldots \; c_{l,Q-1}]^{\mathrm{T}} \in \mathbb{C}^Q$. Since the same spreading sequence is used to spread every data symbols in a packet, the spread signal for the nth data symbol, denoted by $s_{k,n}$, is usually given by [22]

$$x_{k,nQ+q} = c_{l,q} s_{k,n}, \; q = 0, \ldots, Q-1 \text{ and } n = 0, \ldots, B-1,$$

or

$$\mathbf{x}_{k,n} = [x_{k,nQ} \; \ldots \; x_{k,(n+1)Q-1}]^{\mathrm{T}} = \mathbf{c}_l s_{k,n}.$$

Thus, in this subsection, we only consider the received signal for one data symbol. For convenience, the data symbol index n is omitted.

Let $\mathcal{K}$ be the index set of active devices. Then, thanks to the channel reciprocity, the uplink channel vector is the same as the downlink channel vector, $\mathbf{h}_k$, and the received signal during DTP at the BS can be given by

$$\mathbf{R} = \sum_{k \in \mathcal{K}} s_k \mathbf{c}_{l(k)} \mathbf{h}_k^{\mathrm{T}} + \mathbf{N} \in \mathbb{C}^{Q \times M}, \tag{8}$$

where $l(k)$ represents the index of the spreading sequence chosen by device k and $\mathbf{N}$ is the background noise with $[\mathbf{N}]_{q,m} \sim \mathcal{CN}(0, N_0)$. Here, $[\mathbf{R}]_{q,m}$ and $[\mathbf{N}]_{q,m}$ represent the received signal and background noise at the qth symbol duration of the PTP through the mth antenna, respectively. The output of the dth beamformer is given by

$$\begin{aligned} \mathbf{r}_d &= \mathbf{R} \mathbf{w}_d^* \\ &= \left(\sum_{k \in \mathcal{K}} s_k \mathbf{c}_{l(k)} \mathbf{h}_k^{\mathrm{T}} \right) \mathbf{w}_d^* + \tilde{\mathbf{n}}_d \\ &= \sum_{k \in \mathcal{K}_d} s_k \mathbf{c}_{l(k)} \mathbf{h}_k^{\mathrm{T}} \mathbf{w}_d^* + \sum_{k \in \bar{\mathcal{K}}_d} s_k \mathbf{c}_{l(k)} \mathbf{h}_k^{\mathrm{T}} \mathbf{w}_d^* + \tilde{\mathbf{n}}_d, \end{aligned} \tag{9}$$

where $\tilde{\mathbf{n}}_d = \mathbf{N} \mathbf{w}_d^*$, $\mathcal{K}_d$ represents the index set of the active devices that choose the dth sub-pool or the dth beam, and $\bar{\mathcal{K}}_d = \mathcal{K} \setminus \mathcal{K}_d$.

In (9), according to (7), for $k \in \mathcal{K}_d$, the composite-channel gain, $|\mathbf{h}_k^{\mathrm{T}} \mathbf{w}_d^*|^2$, would be high, while it would be low for $k \in \bar{\mathcal{K}}_d$. That is, the spatial selectivity can be exploited using beam selection. As a result, the interference from the other active devices choosing different sub-pools or beams, $\sum_{k \in \bar{\mathcal{K}}_d} \mathbf{c}_{l(k)} \mathbf{h}_k^{\mathrm{T}} \mathbf{w}_d^*$, is expected to be insignificant, which may lead to a reasonably high SINR (unless there are other active devices in the same beam choosing the same spreading sequence). For convenience, this interference is referred to as the inter-pool interference, and the SINR will be discussed in the following section.

IV. SINR ANALYSIS

In this section, we derive the outage probability of the SINR, which is also the cumulative distribution function (cdf), from (9) and explain the advantage of the association between sub-pools and beams in Subsection III-B.

In each spreading sequence sub-pool, there are Q spreading sequences. Thus, in order to despread the transmitted spreading sequences, the output of each beamformer becomes the input to a bank of Q correlators and the output of the qth correlator is given by

$$\begin{aligned} z_{d,q} &= \mathbf{c}_{d,q}^{\mathrm{H}} \mathbf{r}_d \\ &= \mathbf{c}_{d,q}^{\mathrm{H}} \sum_{k \in \mathcal{K}_d} s_k \mathbf{c}_{l(k)} \mathbf{h}_k^{\mathrm{T}} \mathbf{w}_d^* + \mathbf{c}_{d,q}^{\mathrm{H}} (\mathbf{u}_d + \tilde{\mathbf{n}}_d), \end{aligned} \tag{10}$$

where $\mathbf{u}_d = \sum_{k \in \bar{\mathcal{K}}_d} s_k \mathbf{c}_{l(k)} \mathbf{h}_k^{\mathrm{T}} \mathbf{w}_d^*$ is the dth inter-pool interference.

For $k \in \mathcal{K}_d$, $\mathbf{c}_{l(k)}$ is one of the spreading sequences in $\mathcal{C}_d$. Thus, for the signal term in (10), there is one of the following cases:

C0: No active device choosing $\mathbf{c}_{d,q}$: In this case, we have

$$\mathbf{c}_{d,q}^{\mathrm{H}} \sum_{k \in \mathcal{K}_d} \mathbf{c}_{l(k)} \mathbf{h}_k^{\mathrm{T}} \mathbf{w}_d^* = 0.$$

C1: Only one active device choosing $\mathbf{c}_{d,q}$: In this case, we have

$$\mathbf{c}_{d,q}^{\mathrm{H}} \sum_{k \in \mathcal{K}_d} \mathbf{c}_{l(k)} \mathbf{h}_k^{\mathrm{T}} \mathbf{w}_d^* = \mathbf{h}_k^{\mathrm{T}} \mathbf{w}_d. \tag{11}$$

If the SINR is sufficiently high, the BS can decode the signal from the active device.

C2: Multiple active devices choosing $\mathbf{c}_{d,q}$: In this case, a spreading sequence collision happens, which means that the SINR can be low due to the interference from the other active devices that choose the same spreading sequence. Thus, we assume that the BS fails to decode the packets.

Clearly, we are only interested in Case of **C1** to find the SINR, which has the signal power, $|\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_d^*|^2$, for $k \in \mathcal{K}_d$ where $|\mathcal{K}_d| = 1$, according to (11).

In (10), due to the cross-correlation between spreading sequences in different sub-pools, the inter-pool interference becomes

$$\begin{aligned} \mathbf{c}_{d,q}^{\mathrm{H}} \mathbf{u}_d &= \mathbf{c}_{d,q}^{\mathrm{H}} \sum_{k \in \bar{\mathcal{K}}_d} s_k \mathbf{c}_{l(k)} \mathbf{h}_k^{\mathrm{T}} \mathbf{w}_d^* \\ &= \frac{1}{\sqrt{Q}} \sum_{k \in \bar{\mathcal{K}}_d} e^{j\theta_{k,d,q}} s_k \mathbf{h}_k^{\mathrm{T}} \mathbf{w}_d^*, \end{aligned} \tag{12}$$

where $\theta_{k,d,q} \angle(\mathbf{c}_{d,q}^{\mathrm{H}} \mathbf{c}_{l(k)}) \in [0, 2\pi)$.

For convenience, let $K_d = |\mathcal{K}_d|$ and $\bar{K}_d = |\bar{\mathcal{K}}_d|$, and we only focus on the output of the first beamformer to derive the SINR, i.e., $d = 1$, without loss of generality. In addition, let $k(1)$ be the index of the active device that chooses the first beam, assume that s_k is independent and identically distributed and has zero mean. Denote by $E_s = \mathbb{E}[|s_k|^2]$ the symbol energy. Then, from (10) and (12), for Case of **C1**, the SINR becomes

$$\begin{aligned} \mathsf{SINR} &= \frac{E_s |\mathbf{h}_{k(1)}^{\mathrm{H}} \mathbf{w}_1|^2}{\frac{E_s}{Q} \sum_{k \in \bar{\mathcal{K}}_1} |\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_1|^2 + \mathbb{E}[\mathbf{c}_{d,q}^{\mathrm{H}} \mathbf{N} \mathbf{w}_1 \mathbf{w}_1^{\mathrm{H}} \mathbf{N} \mathbf{c}_{d,q}]} \\ &= \frac{|\mathbf{h}_{k(1)}^{\mathrm{H}} \mathbf{w}_1|^2}{\frac{1}{Q} \sum_{k \in \bar{\mathcal{K}}_1} |\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_1|^2 + \frac{1}{\gamma} ||\mathbf{w}_1||^2}, \end{aligned} \tag{13}$$

where $\gamma = \frac{E_s}{N_0}$ is the signal-to-noise ratio (SNR). Since the channel vectors are random, the SINR in (13) is a random variable. For a coded packet, we assume that the BS can succeed to decode it if the SINR is greater than a threshold that is denoted by $\Gamma > 0$. As a result, the outage probability, which is also the cdf of SINR (if $<$ is replaced with $\leq$ in the following equation), is given by

$$P_{out}(\Gamma, K) = \Pr(\mathsf{SINR} < \Gamma \,|\, K), \tag{14}$$

which is the probability that the BS fails to decode under Case of **C1** when there are $K = |\mathcal{K}|$ active devices. That is, although there is no spreading sequence collision, the BS can fail to decode as the SINR is not sufficiently high. In what follows, we will find the outage probability.

For tractable analysis, we consider the following assumptions.

A1: The beams, $\mathbf{w}_d$'s, are orthonormal, i.e., $\mathbf{w}_d^{\mathrm{H}} \mathbf{w}_{d'} = \delta_{d,d'}$. This implies that $D \leq M$ as $\mathbf{w}_d \in \mathbb{C}^M$.

A2: The channels are modeled as independent Rayleigh fading channels so that

$$\mathbf{h}_k \sim \mathcal{CN}(0, \mathbf{I}). \tag{15}$$

Assumption of **A1** is valid when random orthogonal beams are employed. In MTC random access, the BS is unable to know if a device is active before it sends a signal. As a result, the locations of active devices are unknown, and beams cannot be formed towards them. Thus, random orthogonal beams might be natural to use.

The channel gain of an active device is in general dependent on the distance between the BS and the device. Thus, in (15), the large-scale fading term has to be included. However, we drop this under the assumption that each active device is able to perform open-loop power control to compensate the large-scale fading term [29]. With ideal power control, the resulting channel coefficients from the BS's perspective can be expressed by the small-scale fading terms with Rayleigh fading as in (15).

We can derive an upper-bound on the outage probability in (14) under the above conditions and assumptions as follows.

Lemma 1: Suppose that Assumptions of **A1** and **A2** hold. For a given $K = |\mathcal{K}|$, we have the following upper-bound on the outage probability:

$$P_{out}(\Gamma, K) \leq \sum_{d=0}^{D} \binom{D}{d} (-e^{-\frac{\Gamma}{\gamma}})^d \left(\frac{DQ+\Gamma d}{D(Q+\Gamma d)}\right)^{K-1}. \quad (16)$$

Proof: See Appendix A. ■

For given Γ and K, based on (16), an approximate outage probability can also be found as follows.

Lemma 2:

$$P_{out}(\Gamma, K) \approx \sum_{d=0}^{D} \binom{D}{d} (-e^{-\frac{\Gamma}{\gamma}})^d \left(\frac{D\bar{Q}+\Gamma d}{D(\bar{Q}+\Gamma d)}\right)^{K-1}, \quad (17)$$

where $\bar{Q} = \frac{Q}{\phi}$ that is referred to as the effective spreading gain. Here, $\phi = \frac{D - \sum_{n=1}^{D} \frac{1}{n}}{D-1}$.

Proof: See Appendix B. ■

Since $\phi \leq 1$, we can see the effective spreading gain is greater than or equal to the spreading gain, i.e., $\bar{Q} \geq Q$. From this, we see that the approximate outage probability in (17) is lower than the upper-bound in (16).

In order to gain insight into the outage probability, a simple expression for the outage probability can be considered if $\Gamma D < \bar{Q}$ so that it leads to the following approximation:

$$\left(\frac{1+\frac{\Gamma d}{D\bar{Q}}}{1+\frac{\Gamma d}{\bar{Q}}}\right)^{K-1} \approx e^{-\frac{d\Gamma(K-1)}{\bar{Q}}\left(1-\frac{1}{D}\right)}.$$

From this, (17) can be further approximated as

$$\begin{aligned} P_{out}(\Gamma, K) &\approx \sum_{d=0}^{D} \binom{D}{d} (-e^{-\frac{\Gamma}{\gamma}})^d \left(\frac{1+\frac{\Gamma d}{D\bar{Q}}}{1+\frac{\Gamma d}{\bar{Q}}}\right)^{K-1} \\ &\approx \left(1 - e^{-\Gamma\left(\frac{1}{\gamma}+\frac{(K-1)\left(1-\frac{1}{D}\right)}{\bar{Q}}\right)}\right)^{D}. \end{aligned} \quad (18)$$

As shown in (18), thanks to the beam selection, we can have a diversity gain of D in the outage probability [30], while the inter-pool interference is mitigated by a factor of $\bar{Q}$ or Q, which is the spreading gain.

In Fig. 4, we show the cdf of the SINR when $D = M = 4$, $Q = 11$, $K = 10$, and $\gamma = 10$ dB. We can see that the approximation in (17) is quite close to the empirical result. In addition, (18) is also tight when the outage probability is sufficiently low (i.e., ≤ 0.4).

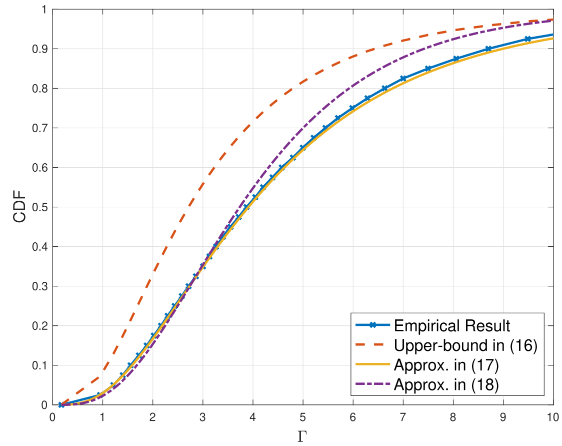


FIGURE 4. **The empirical cdf of the SINR with the upper-bound in (16), the approximations in (17) and (18) when $D = M = 4$, $Q = 11$, $K = 10$, and $\gamma = 10$ dB.**

V. SUCCESS PROBABILITY AND THROUGHPUT

In this section, we focus on the success probability from an active device's perspective and derive the throughput that represents the system performance. We also compare the performance of the proposed approach with the conventional approaches.

A. PERFORMANCE OF THE PROPOSED APPROACH

In the previous sections, the number of active devices, K, is assumed to be fixed. However, in practice, the number of active devices varies and can be seen as a random variable. As discussed in Assumptions of **A1** and **A2**, for random channels and orthogonal beams, we assume that the number of the active devices that choose beam d is independent. Thus, K can be seen as the sum of D independent random variables as follows: $K = \sum_{d=1}^{D} K_d$. Provided that K follows a Poisson distribution with mean λ (here, λ is the traffic intensity or the average number of active devices per slot) [31], we have the following assumption.

A3: The K_d's are independent and

$$K_d \sim \text{Poiss}\left(\frac{\lambda}{D}\right), \quad d = 1, \ldots, D. \quad (19)$$

Since the sum of independent Poisson random variables is also Poisson [32], (19) is valid when K is Poisson and each active device can choose one of D beams equally likely. With (19), we will find the success probability and throughput in this section.

The success probability is the probability that *an active device* can successfully transmit its data packet. The event of successful packet transmission happens if *i)* Case of **C1** holds and *ii)* the SINR is greater than or equal to Γ. Based on the two above conditions, the success probability can be found as follows.

Lemma 3: Under Assumption of **A3**, the success probability is given by

$$P_{succ} = (1 - P_{out}(\Gamma)) \frac{Q}{Q-1} \frac{e^{-\frac{\lambda}{QD}} - e^{-\frac{\lambda}{D}}}{1 - e^{-\frac{\lambda}{D}}}. \quad (20)$$

Here,

$$P_{out}(\Gamma) = \mathbb{E}[\Pr(\mathsf{SINR} < \Gamma \mid \bar{K}_1)], \tag{21}$$

where the expectation is carried out over $\bar{K}_1$, its upper-bound and approximation are given by

$$P_{out}(\Gamma) \leq \sum_{d=0}^{D} \binom{D}{d} (-e^{-\frac{\Gamma}{\gamma}})^d e^{-\lambda \frac{D-1}{D} \frac{\Gamma d}{Q+\Gamma d}} \tag{22}$$

$$P_{out}(\Gamma) \approx \sum_{d=0}^{D} \binom{D}{d} (-e^{-\frac{\Gamma}{\gamma}})^d e^{-\lambda \frac{D-1}{D} \frac{\Gamma d}{Q+\Gamma d}}. \tag{23}$$

Note that if $Q = 1$, it can be shown that

$$P_{succ} = (1 - P_{out}(\Gamma)) \frac{\frac{\lambda}{D} e^{-\frac{\lambda}{D}}}{1 - e^{-\frac{\lambda}{D}}}.$$

Proof: See Appendix C. ∎

In Fig. 5, we show the success probability in (20) for different values of traffic intensity, λ, when $Q = 11$, $M = D = 4$, $\gamma = 10$ dB, and $\Gamma = 6$ dB. It is shown that the success probability decreases with λ as there are more active devices. When the upper-bound in (22) is used, we have a lower-bound on the success probability. An approximate success probability can also be found with (23), which is quite close to simulation results. Thus, in finding the throughput below, we will use (22).

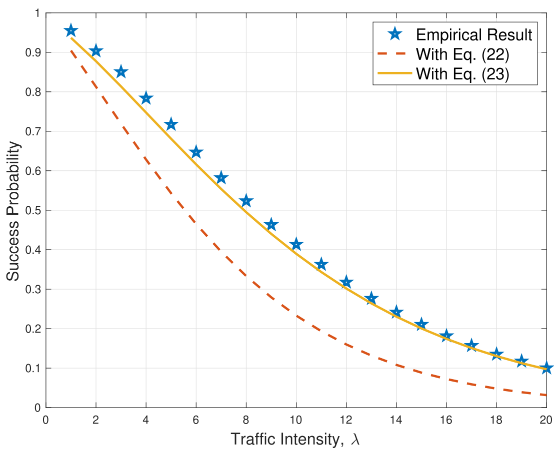


FIGURE 5. Success probability for different values of traffic intensity, λ, when $Q = 11$, $M = D = 4$, $\gamma = 10$ dB, and $\Gamma = 6$ dB.

While the success probability in (20) is a performance metric from an active device's perspective, the throughput is the average number of the active devices that succeed to transmit their packets, which is a performance metric from the system's point of view. In the following result, the throughput is given.

Lemma 4: Under Assumption of **A3**, the throughput is given by

$$\eta_{\text{prop}} = (1 - P_{out}(\Gamma))\, \lambda e^{-\frac{\lambda}{QD}}. \tag{24}$$

Proof: See Appendix D, ∎

B. THROUGHPUT COMPARISON WITH THE CONVENTIONAL APPROACHES

In this subsection, we consider two conventional approaches, i.e., type-1 and type-2 approaches, for comparisons.

We assume that the preambles are orthogonal in type-1 and type-2 approaches and the number of active devices, K, follows the Poisson distribution according to Assumption of **A3**. For type-1 approach, we consider one mini-slot and the SINR when there is only one active device (in this case, the SINR is actually the SNR) without any interference, which has the following outage probability:

$$\Pr(\mathsf{SINR} < \Gamma) = \Pr(\gamma ||\mathbf{h}_k||^2 < \Gamma) = \Pr\left(\chi^2_{2M} < \frac{2\Gamma}{\gamma}\right), \tag{25}$$

where χ^2_n represents a chi-square random variable with n degrees of freedom, according to Assumption of *A*2. This shows that type-1 approach fully exploits the antenna diversity gain to increase the SINR. Then, the throughput of type-1 approach, which is the average number of active devices that can successfully transmit their data packets, becomes

$$\begin{aligned} \eta_1 &= \mathbb{E}\left[(1 - \Pr(\mathsf{SINR} < \Gamma))\, K \left(1 - \frac{1}{L}\right)^{K-1} \right] \\ &= \left(1 - \Pr\left(\chi^2_{2M} < \frac{2\Gamma}{\gamma}\right)\right) \lambda e^{-\frac{\lambda}{L}}. \end{aligned} \tag{26}$$

For comparisons, consider (24) and (26). If the outage probability is sufficiently low, we can have $\eta_{prop} \approx \lambda e^{-\frac{\lambda}{QD}} \leq QDe^{-1}$ and $\eta_1 \approx \lambda e^{-\frac{\lambda}{L}} \leq Le^{-1}$. Thus, if $Q = L$, we can see that the proposed approach can have a higher maximum throughput than type-1 approach by a factor of D.

In type-2 approach, the signals transmitted by K active devices co-exist and the distribution of the SINR obtained in [21] for a small number of antennas can be used. Since the distribution of the SINR depends on K, the throughput of type-2 approach becomes

$$\begin{aligned} \eta_2 &= \mathbb{E}\left[(1 - \Pr(\mathsf{SINR} < \Gamma \mid K))\, K \left(1 - \frac{1}{L}\right)^{K-1} \right] \\ &= \lambda e^{-\frac{\lambda}{L}} - e^{-\lambda} \\ &\quad \times \sum_{k=1}^{\infty} \frac{\Pr(\mathsf{SINR} < \Gamma | k) \left(\lambda \left(1 - \frac{1}{L}\right)\right)^{k-1}}{(k-1)!}, \end{aligned} \tag{27}$$

under the assumption that the BS fails to decode the signals transmitted by the active devices associated with preamble collisions.

For fair comparisons, we need to consider the normalized throughput. As shown in Fig. 1, the length of data packet in data transmission phase becomes $T - L$. Thus, assuming that the slot consists of T unit times, the normalized throughput (i.e., the average number of successfully transmitted packets per unit time) or spectral efficiency is given by

$$\kappa_1 = \frac{T-L}{TL}\eta_1 \tag{28}$$
$$\kappa_2 = \frac{T-L}{T}\eta_2, \tag{29}$$

In κ_1, we have L in the denominator as the DTP has L mini-slots. Clearly, type-2 approach can have a higher normalized throughput than type-1 approach by a factor of L. In particular, in type-1 approach, when M is sufficiently large, we may have $\eta_1 = \lambda e^{-\frac{\lambda}{L}}$. Thus, from (28), the normalized throughput becomes

$$\kappa_1 = \frac{T-L}{T}\frac{\lambda}{L}e^{-\frac{\lambda}{L}} \leq \frac{T-L}{T}e^{-1},$$

where the inequality is obtained using $xe^{-x} \leq e^{-1}$ for $x \geq 0$. This shows that the normalized throughput cannot exceed e^{-1} (although $T \to \infty$), which is also the maximum throughput of ALOHA. On the other hand, in type-2 approach, if $\eta_2 \to \lambda e^{-\frac{\lambda}{L}}$ for a large number of antennas, from (29), we can see that κ_2 becomes

$$\kappa_2 = \frac{T-L}{T}L\frac{\lambda}{L}e^{-\frac{\lambda}{L}} \leq \frac{T-L}{T}Le^{-1} \leq Le^{-1},$$

which just confirms that the normalized throughput of type-2 approach can be L-time higher than that of type-1 approach. However, if M is small and K is relatively large, the SINR of type-2 approach may not be high. In this case, the normalized throughput of type-1 approach can be higher than that of type-2 approach (in Section VI, we will see more details).

The normalized throughput of the proposed approach is given by

$$\kappa_{prop} = \frac{T-D}{TQ}\eta_{prop}. \tag{30}$$

Here, we have Q in the denominator due to the spreading gain Q. By comparing (30) and (28), we can see that the normalized throughput of the proposed approach may be similar to that of type-1 approach if $L = Q$. However, if η_{prop} is higher than η_1, the proposed approach can have a higher normalized throughput than type-1 approach.

VI. SIMULATION RESULTS

In this section, we present simulation results to compare the performance of the three different approaches, namely type-1, type-2, and the proposed approaches, in terms of the normalized throughput. For simulations, Assumptions of **A1**, **A2**, and **A3** are used to generate the channel vectors and the number of active devices. For convenience, we assume that the number of preambles, L, in type-1 and type-2 approaches is equal to the number of spreading sequences per beam, Q, in the proposed approach, i.e., $L = Q$. In addition, the length of slot is fixed as $T = 128$. Note that the outage probability of SINR in (22) is used to find the normalized throughput of the proposed approach.

Fig. 6 shows the normalized throughput curves as functions of traffic intensity, λ, when $T = 128$, $Q = 17$, $M = D = 4$, $\gamma = 10$ dB, and $\Gamma = 4$ dB. Type-2 approach provides a high normalized throughput when λ is small (i.e., ≤ 5). As mentioned earlier, if λ is small, the spatial selectivity with $M = 4$ antennas could be sufficient to support a relatively small number of active devices within the shared channel resource in type-2 approach. However, as λ increases, there are more active devices that result in a low SINR and a low normalized throughput. Thus, for a large λ (i.e.,, $\lambda \geq 5$), we can see that the proposed and type-1 approaches perform better than type-2 approach. Furthermore, the proposed approach outperforms type-1 approach for a wide range of λ.

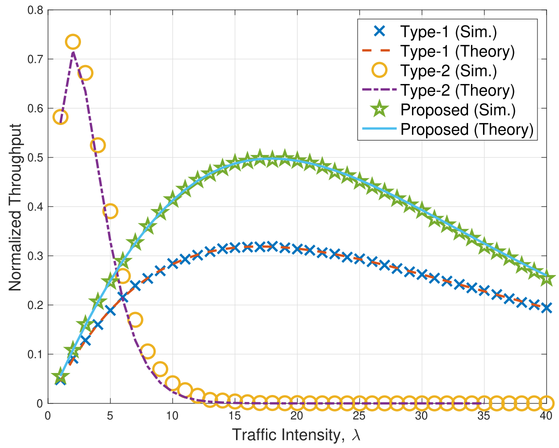


FIGURE 6. Normalized throughput curves as functions of traffic intensity, λ, of type-1, type-2, and proposed approaches when $T = 128$, $Q = 17$, $M = D = 4$, $\gamma = 10$ dB, and $\Gamma = 4$ dB.

It is noteworthy that as mentioned earlier, the normalized throughput of type-1 cannot exceed $e^{-1} \approx 0.367$, while the other approaches can have a higher normalized throughput than e^{-1} as shown in Fig. 6.

Fig. 7 shows the normalized throughput curves as functions of the number of antennas, $M = D$, when $T = 128$, $Q = 17$, $\lambda = 10$, $\gamma = 10$ dB, and $\Gamma = 4$ dB. While the normalized throughput of type-2 increases rapidly with M, its performance is quite limited when M is small (e.g., ≤ 10). As mentioned earlier, this results from a low SINR as multiple signals co-exist in the shared channel resource in type-2 approach. It is also observed that the proposed approach performs better than type-1 approach for a wide range of M. Clearly, this shows that the proposed approach is a better option for two-step random access than the others when M is not large.

As the threshold SINR, Γ, increases, the BS can fail to decode a packet although no collision happens. Thus, we can expect that the normalized throughput decreases with Γ. Fig. 8 shows the normalized throughput curves as functions of target SINR, Γ, when $T = 128$, $Q = 17$, $M = D = 4$, $\lambda = 10$, and $\gamma = 10$ dB. As expected, the normalized throughput decreases with Γ. In particular, a sharp decline can be observed in type-2 approach. On the other hand, in type-1 approach, the decline is negligible as the SINR can be sufficiently high thanks to the antenna diversity gain. In the

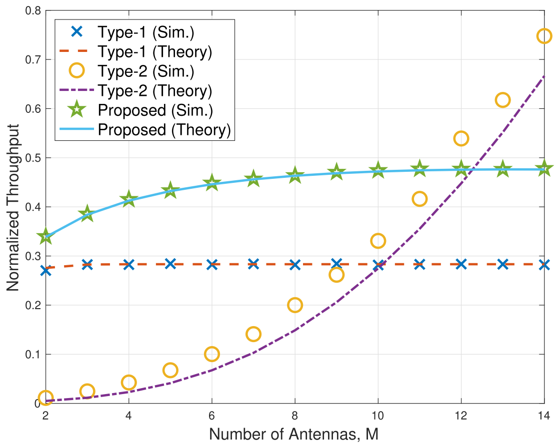


FIGURE 7. **Normalized throughput curves as functions of the number of antennas, $M = D$, of type-1, type-2, and proposed approaches when $T = 128$, $Q = 17$, $\lambda = 10$, $\gamma = 10$ dB, and $\Gamma = 4$ dB.**

proposed approach, due to the spreading gain, the decline is not significant compared to that of type-2 approach.

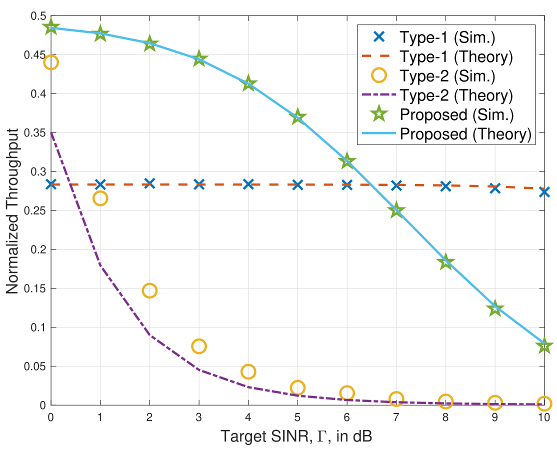


FIGURE 8. **Normalized throughput curves as functions of target SINR, Γ, of type-1, type-2, and proposed approaches when $T = 128$, $Q = 17$, $M = D = 4$, $\lambda = 10$, and $\gamma = 10$ dB.**

Since $T = 128$ is fixed, when Q increases, the duration of DTP decreases, which may result in the decrease in the normalized throughput. Fig. 9 shows the normalized throughput curves as functions of the number of preambles per beam, $Q = L$, when $T = 128$, $M = D = 4$, $\lambda = 10$, $\gamma = 10$ dB, and $\Gamma = 4$ dB. Since the length of Alltop sequences has to be a prime, we only consider primes less than 50 in Fig. 9. In general, as Q increases from 1, the probability of collisions decreases and the normalized throughput increases. However, as mentioned earlier, the increase of Q leads to the decrease of the duration of DTP and eventually decreases the normalized throughput as shown in Fig. 9. In general, for a wide range of Q, we can see that the proposed approach outperforms type-1 approach. Note that type-2 approach does not provide a good performance and is less sensitive to Q in this case, since its SINR is not sufficiently high due to a relatively large number of active devices (i.e., $\lambda = 10 > M = 4$).

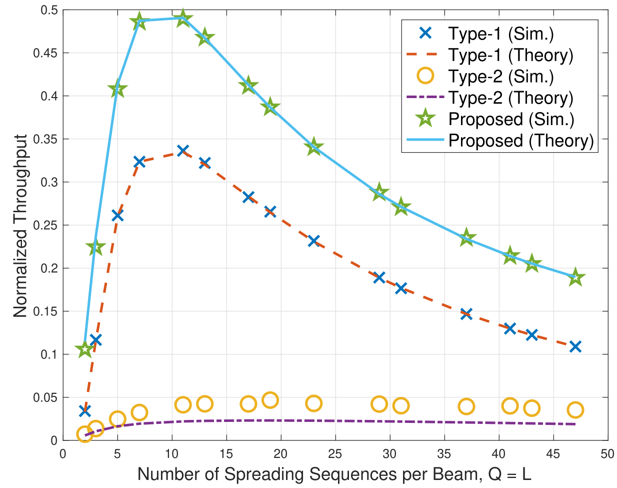


FIGURE 9. **Normalized throughput curves as functions of the number of preambles per beam, $Q = L$, of type-1, type-2, and proposed approaches when $T = 128$, $M = D = 4$, $\lambda = 10$, $\gamma = 10$ dB, and $\Gamma = 4$ dB.**

VII. CONCLUDING REMARKS

Provided that a BS is equipped with a large number of antennas, the spatial selectivity can be exploited to support a large number of devices in two-step random access (where the channel vectors of multiple devices are asymptotically orthogonal) for massive connectivity in IoT applications such as smart farming and smart cities. However, the spatial selectivity becomes limited as the number of antennas decreases. Thus, different approaches to exploit the spatial selectivity would be desirable. In particular, most MTC technologies, such as narrowband (NB)-IoT for cellular IoT applications, use relatively low carrier frequencies (i.e., sub-GHz), making it difficult to equip BSs with large antenna arrays.

In this paper, based on beam selection to effectively exploit the spatial selectivity with a small number of antennas at a BS, we proposed a two-step random access approach that can provide a good performance. Spreading sequences have also been used in conjunction with beam selection to mitigate interference. In order to see the performance of the proposed approach, we derived the outage probability of SINR and obtained the throughput. The proposed approach was compared with conventional approaches in terms of normalized throughput, and it was shown that the proposed approach can perform better than conventional approaches when the number of antennas is small.

There are a number of issues to be investigated in the future. For example, an active device can select a beam based on not only its channel gain, but also the likelihood of collision with other active devices. This may require learning of past beam selection by other devices, and reinforcement learning [33] can be applied.

APPENDIX A
PROOF OF LEMMA 1

For convenience, let $\mathbf{h}_1 = \mathbf{h}_{k(1)}$ and $\mathbf{h}_k = \mathbf{g}_k$ for $k \in \bar{\mathcal{K}}_1$ in (13). According to (7), it can be shown that

$$|\mathbf{h}_1^{\mathrm{H}}\mathbf{w}_1|^2 \geq \max_{d \in \{2,\ldots,D\}} |\mathbf{h}_1^{\mathrm{H}}\mathbf{w}_d|^2, \tag{31}$$

$$|\mathbf{g}_k^{\mathrm{H}}\mathbf{w}_1|^2 \leq \max_{d \in \{2,\ldots,D\}} |\mathbf{g}_k^{\mathrm{H}}\mathbf{w}_d|^2, \tag{32}$$

i.e., (31) is valid as the active device chooses the beam that has the maximum channel gain and (32) is valid as all the active devices in $\bar{\mathcal{K}}_1$ choose different beams than beam 1, $\mathbf{w}_1$. Under Assumptions of **A1** and **A2**, we can see that $|\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_d|^2 \sim \mathrm{Exp}(1)$, i.e., an exponential random variable. Thus, from (31), letting $X = |\mathbf{h}_1^{\mathrm{H}}\mathbf{w}_1|^2$, we can see that $\Pr(X \leq x) = \left(1 - e^{-x}\right)^D$, i.e., X is seen as the largest order statistics for given D independent exponential random variables [34]. Let

$$Y = \sum_{k \in \bar{\mathcal{K}}_1} |\mathbf{g}_k^{\mathrm{H}}\mathbf{w}_1|^2 \tag{33}$$

be the interference term in (13). Then, with $c = \frac{1}{\gamma}$, we have

$$\begin{aligned} P_{out}(\Gamma) &= \Pr\left(\frac{X}{\frac{Y}{Q} + c} < \Gamma\right) \\ &= \mathbb{E}\left[\Pr\left(X < \left(\frac{Y}{Q} + c\right)\Gamma \,\middle|\, Y\right)\right] \\ &= \mathbb{E}\left[\left(1 - e^{-\left(\frac{Y}{Q}+c\right)\Gamma}\right)^D\right]. \end{aligned} \tag{34}$$

According to (32), we see that $|\mathbf{g}_k^{\mathrm{H}}\mathbf{w}_1|^2$ is any order statistics other than the largest order statistics. Thus, we have

$$\Pr(Y \leq y) \geq \Pr\left(\sum_{m=1}^{\bar{K}_1} X_m \leq y\right), \tag{35}$$

where $X_m \sim \mathrm{Exp}(1)$ is iid. Letting $Y' = \sum_{m=1}^{\bar{K}_1} X_m$, we can say that Y' is stochastically larger than Y [35]. Since the function $\left(1 - e^{-\left(\frac{Y}{Q}+c\right)\Gamma}\right)^D$ in (34) is an increasing function of Y, we have

$$\mathbb{E}\left[\left(1 - e^{-\left(\frac{Y}{Q}+c\right)\Gamma}\right)^D\right] \leq \mathbb{E}\left[\left(1 - e^{-\left(\frac{Y'}{Q}+c\right)\Gamma}\right)^D\right] \tag{36}$$

according to [35, Proposition 8.1.2]. Substituting (36) into (34), for a given $\bar{K}_1 = |\bar{\mathcal{K}}_1|$, it can be shown that

$$\begin{aligned} &\Pr(\mathsf{SINR} < \Gamma \,|\, \bar{K}_1) \\ &\quad\leq \sum_{d=0}^{D} \binom{D}{d} (-e^{-\frac{\Gamma}{\gamma}})^d \mathbb{E}[e^{-\frac{\Gamma d}{Q} Y'}] \\ &\quad= \sum_{d=0}^{D} \binom{D}{d} (-e^{-\frac{\Gamma}{\gamma}})^d \left(\frac{1}{1 + \frac{\Gamma d}{Q}}\right)^{\bar{K}_1}. \end{aligned} \tag{37}$$

Under Assumptions of **A1** and **A2**, since beams and active devices can be seen as bins and balls, respectively, the event that $\bar{K}_1$ devices (or balls) choose beams (or bins) $2, \ldots, D$, has the following probability:

$$\Pr(\bar{K}_1 = k) = \binom{K-1}{k} \left(1 - \frac{1}{D}\right)^k \left(\frac{1}{D}\right)^{K-1-k}, \tag{38}$$

as there are $K - 1$ other active devices excluding the active device of interest. Applying (38) to (37), we have

$$\begin{aligned} P_{out}(\Gamma, K) &= \sum_{k=0}^{K} \Pr(\mathsf{SINR} < \Gamma \,|\, \bar{K}_1 = k) \Pr(\bar{K}_1 = k) \\ &\leq \sum_{d=0}^{D} \binom{D}{d} (-e^{-\frac{\Gamma}{\gamma}})^d \left(\frac{DQ + \Gamma d}{D(Q + \Gamma d)}\right)^{K-1}, \end{aligned} \tag{39}$$

which is given in (16). This completes the proof.

APPENDIX B
PROOF OF LEMMA 2

Due to the inequality in (35), we have an upper-bound on the outage probability as in (16). Without using (35), we can have a tight approximation. To this end, we need the following result.

Proposition 1: Let Y in (33) be $Y = \sum_{i=1}^{\bar{K}_1} Y_i$, where Y_i is the ith elements of $\{|\mathbf{g}_k^{\mathrm{H}}\mathbf{w}_1|^2, k \in \bar{\mathcal{K}}_1\}$. Then, the mean of Y_i is given by

$$\mathbb{E}[Y_i] = \frac{D - \sum_{n=1}^{D} \frac{1}{n}}{D - 1} \ (\leq 1). \tag{40}$$

Proof: From Assumptions of **A1** and **A2**, $|\mathbf{g}_k^{\mathrm{H}}\mathbf{w}_d|^2$ is seen as an independent exponential random variable with mean 1. Then, with $d = 1$, according to (32), Y_i becomes an order statistics except the maximum. Let $X_{(1)}, \ldots, X_{(D)}$ be the order statistics of D independent exponential random variables with mean 1, where $X_{(d)}$ represents the dth smallest order statistics, i.e., $X_{(1)} \leq \ldots \leq X_{(D)}$. Then, Y_i can be one of $X_{(d)}$ except $d = D$. From this, we have

$$(D - 1)\mathbb{E}[Y_i] + \mathbb{E}[X_{(D)}] = \sum_{d=1}^{D} \mathbb{E}[X_{(d)}] = D. \tag{41}$$

Since the mean of the largest order statistics is given by $\mathbb{E}[X_{(D)}] = \sum_{d=1}^{D} \frac{1}{d}$ [34], $\mathbb{E}[Y_i]$ can be given as in (40), which completes the proof. ■

For an approximation, we now assume that Y_i is an independent exponential random variable with mean $\phi = \frac{D - \sum_{n=1}^{D} \frac{1}{n}}{D-1}$ as in (40). With this approximation, (37) can be replaced with the following:

$$\Pr(\mathsf{SINR} < \Gamma \,|\, \bar{K}_1) \approx \sum_{d=0}^{D} \binom{D}{d} (-e^{-\frac{\Gamma}{\gamma}})^d \left(\frac{1}{1 + \frac{\phi \Gamma d}{Q}}\right)^{\bar{K}_1}. \tag{42}$$

Then, after some manipulations, we have the following approximation of the outage probability:

$$P_{out}(\Gamma, K) \approx \sum_{d=0}^{D} \binom{D}{d} (-e^{-\frac{\Gamma}{\gamma}})^d \left(\frac{D\bar{Q} + \Gamma d}{D(\bar{Q} + \Gamma d)} \right)^{K-1}, \quad (43)$$

which is given in (17) and completes the proof.

APPENDIX C
PROOF OF LEMMA 3

For given K_1, the conditional probability of the first condition, i.e., **C1**, is the probability that the active device of interest does not experience spreading sequence collision, which is given by

$$P_{\rm nc}(K_1) = \left(1 - \frac{1}{Q}\right)^{K_1 - 1}. \quad (44)$$

If the first condition holds, the probability of the second condition (i.e., $\mathsf{SINR} \geq \Gamma$) for given $\bar{K}_1$ is $1 - \Pr(\mathsf{SINR} < \Gamma | \bar{K}_1)$. As a result, for given K_1 and $\bar{K}_1$, the conditional success probability is given by

$$P_{succ}(K_1, \bar{K}_1) = (1 - \Pr(\mathsf{SINR} < \Gamma | \bar{K}_1)) P_{\rm nc}(K_1). \quad (45)$$

Then, the success probability becomes

$$\begin{aligned} P_{succ} &= \mathbb{E}[P_{succ}(K_1, \bar{K}_1)] \\ &= \mathbb{E}[1 - \Pr(\mathsf{SINR} < \Gamma | \bar{K}_1)] \mathbb{E}[P_{\rm nc}(K_1) \,|\, K_1 \geq 1], \end{aligned} \quad (46)$$

where the second equality is valid, because the K_d's are independent as in (19) according to Assumption of **A3**. That is, since $\Pr(\mathsf{SINR} < \Gamma | \bar{K}_1)$ and $P_{\rm nc}(K_1)$ are functions of $\bar{K}_1 = K_2 + \ldots + K_D$ and K_1, respectively, and the K_d's are independent, the expectation can be carried out for each term separately (as $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ if X and Y are independent). The condition that $K_1 \geq 1$ is necessary for the probability of no spreading sequence collision as there should be at least one active device choosing beam 1. From (19) and (44), we have

$$\begin{aligned} \mathbb{E}[P_{\rm nc}(K_1) \,|\, K_1 \geq 1] &= \sum_{k=1}^{\infty} P_{\rm nc}(k) \frac{\Pr(K_1 = k)}{\Pr(K_1 \geq 1)} \\ &= \sum_{k=1}^{\infty} \left(1 - \frac{1}{Q}\right)^{k-1} \frac{e^{-\frac{\lambda}{D}} \left(\frac{\lambda}{D}\right)^k}{k!(1 - e^{-\frac{\lambda}{D}})} \\ &= \frac{Q}{Q-1} \frac{e^{-\frac{\lambda}{QD}} - e^{-\frac{\lambda}{D}}}{1 - e^{-\frac{\lambda}{D}}}. \end{aligned} \quad (47)$$

From (37), since $\bar{K}_1 \sim \mathrm{Poiss}\left(\lambda \frac{D-1}{D}\right)$, for the the first term in (46), we have

$$\begin{aligned} P_{out}(\Gamma) &\leq \sum_{d=0}^{D} \binom{D}{d} (-e^{-\frac{\Gamma}{\gamma}})^d \mathbb{E}\left[\left(\frac{1}{1 + \frac{\Gamma d}{Q}} \right)^{\bar{K}_1} \right] \\ &= \sum_{d=0}^{D} \binom{D}{d} (-e^{-\frac{\Gamma}{\gamma}})^d e^{-\lambda \frac{D-1}{D} \frac{\Gamma d}{Q + \Gamma d}}, \end{aligned} \quad (48)$$

which is the same as in (22). An approximation can also be found using (42), which leads to (23).

APPENDIX D
PROOF OF LEMMA 4

According to Assumption of **A3**, each beam has identical conditions. Thus, we need to consider only one beam, say beam 1, when finding the throughput.

For given K_1 and $\bar{K}_1$, the number of the active devices that succeed to transmit their packets through beam 1 is given by

$$\eta_1(K_1, \bar{K}_1) = \left(1 - \Pr(\mathsf{SINR} < \Gamma | \bar{K}_1)\right) K_1 P_{\rm nc}(K_1). \quad (49)$$

Since K_1 and $\bar{K}_1$ are independent, the mean becomes

$$\begin{aligned} \eta_1 &= \mathbb{E}[\eta_1(K_1, \bar{K}_1)] = (1 - P_{out}(\Gamma))\, \mathbb{E}\left[K_1 P_{\rm nc}(K_1)\right] \\ &= (1 - P_{out}(\Gamma)) \frac{\lambda}{D} e^{-\frac{\lambda}{QD}}. \end{aligned} \quad (50)$$

Since there are D beams, the throughput becomes

$$\eta = D\eta_1, \quad (51)$$

which is (24). This completes the proof.

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] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.

[3] Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews, and E. Yaacoub, "Toward massive machine type cellular communications," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 120–128, Feb. 2017.

[4] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 615–637, Mar. 2021.

[5] H. Shariatmadari, R. Ratasuk, S. Iraji, A. Laya, T. Taleb, R. Jäntti, and A. Ghosh, "Machine-type communications: Current status and future perspectives toward 5G systems," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 10–17, Sep. 2015.

[6] S. K. Sharma and X. Wang, "Toward massive machine type communications in ultra-dense cellular IoT networks: Current issues and machine learning-assisted solutions," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 426–471, 1st Quart., 2020.

[7] *Study on RAN Improvments for Machine-Type Communications*, document TR 37.868, V11.0, 3GPP, Oct. 2011.

[8] *Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) Protocol Specification*, document TS 36.321, V13.2.0, 3GPP, Jun. 2016.

[9] *Evolved Universal Terrestrial Radio Access (EUTRA) and Evolved Universal Terrestrial Radio Access Network (EUTRAN); Overall Description*, document TS 36.300, V.14.7.0, 3GPP, Jun. 2018.

[10] J. Kim, G. Lee, S. Kim, T. Taleb, S. Choi, and S. Bahk, "Two-step random access for 5G system: Latest trends and challenges," *IEEE Netw.*, vol. 35, no. 1, pp. 273–279, Jan. 2021.

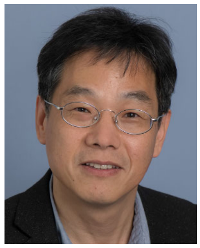
[11] K. Senel and E. G. Larsson, "Grant-free massive MTC-enabled massive MIMO: A compressive sensing approach," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6164–6175, Dec. 2018.

[12] A. T. Abebe and C. G. Kang, "MIMO-based reliable grant-free massive access with QoS differentiation for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 773–787, Mar. 2021.

[13] C. G. Kang, A. T. Abebe, and J. Choi, "NOMA-based grant-free massive access for latency-critical Internet of Things: A scalable and reliable framework," *IEEE Internet Things Mag.*, vol. 6, no. 3, pp. 12–18, Sep. 2023.

[14] J. Choi, J. Ding, N.-P. Le, and Z. Ding, "Grant-free random access in machine-type communication: Approaches and challenges," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 151–158, Feb. 2022.

[15] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
[16] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.
[17] E. de Carvalho, E. Björnson, J. H. Sørensen, E. G. Larsson, and P. Popovski, "Random pilot and data access in massive MIMO for machine-type communications," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 7703–7717, Dec. 2017.
[18] J. Ding, D. Qu, H. Jiang, and T. Jiang, "Success probability of grant-free random access with massive MIMO," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 506–516, Feb. 2019.
[19] J. Ding, D. Qu, and J. Choi, "Analysis of non-orthogonal sequences for grant-free RA with massive MIMO," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 150–160, Jan. 2020.
[20] J. Choi, "An approach to preamble collision reduction in grant-free random access with massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1557–1566, Mar. 2021.
[21] J. Choi, "On the SINR distribution of grant-free random access with moderate MIMO," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1389–1392, Sep. 2020.
[22] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*. Reading, MA, USA: Addison-Wesley, 1995.
[23] A. Yener, R. D. Yates, and S. Ulukus, "Interference management for CDMA systems through power control, multiuser detection, and beamforming," *IEEE Trans. Commun.*, vol. 49, no. 7, pp. 1227–1239, Jul. 2001.
[24] S. Choi, J. Choi, H.-J. Im, and B. Choi, "A novel adaptive beamforming algorithm for antenna array CDMA systems with strong interferers," *IEEE Trans. Veh. Technol.*, vol. 51, no. 5, pp. 808–816, Sep. 2002.
[25] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, Jan. 2018.
[26] W. Alltop, "Complex sequences with low periodic correlations (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 3, pp. 350–354, May 1980.
[27] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Berlin, Germany: Springer, 2013.
[28] D. Chu, "Polyphase codes with good periodic correlation properties (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 531–532, Jul. 1972.
[29] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, Feb. 2016.
[30] E. Biglieri, *Coding for Wireless Channels*. New York, NY, USA: Springer, 2005.
[31] J. Guo, S. Durrani, X. Zhou, and H. Yanikomeroglu, "Massive machine type communication with data aggregation and resource scheduling," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 4012–4026, Sep. 2017.
[32] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probability Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[33] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.
[34] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd ed. New York, NY, USA: Wiley, 2003.
[35] S. Ross, *Stochastic Processes*. New York, NY, USA: Wiley, 1996.



JINHO CHOI (Fellow, IEEE) was born in Seoul, South Korea. He received the B.E. degree (magna cum laude) in electronics engineering from Sogang University, Seoul, in 1989, and the M.S.E. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), in 1991 and 1994, respectively. He is currently a Professor with the School of Information Technology, Deakin University, Burwood, Australia. Prior to joining Deakin University, in 2018, he was with Swansea University, U.K., as a Professor/Chair of wireless, and the Gwangju Institute of Science and Technology (GIST), South Korea, as a Professor. He has authored two books published by Cambridge University Press, in 2006 and 2010, and one book by Wiley-IEEE, in 2022. His research interests include the Internet of Things (IoT), wireless communications, and statistical signal processing. He received several best paper awards, including the 1999 Best Paper Award for Signal Processing from EURASIP. He has been on the list of the World's Top 2% Scientists by Stanford University, since 2020. He is a Senior Editor of IEEE Wireless Communications Letters and an Associate Editor of IEEE Transactions on Mobile Computing. He has also served as a Division Editor for *Journal of Communications and Networks* and an Associate Editor or an Editor for other journals, including IEEE Transactions on Communications, IEEE Communications Letters, *Journal of Communications and Networks*, IEEE Transactions on Vehicular Technology, and *ETRI Journal*.