Low-overhead Multi-antenna-enabled Random Access for Machine-type Communications with Low Mobility

Yihan Zou, Kwang Taik Kim, Xiaojun Lin, Mung Chiang
School of ECE
Purdue University, West Lafayette, IN, USA
Email:{zou59, kimkt, linx, chiang}@purdue.edu

Zhi Ding
ECE
UC Davis, CA, USA
Email: zding@ucdavis.edu

Risto Wichman, Jyri Hämäläinen
School of Electrical Engineering
Aalto University, Finland
Email:{risto.wichman, jyri.hamalainen}@aalto.fi

***Abstract*—A pseudo-random beamforming (PRBF) based random access (RA) system is proposed to enable uplink (UL) machine-type communications (MTC) with ultra low signaling overheads. Specifically, a pseudo random (PR) sequence is used as public information to coordinate the beamforming vectors used at the base station (BS) and the devices. Within the coherence time window, each device distributively determines in advance the "good" time slots and receiving beams for transmission. This UL protocol reduces the overheads due to the feedback of channel state information and the control signals for centralized scheduling. This paper derives the throughput and user scaling of the proposed M-PRBF-CA protocol for achieving spatial multiplexing gain, under both an i.i.d. slow fading channel and a correlated slow fading channel. Our simulation results confirm the analysis in both fading channel models.**

I. Introduction

The upcoming fifth generation (5G) wireless network aims to achieve ubiquitous connectivity of billions of devices, objects and machines, to form the so-called massive internet-of-things (IoT). 3GPP defines that massive IoT consists of at least 1M devices/km^2. To support such massive machine-type communication (MTC) for IoT, wireless networks are required to more efficiently support the simplest devices communicating sporadically and to be highly energy efficient so that massive IoT devices can deliver exceedingly long battery life.

One of the toughest challenges for IoT communication is in the uplink (UL), where massive number of devices try to access shared resources with limited coordination. ALOHA is a common random access (RA) protocol for organizing multiple access in a decentralized manner. However, the original slotted-ALOHA does not consider wireless channel fading. In [1], the authors propose "Channel-aware ALOHA" (CA-ALOHA), which allows only those users with "good" channels to enter the contention. It is shown that the asymptotic throughput of CA-ALOHA is equal to $1/e$ of the optimal throughput of centralized opportunistic scheduling. Recently, a similar channel-aware ALOHA scheme in a multi-cell interference model was studied in [2]. CSMA-type is another common RA protocol, but its gain over ALOHA will be limited when the IoT payload is short. There have also been significant progress on designing non-orthogonal multiple access (NOMA), such as sparse code multiple access (SCMA), to multiplex (overload) users using different codebooks over the same resource block [3]. However, the above studies often overlook the use of MIMO technology either, which has been the key driving force of performance gain in the 4G system. The reason is likely due to the complexity of MIMO. Despite its spatial multiplexing gain provided by MIMO, one has to keep track of the Channel State Information (CSI) for all time and make careful scheduling decisions, which incur significant overhead for IoT multiple access [4]. Thus, it remains an open question how to design UL IoT multiple access schemes that can explore the MIMO gain at low complexity and with low overheads.

In order to obtain MIMO gains for IoT, some research efforts have been made on combining random beamforming (RBF) with opportunistic scheduling [5]. For low-mobility IoT devices with slow channel dynamics, RBF can induce more channel fluctuations, and thus achieve a MUD gain. Traditional RBF architecture requires each device to feedback the effective channel gain to the base station (BS) whenever the RBF vector changes [4]. This requirement is unrealistic in IoT UL communication because the massive number of channel feeckback can easily overwhelm the communication channel. To address this issue, [6] proposes to reduce the overhead by sampling only a small subset of the IoT devices, which will likely lead to a much lower MUD gain. In [7], the authors propose a pseudo-random beamforming (PRBF) approach where each device can determine its channel quality before-hand. However, the system forms only one beam at each time, thus does not provide any spatial multiplexing gain.

The most closely related papers to our work are [8] and [9], which derive the throughput and user scaling for multi-beam RBF. They study the multi-beam RBF downlink (DL) case in the context of millimeter wave (mmWave). Despite achieving the spatial multiplexing gain of MIMO, they rely on centralized scheduling to resolve contention, which incurs high overhead due to the need of CSI feedback. If one removes centralized scheduling, it is unclear how UL multiple access can be organized in a low-overhead way so that the spatial multiplexing gain of MIMO can still be attained. Second, they both focus on the DL. In contrast, we focus on the UL, which is fundamentally more difficult due to massive random access. Specifically, in DL the number of interference sources to a given receiver is equal to the number of other beams. In contrast, in UL the number can be as large as the number of

978-1-7281-0962-6/19/$31.00 ©2019 IEEE

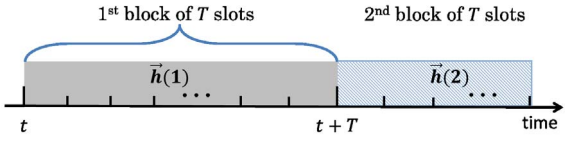


Fig. 1. An illustration for the slow block fading channel model.

all other users. It is thus critical to design a multiple-access scheme that can limit the interference. Third, instead of the classic block fading channel (or fast block fading) in [8] and [9], in this paper we consider a slow block fading channel, which is more suitable for low-mobility IoT devices. We shall see in Remark 1 in Section III that some well-known results from the fast fading case cannot be directly applied in case of slow block fading channel. Therefore, it remains unclear whether similar throughput and user scaling can be attained in the use of a MIMO massive IoT system.

In this paper, we propose a multi-beam PRBF channel-aware (M-PRBF-CA) random access scheme in Section II that can achieve the spatial multiplexing gain with ultra-low overhead. Specifically, in slow fading channel, we show that M-PRBF-CA can achieve a total throughput linear in the number of antennas in Sections III and IV. Moreover, while an exponential user scaling in the number of BS antennas is required to achieve this spatial multiplexing gain in the i.i.d. slow fading channel, only a linear user scaling is required for a correlated channel, which is confirmed by our simulation results in Section V.

II. System Model

We consider a single-cell system, where a BS equipped with $M \geq 3$ antennas serves $N \gg 1$ single-antenna IoT devices. To keep the cost of RF components low, a narrowband channel is allocated for IoT communication purposes. We focus on a time division duplex system and assume that the channel reciprocity property holds. Note that the extension to frequency division duplex is also possible.

The BS can simultaneously use $B = \Theta(M)$ random receiving beamforming vectors. Time is slotted. To coordinate the beamforming vectors at each time, a pseudo random (PR) sequences is known in both BS and IoT devices. This can be done by storing hard-coded PR sequence information in devices' chip sets [7]. Denote the i-th random beamforming vector at time t by $\mathbf{w}_i(t)$ for $1 \leq i \leq B$. The $M \times 1$ channel gain vector for user k is denoted by $\mathbf{h}_k(t)$. Since IoT devices are static or admit very low mobility, the channel gain vector $\mathbf{h}_k(t)$ changes slowly across time, i.e., the coherence time is large (multiple hundreds of milliseconds in practice) compared to the time scale of multiple-access decision (1ms in 4G LTE). Thus, as shown in Fig. 1, we assume static $\mathbf{h}_k$ within a long coherence time period T, where $\mathbf{h}_k$ only changes between blocks. Due to static $\mathbf{h}_k$ (within each block) and known $\mathbf{w}_i(t)$, each device k can compute its effective receiving gain at the BS for each beam i and only wake up when the effective channel gain is high (unless the coherence time has passed, in which case the device has to re-measure the channel).

Due to the short message payload of IoT traffic and massive number of IoT devices, centralized scheduling becomes expensive due to the overhead of exchanging control signals and solving optimal schedule. Hence, we eliminate centralized scheduling or the need of feeding back effective channel gains to the BS. Instead, we propose to use a distributed protocol similar to slotted-ALOHA for UL random access. We assume the infinite backlog case so that each user always has packets to send. At the BS, simultaneous transmissions from multiple IoT devices aiming at the same beam may collide, and thus fail to be decoded by the BS. This contention will be controlled by the channel-aware RA described below. As for inter-beam interference, we assume that the BS can successfully decode for the message at an achievable rate determined by the received signal-to-interference-plus-noice ratio (SINR) by treating the inter-beam interference as noise.

We consider two models of channel fading with and without spatial correlation between antennas. First, we study classic independent and identically distributed (i.i.d.) Rayleigh fading channel, where each element in $\mathbf{h}_k$ follows $\mathrm{CN}(0,1)$. Then, motivated by the fact that channel gain vector $\mathbf{h}_k$ often exhibits correlation among different antennas in practice, we also study a correlated channel model, assuming the uniform random single path (UR-SP) model of [8]. The overall system operation is summarized as follows.

Multi-beam PRBF Channel-aware (M-PRBF-CA) Protocol

1) Channel Estimation Phase: In this phase, BS broadcasts DL pilot signals that devices can use to estimate the channel vector $\mathbf{h}_k$. This phase is carried out during every coherence time period to update the channel vector.

2) Sleep/Wake-up Phase: Since device knows the PRBF vectors that are applied in different time slots, it can go to a sleep mode when the effective beamforming gain is low, and wake up when the good PRBF vectors occur.

3) Random Access Phase: Each device k makes transmission decision individually, and only picks a beam i to attempt transmission when the following conditions are satisfied:

i Target beam condition is good, i.e., $|\mathbf{h}_k^T \mathbf{w}_i|^2 > \Phi_G$.

ii The interference caused on other beams is small.

- For i.i.d. channel: $\sum_{j \neq i} |\mathbf{h}_k^T \mathbf{w}_j|^2 < (B-1)\Phi_I$.
- For correlated channel: $|\mathbf{h}_k^T \mathbf{w}_j|^2 < \Phi_I$ for all $j \neq i$.

By properly setting thresholds Φ_G and Φ_I, we can make sure that each user k sees only one good beam (and there can be multiple transmitting users in each beam). Then, multiple beamforming vectors divide users into groups. In the rest of the paper, we are interested in whether Φ_G and Φ_I can be chosen such that the throughput of the system increases linearly in M, and how large the number of users has to scale in order to achieve such a linear sum-rate in M.

Overhead Comparison with Centralized Scheduling

We note that our M-PRBF-CA scheme incurs much lower channel estimation overhead and control signaling overhead. To see the reduction in overhead, we can consider a time

span of T slots. For the i.i.d. slow block-fading channel, our M-PRBF-CA broadcasts M pilots (one for each antenna) for channel estimation at the beginning of a block. In the following T slots, each device attempts random access using knowledge of the PRBF vectors, and thus no control signals for scheduling are needed. The total overhead for M-PRBF-CA UL access is hence $\Theta(M)$. In contrast, when a centralized scheduling is used, at every time slot, the BS has to send M pilots (one for each PRBF vector). Each device reports the effective channel gain of the best beam back to the BS. After receiving all the channel reports, the BS grants transmission to M selected devices. The total overhead for UL using central scheduling is hence $T \cdot (M + N + M) = \Theta((M + N)T)$. Thus, the overheads differ by a factor of $\Theta((1 + N/M)T)$.

III. Independent Slow Block Fading Channels

Due to the low mobility of IoT devices, the channels vary much more slowly compared to traditional mobile handsets. Compared to the well-studied slot-by-slot fast fading case in [8] and [9], such slow fading will cause large delay if not treated properly. To see this, note that if some users are stuck in poor channels, they will never get a chance to transmit within a long coherence time. To alleviate such fairness issue, it is necessary for each IoT device to adopt some form of power control. Specifically, suppose that the channel vector for user k is $\mathbf{h}_k \sim \mathrm{CN}(0, I)$. We set the transmission power to $P_k = 1/||\mathbf{h}_k||^2$ to maintain a constant expected receiving power at the BS. As we discussed earlier, at each time t, the BS forms M orthogonal beams. The PRBF vectors $\mathbf{w}_i$'s form a random orthonormal matrix $\mathbf{W}(t)$ in $\mathbb{C}^M$, such that every column $\mathbf{w}_i$ is uniform on the unit-sphere in $\mathbb{C}$. After applying the BF vector $\mathbf{w}_i(t)$, we obtain the effective received gain for device k at beam i

$$P_{ki}^{(r)} = \left|(\sqrt{P_k}\mathbf{h}_k^T)\mathbf{w}_i(t)\right|^2, \tag{1}$$

where $(\sqrt{P_k}\mathbf{h}_k^T)$ is a unit-length vector fixed over a coherence time. According to the M-PRBF-CA random access scheme described in Section II, each user will see a good channel at beam i for transmission with probability

$$p_t = \Pr\left\{P_{ki}^{(r)} \geq \Phi_G, \sum_{j\neq i} P_{kj}^{(r)} \leq (M-1)\Phi_I\right\}. \tag{2}$$

We can write the expression (2) for p_t in terms of Φ_G and Φ_I. Before proceeding, we introduce the following lemma.

Lemma 1. *Suppose i.i.d. random variables $X_i \sim \mathrm{CN}(0,1)$ for all i. Then* $\Pr\left\{\frac{|X_i|^2}{\sum_{j=1}^n |X_j|^2} \geq a\right\} = (1-a)^{n-1}$.

Proof. See technical report [10] for details. □

From (1) and (2), we can derive an expression for p_t. In (1), since $\mathbf{h}_k$ and $\mathbf{w}_i$ are uniform, we can use change of coordinates, i.e., we can fix $\mathbf{w}_i(t)$ to be the unit vector on the i-th axis of $\mathbb{C}^M$, and assume that $\sqrt{P_k}\mathbf{h}_k^T$ is rotating uniformly in space. Without loss of generality, we can assume that the beam 1 is the desired beam, and that $\mathbf{w}_i(t) = \mathbf{e}_i$ for $i \in [1:M]$, e.g., $\mathbf{w}_1 = (1, 0, \ldots, 0)^T$. Then, we have, for any beam i,

$$\begin{aligned}
p_t &= \Pr\{P_{ki}^{(r)} \geq \Phi_G; \sum_{j\neq i} P_{kj}^{(r)} \leq (M-1)\Phi_I\} \\
&= \Pr\{P_{k1}^{(r)} \geq \Phi_G; \sum_{j\geq 2} \left|(\sqrt{P_k}\mathbf{h}_k^T)\mathbf{w}_j(t)\right|^2 \leq (M-1)\Phi_I\} \\
&= \Pr\{P_k|\hat{h}_{k1}|^2 \geq \Phi_G; P_k \sum_{j\geq 2} |\hat{h}_{kj}|^2 \leq (M-1)\Phi_I\} \\
&= \Pr\left\{P_k|\hat{h}_{k1}|^2 \geq \max\{\Phi_G, 1-(M-1)\Phi_I\}\right\} \\
&= \left(1 - \max\{\Phi_G, 1-(M-1)\Phi_I\}\right)^{M-1},
\end{aligned} \tag{3}$$

where in the third step we have used $(\sqrt{P_k}\mathbf{h}_k^T)\mathbf{w}_j(t) = \sqrt{P_k}\hat{h}_{kj}$, and in the fourth step we have used $\sum_j P_k|\hat{h}_{kj}|^2 = 1$. In the above equations, $\hat{\mathbf{h}}_k$ is an uniformly rotating vector with unit norm. The last equality comes from Lemma 1.

Within each beam i, the users compete for transmission turn by using slotted-ALOHA. The received signal from device k on beam i at the BS is

$$y_k = \sqrt{P_k}\mathbf{h}_k^T\mathbf{w}_i s_k + \sum_{u=1}^{N_I} \sqrt{P_u}\mathbf{h}_u^T\mathbf{w}_i s_u + n,$$

where $|s_k|^2 = 1$ is the power constraint per transmit symbol, N_I is the number of interfering users from other beams $j \neq i$ and $n \sim \mathrm{CN}(0, \sigma^2)$ is the additive noise. The throughput for beam i is thus $T_i = Np_t(1-p_t)^{N-1} \cdot \mathsf{E}\left[\log\left(1 + \mathrm{SINR}_i\right)\right]$, where SINR_i of the transmitting user k in beam i is

$$\mathrm{SINR}_i = \frac{P_{ki}^{(r)}}{\sum_{u=1}^{N_I} P_{ui}^{(r)} + \sigma^2}, \tag{4}$$

and E refers to the expectation over $\mathbf{h}_k$, $\mathbf{w}_i$ and N_I. Now, we are ready to derive the result for the throughput scaling.

Theorem 1. *If $\log(N) = \Omega(M)$, M-PRBF-CA can set Φ_G and Φ_I such that the asymptotic throughput is $T_{total} = \Omega(M)$.*

Proof. The total sum rate over M beams is

$$T_{\text{total}} = M \cdot Np_t(1-p_t)^{N-1} \cdot \mathsf{E}\left[\log\left(1 + \mathrm{SINR}_i\right)\right], \tag{5}$$

where p_t is given by (3). Under the M-PRBF-CA protocol, the expectation term in (5) is lower-bounded as

$$\begin{aligned}
\mathsf{E}\left[\log(1+\mathrm{SINR}_i)\right] &= \mathsf{E}\left[\log\left(1 + \frac{P_{ki}^{(r)}}{\sum_{u=1}^{N_I} P_{ui}^{(r)} + \sigma^2}\right)\right] \\
&\geq \mathsf{E}\left[\log\left(1 + \frac{\Phi_G}{N_I\Phi_I + \sigma^2}\right)\right] \\
&\geq \log\left(1 + \frac{\Phi_G}{\mathsf{E}[N_I]\Phi_I + \sigma^2}\right),
\end{aligned}$$

where $\mathsf{E}_N[P_{ui}^{(r)}] \leq (M-1)\Phi_I/(M-1) = \Phi_I$. The third step follows from Jensen's inequality, as $f(x) = \log\left(1 + \frac{a}{bx+c}\right)$ is convex for $a, b, c > 0$. If we set $p_t = 1/N$, we have $\mathsf{E}[N_I] = Np_t(M-1) = (M-1)$. Then, the total throughput admits the lower bound

$$T_{\text{total}} \geq M\left(1 - \frac{1}{N}\right)^{N-1} \cdot \log\left(1 + \frac{\Phi_G}{(M-1)\Phi_I + \sigma^2}\right).$$

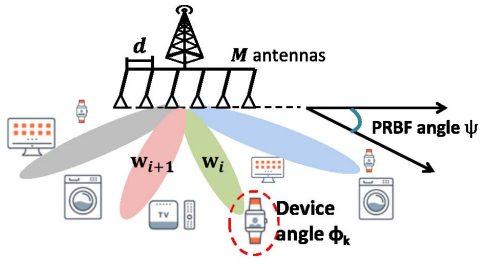


Fig. 2. M-PRBF-CA system diagram for the correlated slow fading channel.

Since $\log(N) = \Omega(M)$, $(1-1/N)^{N-1}$ approaches e^{-1} when M is sufficiently large. Then, to show $T_{\text{total}} = \Omega(M)$, it is sufficient to check that we can choose Φ_G and Φ_I such that

1) $p_t = 1/N$; 2) $\frac{\Phi_G}{(M-1)\Phi_I+\sigma^2} = \Theta(1)$.

Thereby, we verify that the two conditions above are satisfied. Suppose $N \geq a^{M-1}$ with $a > 1$. We can set the thresholds of M-PRBF-CA as $\Phi_G = 1 - N^{-\frac{1}{M-1}} \geq 1 - \frac{1}{a} > 0$ and $\Phi_I = \frac{b}{M-1}$ with $1 > b > \frac{1}{a}$. Under this parameterization, we have $1-(M-1)\Phi_I = 1-b \leq \Phi_G$. By (3), we have $p_t = (1-\Phi_G)^{M-1} = 1/N$, which is precisely condition 1). Also, we have

$$\frac{\Phi_G}{(M-1)\Phi_I+\sigma^2} \geq \frac{1-1/a}{b+\sigma^2} \triangleq c.$$

Hence, we have shown the condition 2). The asymptotic throughput thus satisfies $T_{\text{total}} \geq \frac{M}{e}\log(1+c) = \Omega(M)$. □

Remark 1. *Here we highlight a key difference between the case with power control and the one without in [8], [9]. In fast fading, $\mathbf{h}_k^T(t)\mathbf{w}_i(t)$ is an i.i.d. complex Gaussian r.v.. In slow block fading, due to static $\mathbf{h}_k$ and power control, the norm of $\sqrt{P_k}\mathbf{h}_k$ is always 1. As a result, from* (3)*, we can see that the channel's degree of freedom is $M-1$, instead of M in fast fading case. Thus, as we discussed earlier, the results in [8] based on fast fading cannot be directly applied.*

Remark 2. *Note that the condition $\log N = \Omega(M)$ for M-PRBF-CA to achieve linear sum-rate scaling in Theorem 1 is also necessary. The necessity of $\Theta(\log N) = M$ for i.i.d. Rayleigh block fading and centralized scheduling is shown in [11, Theorem 2]. The same can be shown for UL random access as well.*

IV. Correlated Slow Block Fading Channel

The i.i.d. fading assumption is a simplification. In practice, correlation can exist among users, or even within same user's subchannels, especially when a dominant path exists between senders and receivers, e.g., in mmWave communications. Next, we consider a channel model similar to the UR-SP model in [4] and [8]. However, in our model the low-mobility IoT devices experience the slow block fading channel. The channel vectors between two devices are still assumed to be independent.

A. Device Scaling for a Linear Sum Rate

We first define the correlated channel model. Each device k is associated with an angle-of-arrival (AoA) ϕ_k. The channel

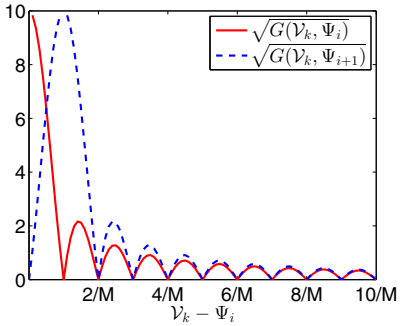


Fig. 3. The square root of beamforming gain $\sqrt{G(\mathcal{V}_k, \Psi_i)}$ as a function of $\mathcal{V}_k - \Psi_i$, where $M = 100$ and $d/\lambda = 1$.

of device k to antenna m at BS is $h_{km} = l_k \exp(j\phi_{km})$, where $\{l_k\}$ is a device-specific path loss term and ϕ_{km} is the phase shift w.r.t. antenna m. Specifically,

$$\phi_{km} = \frac{2(m-1)d\pi\cos\phi_k}{\lambda},$$

where d is the distance between two adjacent antennas and λ is the carrier wavelength. Let $D = \frac{2d}{\lambda}$ and $\mathcal{V}_k = \cos\phi_k$. Then $\phi_{km} = (m-1)D\pi\mathcal{V}_k$, where $\mathcal{V}_k \in [-1,1]$ is the equivalent angle of ϕ_k in a cosine-domain.

Our proposed M-PRBF-CA system diagram is shown in Fig. 2. Similar to Section III, we consider the power control such that user k transmits at power $P_k = 1/(M|l_k|^2)$. The BS forms M beams at each time. Each beam i is determined by a parameter ψ_i to be chosen later. For beam i, its beamforming vector is given by $\mathbf{w}_i = [w_{im}]$ for $m = [1:M]$, where

$$w_{im} = \frac{1}{\sqrt{M}}\exp\{-j(m-1)D\pi\Psi_i\}$$

and $\Psi_i = \cos\psi_i$. Then, the overall channel gain of user k on beam i is given by $\left|\sqrt{P_k}\mathbf{h}_k^T\mathbf{w}_i\right|^2$ as

$$G(\mathcal{V}_k,\Psi_i) = P_k|l_k|^2\Big|\sum_{m=1}^{M}\frac{1}{\sqrt{M}}e^{j(m-1)D\pi(\mathcal{V}_k-\Psi_i)}\Big|^2 = \frac{1}{M^2}\Big|\frac{\sin(\pi\frac{d}{\lambda}M(\mathcal{V}_k-\Psi_i))}{\sin(\pi\frac{d}{\lambda}(\mathcal{V}_k-\Psi_i))}\Big|^2, \quad (6)$$

where the last step above can be obtained using a half-angle trick. Note that (6) only depends on the difference between $\mathcal{V}_k$ and Ψ_i. For simplicity of exposition, we assume $\frac{d}{\lambda} = 1$ in the rest of this section (but similar results also hold for other values of $\frac{d}{\lambda}$). Then, zeros of $G(\cdot,\cdot)$ occur where $\mathcal{V}_k - \Psi_i$ equals to a multiple of $1/M$, as shown by red solid curve in Fig. 3. We choose Ψ_i's such that

$$\Psi_i = \Psi + \frac{(i-1)}{M}, \quad i = 1,2,\ldots,M,$$

where Ψ is uniformly distributed in $[0, \frac{1}{M}]$. We can place the adjacent beam as the blue dashed curve in Fig. 3. This ensures that the separation of beams is at least $\frac{1}{M}$, and implies that the cross-interference between beams is small. Since $\Psi_i \in [0,1]$, the BS can form M orthogonal beams.

To obtain a high SINR and low cross-interference, our M-PRBF-CA scheme only assigns user k to beam i if

$$|\mathcal{V}_k - \Psi_i| \leq \frac{1}{3M}. \tag{7}$$

Let $\Omega_{kj} = \mathcal{V}_k - \Psi_j$. The interference from user k to the j-th nearest beam can be upper-bounded by the value of $G(\cdot)$ at $\Omega_{kj} = \frac{2j-1}{2M}$. For notation purposes, we define the peak of 0-th nearest side-lobe of beam i is at $\Omega_{k1} = 1/2M$ (which locates in main lobe). Then, the j-th nearest side-lobe is determined by $\pi \frac{d}{\lambda} M \Omega_{kj} = \frac{2j+1}{2}\pi$. From (6), the cross-interference by device k to beam j is upper-bounded by the peak of the $(j-1)$-th nearest side-lobe

$$I_{kj} \leq \frac{1}{M^2}\left|\frac{\sin(\frac{2j-1}{2}\pi)}{\sin(\frac{2j-1}{2M}\pi)}\right|^2 = \frac{1}{M^2}\frac{1}{\sin^2(\frac{2j-1}{2M}\pi)} \triangleq I_{kj}^{\max}.$$

In particular, the 0-th side-lobe satisfies

$$I_{k1} \leq \frac{1}{M^2}\frac{1}{\sin^2(\frac{1}{2M}\pi)}. \tag{8}$$

Similar to the i.i.d. slow fading case in Section III, our threshold-based random access scheme allows user transmission if (1) $G(\mathcal{V}_k, \Psi_i) > \Phi_G$ and (2) $I_{kj} = G(\mathcal{V}_k, \Psi_j) < \Phi_I$ for all $j \neq i$. The transmission probability is

$$p_t = \Pr\{G(\mathcal{V}_k, \Psi_i) > \Phi_G; I_{kj} < \Phi_I \text{ for all } j \neq i\}. \tag{9}$$

By (6) and (8), for transmission event (9) to happen, it is sufficient to have (noting that $\frac{d}{\lambda} = 1$)

$$\left\{\frac{1}{M^2}\left|\frac{\sin(\pi M \Omega_{ki})}{\sin(\pi \Omega_{ki})}\right|^2 > \Phi_G,\ I_{k1} \leq \Phi_I\right\}. \tag{10}$$

Before we derive the asymptotic throughput for the correlated slow fading channel, we first give a lemma.

Lemma 2. *For positive integer M, the inequality holds that $\sum_{j=1}^{M-1} \frac{1}{\sin^2(\frac{2j-1}{2M}\pi)} \leq M^2$.*

Proof. See technical report [10] for details. □

Theorem 2. *For the correlated slow fading channel, if $N = \Omega(M)$, our M-PRBF-CA scheme can set Φ_G and Φ_I such that the asymptotic system throughput scales as $T_{total} = \Omega(M)$.*

Proof. Similar to the i.i.d. slow fading case in Section III, the system sum rate is given by (5). To ensure $T_{\text{total}} = \Omega(M)$ under the M-PRBF-CA scheme, it is sufficient to check that we can choose Φ_G and Φ_I such that

1) $I_{kj} \leq \Phi_I$ for all $\mathcal{V}_k$ satisfying (7);
2) If $G(\mathcal{V}_k, \Psi_i) > \Phi_G$, then (7) holds for beam i only;
3) In (9), $p_t = 1/N$ can be achieved;
4) In (5), SINR $\geq \Theta(1)$.

Towards this end, we can set $\Phi_I = b$, where $b = \frac{25}{4\pi^2}$. From (8), we have

$$I_{kj} \leq \frac{1}{M^2}\frac{1}{\sin^2(\frac{1}{2M}\pi)} \leq \frac{1}{M^2}\frac{1}{(\frac{4}{5}\frac{1}{2M}\pi)^2} = \frac{25}{4\pi^2} \text{ for all } j \neq i,$$

since $\sin x \geq \frac{4x}{5}$ when $x \in [0, \pi/6]$. Thus, condition 1) is satisfied. Suppose $N \geq aM$ where $a > 1.5$. Setting $\Delta\phi = 2 - 2\sqrt{1 - \frac{1}{2N}}$, we have $\Delta\phi \leq \frac{1}{3M}$. Since $\Delta\phi \leq \frac{1}{3M}$, by setting

$$\Phi_G = \frac{1}{M^2}\left|\frac{\sin(\pi M \Delta\phi)}{\sin(\pi \Delta\phi)}\right|^2, \tag{11}$$

we have $G(\mathcal{V}_k, \Psi_i) > \Phi_G$ if $|\mathcal{V}_k - \Psi_i| < \Delta\phi$. We can verify that $\Phi_G \geq 27/(4\pi^2)$ (see [10] for details). Thus, for $j \neq i$, $G(\mathcal{V}_k, \Psi_j) < I_{k1} < \Phi_G$. Thus, condition 2) holds. The event (10) occurs with probability $\Pr\{|\mathcal{V}_k - \Psi_i| < \Delta\phi\} = \frac{1}{N} = p_t$, which gives condition 3). Finally, by (5), we can write the lower bound for expected SINR as

$$\begin{aligned}\mathsf{E}[\log(1+\text{SINR}_i)] &\geq \mathsf{E}\left[\log\left(1 + \frac{\Phi_G}{\sum_{u=1}^{N_I} I_{ui} + \sigma^2}\right)\right] \\ &\geq \log\left(1 + \frac{\Phi_G}{\mathsf{E}[\sum_{u=1}^{N_I} I_{ui}] + \sigma^2}\right),\end{aligned} \tag{12}$$

where N_I is the number of cross-interfering devices. The total interference term in (12) can be bounded by

$$\begin{aligned}\mathsf{E}\left[\sum_{u=1}^{N_I} I_{ui}\right] &\leq \sum_{j}^{M-1} I_{kj}^{\max}\, \mathsf{E}[N_I(j)] \\ &= \sum_{j}^{M-1} \frac{2}{M^2}\frac{1}{\sin^2(\frac{2j-1}{2M}\pi)} \overset{\text{(Lemma 2)}}{\leq} 2.\end{aligned} \tag{13}$$

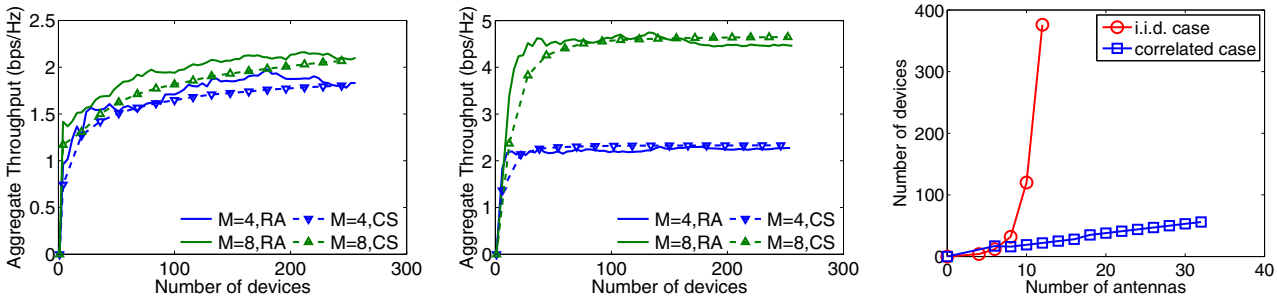
where $N_I(j)$ is the number of interfering devices in beam j. Hence, we have $\text{SINR}_i \geq \frac{27/(4\pi^2)}{2+\sigma^2} \geq \Theta(1)$. This gives condition 4) and completes the proof. □

B. Access Delay Scaling: IID vs Correlated Channel

One way to define the delay is the average waiting time for a device to attempt a transmission. In other words, the probability of coverage is the same as the transmission event in (2). The expected delay is therefore $\mathsf{E}[D] = \frac{1}{p_t}$. If the transmission probability $p_t = \frac{1}{N}$, the expected delay $\mathsf{E}[D]$ scales as the number of devices N. Thus, for the i.i.d. slow fading channel in Section III, if $N = \Theta(e^M)$, the expected delay also scales as $O(e^M)$, i.e., exponentially as M. However, as p_t depends on N, it may suggest that the delay could be improved at smaller values of N (i.e., smaller load). This is unfortunately not the case. Even for smaller N, the delay is still of this order. To see this, note that even if N is not sufficiently large, the i.i.d. interference term $|\mathbf{h}_k^T \mathbf{w}_j|^2$ still needs to be satisfied for every $j \neq i$. In other words, even if only a few users are in the system, each user has to wait till its interference to other beams becomes small. Therefore, p_t still decays exponentially with M, and the delay scales at least exponentially w.r.t. M. On the other hand, for the correlated channel in Section IV, if $N = \Theta(M)$, the access delay scales linearly with N. For small N, as $\Delta\phi < \frac{1}{3M}$, the delay also scales linearly with M.

V. Experimental Results

In this section, we present the numerical analysis of our proposed M-PRBF-CA protocol for UL massive IoT communications. As we discussed earlier in Sections III and IV,



(a) Aggregate throughput scales with N under i.i.d. slow fading. "RA" for random access, and "CS" for centralized scheduling scaled by $1/e$.

(b) Aggregate throughput scales with N under correlated slow fading. "CS" for centralized scheduling scaled by $1/e$.

(c) User scaling for linear sum-rate growth. The target per-beam throughput for the i.i.d. case and the correlated case is 0.2 and 0.5, respectively.

Fig. 4. Throughput and user scaling under the M-PRBF-CA scheme. The interference threshold is $1/M$ and $\sigma^2 = 0.5$ (normalized by received signal power).

we simulate the performance of M-PRBF-CA under the i.i.d. slow fading channel and the correlated slow fading channel. In both cases, we run our M-PRBF-CA protocol using the corresponding PRBF vectors for $T = 500$ time slots, and compare the time-average performance.

In Fig. 4(a), we show, based on our choice of Φ_G and Φ_I, how the time-average of aggregate throughput in the i.i.d. slow fading channel in Section III scales with N under different numbers of the BS antennas ($M = 4$ and 8). First, the throughput (per Hz) increases sub-linearly with the device population. Second, the throughputs under M-PRBF-CA (solid curves) are close to $1/e$ of those in centralized scheme (dashed curves, where the BS also uses PRBF and selects the user with the best effective gain in each beam). The above result illustrates that our proposed M-PRBF-CA architecture achieves the MUD gain by inducing more fluctuations in the channel dynamics for fixed-location IoT devices. Similar observations can be made in Fig. 4(b) for the correlated slow-fading model. However, the increase in aggregate throughput from $M = 4$ to $M = 8$ is more significant in Fig. 4(b). This is because the number of users to achieve a linear sum-rate scaling (in M) is much lower for the correlated fading model of Fig. 4(b) than for the i.i.d. fading model of Fig. 4(a).

Next, we verify the user scaling result derived in Sections III and IV. In Fig. 4(c), we plot the number of devices to achieve a target per-beam throughput for different numbers of the BS antennas M. Note that a higher target for the per-beam throughput is used in the correlated fading model because the interference tends to be lower in the correlated fading model than that in the i.i.d. fading model. We observe that, for both channel models, the lower bound on N increases as M increases. For i.i.d. slow fading, as we discussed in Section III, the lower bound of N increases exponentially to more than 300 when M increases from 4 to 12. Clearly, such exponential scaling under i.i.d. slow fading is undesired, especially for the modern communication system where the BS's are equipped with a large number of antennas. In contrast, for the correlated channel, we can see that the lower bound for N increases only linearly with M. Compared to the exponential user scaling for i.i.d. slow fading, the above result in the correlated slow fading case is more desirable.

VI. Conclusion

We propose M-PRBF-CA for massive IoT with ultra-low overhead. We derive the user scaling to achieve the spatial multiplexing gain of MIMO under i.i.d. and correlated slow fading, which implies that the channel correlation may significantly reduce the number of devices to achieve the MIMO gain. In the future, we plan to extend our results to more general correlation models.

Acknowledgement

This work has been partially supported by NSF grant CNS-1703014, NSF grant 1702752, NSF Alan T. Waterman Award, and Academy of Finland under grant 311752.

References

[1] X. Qin and R. A. Berry, "Distributed approaches for exploiting multiuser diversity in wireless networks," *IEEE Trans. on Inform. Th.*, vol. 52, no. 2, pp. 392–413, Feb 2006.

[2] H. Lin and W. Shin, "Multi-cell aware opportunistic random access," in *2017 IEEE ISIT*, June 2017, pp. 2533–2537.

[3] K. Au, L. Zhang, H. Nikopour, E. Yi, A. Bayesteh, U. Vilaipornsawai, J. Ma, and P. Zhu, "Uplink contention based scma for 5G radio access," in *2014 IEEE Globecom Workshops*, Dec 2014, pp. 900–905.

[4] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. on Inform. Th.*, vol. 48, no. 6, pp. 1277–1294, June 2002.

[5] M. Hasan, E. Hossain, and D. Niyato, "Random access for machine-to-machine communication in LTE-advanced networks: issues and approaches," *IEEE Comm. Mag.*, vol. 51, no. 6, pp. 86–93, June 2013.

[6] B. Li, B. Ji, and J. Liu, "Efficient and low-overhead uplink scheduling for large-scale wireless internet-of-things," in *WiOpt '18*, May 2018.

[7] A. A. Dowhuszko, G. Corral-Briones, J. Hämäläinen, and R. Wichman, "Performance of quantized random beamforming in delay-tolerant machine-type communication," *IEEE Trans. on Wireless Comm.*, vol. 15, no. 8, pp. 5664–5680, Aug 2016.

[8] G. Lee, Y. Sung, and J. Seo, "Randomly-directional beamforming in millimeter-wave multiuser MISO downlink," *IEEE Trans. on Wireless Comm.*, vol. 15, no. 2, pp. 1086–1100, Feb 2016.

[9] G. Lee, Y. Sung, and M. Kountouris, "On the performance of random beamforming in sparse millimeter wave channels," *IEEE JSTSP*, vol. 10, no. 3, pp. 560–575, April 2016.

[10] Y. Zou, K. T. Kim, X. Lin, M. Chiang, Z. Ding, R. Wichman, and J. Hämäläinen, "Low-overhead multi-antenna-enabled uplink random access for massive machine-type communications with low mobility," https://engineering.purdue.edu/%7elinx/papers.html, Tech. Rep., 2019.

[11] M. Sharif and B. Hassibi, "On the capacity of mimo broadcast channels with partial side information," *IEEE Trans. on Inform. Th.*, vol. 51, no. 2, pp. 506–522, Feb 2005.