Performance Analysis of Multi-Service Oriented Multiple Access Under General Channel Correlation

Nassar Ksairi and Mérouane Debbah
Mathematical and Algorithmic Sciences Lab, France Research Center,
Huawei Technologies France SASU, 92100 Boulogne-Billancourt, France.
Emails: {nassar.ksairi, merouane.debbah}@huawei.com

***Abstract*—In this article, we provide both analytical and numerical performance analysis of multi-service oriented multiple access (MOMA), a recently proposed non-orthogonal multiple-access scheme for scenarios with a massive number of concurrent connections originating from separate service classes with diverse quality-of-service (QoS) profiles and running on both handheld terminals and Internet-of-Things (IoT) devices. MOMA is based on both class dependent hierarchical-spreading transmission scheme and per-class reception structure. The performance analysis presented in this article is based on realistic channel and signal models for scenarios where the base station is equipped with a large number of antennas. It provides asymptotically exact approximations of the ergodic rates achievable by MOMA.**

I. Introduction

Recent radio access proposals for machine type communications (MTC), whether intended as an *in-band* part of future Long-Term Evolution (LTE) and fifth-generation (5G) networks [1] or as an *out-of-band* standalone system such as LoRa™ and SIGFOX™ [2], offer improved IoT deployment capabilities than previous cellular technologies. However, they suffer from limitations in their support for at least one of the following qualities: flexibility in resource assignment, efficiency in resource utilization and differential treatment of diverse MTC QoS and traffic profiles. Indeed, while all of the existing in-band IoT compatible radio access solutions are based on reserving a region in the resource grid for MTC, they have limited flexibility in resource allocation within that region. For instance, resource assignment in eMTC [1] is limited to choosing for each MTC transmission one (or several) resource blocks (RBs) out of a total of up to only 6 RBs. The higher resource granularity of narrow-band IoT (NB-IoT) [1] comes at the price of using a different numerology in the MTC region than the one used in the rest of the resource grid by means of narrow-band frequency division multiple access (FDMA). Achieving this numerology requires in practice the insertion of wasteful guard bands. Moreover, orthogonal multiple-access schemes (such as FDMA in NB-IoT and SIGFOX™ and quasi-orthogonal code division multiple access (CDMA) in LoRa™) are not the best in terms of scalability [3] with respect to increasing device densities. Furthermore, existing proposals achieve MTC coverage enhancement by means of signal repetition thanks to transmission time interval (TTI) bundling [1]. However, this transmission technique is far from optimal in terms of resource utilization efficiency [4]. Another issue is that MTC transmissions do not all have the same QoS and traffic characteristics. Indeed, while some IoT emissions will be mainly small packets of data from simple sensors, other transmissions will carry video streaming from unmanned aerial vehicles and other IP-enabled cameras [5]. Moreover, some of the services running on handheld devices, such as messaging, have traffic characteristics and data rate requirements that resemble IoT services. One step towards supporting diverse traffic and QoS profiles across both the IoT and the handheld categories is the *prioritized random access* scheme (see [6] and [7]) which introduces different treatment in the random-access step for five classes of services/traffic profiles, namely broadband communications, high-priority MTC, low-priority MTC, scheduled MTC and emergency services. However, this support is not yet extended to multiple access.

MOMA is a in-band multiple-access scheme that has been conceived for scenarios with multiple service classes. It is based on a novel code domain multiplexing method that we dubbed *service dependent hierarchical spreading* ([8], [9]) and which guarantees support for both inter-class (quasi) orthogonality and different degrees of resource overloading across different classes. Thanks to its use of the code domain, MOMA offers more flexibility in assigning resources to different service classes and to different connections within each class without being subject to the constraints of narrow-band FDMA and while offering higher efficiency in resource utilization. These advantages were partially analyzed in [8] under an ideal array correlation model and only one linear combining method at the base station. In this article, we present a more thorough performance analysis of MOMA under more general channel correlation model and receiver structure using both theoretical tools from large random matrix theory and simulations.

Notations: The $N \times N$ identity matrix is denoted by $\mathbf{I}_N$ while notation $\mathbf{1}_{N\times M}$ stands for the $N \times M$ matrix with all its entries set to one. The spectral norm of matrix $\mathbf{M}$ is denoted as $\|\mathbf{M}\|$ while $\mathrm{tr}(\mathbf{M})$ designates the trace of $\mathbf{M}$.

II. System Model

Consider a wireless system consisting of one base station (BS) equipped with M antennas (indexed using $m \in \{1, \ldots, M\}$) serving K single-antenna user terminals (UTs) that have *uplink* data to transmit and which are grouped into $C \geq 2$ service classes. Each service class $c \in \{1, \ldots, C\}$ is characterized with a target uplink data rate r_c that is common

to all the class members. The C classes are indexed in a way that their respective target data rates satisfy $r_1 > r_2 > \ldots > r_C$. In the sequel, we use $\mathcal{K}_c \subset \{1, 2, \ldots, K\}$ to designate the indexes of the users of the c-th service class. We also define $K_c \stackrel{\text{def.}}{=} |\mathcal{K}_c|$. Signal transmission is done using orthogonal frequency division multiplexing (OFDM) with N_{FFT} subcarriers, N_{CP}-long cyclic prefix and a total bandwidth of W Hz. Assume that the OFDM resource grid is structured into RBs and focus on a transmission that is taking place on N RBs which are adjacent in the frequency, the time or in both the frequency and the time domains as shown in Figure 1. Note that RB adjacency in the time domain could be the result of TTI bundling that will be supported for several transmission scenarios in next-generation cellular systems [1]. The wireless link from UT k to BS antenna m is a multi-path channel with $L \geq 1$ dominant paths that remain constant within the duration of an OFDM symbol but which might vary from one symbol to another. Denote by $t \in \{1, \ldots, T\}$ the index of the $T \geq N$ OFDM symbols that fall within the N RBs and let $h_{k,m,t}(.)$ designate the continuous-time impulse response of this channel during the t-th OFDM symbol, then

$$h_{k,m,t}(\tau) = \sum_{l=0}^{L-1} \alpha_{k,m,l,t} \delta(\tau - \tau_l), \; \forall \tau \in \mathbb{R}, \tag{1}$$

where τ_l is the delay of the l-th path and $\alpha_{k,m,l,t}$ is its complex-valued gain. Path gains $\{\alpha_{k,m,l,t}\}_{l \in \{0 \cdots L-1\}}$ are modeled, as in [10], as mutually independent zero-mean random variables with variances $\sigma_l^2 \stackrel{\text{def.}}{=} \mathbb{E}\left[|\alpha_{k,m,l,t}|^2\right]$ that satisfy $\sum_{l=0}^{L-1} \sigma_l^2 = 1$ and

$$\begin{aligned}\mathbb{E}\left[\alpha_{k,m,l,t} \alpha^*_{k,m,l,s}\right] &= \sigma_l^2 J_0\left(2\pi f_k^D \left(N_{\text{FFT}} + N_{\text{CP}}\right) T_s (t-s)\right) \\ &\stackrel{\text{def.}}{=} \sigma_l^2 r_k^\alpha (t-s),\end{aligned} \tag{2}$$

where f_k^D is the maximum Doppler frequency shift and where $T_s \stackrel{\text{def.}}{=} 1/W$. Now, define $\mathbf{h}_{k,l,t} \stackrel{\text{def.}}{=} [h_{k,1,l,t} \cdots h_{k,M,l,t}]^{\text{T}}$, $\boldsymbol{\alpha}_{k,l,t} \stackrel{\text{def.}}{=} [\alpha_{k,1,l,t} \cdots \alpha_{k,M,l,t}]^{\text{T}}$ and assume $\boldsymbol{\alpha}_{k,l,t} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{R}_{k,l}^\alpha\right)$, where $\mathbf{R}_{k,l}^\alpha \stackrel{\text{def.}}{=} \mathbb{E}\left[\boldsymbol{\alpha}_{k,l,t} \boldsymbol{\alpha}_{k,l,t}^{\text{H}}\right]$. Finally, assume as in [11] and [12] that the correlation between the two vectors $\boldsymbol{\alpha}_{k,l,t_1}$ and $\boldsymbol{\alpha}_{k,l,t_2}$ where $t_1 \neq t_2$ is separable into time domain and space domain multiplicative terms. More precisely, $\mathbb{E}\left[\boldsymbol{\alpha}_{k,l,t_1} \boldsymbol{\alpha}_{k,l,t_2}^{\text{H}}\right] = r_k^\alpha (t_1 - t_2) \mathbf{R}_{k,l}^\alpha$. From (1), the frequency domain channel coefficient between UT k and the m-th BS antenna at subcarrier n during OFDM symbol t, denoted as $H_{k,m,t}$, is given by

$$H_{k,m,t,n} = \sum_{l=0}^{L-1} \alpha_{k,m,l,t} e^{-2\pi\imath \frac{\tau_l n}{N_{\text{FFT}} T_s}}. \tag{3}$$

Now, define $\mathbf{H}_{k,t,n} \stackrel{\text{def.}}{=} [H_{k,1,t,n} \cdots H_{k,M,t,n}]^{\text{T}} \in \mathbb{C}^{M \times 1}$ and denote by $x_{k,t,n}$ the sample transmitted on the n-th subcarrier during the t-th OFDM symbol. The corresponding $M \times 1$ vector of samples received at the BS array then writes as

$$\mathbf{y}_{t,n} = \sum_{k=1}^{K} \sqrt{g_k} \mathbf{H}_{k,t,n} x_{k,t,n} + \mathbf{z}_{t,n}, \tag{4}$$

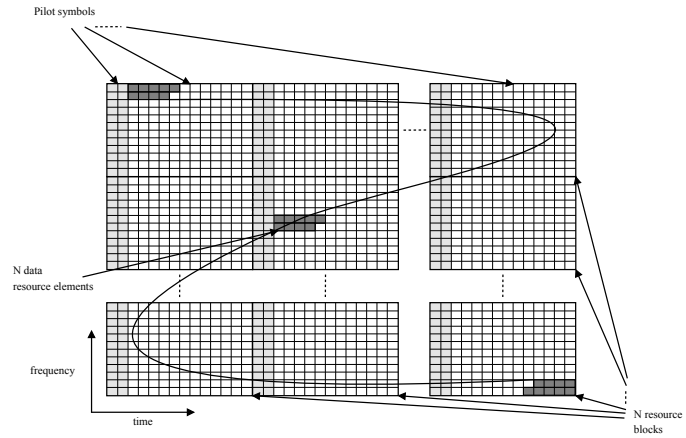

Fig. 1. An example of a resource block assignment and of a possible associated pilot symbol pattern

where $\mathbf{z}_{t,n}$ is a $M \times 1$ vector of $\mathcal{CN}(0, \sigma^2)$ noise samples and g_k is the large-scale fading factor. Next, we show how samples $x_{k,t,n}$ are constructed from the data symbols of UT k depending on the service class to which UT k belongs.

III. Multi-Service Oriented Multiple Access

A. MOMA Transmitter

Let $\mathbf{U}$ be a $N \times N$ orthogonal code matrix, e.g. a Walsh-Hadamard matrix, a discrete Fourier transform (DFT) matrix, etc. The u-th column of $\mathbf{U}$ is normalized such that

$$\|[\mathbf{U}]_u\|^2 = 1, \; \forall u \in \{1, \ldots, N\}. \tag{5}$$

Next, split $\mathbf{U}$ into C sub-matrices $\{\mathbf{U}_c\}_{c=1 \cdots C}$ with the c-th sub-matrix having dimensions $N \times N_c$ that satisfy

$$\frac{K_C}{N_C} > \cdots \frac{K_2}{N_2} > \frac{K_1}{N_1}. \tag{6}$$

Now, let $\mathbf{W}_c \in \mathbb{C}^{N_c \times K_c}$ be a set of predefined C matrices, each having mutually distinct columns. In MOMA, the N-long spreading code $\mathbf{c}_k$ for $k \in \mathcal{K}_c$ is defined as

$$\mathbf{c}_k = [\mathbf{U}_c \mathbf{W}_c]_{i_k} = \sum_{u=1}^{N_c} [\mathbf{w}_k]_u [\mathbf{U}_c]_u$$

where $\mathbf{w}_k \stackrel{\text{def.}}{=} [\mathbf{W}_c]_{i_k}$ is the i_k-th column of $\mathbf{W}_c$ (for some $i_k \in \{1, \ldots, K_c\}$) and where $[\mathbf{w}_k]_u$ is the u-th entry of $\mathbf{w}_k$. Let s_k designate one of the unit-variance data symbols from UT k. Any such data symbol s_k is then multiplied with vector $\mathbf{c}_k$ resulting in N spread samples $\mathbf{c}_k s_k$. To limit as much as possible the effect of channel gain variations on the desired properties of the code, we map the samples resulting from spreading one data symbol to a group of N neighboring resource elements (REs). Let $\mathcal{N}$ be the group of REs used to transmit the spread samples $\mathbf{c}_k s_k$ and denote by $x_{k,t,n}$ the scaled spread sample transmitted on $(t, n) \in \mathcal{N}$, then

$$x_{k,t,n} = \sqrt{P_c} [\mathbf{c}_k]_{(t,n)} s_k, \tag{7}$$

where $[\mathbf{c}_k]_{(t,n)}$ is the entry of $\mathbf{c}_k$ mapped to (t, n). Vector $\mathbf{w}_k$ will help the BS distinguish the signal of UT k from the

interfering signals transmitted by the other UTs that belong to the same service class as k. In principle, $\mathbf{W}_c$ could be any $N_c \times K_c$ complex-valued matrix with entries chosen that the following average power constraint is satisfied

$$\mathbb{E}\left[|x_{k,t,n}|^2\right] = P_c \quad (k \in \mathcal{K}_c) \tag{8}$$

Remark 1. *The following shorter-term average transmit power constraint might be preferred in order to ease the requirements on the transmit power amplifiers:*

$$\|\mathbf{w}_k\|^2 = 1, \forall k \in \{1, \ldots, K_c\}. \tag{9}$$

Indeed, if (5) *and* (9) *are respected, then* $\|\mathbf{c}_k\|^2 = 1$ *and the short-term constraint* $\frac{1}{N}\sum_{(t,n)\in\mathcal{N}} |x_{k,t,n}|^2 = P_c$ *replaces the long-term constraint in* (8)*. One practical way to obtain matrices* $\mathbf{W}_c$ *that satisfy* (9) *is to use pseudo-noise (PN) generators [13] with different initializations to generate the different columns of* $\mathbf{W}_c$*. However, since PN sequences are binary, the number of different* N*-long sequences we can get will be upper bounded by* 2^N*. A more systematic method (left for future work) is to select* $\{\mathbf{w}_k\}_{k\in\mathcal{K}_c}$ *as a collection of* K_c *points on the surface of a unit-radius* K_c*-dimensional complex-valued sphere with a minimal pairwise angle that is as large as possible. This problem can be solved in advance for different configurations of* (N_c, K_c) *[14] and the outcome of this offline optimization can be stored in look-up tables.*

B. MOMA Receiver

1) Channel Estimation: Assume that N_p REs are reserved for pilot transmission within the N RBs and denote by $\mathcal{P}$ the set of positions of these REs. Let $\mathcal{P}_k \subset \mathcal{P}$ denote the subset of pilot resource elements used in estimating the channel from UT $k \in \mathcal{K}_c$. Let p_{k,t_i^p,n_i^p} designate the pilot symbol transmitted by UT k at $(t_i^p, n_i^p) \in \mathcal{P}_k$ with power P_c. Pilot sequences from different UTs are assumed here to be multiplexed using some combination of frequency division multiplexing (FDM) and time division multiplexing (TDM) resulting in $N_k^p = N_p/K$.

Remark 2. *To reduce pilot overhead, the same service dependent hierarchical spreading principle that is used for data transmission in MOMA can in principle be applied to pilots. Performance analysis of such a non-orthogonal pilot multiplexing scheme is left for future research works.*

The estimate $\hat{\mathbf{H}}_{k,t_i^p,n_i^p}$ of $\mathbf{H}_{k,t_i^p,n_i^p}$ can be obtained from $\mathbf{p}_k \overset{\text{def.}}{=} [p_{k,t_1,n_1} \cdots p_{k,t_{N_k^p},n_{N_k^p}}]^{\mathrm{T}}$ using least squares (LS) or linear minimum mean squared error (LMMSE) estimation as

$$\hat{\mathbf{H}}_{k,t_i^p,n_i^p} = \mathbf{H}_{k,t_i^p,n_i^p} + \frac{1}{\sqrt{\gamma_k^{\mathrm{CE}}}}\mathbf{n}_{k,t_i^p,n_i^p}, \tag{10}$$

where $\mathbf{n}_{k,t_i^p,n_i^p} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ is an estimation error vector that is uncorrelated with $\mathbf{H}_{k,t_i^p,n_i^p}$ and where γ_k^{CE} is the effective channel estimation signal-to-noise ratio (SNR) given by $\gamma_k^{\mathrm{CE}} = \frac{g_k P_c}{\sigma^2}$. Define $\hat{\mathbf{H}}_k^p$ as the $N_p M \times 1$ vector made by concatenating the N_k^p channel estimates $\hat{\mathbf{H}}_{k,t_i^p,n_i^p}$:

$$\hat{\mathbf{H}}_k^p \overset{\text{def.}}{=} \left[\hat{\mathbf{H}}_{k,t_1^p,n_1^p}^{\mathrm{T}} \cdots \hat{\mathbf{H}}_{k,t_{N_k^p},n_{N_k^p}^p}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{N_p M \times 1}. \tag{11}$$

Now, define $\mathbf{H}_k$ as the vector made by concatenating the N channel vectors $\mathbf{H}_{k,t_i,n_i}$ at the position $(t_i, n_i) \in \mathcal{N}$ to which the spread samples of data symbol s_k are mapped, i.e.,

$$\mathbf{H}_k \overset{\text{def.}}{=} \left[\mathbf{H}_{k,t_1,n_1}^{\mathrm{T}} \cdots \mathbf{H}_{k,t_N,n_N}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{NM \times 1}. \tag{12}$$

We propose to estimate $\mathbf{H}_k$ using LMMSE interpolation based on $\hat{\mathbf{H}}_k^p$. The resulting estimate (denoted $\hat{\mathbf{H}}_k$) is given by

$$\hat{\mathbf{H}}_k = \mathbf{R}_k^{\mathcal{NP}} \mathbf{Q}_k^{\mathcal{P}} \hat{\mathbf{H}}_k^p \in \mathbb{C}^{NM \times 1}, \tag{13}$$

with $\mathbf{R}_k^{\mathcal{NP}} = \mathbb{E}\left[\mathbf{H}_k \left(\hat{\mathbf{H}}_k^p\right)^{\mathrm{H}}\right]$, $\mathbf{Q}_k^{\mathcal{P}} = \left(\frac{1}{\gamma_k^{\mathrm{CE}}}\mathbf{I}_{MN_k^p} + \mathbf{R}_k^{\mathcal{P}}\right)^{-1}$ and $\mathbf{R}_k^{\mathcal{P}} = \mathbb{E}\left[\hat{\mathbf{H}}_k^p \left(\hat{\mathbf{H}}_k^p\right)^{\mathrm{H}}\right]$. Using the basic properties of LMMSE estimation, it is straightforward to show that $\mathbf{H}_k = \hat{\mathbf{H}}_k + \tilde{\mathbf{H}}_k$, where $\tilde{\mathbf{H}}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_k - \mathbf{\Phi}_k)$ and

$$\mathbf{\Phi}_k \overset{\text{def.}}{=} \mathbf{R}_k^{\mathcal{NP}} \mathbf{Q}_k^{\mathcal{P}} \left(\mathbf{R}_k^{\mathcal{NP}}\right)^{\mathrm{H}}, \tag{14}$$

$$\mathbf{R}_k \overset{\text{def.}}{=} \left[\sum_{l=0}^{L-1} r_k^{\alpha}(t_i - t_j)\mathbf{R}_{k,l}^{\alpha} e^{-2\pi\imath\frac{\tau_l(n_i - n_j)}{N_{\mathrm{FFT}}T_s}}\right]_{1\le i,j\le N}. \tag{15}$$

2) Data Detection: It is useful to define

$$\mathbf{C}_k \overset{\text{def.}}{=} \operatorname{diag}(\mathbf{c}_k) \otimes \mathbf{I}_M. \tag{16}$$

Data detection is done using $\sqrt{g_k P_c}\mathbf{C}_k\hat{\mathbf{H}}_k$ (the effective NM-long channel vector when the spreading code is incorporated) as the signature of UT k. Moreover, we assume that the BS employs minimum mean squared error (MMSE) detection for the higher-rate service classes and matched-filtering (MF) for the lower-rate ones using the following vector

$$\mathbf{r}_k = \begin{cases} \mathbf{r}_k^{\mathrm{MF}}, & \text{MF detection,} \\ \mathbf{r}_k^{\mathrm{MMSE}}, & \text{MMSE detection,} \end{cases} \tag{17}$$

where $\mathbf{r}_k^{\mathrm{MF}} = \sqrt{g_k P_c}\mathbf{C}_k\hat{\mathbf{H}}_k$ and $\mathbf{r}_k^{\mathrm{MMSE}} = \left(\sum_{j\in\mathcal{K}_c} g_j P_c \mathbf{C}_j \hat{\mathbf{H}}_j \hat{\mathbf{H}}_j^{\mathrm{H}} \mathbf{C}_j^{\mathrm{H}} + \sigma^2 \mathbf{I}_{NM}\right)^{-1} \sqrt{g_k P_c}\mathbf{C}_k\hat{\mathbf{H}}_k$.

IV. Achievable Rates and Asymptotic Analysis

The ergodic rate R_k in bits/s/Hz achieved in the uplink by UT k from the c-th service class is given by $R_k = \mathbb{E}[\log_2(1+\gamma_k)]$ where the expectation is taken with respect to the joint distribution of $\{\hat{\mathbf{H}}_k, \mathbf{H}_k\}_{k\in\{1;\ldots,K\}}$ and where γ_k is the signal-to-interference-plus-noise ratio given by (18). To derive asymptotic deterministic equivalents of the SINR in (18), the following assumption is needed.

Assumption 1. *For all* $k \in \{1, \ldots, K\}$,$l \in \{0, \ldots, L-1\}$*:*

$$\limsup_M \|\mathbf{R}_{k,l}^{\alpha}\| < \infty, \quad \liminf_M \frac{1}{M}\mathrm{tr}\mathbf{R}_{k,l}^{\alpha} > 0. \tag{20}$$

Theorem 1. *If* $\mathbf{r}_k = \mathbf{r}_k^{\mathrm{MF}}$*, then under Assumption 1* $\gamma_k - \bar{\gamma}_k^{\mathrm{MF}} \xrightarrow[M,K_c\to\infty,\lim K_c/M<\infty]{a.s.} 0$ *where* $\bar{\gamma}_k^{\mathrm{MF}}$ *is defined in* (19)*.*

Sketch of the proof. The proof follows from [15, Theorem 3] if matrices $\mathbf{S}_k = g_k P_c \mathbf{C}_k \mathbf{R}_k \mathbf{C}_k^{\mathrm{H}}$ and $\mathbf{F}_k = g_k P_c \mathbf{C}_k \mathbf{\Phi}_k \mathbf{C}_k^{\mathrm{H}}$ satisfy $\limsup_M \|\mathbf{S}_k\| < \infty$, $\liminf_M \frac{1}{NM}\mathrm{tr}\mathbf{S}_k > 0$ and

$$\gamma_k = \frac{P_c g_k \left|\mathbf{r}_k^{\mathrm{H}} \mathbf{C}_k \hat{\mathbf{H}}_k\right|^2}{\mathbf{r}_k^{\mathrm{H}} \left(\sigma^2 \mathbf{I}_{MN} + P_c g_k \mathbf{C}_k \tilde{\mathbf{H}}_k \tilde{\mathbf{H}}_k^{\mathrm{H}} \mathbf{C}_k^{\mathrm{H}} - P_c g_k \mathbf{C}_k \mathbf{H}_k \mathbf{H}_k^{\mathrm{H}} \mathbf{C}_k^{\mathrm{H}} + \sum_{\tilde{c}=1}^{C} \sum_{j \in \mathcal{K}_{\tilde{c}}} P_{\tilde{c}} g_j \mathbf{C}_j \mathbf{H}_j \mathbf{H}_j^{\mathrm{H}} \mathbf{C}_j^{\mathrm{H}}\right) \mathbf{r}_k}, \quad \forall k \in \mathcal{K}_c \tag{18}$$

$$\bar{\gamma}_k^{\mathrm{MF}} = \frac{P_c g_k \left(\frac{1}{NM} \mathrm{tr} \mathbf{C}_k \mathbf{\Phi}_k \mathbf{C}_k^{\mathrm{H}}\right)^2}{\frac{\sigma^2}{(NM)^2} \mathrm{tr} \mathbf{C}_k \mathbf{\Phi}_k \mathbf{C}_k^{\mathrm{H}} + \frac{1}{NM} \sum_{\tilde{c}=1}^{C} \sum_{j \in \mathcal{K}_{\tilde{c}}} P_{\tilde{c}} g_j \frac{1}{NM} \mathrm{tr} \mathbf{C}_j \mathbf{R}_j \mathbf{C}_j^{\mathrm{H}} \mathbf{C}_k \mathbf{\Phi}_k \mathbf{C}_k^{\mathrm{H}}}, \quad \forall k \in \mathcal{K}_c \tag{19}$$

$\limsup_M \left(\frac{1}{K} \sum_{j \in \{1,\ldots,K\}} \frac{1}{NM} \mathrm{tr} \mathbf{F}_j\right)^{-1} < \infty$. These conditions can be shown to hold under Assumptions 1 by using basic properties of the trace function and of the spectral norm. □

Now, let $\left\{\{\mathbf{S}_j\}_{j \in \{1 \cdots J\}}, \mathbf{F}\right\}$ be a set of $J+1$ Hermitian non-negative definite $I \times I$ matrices, each with a uniformly bounded spectral norm with respect to I. We define for any $\rho > 0$ the matrix $\mathbf{T}\left(\rho, I, \{\mathbf{S}_j\}_{j \in \{1 \cdots J\}}\right) \in \mathbb{C}^{I \times I}$ as

$$\mathbf{T} \overset{\mathrm{def.}}{=} \left(\frac{1}{I} \sum_{j \in \{1 \cdots J\}} \frac{\mathbf{S}_j}{1+\delta_j} + \rho \mathbf{I}_I\right)^{-1} \tag{21}$$

where for all $k \in \{1, \ldots, J\}$ and $t \geq 1$

$$\delta_k\left(\rho, I, \{\mathbf{S}_j\}_{j \in \{1 \cdots J\}}\right) \overset{\mathrm{def.}}{=} \lim_{t \to \infty} \delta_k^{(t)}, \tag{22}$$

$$\delta_k^{(t)} \overset{\mathrm{def.}}{=} \frac{1}{I} \mathrm{tr} \mathbf{S}_k \left(\frac{1}{I} \sum_{j \in \{1 \cdots J\}} \frac{\mathbf{S}_j}{1+\delta_j^{(t-1)}} + \rho \mathbf{I}_I\right)^{-1}. \tag{23}$$

If we set $\delta_k^{(0)} = \frac{1}{\rho}$ for all $k \in \{1, \ldots, J\}$, then the limit in (22) exists due to [16, Theorem 1]. We also define

$$\mathbf{T}' \overset{\mathrm{def.}}{=} \mathbf{T}\mathbf{F}\mathbf{T} + \frac{1}{I} \mathbf{T} \sum_{j \in \{1 \cdots J\}} \frac{\mathbf{S}_j \delta_j'}{(1+\delta_j)^2} \mathbf{T}, \tag{24}$$

where $\{\delta_j'\left(\rho, I, \{\mathbf{S}_j\}_{j \in \{1 \cdots J\}}, \mathbf{F}\right)\}_{j \in \{1 \cdots J\}}$ are the entries of the J-long vector $\boldsymbol{\delta}'$ defined as $\boldsymbol{\delta}' \overset{\mathrm{def.}}{=} (\mathbf{I}_J - \mathbf{J})^{-1} \mathbf{v}$ with

$$[\mathbf{J}]_{kj} \overset{\mathrm{def.}}{=} \frac{\frac{1}{I} \mathrm{tr} \mathbf{S}_k \mathbf{T} \mathbf{S}_j \mathbf{T}}{I(1+\delta_j)^2}, \qquad [\mathbf{v}]_k \overset{\mathrm{def.}}{=} \frac{1}{I} \mathrm{tr} \mathbf{S}_k \mathbf{T} \mathbf{F} \mathbf{T} \tag{25}$$

for any $(k, j) \in \{1, \ldots, J\}^2$. MOMA's asymptotic ergodic rates for classes with MMSE detection can be characterized with the help of the above definitions by the following theorem. Its proof is omitted here due to page count limitations.

Theorem 2. *If $\mathbf{r}_k = \mathbf{r}_k^{\mathrm{MMSE}}$, then under Assumption 1 $\gamma_k - \bar{\gamma}_k^{\mathrm{MMSE}} \xrightarrow[M, K_c \to \infty, \lim K_c/M < \infty]{a.s.} 0$ where $\forall k \in \mathcal{K}_c$*

$$\bar{\gamma}_k^{\mathrm{MMSE}} = \frac{\delta_{ck}^2}{\frac{\sigma^2}{(NM)^2} \mathrm{tr} P_c g_k \mathbf{C}_k \mathbf{\Phi}_k \mathbf{C}_k^{\mathrm{H}} \bar{\mathbf{T}}_c' + \frac{1}{NM} \sum_{\tilde{c}=1}^{C} \sum_{j \in \mathcal{K}_{\tilde{c}}} \mu_{c\tilde{c}kj}}, \tag{26}$$

and

$$\mu_{c\tilde{c}kj} \overset{\mathrm{def.}}{=} \frac{P_{\tilde{c}} g_j}{NM} \mathrm{tr} \mathbf{C}_j \mathbf{R}_j \mathbf{C}_j^{\mathrm{H}} \mathbf{T}_{ck}' - \mathcal{K}_c(j) \frac{2\mathrm{Re}\left(\vartheta_{c\tilde{c}j}^* \vartheta_{c\tilde{c}kj}'\right)(1+\delta_{cj}) - |\vartheta_{c\tilde{c}j}|^2 \delta_{ckj}'}{(1+\delta_{cj})^2}, \tag{27}$$

$$\vartheta_{c\tilde{c}j} \overset{\mathrm{def.}}{=} \frac{P_{\tilde{c}} g_j}{NM} \mathrm{tr} \mathbf{C}_j \mathbf{\Phi}_j \mathbf{C}_j^{\mathrm{H}} \mathbf{T}_c, \tag{28}$$

$$\vartheta_{c\tilde{c}kj}' \overset{\mathrm{def.}}{=} \frac{P_{\tilde{c}} g_j}{NM} \mathrm{tr} \mathbf{C}_j \mathbf{\Phi}_j \mathbf{C}_j^{\mathrm{H}} \mathbf{T}_{ck}'. \tag{29}$$

Here, we defined $\delta_{ck} = \delta_k\left(\frac{\sigma^2}{M}, NM, \{P_c g_j \mathbf{C}_j \mathbf{\Phi}_j \mathbf{C}_j^{\mathrm{H}}\}_{j \in \mathcal{K}_c}\right)$, $\delta_{ck}' = \delta_k'\left(\frac{\sigma^2}{M}, NM, \{P_c g_j \mathbf{C}_j \mathbf{\Phi}_j \mathbf{C}_j^{\mathrm{H}}\}_{j \in \mathcal{K}_c}, \mathbf{\Phi}_k\right)$, $\mathbf{T}_c = \mathbf{T}\left(\frac{\sigma^2}{M}, NM, \{P_c g_j \mathbf{C}_j \mathbf{\Phi}_j \mathbf{C}_j^{\mathrm{H}}\}_{j \in \mathcal{K}_c}\right)$, $\mathbf{T}_{ck}' = \mathbf{T}'\left(\frac{\sigma^2}{M}, NM, \{P_c g_j \mathbf{C}_j \mathbf{\Phi}_j \mathbf{C}_j^{\mathrm{H}}\}_{j \in \mathcal{K}_c}, \mathbf{\Phi}_k\right)$, *and* $\bar{\mathbf{T}}_c' = \mathbf{T}'\left(\frac{\sigma^2}{M}, NM, \{P_c g_j \mathbf{C}_j \mathbf{\Phi}_j \mathbf{C}_j^{\mathrm{H}}\}_{j \in \mathcal{K}_c}, \mathbf{I}_{NM}\right)$.

Let us now consider the case where the time and frequency domain correlation functions are flat over the RE set $\mathcal{N}$, i.e., $r_k^{\alpha}(t_i - t_j) = 1$ and $e^{-2\pi\imath \frac{\tau_l (n_i - n_j)}{N_{\mathrm{FFT}} T_s}} = 1$. In this case, $\mathbf{R}_k = \mathbf{1}_{N \times N} \otimes \check{\mathbf{R}}_k$, $\mathbf{R}_k^{\mathcal{NP}} = \mathbf{1}_{N \times 1} \otimes \left[\check{\mathbf{R}}_{k,1}^{\mathcal{NP}} \cdots \check{\mathbf{R}}_{k,N_p}^{\mathcal{NP}}\right]$, $\mathbf{\Phi}_k = \mathbf{1}_{N \times N} \otimes \check{\mathbf{\Phi}}_k$, where $\forall j \in \{1, \ldots, N_p\}$

$$\check{\mathbf{R}}_k \overset{\mathrm{def.}}{=} \sum_{l=0}^{L-1} \mathbf{R}_{k,l}^{\alpha} \tag{30}$$

$$\check{\mathbf{R}}_{k,j}^{\mathcal{NP}} \overset{\mathrm{def.}}{=} \sum_{l=0}^{L-1} r_k^{\alpha}(t_1 - t_j^p) \mathbf{R}_{k,l}^{\alpha} e^{-2\pi\imath \frac{\tau_l (n_1 - n_j^p)}{N_{\mathrm{FFT}} T_s}}, \tag{31}$$

$$\check{\mathbf{\Phi}}_k \overset{\mathrm{def.}}{=} \sum_{i=1}^{N_p} \sum_{j=1}^{N_p} \check{\mathbf{R}}_{k,i}^{\mathcal{NP}} \left[\mathbf{Q}_k^{\mathcal{P}}\right]_{i,j} \left(\check{\mathbf{R}}_{k,j}^{\mathcal{NP}}\right)^{\mathrm{H}} \tag{32}$$

The following assumption is a relaxation of strict equality to 1 of the time and frequency domain correlation coefficients.

Assumption 2. *For all $k \in \{1, \ldots, K\}$:*

$$\mathbf{R}_k = \mathbf{1}_{N \times N} \otimes \check{\mathbf{R}}_k + \tilde{\mathbf{R}}_k,$$
$$\mathbf{R}_k^{\mathcal{NP}} = \mathbf{1}_{N \times 1} \otimes \left[\check{\mathbf{R}}_{k,1}^{\mathcal{NP}} \cdots \check{\mathbf{R}}_{k,N_p}^{\mathcal{NP}}\right] + \tilde{\mathbf{R}}_k^{\mathcal{NP}},$$

with $\tilde{\mathbf{R}}_k \in \mathbb{C}^{NM \times NM}$ and $\tilde{\mathbf{R}}_k^{\mathcal{NP}} \in \mathbb{C}^{NM \times N_p M}$ satisfying $\lim_{M \to \infty} \left\|\tilde{\mathbf{R}}_k\right\| = \lim_{M \to \infty} \left\|\tilde{\mathbf{R}}_k^{\mathcal{NP}}\right\| = 0$.

Note that in practice, if we choose $\mathcal{N}$ to consist of at most few tens of REs, Assumption 2 becomes a mild one.

Corollary 1. *Under Assumptions 1 and 2, we have $\forall k \in \mathcal{K}_c$*

$$\bar{\gamma}_k^{\mathrm{MF}} = \frac{P_c g_k \left(\frac{1}{M} \mathrm{tr} \check{\mathbf{\Phi}}_k\right)^2}{\frac{\sigma^2}{M^2} \mathrm{tr} \check{\mathbf{\Phi}}_k + \frac{1}{M} \sum_{j \in \mathcal{K}_c} P_c g_j \frac{1}{M} \left(\mathbf{w}_j^{\mathrm{H}} \mathbf{w}_k\right)^2 \mathrm{tr} \check{\mathbf{R}}_j \check{\mathbf{\Phi}}_k}. \tag{33}$$

Sketch of the proof. Note that under Assumption 2 and due to (14), $\mathbf{\Phi}_k = \mathbf{1}_{N \times N} \otimes \check{\mathbf{\Phi}}_k + \tilde{\mathbf{\Phi}}_k$ where $\lim_{M \to \infty} \left\|\tilde{\mathbf{\Phi}}_k\right\| = 0$ for

all $k \in \{1, \ldots, K\}$. Also note that for any pair $(k, j) \in \mathcal{K}_c^2$, $\mathbf{c}_j^{\mathrm{H}} \mathbf{c}_k = \mathbf{w}_j^{\mathrm{H}} \mathbf{w}_k$ due to (7). Finally, use basic properties of the trace function and of the Kronecker product to conclude, after some tedious derivations, that (33) holds. □

Corollary 2. *Under Assumptions 1 and 2, we have* $\forall k \in \mathcal{K}_c$

$$\bar{\gamma}_k^{\mathrm{MMSE}} = \frac{\delta_{ck}^2}{\frac{\sigma^2}{(NM)^2} \mathrm{tr} P_c g_k \mathbf{C}_k \mathbf{\Phi}_k \mathbf{C}_k^{\mathrm{H}} \bar{\mathbf{T}}_c' + \frac{1}{NM} \sum_{j \in \mathcal{K}_c} \mu_{cckj}}. \tag{34}$$

Sketch of the proof. First, note that for any $j \in \mathcal{K}_{\tilde{c}}$ with $\tilde{c} \neq c$ the term $\mu_{c\tilde{c}kj}$ in (26) can be written due to (27) as

$$\mu_{c\tilde{c}kj} = \frac{P_{\tilde{c}} g_j}{NM} \mathrm{tr} \mathbf{C}_j \mathbf{R}_j \mathbf{C}_j^{\mathrm{H}} \mathbf{T}_c \mathbf{C}_k \mathbf{\Phi}_k \mathbf{C}_k^{\mathrm{H}} \mathbf{T}_c + \frac{P_{\tilde{c}} g_j}{(NM)^2} \sum_{i \in \mathcal{K}_c} \frac{\delta_{ci}^{'}}{(1+\delta_{ci})^2} \mathrm{tr} \mathbf{C}_j \mathbf{R}_j \mathbf{C}_j^{\mathrm{H}} \mathbf{T}_c \mathbf{C}_i \mathbf{\Phi}_i \mathbf{C}_i^{\mathrm{H}} \mathbf{T}_c \,. \tag{35}$$

Second, refer to (21) and use the Cayley-Hamilton theorem [17] to write $\mathbf{T}_c$ as a weighted sum of the first NM powers of the matrix $\frac{1}{NM} \sum_{j \in \mathcal{K}_c} \frac{\mathbf{\Phi}_j}{1+\delta_{cj}} + \frac{\sigma^2}{M} \mathbf{I}_{NM}$. Finally, plug this weighted sum into (35) and use Assumption 2 along with the fact that $\mathbf{c}_j^{\mathrm{H}} \mathbf{c}_k = 0$ for any $(k, j) \in \mathcal{K}_c \times \mathcal{K}_{\tilde{c}}$ satisfying $\tilde{c} \neq c$ to conclude after some tedious calculation that $\mu_{c\tilde{c}kj} \xrightarrow[M \to \infty]{a.s.} 0$ for any $j \in \mathcal{K}_{\tilde{c}}$, thus completing the proof of Corollary 2. □

Asymptotic inter-class orthogonality is thus achieved by MOMA without requiring the actual channel gains over the different N REs used for spreading one data symbol to be equal. This relaxation can be thought of as a consequence of the *massive MIMO effect* which restores the orthogonality of the original $N \times N$ code matrix. Now, to get more insight into the intra-class multiuser interference performance achieved by MOMA, we focus on MF detection and we consider a simplified channel model by assuming

$$\mathbf{R}_{k,l}^{\alpha} = \sigma_l^2 \mathbf{I}_M, \; \forall k \in \{1 \ldots, K\}, l \in \{0, \ldots, L-1\}. \tag{36}$$

Corollary 3. *Assume that the empirical distribution of the large-scale fading coefficients* $\{g_k\}_{k \in \mathcal{K}_c}$ *converges as* $K_c \to \infty$ *to the distribution of a random variable with mean* $\bar{g}$*. If the columns of* $\mathbf{W}_c$ *are further chosen as realizations of i.i.d. zero-mean random vectors with independent entries that satisfy the average power constraint in* (8)*, then under the simplified channel model in* (36) *and Assumptions 1 and 2 the deterministic equivalent* $\bar{\gamma}_k^{\mathrm{MF}}$ *defined in* (19) *is given as*

$$\bar{\gamma}_k^{\mathrm{MF}} = \frac{P_c g_k \frac{\mathrm{tr} \check{\mathbf{\Phi}}_k}{M}}{\frac{\sigma^2}{M} + \frac{\bar{g} P_c K_c}{N_c M}}, \forall k \in \mathcal{K}_c \,. \tag{37}$$

Note that the assumption on the empirical distribution of $\{g_k\}_{k \in \mathcal{K}_c}$ is a mild technical assumption which holds, for instance, if the positions of the UTs of each service class are scattered diversely enough within the cell area. Associating massive MIMO and MOMA thus amounts to an effective spreading gain equal to $N_c M$, thus validating (6).

Sketch of the proof. Plugging (36) into (30) gives $\check{\mathbf{R}}_k = \sum_{l=0}^{L-1} \sigma_l^2 \mathbf{I}_M = \mathbf{I}_M$. Since the entries of $\mathbf{W}_c$ are zero-mean i.i.d., then $\frac{1}{M} \sum_{j \in \mathcal{K}_c} g_j P_c \left(\mathbf{w}_j^{\mathrm{H}} \mathbf{w}_k\right)^2 \mathrm{tr} \check{\mathbf{R}}_j \check{\mathbf{\Phi}}_k$ in (33) is asymptotically equivalent to $\mathrm{tr} \check{\mathbf{\Phi}}_k \frac{\bar{g} P_c K_c}{N_c M}$ as $K_c \to \infty$ due to the *central limit theorem* and to the assumption made about the empirical distribution of $\{g_k\}_{k \in \mathcal{K}_c}$. Putting things together completes the proof that (37) holds. □

V. Numerical Results

In this section, we consider a LTE-like system with a total bandwidth $W = 10$ MHz out of which MOMA occupies a sub-band covering one RB. On this RB, TTI bundling is applied on the basis of 6 consecutive TTIs i.e., $N = 6$. We also assume $C = 2$, i.e., the terminals scheduled on the considered resource region belong to two service classes. The noise power spectral density is equal to $N_0 = -174$ dBm/Hz and users' channels follow the Extended Type Urban (ETU) model [18] with $f_k^D = 70$ Hz and $\max_l \tau_l = 5$ μs. Channel vector spatial correlation is simulated using the *physical channel model* [15] for uniform linear arrays with a number of dimensions equal to half the number of BS antennas. The number N_p/N of pilot symbols per RB is equal to the total number of UTs, i.e., $N_p = N(K_1 + K_2)$. Devices of the first class emit with power $P_1 = 23$ dBm while the transmit power of the second class is $P_2 = 17$ dBm. The SNR $P_c g_k / \sigma^2$ of the signals received from the terminals of the first class is equal to 15 dB while the SNR of the second class is equal to 1.5 dB. This difference in SNR values is intended to simulate the effects of different transmit power levels and different propagation conditions that differentiate in practice among the devices from different service classes. All the following results have been obtained by averaging over 100 realizations of the wireless channels.

The orthogonal code matrix $\mathbf{U}$ is a 6×6 DFT matrix and $N_1 = N_2 = 3$. The columns of matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ are random vectors with i.i.d. modulus-one complex entries. In the following, we compare MOMA to three multiple-access schemes implemented on the same resource region, namely "MIMO-eMTC with repetitions", "MIMO-eMTC with spreading", and NB-IoT. By "MIMO-eMTC with repetitions" we mean an eMTC system where several UTs are allowed to be scheduled on the same RB (hence the MIMO prefix) while there signals are repeated on 6 consecutive TTIs. In "MIMO-eMTC with spreading", signal repetition is replaced with random binary spreading (using the same resource mapping as in MOMA). Finally, NB-IoT designates a pure FDMA system where the available bandwidth is split into K sub-channels, each assigned to a single UT, with inter-sub-channel guard bands that amount to 10% of the bandwidth of the RB.

First, assume that $K_1 = 6$ and $K_2 = 18$. Figure 2 shows the maximum ergodic rate that can be achieved for the two service classes under this assumption. The deterministic equivalent of this rate as given by Corollaries 1 and 2 is also plotted in the same figure and is shown to be close to the simulated value even for numbers of BS antennas as small as 16. The approximation is slightly optimistic due to the fact that residual

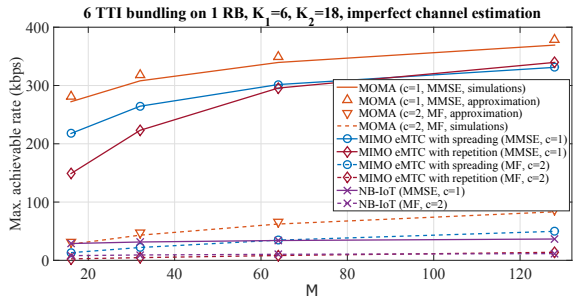

Fig. 2. Maximum achievable data rate vs. BS array size

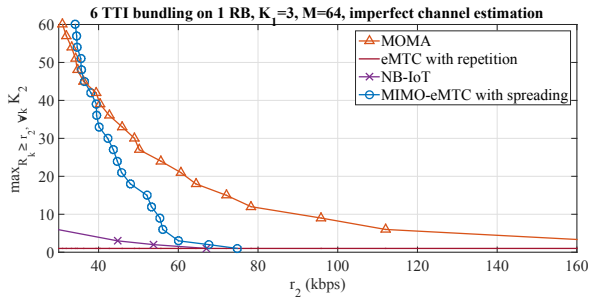

Fig. 3. Maximum number of simultaneous transmissions from the second class vs. target data rate

inter-class interference at finite values of M is not accounted for. Also note that the schemes using spreading, namely MOMA and "MIMO-eMTC with spreading", outperform as expected "MIMO-eMTC with repetition". Moreover, MOMA's service dependent hierarchical spreading achieves higher data rates for both service classes than conventional random spreading. This is thanks to inter-class quasi-orthogonality achieved by the former spreading technique. The smaller bandwidth per transmission in NB-IoT limits the achievable data rate, especially for the UTs of the first class.

Now, assume that $M = 64$ and $K_1 = 3$. Figure 3 shows in this case the maximum value of K_2 that can be supported while $R_k \geq r_2$ for all $k \in \mathcal{K}_2$. Again, MOMA outperforms all the other multiple-access schemes on most of the considered range of target rate values. It is worth pointing out that the bundling frame structure needs not be aligned in order for MOMA's spreading to provide the multiplexing gain plotted in Figure 3. In other words, if one device's first bundled sub-frame coincides with other devices' second (or third or ...) bundled sub-frames, their respective MOMA's codewords will still allow the BS to distinguish them from each other. This is thanks to the particular resource mapping we propose for MOMA (see Subsection III-A) and which consists in mapping the spread samples to neighboring REs within one RB/sub-frame rather than to neighboring RBs or sub-frames.

VI. Conclusions

MOMA can create high resource granularity on any sub-region of the resource grid without the need to change the numerology used in that region. This results in an intrinsic flexibility in resource assignment to different classes and concurrent transmissions within them. Moreover, using class dependent hierarchical spreading in MOMA instead of signal repetition, FDMA or conventional CDMA allows for more simultaneous transmissions to be multiplexed at higher target data rates, thus offering better system scalability with respect to increasing device densities. In this article, MOMA's scalability, efficiency and flexibility in resource utilization have been validated using both theoretical analysis and simulations.

References

[1] A. Rico-Alvariño, M. Vajapeyam, H. Xu, X. Wang, Y. Blankenship, J. Bergman, T. Tirronen, and E. Yavuz, "An Overview of 3GPP Enhancements on Machine to Machine Communications," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 14-21, June 2016.

[2] L. Vangelista, A. Zanella, and M. Zorzi, *Long-Range IoT Technologies: The Dawn of LoRa^{TM}*, in *FABULOUS*, Ohrid, pp. 51-58., Sep. 2015.

[3] H. S. Dhillon, H. C. Huang, H. Viswanathan, and R. A. Valenzuela, "Power-Efficient System Design for Cellular-Based Machine-to-Machine Communications," *IEEE Tran. Wireless Commun.*, vol. 12, no. 11, pp. 5740-5753, Nov. 2013.

[4] G. Naddafzadeh-Shirazi, L. Lampe, G. Vos, and S. Bennett, "Coverage Enhancement Techniques for Machine-to-Machine Communications over LTE," *IEEE Commun. Mag.*, vol. 53, no. 7, July 2015.

[5] A. Bader, M.-S. Alouini, A. Adinoyi, and F. Alshaalan, "First Mile Challenges for Large-Scale IoT," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 138-144, Mar. 2017.

[6] A. Lo, Y.W. Law, M. Jacobsson, and M. Kucharzak, *Enhanced LTE-Advanced Random-Access Mechanism for Massive Machine-to-Machine Communications*, in *WWRF*, Düsseldorf, Oct. 2011.

[7] M. S. Ali, E. Hossain, and D. I. Kim, "LTE/LTE-A Random Access for Massive Machine-Type Communications in Smart Cities," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 76-83, Jan. 2017.

[8] N. Ksairi, S. Tomasin, and M. Debbah, *A Multi-Service Oriented Multiple-Access Scheme For Next-Generation Mobile Networks*, in *EUCNC*, Athens, June 2016.

[9] N. Ksairi, S. Tomasin, and M. Debbah, "A Multi-Service Oriented Multiple-Access Scheme For M2M Support in Future LTE," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 218-224, Jan. 2017.

[10] A. K. Sadek, W. Su, and K. J. R. Liu, *Maximum Achievable Diversity for MIMO-OFDM Systems with Arbitrary Spatial Correlation*, in *GLOBECOM*, Dallas, Dec. 2004.

[11] A. Hedayat, H. Shah, and A. Nosratinia, "Analysis of Space-Time Coding in Correlated Fading Channels," *IEEE Trans. on Wireless Commun.*, vol. 4, no. 6, pp. 2882-2891, Nov. 2005.

[12] C.-X. Wang, X. Hong, H. Wu, and W. Xu, "Spatial-Temporal Correlation Properties of the 3GPP Spatial Channel Model and the Kronecker MIMO Channel Model," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, no. 1, Jan. 2007.

[13] F. Khan, *LTE for 4G Mobile Broadband: Air Interface Technologies and Performance*. Cambridge University Press, 2009.

[14] A. Decurninge and M. Guillaud, *Cube-Split: Structured Quantizers on the Grassmannian of Lines*, in *WCNC*, San Francisco, Mar. 2017.

[15] J. Hoydis, S. t. Brink, M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160-171, Feb. 2013.

[16] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large System Analysis of Linear Precoding in Correlated MISO Broadcast Channels Under Limited Feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509-4537, Jul. 2012.

[17] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed., New York, NY, USA: Cambridge University Press, 2013.

[18] The 3rd Generation Partnership Project (3GPP), *Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception*, Sep. 2015. Available: http://www.3gpp.org/.