

# On the Application of Massive MIMO Systems to Machine Type Communications

Felipe A. P. de Figueiredo\*, Fabbryccio A. C. M. Cardoso<sup>†</sup>, Ingrid Moerman\* and Gustavo Fraidenraich<sup>‡</sup>

\*Ghent University - imec, IDLab, Department of Information Technology, Ghent, Belgium.

<sup>†</sup>CPqD – Research and Development Center on Telecommunications, Brazil.

<sup>‡</sup>DECOM/FEEC – State University of Campinas (UNICAMP), Brazil.

Email: \*[felipe.pereira, ingrid.moerman]@ugent.be, <sup>†</sup>fcardoso@cpqd.com.br, <sup>‡</sup>gf@decom.fee.unicamp.br

**Abstract**—This paper evaluates the feasibility of applying Massive MIMO to tackle the uplink mixed-service communication problem. Under the assumption of an available physical narrow-band shared channel (PNSCH), devised to exclusively consume data traffic from Machine Type Communications (MTC) devices, the capacity (*i.e.*, number of connected devices) of MTC networks and, in turn, that of the whole system, can be increased by clustering such devices and letting each cluster share the same time-frequency physical resource blocks. Following this research line, we study the possibility of employing sub-optimal linear detectors to the problem and present a simple and practical channel estimator that works without previous knowledge of the large-scale channel coefficients. Our simulation results suggest that the proposed channel estimator performs asymptotically as well as the MMSE estimator with respect to the number of antennas and the uplink transmission power. Furthermore, the results also indicate that, as the number of antennas is made progressively larger, the performance of sub-optimal linear detection methods approaches the perfect interference-cancellation bound. The findings presented in this paper shed light on and motivate for new and exciting research lines towards a better understanding of the use of massive MIMO in MTC networks.

**Index Terms**—Large-scale antenna systems, 5G networks, machine type communications, channel estimation, linear detection.

## I. INTRODUCTION

RECENT technological developments taking place in our society have been drastically changing the way we use communications systems. These changes are in their great part due to the huge (and also foreseen [1]) increase in on-demand data consumption over both wireless and mobile networks. In order to support such changes, it is mandatory to devise solutions that can meet the different requirements of use cases regarded as the market drivers for next generation wireless networks (5G). ITU-R has defined the following three main 5G use cases: Enhanced Mobile Broadband (eMBB); Ultra-Reliable and Low Latency Communications (URLLC); and Massive Machine Type Communications (mMTC) [2]. They aim at significantly improving performance, scalability and (cost/energy) efficiency of the current wireless networks such as LTE, LTE-A and LTE-A Pro. These use cases and their direct requirements will demand huge improvements in comparison with the previous generation of IMT systems [2]. A non-exhaustive list of 5G applications grouped by use case and a brief explanation about them follows next.

- **eMBB**: focus on improvements to data rate, user density, latency, capacity and coverage of the current wireless networks [3], [4]. Some applications are: high-speed mobile broadband, augmented and virtual realities (e.g., gaming), smart office environments, pervasive video (*i.e.*, high-resolution video everywhere), etc.
- **URLLC**: aims at allowing devices and machines to communicate with ultra-reliability, high availability and very low latency, which make it ideal for real-time applications [5]–[7]. Some applications are: wireless industrial control, factory automation, remote surgery, cellular vehicle-to-everything (C-V2X) communications, self-driving cars, smart grids, public safety, etc.
- **mMTC**: focus on enabling machine-centered communications among devices that are massive in number, battery-driven, generate bursty traffic and have low-cost, *i.e.*, Internet of Things (IoT) devices [3], [8], [9]. This use case is intended to support applications like: smart metering, smart cities, asset tracking, remote monitoring (e.g., field and body sensors), etc.

Applications within the scope of the MTC driver range from smart cities and smart grid to critical infrastructure monitoring [10]–[12], and from Advanced Driver Assistance Systems (ADAS) to mobile health, which includes sports/fitness and telemedicine [13], [14]. Reliability in critical infrastructure monitoring and smart grid, for example, is often achieved only through dedicated land-line connections (*i.e.*, wired connections) [15]–[17]. Telemedicine makes use of telecommunications and information technology systems in order to provide remote clinical health care. It involves, for example, diagnostics realized through medical data stored in cloud servers, which requires low-latency, real-time access and high capacity servers capable of dealing with massive amounts of data, *e.g.*, computerized axial tomography and magnetic resonance imaging [18]–[20]. Automotive infotainment, vehicular cooperation in ADAS, and pre-crash sensing and mitigation applications also require high-speed, low-latency car-to-infrastructure and car-to-car communications [21]–[23].

Reliability and power consumption are of huge importance for wireless sensor networks (WSN), where a few to several hundreds or even thousands of low-cost and power-constrained sensor devices (in most of the WSNs, the sensors are battery-powered) need to measure environmental conditions like tem-

perature, noise level, air pollution levels, humidity, wind, etc. and reliably transmit them to a central location over harsh channel conditions [24], [25]. Most of the WSN use cases require the deployment of battery-powered sensors for ten years without any maintenance, meaning that the battery is expected to last a decade without being recharged [26].

As can be noticed from the previous discussion, the requirements necessary for the implementation of next generation wireless networks (*i.e.*, 5G) are quite diverse, even within the same market driver. Scalability is yet another issue posed by IoT, as the main assumption behind it is that hundreds to hundreds of thousands of low-cost MTC devices shall be served by a single Base Station (BS) [27]. Scalability issues have been mainly tackled by adopting different and sometimes complementary approaches, such as sparse signal processing techniques [28], techniques brought from duty-cycled Wireless Sensor Networks [29] and new waveforms specially designed for bursty and asynchronous data transmissions [30], [31], however, until now, the use of Multiple Input Multiple Output (MIMO) techniques in the context of MTC networks and the scalability issue are less understood.

The sentiment shared by most researchers nowadays is that the foreseen increase in data rate will be achieved by *combined gains* [32] provided by (i) increasing the network density, *i.e.*, the addition of more radio sites with smaller cell coverage areas to the same region (extreme network densification), which consequently improves the area spectral efficiency [33] (ii) increasing spectrum availability such as the introduction of new spectrum bands like mmWaves [34], [35], (iii) improving the use of licensed, unlicensed and licensed-shared spectrum bands [36] with more efficient and intelligent sharing techniques, (iv) and increasing spectral efficiency of digital communications systems through advances in MIMO techniques. One of the benefits resulting directly from the powerful processing gains provided by the use of large arrays of antennas (*i.e.*, massive MIMO systems) is that the majority of the physical layer signal processing and resource allocation (*i.e.*, scheduling) issues are simplified, if not solved, which is clearly not the case for systems employing only a moderate to small number of antennas [37].

Massive MIMO has been gaining significant attention and strength as a very promising candidate to improve spectral efficiency and consequently increase the channel capacity in multi-user networks. Massive MIMO is a scalable technology through which large numbers of devices can simultaneously communicate through the entire allocated spectrum, *i.e.*, thanks to its many spatial degrees of freedom, the same allocated frequency band can be reused by many users at the same time [37]. In the limit, as the number of antennas,  $M$ , deployed at the BS increases, the system processing gain also increases, *i.e.*, as  $M$  tends to infinity, the processing gain tends to infinity as well. Massive MIMO not only provides high spectral efficiency in a cell, but also provides a good and uniform service to a large number of devices simultaneously [37]. A consequence of this powerful processing gain is that the effects of small-scale fading and frequency dependence disappear. In [38] it is indicated that, due to the law of large numbers, the channel becomes reliable (*i.e.*, it becomes

deterministic) so that each one of the subcarriers in an OFDM-based massive MIMO system considerably experiences the same channel gain. This phenomenon is known as *channel hardening* [39]. *Channel hardening* renders frequency-domain scheduling unnecessary as all subcarriers are considered equally good, and consequently, makes most of the physical layer control signaling no longer needed [40]. Additionally, the adoption of massive MIMO systems also improve frequency reuse (due to the reduced radiated power), simplifies power control (power control coefficients depend only on the large-scale fading coefficients) and decreases multi-user interference (due to the possibility of having very narrow beams as  $M$  increases) [39], [41].

On the other hand, massive MIMO also presents some challenges that need to be studied and addressed in order to fully reap its benefits. Next, some of the issues regarded as the most challenging in the Massive MIMO literature are briefly discussed.

- In multi-cell scenarios, the use of non-orthogonal pilot signals by different users in different cells during the training-phase brings about a phenomenon known as *pilot-contamination*, which makes the target user's channel estimate contaminated by other user's channels using the same pilot. This phenomenon degrades the quality of the channel estimates and causes coherent interference that does not vanish by increasing the number of antennas [38], [42].
- Cost-efficient massive MIMO systems are expected to be constructed making use of low-cost components, however, this leads to the appearance of non-negligible signal distortion caused by hardware impairments. Hardware impairments cause channel estimation errors and limit the system's achievable capacity, which theoretically should be unlimited as the number of antennas increases, due to phase-noise, I/Q imbalance, power-amplifier nonlinearities, and quantization errors generally intrinsic to low-cost components [43].
- The digital signal processing in massive MIMO systems is inherently more computationally challenging when compared to the processing required by systems with single or small number of antennas. The signal processing in massive MIMO systems generally involves the following tasks: fast Fourier transform (FFT), channel estimation, precoding/detection and computation of precoding/detection matrices. The complexity of these digital signal processing operations increases linearly with the number of antennas, and everything but the FFT processing complexity also scales with the number of devices. Therefore, low-complexity digital signal processing techniques have to be devised in order to deal with this massive computational complexity expected to be created by these systems [38], [40].
- Massive MIMO systems are expected to mainly operate in time division duplex (TDD) mode due to the fact that the propagation channels can be assumed reciprocal during a coherence interval, *i.e.*, both downlink and uplink channel responses are approximately equal during

TABLE I  
BENEFITS AND CHALLENGES OF MASSIVE MIMO SYSTEMS

	Characteristic	Explanation	References
Benefits	High spectrum efficiency	Huge multiplexing and array gains	[38], [41], [49]
	High energy efficiency	Concentration of radiated energy on specific UEs	[41]
	High reliability	Increased diversity gain	[41], [49]
	Efficient linear precoding and detection	Favorable propagation condition for i.i.d. Rayleigh channels	[39], [41]
	Weak inter-user interference and enhanced physical layer security	With large $M$ , channels become orthogonal and with extremely narrow beams	[41], [43]
	Simpler scheduling scheme	Channel hardening effect averages out fast fading and receiver noise	[40], [41], [50]
	Robustness to failures of individual antenna elements	Massive number of antenna elements	[38], [41]
Challenges	Pilot-contamination	Limited number of orthogonal pilots due to coherence interval	[38], [42]
	Hardware impairments	Cost-efficient massive MIMO deployments use low-cost components, which suffer from hardware impairments	[43]
	High digital signal processing complexity	Digital signal processing scales with number of antennas and devices	[38], [40]
	Non-reciprocal transceiver characteristics	Transfer response of RF transceiver chains (amplifiers, filters, etc.) are different	[44]

a time interval. Based on the reciprocity assumption, TDD massive MIMO systems use the uplink channel as an estimate of the downlink channel. However, an issue that arises from this approach is that the transfer characteristics of transmit and receive RF (*i.e.*, RF transceiver) chains are different (amplifiers, filters, local oscillators, etc. have different characteristics), which directly impacts the calculation of precoding matrices. Therefore, effective and efficient reciprocity calibration techniques are needed to exploit the channel reciprocity in practice [44].

Although the deployment of massive MIMO systems still poses several challenges (*i.e.*, open questions), theoretical and measurement results demonstrate that its adoption can tremendously improve the spectral efficiency of wireless communications systems, [39], [41]. In Table I we summarize the most important benefits and challenges brought about by massive MIMO technologies.

The adoption of massive MIMO technology can specially help leveraging and simplifying the deployment of mMTC systems in cellular networks, which are potential candidates to accommodate the emerging MTC data traffic thanks to the existing infrastructure and wide-area coverage [45]. Massive MIMO has the potential to enable the multiplexing of a myriad of devices in the same time/frequency resources along with an extension in range due to the coherent beamforming gain inherent to this technology [37], [45].

The main contribution of this paper is the proposal of a data transmission scheme employing massive MIMO technology as a way to address the uplink mixed-service communication problem. In the uplink mixed-service communication problem, a BS has to serve not only Human Type Communications (HTC) devices but also a possible massive number of MTC devices. In order to be addressed properly, the problem can be split into two subproblems, namely, random access and data transmission problems. During the random access phase, a huge number of MTC along with HTC devices might simultaneously try access the network, which results in congestion and overloading [46]. On the other hand, after the MTC devices are granted access to the network, the BS has to allocate dedicated physical resource blocks to these devices [47]. With the foreseen number of connected devices raising up to tens of

thousands per cell [46], a BS might easily run out of available physical resource blocks (*i.e.*, congestion due to user data packets) to accommodate the data transmissions of this huge number of devices, tremendously impacting on the operations and quality of the provided services of a mobile network.

We focus our work on the data transmission phase by proposing a massive MIMO-based scheme where the data transmissions of a great number of MTC devices are served through the same time-frequency resources by a BS equipped with a large number of antennas. The proposed scheme has the potential to mitigate the congestion due to the large (and possibly massive) number of user data packets and additionally, it offers scalability, as the number of served devices can easily grow by increasing the number of deployed antennas at the BS [37]. The proposed scheme deals with data transmission, channel estimation and detection of the many data streams simultaneously transmitted by multiple MTC devices using the PNSCH's shared time-frequency resources.

We employ the maximum likelihood (ML) method to find an estimator for the large-scale fading coefficients present in the MMSE channel estimator. We show that this estimator is not only unbiased but it also achieves the Crámer-Rao lower bound. The estimated large-scale fading coefficients are replaced into the MMSE channel estimator, giving rise to a new channel estimator, which asymptotically approaches the performance of the MMSE channel estimator as the number of antennas and the uplink transmit power increase. Additionally, we derive closed-form and approximate expressions for the mean square error (MSE) of the proposed estimator. Moreover, we find lower bounds on the achievable rate for each one of the studied linear detectors. We also show that even for simple linear receivers (*i.e.*, MRC, ZF and MMSE), the transmitted power of each MTC device can be reduced as the number of antennas,  $M$ , grows without bound, which is very beneficial for power-constrained devices running on batteries.

This paper is an extension of a previous paper [48]. Differently from [48], where we have only considered Bit Error Rate (BER) analysis for perfect channel estimation (*i.e.*, full channel knowledge) and some linear detectors, the current paper not only deals with imperfect channel estimation, proposing and assessing the performance of a channel estimator in terms

of mean squared error (MSE) and BER, but also proposes a scheme to tackle the problem posed by the simultaneous data transmission of a large number of MTC devices connected to the base station (*i.e.*, user data packet congestion). Additionally, we also analyze the achievable rates of the studied linear detectors with the proposed data transmission scheme.

The remainder of the paper is as follows. Section II provides a brief discussion on related works. Section III presents a study case, where the feasibility of Massive MIMO for MTC networks is investigated as means to address the uplink mixed-service communication problem. Section IV outlines one possible approach to estimate the large-scale fading coefficients. Section V proposes a channel estimator that takes into account the estimation of the large-scale fading coefficients. Section VI presents simulation results and discussions on the outcomes. Section VII wraps up the paper with concluding remarks.

### A. Notations

Vectors and matrices are denoted by bold lower-case and upper-case letters, respectively. The matrix/vector conjugate-transpose is denoted by  $(\cdot)^H$ . We use  $\mathbb{E}[\cdot]$ ,  $\text{var}(\cdot)$  and  $\text{Cov}[\cdot]$  to denote the expectation, variance and covariance operators. The circularly-symmetric Gaussian distribution is denoted by  $\mathcal{CN}$ . We denote equality in asymptotic sense by  $\overset{\approx}{\sim}$ .  $\Gamma(\cdot)$  and  $B(\cdot, \cdot)$  denote the Gamma and Beta functions respectively.  $\mathbf{I}_K$  is the  $K \times K$  identity matrix and  $\mathbf{0}_N$  is the  $N \times 1$  zero vector.  $\|\cdot\|$ ,  $\mathbb{P}\{\cdot\}$ ,  $\Re\{\cdot\}$  and  $\cos^{-1}(\cdot)$  denote Euclidean norm, probability, real part and arc-cosine respectively. The big- $\mathcal{O}$  notation  $\mathcal{O}(Mx)$  describes that the complexity is bounded by  $CMx$  for some  $0 < C < \infty$ .

## II. RELATED WORK

In the literature, there are a myriad of works proposing solutions exclusively tailored to increase the capacity of the random access channel of LTE/LTE-A networks. In those networks, the MTC devices compete for resource blocks for their data transmission using a random access scheme. The works [46], [51]–[54] and the vast number of papers

therein mentioned, review and propose solutions to tackle the scalability issue posed by the random access of tens to hundreds of thousands of MTC devices during the random access/synchronization phase. The solutions presented in these works can accommodate from 30000 to more than 78000 MTC devices per cell with low collision probabilities. However, these works do not deal with the problem involving simultaneous data transmissions coming from a possible massive number of MTC devices during the Radio Resource Control (RRC) connected state [47].

Although there exist numerous studies on the massive random access problem, there are relatively few publications addressing the massive data transmission problem (*i.e.*, congestion due to user data packets) [55], which arises from the simultaneous data transmission of a huge number of devices during the RRC connected state (*i.e.*, during the data transmission phase).

The work presented in [56] tackles the problem of device activity detection and joint channel estimation when non-orthogonal pilot sequences are used by the devices. The authors use approximate message passing (AMP) algorithm in compressed sensing to exploit the sparsity in device activity detection. The work considers a grant-free multiple-access scheme and that the devices are already synchronized to the BS. The drawbacks of the proposed solution are the lack of information on how the devices stay synchronized to the BS and the analysis for multi-cell scenarios. In [57] the authors study the coexistence of HTC and MTC devices under a single-cell massive MIMO setup and assess their joint spectral efficiency, however they do not deal with channel estimation, linear decoding problems and multi-cell scenarios. The authors of [58] develop a stochastic geometry model for dense MTC systems adopting massive MIMO setups however, their focus is on providing a random access solution for such networks, failing to analyze the impact of massive MIMO during the data transmission phase. Additionally, it is worth noticing that all these works assume that all devices are synchronized to the uplink of the base station.

Therefore, we decided to focus our work on the data transmission phase, by proposing a solution where clusters of MTC devices share exclusive and periodic time-frequency resources and simultaneously transmit their data with massive MIMO technology being deployed at the BS to retrieve each one of the device's transmissions. By using massive MIMO at the BS, a great number of MTC devices can be assigned to the same time-frequency resources, consequently, mitigating the negative effects on human type communications (HTC), *e.g.*, data congestion. The proposed solution allows the addition of MTC services to wireless cellular networks without the necessity of additional time-frequency resources.

## III. THE UPLINK MIXED-SERVICE COMMUNICATIONS PROBLEM

In this section, we investigate the feasibility of massive MIMO as a means to address the so-called uplink mixed-service communication problem, where a *single* BS simultaneously delivers services to *both* narrowband MTC devices *and*

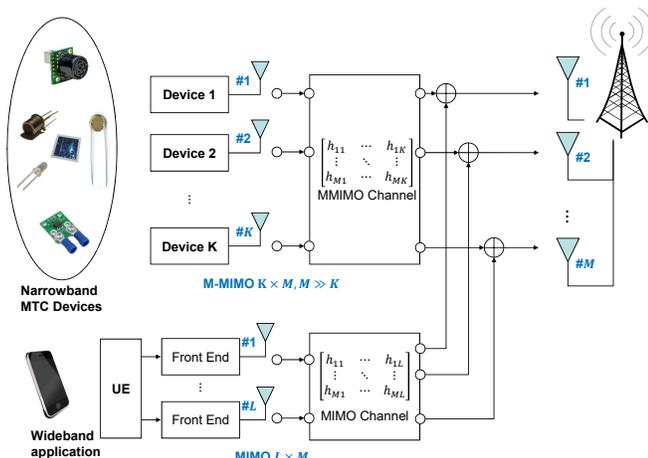


Fig. 1. Context application: Enabling a great number of low rate MTC devices in a cell.

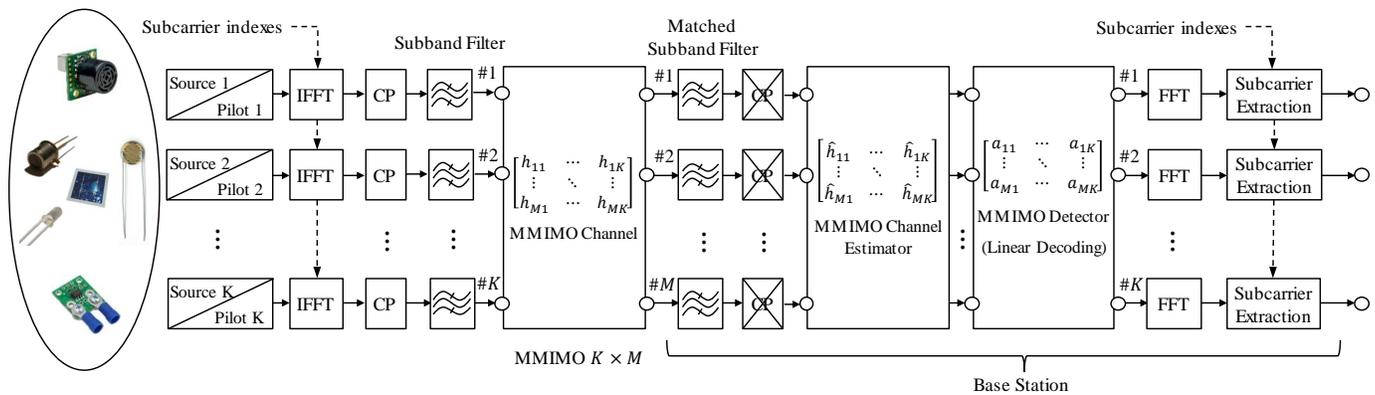


Fig. 2. Block diagram of a Massive MIMO uplink for mixed networks, where the BS simultaneously serves narrowband MTC devices and wideband UEs. The cluster of MTC devices seen at the transmit side share the same time-frequency PRBs, while the sole BS at the receive side is equipped with an antenna array at least one order of magnitude larger than the number of MTC devices.

Fourth Generation (4G) wideband User Equipment (UEs). We propose an approach that enables a huge number of bursty and low rate devices in a cell without compromising the Radio Access Network (RAN) as depicted in Figure 1. Our proposal is in line with the set of MTC features considered in 3GPP [55], [59]: (i) low mobility: the devices rarely move or only move within a certain region; (ii) time controlled: MTC data delivery only occurs during predefined time intervals; (iii) time tolerant: MTC data transfer can be delayed; (iv) small data transmissions: only small amounts of data are exchanged between the device and the BS, *i.e.*, bursty transmissions; (v) mobile originated only: MTC devices utilizing only mobile originated communications; (vi) infrequent transmission: long period between two data transmissions. Treating MTC devices as regular UEs turns out to be an issue, as scheduling Physical Resource Blocks (PRBs) in extremely dense networks is a nontrivial task made harder in the presence of retransmissions and intrinsic uplink synchronization procedures [60]–[63].

Assuming the availability of a Physical Narrowband Shared Channel (PNSCH), exclusively devised to consume data traffic generated by MTC devices, the capacity of the MTC network – and, in turn, that of the mixed-service system – can be increased by clustering MTC devices and letting clusters share the same time-frequency PRBs. The idea behind the PNSCH is to allow the exploitation of the channel’s geometric scattering characteristics to spread MTC signals in the spatial domain. The individual data streams conveyed by spatially spread MTC signals can be separated thanks to the inherent spatial multiplexing properties of massive MIMO technology [39], where the antenna array size at the BS is at least one order of magnitude larger than the number of served MTC devices. Next, we describe the system depicted in Figure 2 in terms of its underlying functional blocks.

#### A. Signal Generation & Transmission

We assume the transmitted signals of a cluster with  $K$  single-antenna MTC devices are detected by a Massive MIMO BS equipped with  $M$  receive antennas,  $M \gg K$ . All the  $K$  MTC sources map data into a set of continuous PRBs in the

frequency domain, with the subcarrier indexes providing the spectral position of the PNSCH at the physical layer level.

As the focus of our work is on the data transmission phase (*i.e.*, during the RRC\_CONNECTED state [47]), we, therefore, assume that all MTC devices being served by a BS are already synchronized and connected to it before accessing the PNSCH, *i.e.*, the MTC devices have already performed the random access and attach procedures before any data is sent through the PNSCH. Before any transmission, in order to align its uplink transmissions to the BS timing, each one of the MTC devices must perform a random access procedure through the physical random access channel (PRACH) [47], [62], [64]. Upon successful random access procedure, a MTC device holds a Cell-Radio Temporary Identifier (C-RNTI) that is then mapped to a pilot sequence, which will be used uniquely by that device while it is connected to the BS. The MTC device will use the same pilot sequence whenever it needs to transmit data towards the BS. This unique correspondence between a MTC device and a pilot sequence guarantees orthogonality among all the MTC devices being served by the same BS, which is of utmost importance to massive MIMO systems due to the pilot-contamination problem that might arise when pilot sequences are reused [65]. The interested reader is referred to [46], [51], [52] for a list of solutions to the random access problem posed by the large number of random access attempts coming from a massive number of MTC devices.

The BS broadcasts system information blocks (SIB), just like it is done for the PRACH used in current 4G systems (see, *e.g.* [64] and the references therein), in order to configure the PNSCH at the MTC devices. This allows the number of PNSCH transmission opportunities in the uplink to be scheduled while taking into consideration discrepancies between the (likely different) capacities of MTC devices and UEs. PNSCH time-frequency resources are semi-statically allocated by the BS, and repeat periodically. Additionally, the SIB messages can carry, for instance, information about the pilot sequence length, which in turn, dictates the capacity of the PNSCH as it will determine the remaining period of time destined to data symbols. The pilot sequence length can be varied so that more MTC devices can be simultaneously served by the BS at the

cost of smaller data capacity. Figure 3 depicts the uplink frame structure devised for the PNSCH. As can be seen in the figure, we assume 1 ms long PNSCH transmission opportunities. The PNSCH is time- and frequency-multiplexed with Physical Uplink Shared Channel (PUSCH), Physical Uplink Control Channel (PUCCH) and PRACH as illustrated in the figure.

In this work we assume that inter-cell interference is negligible. Inter-cell interference can be heavily mitigated, and therefore, considered insignificant, if less-aggressive frequency-reuse (*e.g.*, reuse of 3 or 7) is adopted [74]. Inter-cell interference manifests itself in two ways, namely, coherent and non-coherent interference, being the former caused by contaminating cells (*i.e.*, cells that use the same set of pilots as the home cell, causing pilot-contamination) and the latter caused by non-contaminating cells (*i.e.*, cells that do not use the same pilots as the home cell) [37]. In multi-cell scenarios, pilot-contamination, and consequently, coherent interference, can be disregarded once the PNSCH time-frequency resource intervals in each one of the neighbor cells can be configured to refrain them from overlapping with the intervals chosen for the target cell. This kind of configuration can be implemented in order to improve the overall system performance as pilot-contamination results in degradation of the channel estimate quality, which directly impacts on the spectral efficiency [37]. Figure 4 illustrates one possible configuration for the PNSCH intervals of neighbor cells so that pilot-contamination is mitigated.

We assume the utilization of OFDM block-based transmissions. We denote the OFDM symbol interval by  $T_s$ , the subcarrier spacing by  $\Delta f$ , the useful symbol duration by  $T_u = 1/\Delta f$ , and the guard interval (duration of the cyclic prefix) by  $T_g = T_s - T_u$ . As in [39], we call the reciprocal of the guard interval, when measured in subcarrier spacings, the "frequency smoothness interval",

$$N_{\text{smooth}} = \frac{1}{T_g \Delta f} = \frac{T_u}{T_g}, \quad (1)$$

where  $N_{\text{smooth}}$  represents the number of subcarriers over which the channel frequency response is considered smooth, *i.e.*, approximately constant [37].

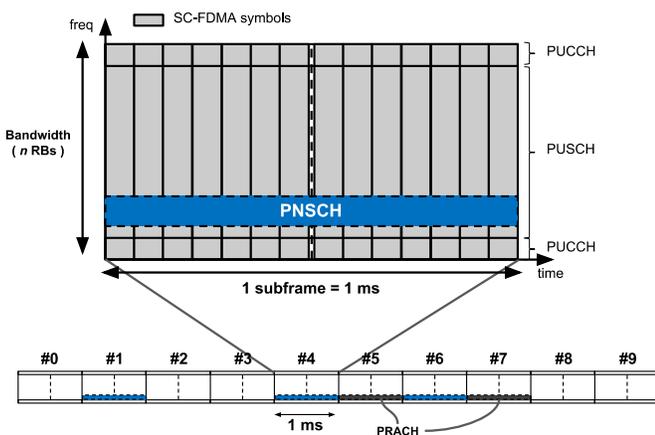


Fig. 3. Uplink Frame Structure with PNSCH.

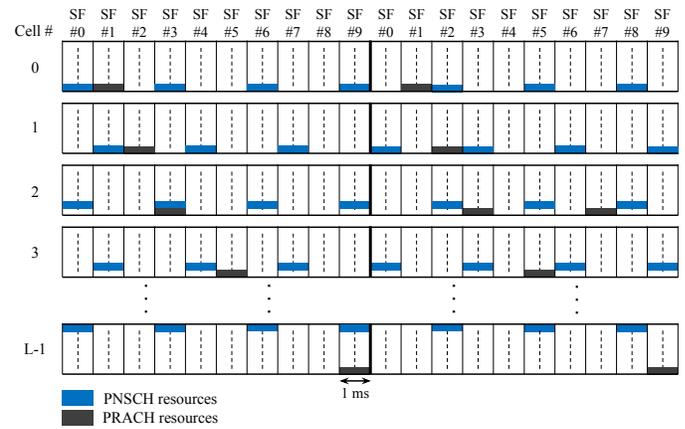


Fig. 4. Possible PNSCH resource configuration throughout different neighbor cells.

A total of  $\tau_p$  OFDM symbols are used entirely for pilot sequence transmission. The remaining symbols,  $\tau_u$ , within the same coherence block are used for data transmission. In general, the response is constant over  $N_{\text{smooth}}$  consecutive subcarriers and, therefore, the BS can estimate the channel for a total of  $K_{\text{max}} = \tau_p N_{\text{smooth}}$  terminals. We assume that a coherence block consists of  $N_{\text{smooth}}$  subcarriers and  $\tau_p + \tau_u$  OFDM symbols, *i.e.*,  $N_{\text{smooth}} \times (\tau_p + \tau_u)$  subcarriers, over which the channel response is approximated as constant and flat-fading [41].

1) *Pilot Transmission*: as widely used in LTE systems [47], we adopt Zadoff-Chu sequences to design the mutually orthogonal pilot sequences allocated to the MTC devices. These sequences present unit-norm elements but also the additional feature that each sequence is the cyclic shift of another sequence [66], [67]. However, any other set of sequences exhibiting the mutual orthogonality property could be used as pilot sequence, *e.g.*, Walsh Hadamard sequences [68]. Within a cell, each terminal is assigned a  $\tau_p N_{\text{smooth}}$  pilot sequence, which is orthogonal to the pilot sequences that are assigned to other terminals in the cell. Collectively, the  $K \leq \tau_p N_{\text{smooth}}$  terminals in the cell have the pilot book represented by  $\Phi$  - a  $\tau_p N_{\text{smooth}} \times K$  unitary matrix such that  $\Phi^H \Phi = I_K$ . The pilot sequence assigned to the  $k$ -th MTC device is represented by the column vector  $\phi_k$ .

2) *Data transmission*: we assume that the modulated symbols (carrying data of a MTC device) are randomly and independently drawn from a digital modulation alphabet (*e.g.*, PSK, 16QAM, etc.) with normalized average energy. The modulated symbols are mapped into  $\tau_u$  OFDM symbols.

Each MTC device transmits its signal (*i.e.*, allocated pilot sequence and data) by taking the Inverse Fast Fourier Transform (IFFT) of the mapped pilot sequence and data, and subsequently adding a CP.

Figure 5 shows the time-frequency plane for one possible configuration of the PNSCH. It shows how pilots and data symbols are mapped into the frequency/time domain of a coherence block. In that configuration, half of the OFDM symbols are used for pilots and the other half for data transmission. The time-frequency plane is divided into coherence

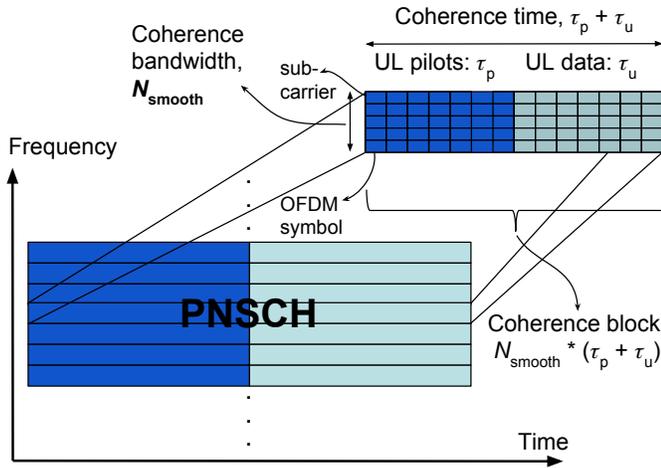


Fig. 5. Time-frequency plane.

blocks in which each channel is time-invariant and frequency-flat. The fraction of pilot symbols and UL data can be selected based on the network traffic characteristics, *i.e.*, the PNSCH configuration can be modified to increase the number of served MTC devices or the data rate per MTC device by increasing the number of OFDM symbols used for pilot or data transmission accordingly.

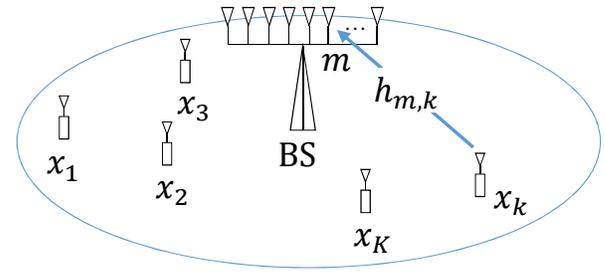
As can be seen in Figure 2, filters are added to both transmission and reception chains. These filters are added to the processing chains so that out-of-band emissions (OOBE), which are intrinsic to the OFDM waveform due to the discontinuities at its edges, do not interfere with adjacent channels, *i.e.*, Physical Random Access Channel (PRACH), Physical Uplink Shared Channel (PUSCH) and Physical Uplink Control Channel (PUCCH). Additionally, the filters help to mitigate inter-symbol interference (ISI) and inter-carrier interference (ICI) caused by asynchronous transmissions coming random access attempts happening at the PRACH [69]. Note that the received signal is passed through a matched filter, which maximizes the signal-to-noise ratio (SNR). The filter is applied to each time domain OFDM symbol (*i.e.*, after IFFT and CP insertion) to mitigate the OOBE of the PNSCH transmissions. The filters should be carefully designed to (i) maintain the complex-domain orthogonality of OFDM symbols, (ii) exhibit flat passband over the subcarriers in the PNSCH, (iii) have sharp transition band in order to reduce the guard-bands, and (iv) present sufficient stop-band attenuation [70].

### B. The Massive MIMO Channel

Let  $h_{m,k,n}$  denote the complex propagation coefficient between the  $m$ -th BS antenna and the  $k$ -th MTC device in the  $n$ -th subcarrier

$$h_{m,k,n} = g_{m,k,n} \sqrt{d_k}, \quad (2)$$

where  $g_{m,k,n}$  is a complex small-scale fading coefficient, and  $d_k$  is an amplitude coefficient that accounts for geometric attenuation and shadowing, *i.e.* large-scale fading [39]. The large-scale fading coefficients are assumed constant with respect to both subcarrier number and with respect to the index



$K$  - terminals per cell.

$M$  - BS co-located antennas.

← Signal from  $k$ -th user to the  $m$ -th BS antenna.

Fig. 6. Propagation model.

of the BS antenna since the geometric and shadow fading change slowly over space [39]. Therefore, between any given BS and a MTC device, there is only one large-scale fading coefficient. Additionally, these coefficients change only when a MTC device significantly change its position. It is normally assumed that in the radius of 10 wavelengths, the large-scale fading coefficients are approximately constant. On the other hand, small-scale fading coefficients significantly change as soon as the MTC device moves by a quarter of the wavelength. Therefore, the large-scale fading coefficients change about 40 times slower than the small-scale fading coefficients [71].

The random channel responses in one coherence block are statistically identical to the ones in any other coherence block, irrespective of whether they are separated in time and/or frequency. Hence, the channel fading is described by a stationary ergodic random process. Therefore, hereafter, our analysis is carried out by studying a single statistically representative coherence block. We assume that the channel realizations are independent between any pair of blocks, which is known as a block fading assumption. Consequently, for notational simplicity we suppress the dependency of  $h_{m,k,n}$  on the subcarrier index and write it as  $h_{m,k}$  (see Figure 6).

The elements  $h_{m,k}$  of the  $M \times K$  channel matrix

$$\mathbf{H} = [\mathbf{h}_1 \ \mathbf{h}_2 \ \dots \ \mathbf{h}_K] = \underbrace{\begin{pmatrix} g_{1,1} & \dots & g_{1,K} \\ \vdots & \ddots & \vdots \\ g_{M,1} & \dots & g_{M,K} \end{pmatrix}}_{\mathbf{G}} \cdot \underbrace{\begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_K \end{pmatrix}}_{\mathbf{D}}^{1/2}, \quad (3)$$

correspond to the complex channel gains from the transmit antennas to the receive antennas, where  $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}_M, b_k \mathbf{I}_M)$ ,  $\forall k$ . The channel model in (3) is called uncorrelated Rayleigh fading or independent and identically distributed (i.i.d.) Rayleigh fading, because the elements of  $\mathbf{h}_k$ , *i.e.*,  $h_{m,k}$ , are uncorrelated (and also independent) and have Rayleigh distributed magnitudes.

Under the assumption of large  $M$  and that the small-scale fading coefficients experienced by each MTC device are i.i.d. random variables with zero mean and unit variance, the column channel vector from different MTC devices becomes

asymptotically orthogonal as the number of receive antennas at the BS grows without bound [39]

$$\mathbf{H}^H \mathbf{H} = \mathbf{D}^{1/2} \mathbf{G}^H \mathbf{G} \mathbf{D}^{1/2} \stackrel{M \rightarrow \infty}{\approx} \mathbf{M} \mathbf{D}^{1/2} \mathbf{I}_K \mathbf{D}^{1/2} = \mathbf{M} \mathbf{D}, \quad (4)$$

where  $(\cdot)^H$  denotes transpose-conjugate (Hermitian) operation. As can be noticed, the small-scale fading vanishes and only the large-scale fading remains, however, it can be mitigated with power control techniques [72], [73].

One of the key assumptions exploited by massive MIMO systems is that the channel vectors between the BS and the terminals should be nearly orthogonal. This condition, which is shown in (4), is known in the literature as *favorable propagation*. With the assumption of favorable propagation, linear processing techniques can achieve optimal performance. For instance, on the uplink side, simple linear detectors can be used to cancel out noise and interference. On the downlink side, by adopting linear beamforming techniques, the BS can simultaneously beamform multiple data streams to multiple terminals without causing mutual interference [74], [75]. We refer the eager reader to [76] for a discussion on this condition, and to [49] for some experimental evidence supporting the assumption of i.i.d. small-scale fading coefficients in Massive MIMO.

### C. Linear MMSE Channel Estimation

Here we consider the case where CSI, *i.e.*,  $\mathbf{H}$  is estimated from the received pilot sequences at the BS. As mentioned earlier, we do not consider the existence of pilot-contamination during the channel estimation phase once the PNSCH intervals in all cells (target and neighbor ones) can be configured to avoid transmission overlapping.

1) *De-Spreading of the Received Pilot Signal*: the pilot sequences propagate through the uplink channel and are received by the BS as a  $M \times \tau_p N_{\text{smooth}}$  signal,

$$\mathbf{Y} = \sqrt{\rho} \mathbf{H} \Phi^H + \mathbf{W}, \quad (5)$$

where  $\rho$  is the uplink transmit power (UL Tx power) and the elements of the  $M \times \tau_p N_{\text{smooth}}$  receiver noise matrix,  $\mathbf{W}$ , are i.i.d.  $\mathcal{CN}(0, 1)$ . The BS performs a de-spreading operation of the received signal by correlating it with each one of the  $K$  pilot sequences. This is the equivalent of right-multiplying the received signal matrix by the  $k$ -th pilot sequence,  $\phi_k$ , resulting in

$$\begin{aligned} \mathbf{y}_k &= \frac{\mathbf{Y} \phi_k}{\sqrt{\rho}} \\ &= \mathbf{H} \Phi^H \phi_k + \frac{\mathbf{W} \phi_k}{\sqrt{\rho}} \\ &= \mathbf{h}_k + \mathbf{w}', \end{aligned} \quad (6)$$

where  $\mathbf{w}' = \frac{\mathbf{W} \Phi_k}{\sqrt{\rho}}$  is an  $M \times 1$  noise vector, whose elements are i.i.d.  $\mathcal{CN}(0, 1/\rho)$  because they are related to the original Gaussian noise matrix by a unitary multiplication scaled by  $1/\sqrt{\rho}$ . The de-spread signal,  $\mathbf{y}_k$ , is Gaussian distributed as follows

$$\mathbf{y}_k \sim \mathcal{CN} \left( \mathbf{0}_M, \left[ d_k + \frac{1}{\rho} \right] \mathbf{I}_M \right). \quad (7)$$

*Remark 1*: As  $\rho \rightarrow \infty$ , the variance of  $\mathbf{y}_k \rightarrow d_k$ .

Equation (6) is also known as the Least Squares (LS) estimator [77],

$$\hat{\mathbf{h}}_k^{\text{LS}} = \mathbf{y}_k, \quad (8)$$

and its mean-square estimation error per antenna is denoted by

$$\eta_k^{\text{LS}} = \frac{1}{M} \mathbb{E}[\|\hat{\mathbf{h}}_k^{\text{LS}}\|^2] = \frac{1}{M} \mathbb{E}[\|\hat{\mathbf{h}}_k^{\text{LS}} - \mathbf{h}_k\|^2] = \frac{1}{\rho}. \quad (9)$$

The LS channel estimation error is correlated with both the channel estimate and the de-spread received signal,

$$\frac{1}{M} \mathbb{E}[(\hat{\mathbf{h}}_k^{\text{LS}})^H \hat{\mathbf{h}}_k^{\text{LS}}] = \frac{1}{M} \mathbb{E}[(\hat{\mathbf{h}}_k^{\text{LS}})^H \mathbf{y}_k] = \frac{1}{\rho}. \quad (10)$$

As the LS channel estimate,  $\hat{\mathbf{h}}_k^{\text{LS}}$ , the estimation error  $\tilde{\mathbf{h}}_k^{\text{LS}}$  is also Gaussian distributed as follows,

$$\tilde{\mathbf{h}}_k^{\text{LS}} \sim \mathcal{CN} \left( \mathbf{0}_M, \frac{1}{\rho} \mathbf{I}_M \right). \quad (11)$$

Next, we present the MMSE estimator, which exhibits better performance than that of the LS estimator [65].

2) *Channel Estimation*: After de-spreading, the BS has a noisy version of the channel vector, which is defined by (6). Under the assumption of independent Rayleigh fading, the elements of the channel vector and the noise vector are statistically independent. By assumption, the large-scale fading coefficients are considered known at the BS, so the prior distribution of  $h_{m,k} \sim \mathcal{CN}(0, d_k)$ , is also known. In section IV we outline a possible approach for estimation of the large-scale fading coefficients at the BS. The linear MMSE estimator is the estimator achieving minimum MSE among all estimators of such form [41]. That is, it solves the following the optimization problem

$$\hat{\mathbf{h}}_k^{\text{MMSE}} = \arg \min_{\mathbf{B} \in \mathbb{R}^{M \times M}} \|\mathbf{h}_k - \mathbf{B} \mathbf{y}_k\|^2, \quad (12)$$

where  $\mathbf{y}_k$  is defined in (6),  $\mathbf{B}$  is a matrix that minimizes the mean square error (MSE). After solving (12), we find that the linear MMSE channel estimator is given by

$$\begin{aligned} \hat{\mathbf{h}}_k^{\text{MMSE}} &= \frac{d_k}{d_k + \frac{1}{\rho}} \mathbf{y}_k \\ &= \frac{\rho d_k}{\rho d_k + 1} \mathbf{y}_k. \end{aligned} \quad (13)$$

Note that as  $\rho \rightarrow \infty$ , the MMSE estimator becomes the LS estimator. The mean-square per antenna of the channel estimate is denoted by  $\gamma_k$  and given by

$$\gamma_k = \frac{1}{M} \mathbb{E}[\|\hat{\mathbf{h}}_k^{\text{MMSE}}\|^2] = \frac{\rho d_k^2}{1 + \rho d_k}. \quad (14)$$

The channel estimation error is denoted by

$$\tilde{\mathbf{h}}_k^{\text{MMSE}} = \hat{\mathbf{h}}_k^{\text{MMSE}} - \mathbf{h}_k, \quad (15)$$

and the mean-square estimation error per antenna is

$$\begin{aligned} \eta_k^{\text{MMSE}} &= \frac{1}{M} \mathbb{E}[\|\tilde{\mathbf{h}}_k^{\text{MMSE}}\|^2] = \frac{1}{M} \mathbb{E}[\|\hat{\mathbf{h}}_k^{\text{MMSE}} - \mathbf{h}_k\|^2] \\ &= \frac{d_k}{1 + \rho d_k} \\ &= d_k - \gamma_k. \end{aligned} \quad (16)$$

As  $\rho \rightarrow \infty$ , the mean-square error (MSE) per antenna,  $\eta_k^{\text{MMSE}} \rightarrow 0$ .

The channel estimation error is uncorrelated with both the channel estimate and the de-spread received signal,

$$\frac{1}{M} \mathbb{E}[(\tilde{\mathbf{h}}_k^{\text{MMSE}})^H \hat{\mathbf{h}}_k^{\text{MMSE}}] = 0. \quad (17)$$

$$\frac{1}{M} \mathbb{E}[(\tilde{\mathbf{h}}_k^{\text{MMSE}})^H \mathbf{y}_k] = 0. \quad (18)$$

The estimation error  $\tilde{\mathbf{h}}_k^{\text{MMSE}}$  and the estimate  $\hat{\mathbf{h}}_k^{\text{MMSE}}$  are jointly Gaussian distributed as follows,

$$\hat{\mathbf{h}}_k^{\text{MMSE}} \sim \mathcal{CN}(\mathbf{0}_M, \gamma_k \mathbf{I}_M). \quad (19)$$

$$\tilde{\mathbf{h}}_k^{\text{MMSE}} \sim \mathcal{CN}(\mathbf{0}_M, (d_k - \gamma_k) \mathbf{I}_M). \quad (20)$$

Therefore, the fact that they are uncorrelated implies that they are also statistically independent.

*Remark 2:* As  $\rho \rightarrow \infty$ , the variance of  $\hat{\mathbf{h}}_k^{\text{MMSE}} \rightarrow d_k$ .

One final remark about the MMSE estimation is that the MMSE estimator of a Gaussian random variable,  $\mathbf{h}_k$ , that is observed in independent Gaussian noise,  $\mathbf{w}'$ , is a linear estimator and thus equals the linear MMSE channel estimator, *i.e.*, there exist no better non-linear Bayesian estimator in this special case [41].

#### D. Signal Detection

Here we consider the scenario where the  $K$  MTC devices simultaneously transmit signals to the BS. Let  $x_k$ , where  $\mathbb{E}[|x_k|^2] = 1, \forall k$ , be the signal transmitted from the  $k$ -th device to the BS. Since  $K$  devices share the same time-frequency resource, the  $M \times 1$  received signal vector at the BS is the combination of all signals transmitted from all  $K$  devices [38], [74]:

$$\begin{aligned} \mathbf{y} &= \sqrt{\rho} \mathbf{H} \mathbf{x} + \mathbf{w} \\ &= \sqrt{\rho} \sum_{k=1}^K \mathbf{h}_k x_k + \mathbf{w}, \end{aligned} \quad (21)$$

where  $\mathbf{y} \in \mathbb{C}^{M \times 1}$ ,  $\sqrt{\rho} \mathbf{x}$  is the  $K \times 1$  vector of symbols simultaneously transmitted by the  $K$  devices (the average transmitted power of each device is  $\rho$ ) and  $\mathbf{w} \in \mathbb{C}^{M \times 1}$  is a zero-mean noise vector with complex Gaussian distribution and identity covariance matrix, *i.e.*,  $\mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$ . The noise variance is made unitary in order to minimize notation, but without any loss of generality. With this convention,  $\rho$  can be interpreted as normalized uplink transmit SNR or as introduced in subsection III-C, UL Tx power, and is therefore dimensionless [78]. There exist  $M$  PNSCH signal versions in (21) for each of the  $K$  MTC devices. Hence, the task of the BS consists in detecting  $K$  simultaneous MTC transmissions on the basis of estimates of the channel coefficients in (3). Detection techniques need to be employed in order to separate each one of the data streams transmitted by the various devices in a Massive MIMO system.

When it comes to separation of data streams in conventional systems, the complex signal processing technique known as Maximum Likelihood (ML) detection provides the optimal

performance. With ML detection, the BS has to search all possible transmitted signal vectors  $\mathbf{x}$ , and choose the best one as follows:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in X^K} \|\mathbf{y} - \sqrt{\rho} \mathbf{H} \mathbf{x}\|^2, \quad (22)$$

where  $X$  is the finite alphabet of  $x_k, k = 1, 2, \dots, K$ . The problem (22) is a least squares (LS) problem with a finite-alphabet constraint. The BS has to search over  $|X|^K$  vectors, where  $|X|$  denotes the cardinality of the set  $X$ . Hence, ML has a complexity which is exponential in the number of MTC devices,  $K$ . Therefore, although being the optimal solution for detection, ML is a highly complex solution to be implemented in our case, where hundreds to thousands of MTC devices are envisioned. This is the reason why signal detection is a key problem in Massive MIMO systems. To circumvent this limitation, we discuss in the next section a couple of sub-optimal alternatives with reduced computational complexity [77]. However, when the number of BS antennas is large, it is shown in [39], [49] that linear processing is nearly-optimal. We justify our choices for the detectors adopted in this work in the simulation work presented later on in Section VI.

#### E. Linear Decoding

Linear decoders (also known as linear detectors) work by spatially decoupling the effects of the channel by a process known as MIMO equalization. This involves multiplying  $\mathbf{y}$  with a MIMO equalization matrix  $\mathbf{A} \in \mathbb{C}^{M \times K}$  to get  $\hat{\mathbf{x}}(\mathbf{y}) \in \mathbb{C}^{M \times 1}$  [77]. Let  $\mathbf{A}$  be an  $M \times K$  linear detector matrix that depends on the estimated channel  $\hat{\mathbf{H}}$ . By using a linear detector, the received signal can be separated into different data streams using  $\mathbf{A}$  as follows

$$\mathbf{r} = \mathbf{A}^H \mathbf{y}, \quad (23)$$

where the vector  $\mathbf{r}$  collects the data streams received at the BS, *i.e.* the symbols of all  $K$  single-antenna MTC devices, and  $\mathbf{A}$  is a receive matrix that depends on the specific linear detector used at the BS. After linear detection, as seen in Figure 2, each data stream undergoes FFT processing and subcarrier extraction in order to retrieve data symbols.

Inspection of (3) reveals that  $\mathbf{D}$  is a diagonal matrix, so we can use MRC in the uplink to separate the signals from different MTC devices into different streams with asymptotic no inter-user interference [39]. Thereby each MTC device's transmission can be seen as signals of a single device passing through a single input single output channel. In the limit, this implies that MRC is optimal when the number of receive antennas is much larger than the number of transmit antennas, *i.e.*  $M \gg K, M \rightarrow \infty$  – as can be seen from (4). In MRC the linear detection matrix  $\mathbf{A}$  is chosen using

$$\mathbf{A}_{\text{MRC}} = \hat{\mathbf{H}} \quad (24)$$

where the dominant computation is due to matrix transposition. With MRC, the BS aims to maximize the received signal-to-noise ratio (SNR) of each MTC device, *i.e.*, stream, ignoring the effect of multiuser interference and therefore, one of its drawbacks is that it performs poorly in interference-limited

scenarios [37]. The associated complexity is of only  $\mathcal{O}(MK)$  multiplications.

In contrast to the MRC decoder, ZF decoders take the inter user interference into account, but neglect the effect of noise, *i.e.*, it chooses  $\mathbf{A}$  with the objective of completely eliminating interference, regardless of noise enhancement. With ZF, the multiuser interference is completely nulled out by projecting each stream onto the orthogonal complement of the inter user interference [37]. Specifically, the ZF detector chooses  $\mathbf{A}$  constrained to  $\mathbf{A}\mathbf{H} = \mathbf{I}$

$$\mathbf{A}_{\text{ZF}} = \hat{\mathbf{H}}(\hat{\mathbf{H}}^H \hat{\mathbf{H}})^{-1}. \quad (25)$$

The advantages of ZF are that the signal processing is simple and it works well in interference-limited scenarios. The drawback is that since ZF neglects the effect of noise, it works poorly under noise-limited scenarios. Furthermore, if the channel is not well conditioned then the pseudo-inverse amplifies the noise significantly, and therefore, the performance is very poor. Compared with MRC, ZF has a higher implementation complexity due to the computation of the pseudo-inverse of the channel gain matrix [78]. ZF exhibits a complexity of  $\mathcal{O}(MK + 2MK^2 + K^3)$  [49]. A better strategy is to choose  $\mathbf{A}$  so as to balance the signal energy lost with the increased interference. From this point of view, it is better to accept some residual interference provided that this allows the detector to capture more of the desired signal's energy [77].

One last linear detector that, together with MRC and ZF, poses complexity costs that do not depend on the modulation order is the MMSE. As the name suggests, the MMSE detector chooses the  $\mathbf{A}$  that minimizes  $e = E[\|\mathbf{A}^H \mathbf{y} - \mathbf{x}\|^2 | \hat{\mathbf{H}}]$  without any additional constraints

$$\begin{aligned} \mathbf{A}_{\text{MMSE}} &= \hat{\mathbf{H}} \left[ \hat{\mathbf{H}}^H \hat{\mathbf{H}} + \frac{1}{\rho} \text{cov}(\mathbf{w} - \sqrt{\rho} \tilde{\mathbf{H}} \mathbf{x}) \right]^{-1} \\ &= \hat{\mathbf{H}} \left[ \hat{\mathbf{H}}^H \hat{\mathbf{H}} + \frac{1}{\rho} \left( 1 + \rho \sum_{l=1}^K \mathbb{E}[\tilde{\mathbf{h}}_l \tilde{\mathbf{h}}_l^H] \right) \mathbf{I}_K \right]^{-1}, \end{aligned} \quad (26)$$

where  $\text{cov}(a)$  denotes the covariance matrix of a random variable  $a$  and  $\tilde{\mathbf{H}}$  is the estimation error,  $\mathbf{H} = \hat{\mathbf{H}} - \tilde{\mathbf{H}}$ .

In contrast to ZF, which minimizes interference but fails to treat noise, and to MRC, which minimizes noise but fails to treat interference, MMSE achieves an optimal balance between interference suppression and noise enhancement (it maximizes the received SINR [78]) at the same cost (*i.e.*, complexity) of ZF [49], [79].

The shortcomings listed in Table I of [48] under iterative filtering, random step search, and tree-based methods suggest that these detection classes perform well but are still too complex to be practical. This indicates that more work is needed on this matter, perhaps towards turbo codes or Low-density Parity-check (LDPC) codes in iterative detection and decoding settings [80]. On the other hand, linear filtering methods that are non-iterative, such as MRC, ZF, and MMSE, seem more feasible candidates for Massive MIMO systems. For  $1 \ll K \ll M$ , it is known that linear detection performs fairly well, and asymptotically achieves capacity when  $M \rightarrow \infty$ . [39]. We therefore consider only such linear methods in the simulation work section.

## F. Achievable Rates

In this subsection we derive lower bounds on the achievable rate for each one of the studied linear detectors when MMSE channel estimation is considered. Considering that during one PNSCH transmission interval at the target cell there are no other PNSCH transmissions being originated at other cells or that a less-aggressive frequency-reuse factor (*e.g.*, reuse factor of 3 or 7 for instance) is employed, we can analyze and derive the achievable rates as if the target cell was a single-cell system, emphasizing the fact that inter-cell interference is inexistent or negligible and therefore, do not need to be accounted for. The rationale behind the single-cell Multi User (MU)-MIMO analysis is that its results are readily comprehended, they bound the performance of multicell systems and that the single-cell performance can be actually achieved if less-aggressive frequency-reuse is adopted. For the following derivations we use standard capacity bounding techniques from the massive MIMO literature [37], [74], [81].

The received signal vector at the BS can be rewritten as

$$\mathbf{r} = \mathbf{A}^H \left( \sqrt{\rho} \hat{\mathbf{H}} \mathbf{x} - \sqrt{\rho} \tilde{\mathbf{H}} \mathbf{x} + \mathbf{w} \right), \quad (27)$$

where  $\tilde{\mathbf{H}} = \hat{\mathbf{H}} - \mathbf{H}$ . Due to the properties of the MMSE estimation,  $\tilde{\mathbf{H}}$  is independent of  $\hat{\mathbf{H}}$ , see (17). Therefore, the received signal associated with the  $k$ -th MTC device is given by

$$\begin{aligned} r_k &= \sqrt{\rho} \mathbf{a}_k^H \hat{\mathbf{H}} \mathbf{x} - \sqrt{\rho} \mathbf{a}_k^H \tilde{\mathbf{H}} \mathbf{x} + \mathbf{a}_k^H \mathbf{w} \\ &= \underbrace{\sqrt{\rho} \mathbf{a}_k^H \hat{\mathbf{h}}_k x_k}_{\text{desired signal}} + \underbrace{\sqrt{\rho} \sum_{l=1, l \neq k}^K \mathbf{a}_k^H \hat{\mathbf{h}}_l x_l}_{\text{intra-cell interference}} - \underbrace{\sqrt{\rho} \sum_{l=1}^K \mathbf{a}_k^H \tilde{\mathbf{h}}_l x_l}_{\text{effective noise}} + \mathbf{a}_k^H \mathbf{w}, \end{aligned} \quad (28)$$

where  $\mathbf{a}_k$ ,  $\hat{\mathbf{h}}_l$  and  $\tilde{\mathbf{h}}_l$  are the  $l$ -th columns of  $\mathbf{A}$ ,  $\hat{\mathbf{H}}$ , and  $\tilde{\mathbf{H}}$ , respectively. As  $\tilde{\mathbf{H}}$  and  $\hat{\mathbf{H}}$  are independent,  $\mathbf{A}$  and  $\tilde{\mathbf{H}}$  are also independent. As shown in (28), the BS treats the channel estimate as the true channel, and the last three terms in the equation are considered as intra-cell interference and effective noise respectively. Therefore, an achievable rate of uplink transmission for the  $k$ -th MTC device is defined by

$$R_k = \mathbb{E} \{ \log_2 (1 + \text{SINR}_k) \}, \quad (29)$$

$$\text{SINR}_k = \frac{\rho |\mathbf{a}_k^H \hat{\mathbf{h}}_k|^2}{\rho \sum_{l=1, l \neq k}^K |\mathbf{a}_k^H \hat{\mathbf{h}}_l|^2 + \rho \|\mathbf{a}_k\|^2 \sum_{l=1}^K \eta_l^{\text{MMSE}} + \|\mathbf{a}_k\|^2}, \quad (30)$$

where  $\eta_l^{\text{MMSE}}$  is defined in (16). By following a similar line of reasoning as in [37], [74] we obtain lower bounds on the achievable rate for the MRC, ZF and MMSE linear detectors.

1) *MRC detector*: with MMSE channel estimation, Rayleigh fading, and MRC processing, the achievable rate for the  $k$ -th MTC device is lower bounded by

$$\tilde{R}_k^{\text{MRC}} = \log_2 \left( 1 + \frac{\rho(M-1)\gamma_k}{1 + \rho \sum_{l=1}^K d_l - \rho\gamma_k} \right). \quad (31)$$

2) *ZF detector*: with MMSE channel estimation, Rayleigh fading, ZF processing, and for  $M \geq K+1$  the achievable rate for the  $k$ -th MTC device is lower bounded by

$$\tilde{R}_k^{\text{ZF}} = \log_2 \left( 1 + \frac{\rho\gamma_k(M-K)}{1 + \rho \sum_{l=1}^K \eta_l^{\text{MMSE}}} \right). \quad (32)$$

3) *MMSE detector*: with MMSE channel estimation, Rayleigh fading, MMSE processing, the achievable rate for the  $k$ -th MTC device is approximately lower bounded by

$$\tilde{R}_k^{\text{MMSE}} = \log_2 (1 + (\alpha_k - 1)\theta_k), \quad (33)$$

where

$$\alpha_k = \frac{(M-K+1 + (K-1)\mu)^2}{M-K+1 + (K-1)\kappa}, \quad (34)$$

$$\theta_k = \frac{M-K+1 + (K-1)\kappa}{M-K+1 + (K-1)\mu} w\gamma_k, \quad (35)$$

where  $w = \left[ \frac{1}{\rho} + \sum_{l=1}^K \eta_l^{\text{MMSE}} \right]^{-1}$ ,  $\mu$  and  $\kappa$  are obtained by solving the following two equations:

$$\mu = \frac{1}{K-1} \sum_{l=1, l \neq k}^K \frac{1}{Mw\gamma_l \left( 1 - \frac{K-1}{M} + \frac{K-1}{M}\mu \right) + 1} \quad (36)$$

$$\begin{aligned} \kappa & \left( 1 + \sum_{l=1, l \neq k}^K \frac{w\gamma_l}{\left( Mw\gamma_l \left( 1 - \frac{K-1}{M} + \frac{K-1}{M}\mu \right) + 1 \right)^2} \right) \\ & = \sum_{l=1, l \neq k}^K \frac{w\gamma_l \mu + 1}{\left( Mw\gamma_l \left( 1 - \frac{K-1}{M} + \frac{K-1}{M}\mu \right) + 1 \right)^2} \end{aligned} \quad (37)$$

#### IV. ESTIMATION OF LARGE-SCALE FADING COEFFICIENTS

In this section we describe how the large-scale fading coefficients,  $d_k, \forall k$ , can be estimated based on the de-spread signal,  $\mathbf{y}_k$ . We employ the Maximum Likelihood (ML) method to estimate the the large-scale fading coefficients [82]. Applying the ML method to  $f(\mathbf{y}_k; d_k)$  with distribution defined in (7), we find the following estimator for  $d_k$  given the observation  $\mathbf{y}_k$

$$\hat{d}_k = \frac{\|\mathbf{y}_k\|^2}{M} - \frac{1}{\rho}. \quad (38)$$

This estimator exhibits a central Chi-square distribution with  $2M$  degrees of freedom. It has  $\mathbb{E}[\hat{d}_k] = d_k$ , which shows that the ML estimator is unbiased, and  $\text{var}(\hat{d}_k) = \frac{(d_k + \frac{1}{\rho})^2}{M}$ .

The mean-square error of the proposed large-scale fading coefficient estimator is defined as

$$\mathbb{E}[(d_k - \hat{d}_k)^2] = \frac{\left( d_k + \frac{1}{\rho} \right)^2}{M}. \quad (39)$$

Note that the mean-square error of the proposed estimator is also equal to its variance.

*Remark 3*: As  $M \rightarrow \infty$ ,  $\mathbb{E}[(d_k - \hat{d}_k)^2] = \text{var}(\hat{d}_k) \rightarrow 0$ . Remark 3 shows that as  $M$  increases, the estimator defined in (38) becomes a deterministic value and that it tends, in the limit, to the actual  $d_k$  value once the mean-square error vanishes.

In order to assess the efficiency of the estimator we derive the Cramér-Rao bound as [82]

$$\text{var}(\hat{d}_k) \geq \frac{(d_k + \frac{1}{\rho})^2}{M}. \quad (40)$$

Therefore, as can be noticed, the ML estimator derived for  $d_k$  is the minimum variance unbiased estimator (MVUE), *i.e.*, it is an unbiased estimator that has the lowest variance among all other possible unbiased estimators [82].

Additionally, we show that the proposed estimator approaches  $d_k$  as the number of antennas  $M$  increases.

$$\lim_{M \rightarrow \infty} \hat{d}_k = \lim_{M \rightarrow \infty} \frac{\|\mathbf{y}_k\|^2}{M} - \frac{1}{\rho} \stackrel{\text{a.s.}}{=} d_k. \quad (41)$$

The proof of (41) is provided in Appendix A. This is an example of the strong law of large numbers and can be interpreted as the variations of  $\frac{\|\mathbf{y}_k\|^2}{M}$  becoming increasingly concentrated around its mean value  $\mathbb{E} \left[ \frac{\|\mathbf{y}_k\|^2}{M} \right] = d_k + \frac{1}{\rho}$  as more antennas are added.

#### V. PROPOSED CHANNEL ESTIMATOR

In this section we propose a simple and practical channel estimator based on the estimator for the large-scale fading coefficients defined in Section IV. Our proposed approach estimates the large-scale fading coefficients,  $d_k$ , and replaces it into the MMSE channel estimator defined in (13), resulting in the following channel estimator

$$\hat{\mathbf{h}}_k^{\text{prop}} = \left( 1 - \frac{M}{\rho} \frac{1}{\|\mathbf{y}_k\|^2} \right) \mathbf{y}_k. \quad (42)$$

*Remark 4*: The proposed estimator asymptotically approaches the ideal MMSE channel estimator in (13) as  $M \rightarrow \infty$ ,

$$\lim_{M \rightarrow \infty} \left( 1 - \frac{1}{\rho} \frac{M}{\|\mathbf{y}_k\|^2} \right) \mathbf{y}_k = \left( \frac{\rho d_k}{\rho d_k + 1} \right) \mathbf{y}_k. \quad (43)$$

This can be easily proved by applying Lemma 4 to (42). The proposed estimator has  $\mathbb{E}[\hat{\mathbf{h}}_k^{\text{prop}}] = \mathbf{0}_M$  and covariance matrix given by

$$\begin{aligned} \text{Cov}[\hat{\mathbf{h}}_k^{\text{prop}}] & = \mathbb{E}[\hat{\mathbf{h}}_k^{\text{prop}} (\hat{\mathbf{h}}_k^{\text{prop}})^H] \\ & = \frac{(\rho d_k + 1)(\rho d_k - 1)(M-1) + M}{\rho(\rho d_k + 1)(M-1)} \mathbf{I}_M \\ & = \frac{\rho d_k^2}{(\rho d_k + 1)} + \frac{1}{\rho(\rho d_k + 1)(M-1)} \mathbf{I}_M \\ & = \gamma_k + \epsilon_k, \end{aligned} \quad (44)$$

where  $\epsilon_k = \frac{1}{\rho(\rho d_k + 1)(M-1)}$ , which is the error introduced by the estimation of  $d_k$ .

*Remark 5:* As  $\rho$  and  $M \rightarrow \infty$ , the variance of  $\hat{\mathbf{h}}_k^{\text{prop}} \rightarrow d_k$ .

As can be seen by analyzing equation (44), as  $M \rightarrow \infty$ ,  $\text{Cov}[\hat{\mathbf{h}}_k^{\text{prop}}] \rightarrow \frac{\rho d_k^2}{1 + \rho d_k} \mathbf{I}_M$ , which is the covariance matrix of the MMSE estimator. The mean-square estimation error per antenna of the proposed channel estimator is defined as

$$\begin{aligned} \eta_k^{\text{prop}} &= \frac{1}{M} \mathbb{E}[\|\tilde{\mathbf{h}}_k^{\text{prop}}\|^2] = \frac{1}{M} \mathbb{E}[\|\hat{\mathbf{h}}_k^{\text{prop}} - \mathbf{h}_k\|^2] \\ &= \frac{1}{\rho} \left( \frac{M}{M-1} \frac{1}{1 + \rho d_k} - 1 + 2\theta_k \right), \end{aligned} \quad (45)$$

where  $\theta_k = \int_0^1 \int_{-1}^1 \frac{\kappa_k^2 (1-t) + \kappa_k w \sqrt{t(1-t)}}{\kappa_k^2 (1-t) + 2\kappa_k w \sqrt{t(1-t)} + t} f_W(w) f_T(t) dw dt$  and  $\kappa_k \triangleq \sqrt{\rho d_k}$  and  $f_W(w)$  and  $f_T(t)$  are defined by

$$f_T(t) = \frac{\Gamma(2M)}{(\Gamma(M))^2} [t(1-t)]^{M-1}, \quad 0 < t < 1, \quad (46)$$

$$f_W(w) = \frac{M}{\pi} B\left(\frac{1}{2}, M\right) (1-w^2)^{M-\frac{1}{2}}, \quad |w| < 1. \quad (47)$$

The proof of the mean-square estimation error per antenna is given in Appendix B. The analytical mean-square estimation error expression (45) is useful for system design and performance evaluation purposes [83].

*Remark 6:* The mean-square error between  $\hat{\mathbf{h}}_k^{\text{prop}}$  and  $\hat{\mathbf{h}}_k^{\text{MMSE}}$  is defined as

$$\frac{1}{M} \mathbb{E}[\|\hat{\mathbf{h}}_k^{\text{prop}} - \hat{\mathbf{h}}_k^{\text{MMSE}}\|^2] = \frac{1}{\rho(\rho d_k + 1)(M-1)} = \epsilon_k. \quad (48)$$

The proof of (48) is given in Appendix C. From (48) we observe that the mean-square error decreases when  $M$ ,  $\rho$  and  $d_k$  increase, which consequently shows that the proposed channel estimator asymptotically approaches the performance of MMSE channel estimator.

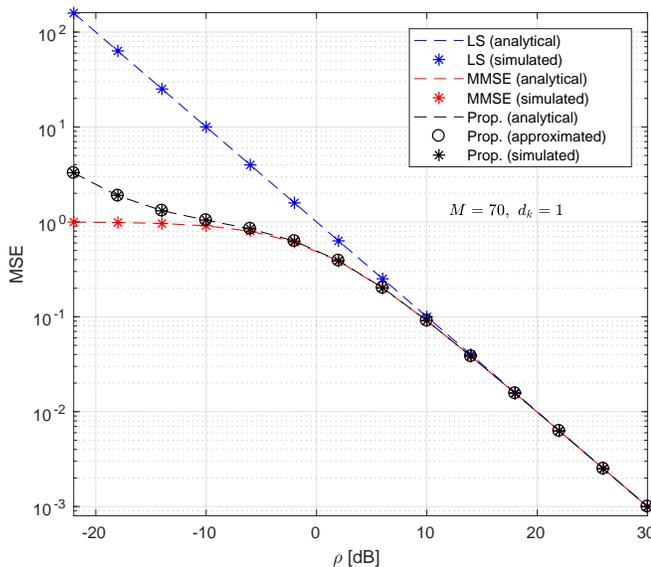


Fig. 7. Channel estimation MSE versus uplink pilot power,  $\rho$ .

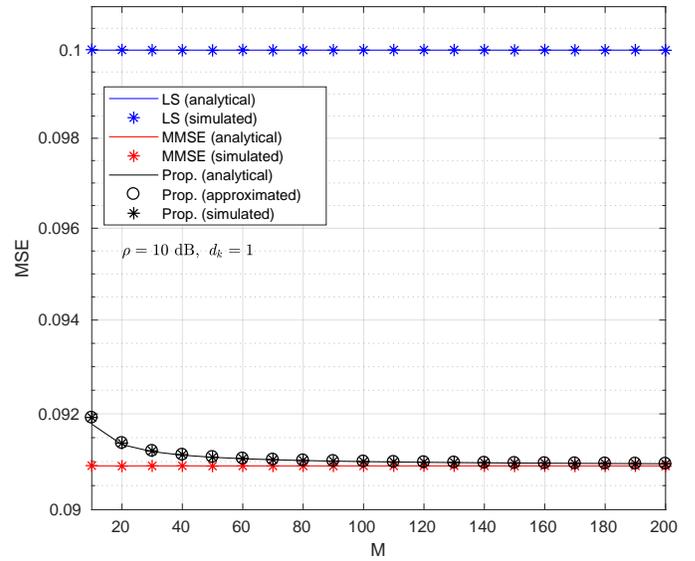


Fig. 8. MSE performance versus number antennas,  $M$ .

Next, we also present a simpler and more tractable expression for the mean-square estimation error per antenna, which is defined as

$$\begin{aligned} \frac{1}{M} \mathbb{E}[\|\tilde{\mathbf{h}}_k^{\text{prop}}\|^2] &= \frac{1}{M} \mathbb{E}[\|\hat{\mathbf{h}}_k^{\text{prop}} - \mathbf{h}_k\|^2] \approx \eta_k^{\text{prop(approx.)}} = \\ &= \frac{1}{\rho} \left[ \frac{\rho d_k}{1 + \rho d_k} + \frac{1}{(M-1)(1 + \rho d_k)} \right] \\ &= d_k - \gamma_k + \epsilon_k. \end{aligned} \quad (49)$$

*Remark 7:* As  $M$  or  $\rho \rightarrow \infty$ ,  $\eta_k^{\text{prop(approx.)}} = d_k / (1 + \rho d_k) = d_k - \gamma_k$ .

Remark 7 clearly shows that the approximated mean-square estimation error per antenna of the proposed estimator tends to that of the MMSE estimator when  $M \rightarrow \infty$ . The proof for the approximated mean-square estimation error per antenna of the proposed channel estimator is given in Appendix D.

The channel estimation error is correlated with the channel estimate and uncorrelated with the de-spread received signal,

$$\frac{1}{M} \mathbb{E}[(\tilde{\mathbf{h}}_k^{\text{prop}})^H \hat{\mathbf{h}}_k^{\text{prop}}] = \epsilon_k. \quad (50)$$

$$\frac{1}{M} \mathbb{E}[(\tilde{\mathbf{h}}_k^{\text{prop}})^H \mathbf{y}_k] = 0. \quad (51)$$

*Remark 8:* As  $M$  or  $\rho \rightarrow \infty$ , then  $\frac{1}{M} \mathbb{E}[(\tilde{\mathbf{h}}_k^{\text{prop}})^H \hat{\mathbf{h}}_k^{\text{prop}}] = 0$ .

The estimation error,  $\tilde{\mathbf{h}}_k^{\text{prop}}$ , have the following mean vector and covariance matrix,

$$\mathbb{E}[\tilde{\mathbf{h}}_k^{\text{prop}}] = \mathbf{0}_M, \quad (52)$$

$$\text{Cov}[\tilde{\mathbf{h}}_k^{\text{prop}}] = (d_k - \gamma_k + \epsilon_k) \mathbf{I}_M. \quad (53)$$

### A. Complexity Analysis

The computational complexity is an important factor indicating the efficiency of a channel estimator. In this section we assess the computational complexity of the studied channel estimators. Table II summarizes the complexities involved in the calculation of the estimators in terms of number of floating-point operations (flops) and the Big- $\mathcal{O}$  notation. The Big- $\mathcal{O}$  notation, also called Landau's symbol, which is a well-known symbolism widely used in complexity theory to describe the asymptotic behavior of algorithms [84]. It basically indicates how fast an algorithm grows or declines. In the table  $P$  is the length of the pilot sequence (*i.e.*,  $\tau_p N_{\text{smooth}}$ ) and  $\mathbf{D}$  is a  $K \times K$  identity matrix with the diagonal elements equal to  $\frac{\rho d_k}{\rho d_k + 1} \forall k$  and  $\left(1 - \frac{M}{\rho} \frac{1}{\|y_k\|^2}\right) \forall k$  for the MMSE and proposed channel estimators respectively. It is important to notice that in this work and in the great majority of works in the literature [65], [85]–[87] the large scale fading coefficients,  $\{d_k\}$ , are assumed perfectly known for the MMSE channel estimation.

The LS estimator is the most computational cost efficient among all of them, presenting a complexity of  $\mathcal{O}(MPK)$ , however, as will be shown later, this is the least efficient estimator in terms of MSE. The MMSE estimator is the most efficient in terms of MSE, exhibiting a complexity of  $\mathcal{O}(MK(P+K))$ , however, as mentioned earlier, the complexity involved in the calculation of large-scale fading coefficients (*i.e.*, the elements of  $\mathbf{D}$ ) is not taken into account. On the other hand, the proposed estimator, presents MSE efficiency that asymptotically approaches that of the MMSE estimator and has a complexity of  $\mathcal{O}(MK(P+K+1))$ , where the calculation (*i.e.*, estimation) of the large-scale fading coefficients is already considered in the presented complexity.

## VI. SIMULATION RESULTS

In this section, we compare the performance of the proposed channel estimator with that of LS and MMSE channel estimators. Additionally, we assess the performance of MRC, ZF, and MMSE linear decoders when the MMSE and the proposed channel estimators are employed. The performance of each linear decoder is quantified in terms of its Bit Error Rate (BER) over a range of UL Tx power (*i.e.*,  $\rho$ ) values.

We consider two different types of simulation setups for the large-scale fading coefficient,  $d_k$ , one with fixed values and other with random values. For the fixed case, we set  $d_k = 1$ .

For the random case, the MTC devices in the cell (see Figure 6) are uniformly distributed within a ring with radii  $r_0 = 100$  m and  $r_1 = 1000$  m respectively. The large-scale fading coefficients  $\{d_k\}$  are independently generated by  $d_k = \psi / \left(\frac{r_k}{r_0}\right)^v$ , where  $v = 3.8$ ,  $10 \log_{10}(\psi) \sim \mathcal{N}(0, \sigma_{\text{shadow,dB}}^2)$  with  $\sigma_{\text{shadow,dB}} = 8$ , and  $r_k$  is the distance of the  $k$ -th MTC device to the BS. Both, the path loss exponent,  $v$ , and the standard deviation of the log-normal shadowing,  $\sigma_{\text{shadow,dB}}$ , are common values for outdoor shadowed urban cellular radio environments [88], [89]. For all simulations we assume  $K = 10$  MTC devices.

Figure 7 shows the MSE versus UL Tx power results ( $\rho$ ) for the case when  $d_k = 1$  and  $M = 70$ . As can be noticed, the

analytical, approximated and simulated MSE curves match for all the studied channel estimators. As expected, the MSE of all estimators decreases as  $\rho$  increases. At low  $\rho$  values (values lower than -10 dB), the MSE of the proposed estimator is higher than that of the MMSE estimator, however, it is still smaller than that of the LS estimator. On the other hand, with the increase of  $\rho$  (for values higher than -10 dB), the gap between the MMSE and proposed estimators decreases.

The MSE versus the number of BS antennas,  $M$ , is compared in Figure 8 for the case when  $d_k = 1$  and  $\rho = 10$  dB. The MSE of the proposed estimator decreases as  $M$  increases, approaching the MSE of the MMSE estimator, while the MSE of the LS and MMSE channel estimators stay constant. The result showed in the figure is in line with Remark 4. Additionally, it is also worth mentioning that the approximated MSE expression for the proposed channel estimator matches the values of the closed-form expression.

In Figure 9, we assess the variation of the MSE under random large-scale fading  $\{d_k\}$  for  $M = 30$  and  $M = 70$  respectively. The results are obtained by averaging the MSE values over  $10 \times 10^3$  realizations of  $\{d_k\}$ . As can be noticed, the simulated MSE values match the values of the analytical and approximated MSE expressions. It can be also noticed that the MSE of the proposed channel estimator approaches that of the MMSE channel estimator as  $M$  increases. The proposed channel estimator outperforms the LS channel estimator significantly at low UL Tx power values.

Figures 10 and 11 present the averaged variance and error under random large-scale fading coefficient,  $\{d_k\}$ , respectively. The results are obtained by averaging the variance and error values over  $10 \times 10^6$  different realizations of  $\{d_k\}$ . As can be seen, the variance and consequently the error of the LS and MMSE channel estimators do not depend on  $M$  (both of them depend only on  $\rho$ ), however, the variance of the proposed channel estimator depends on both  $M$  and  $\rho$ . As  $M$  increases, the variance curve for the the proposed channel

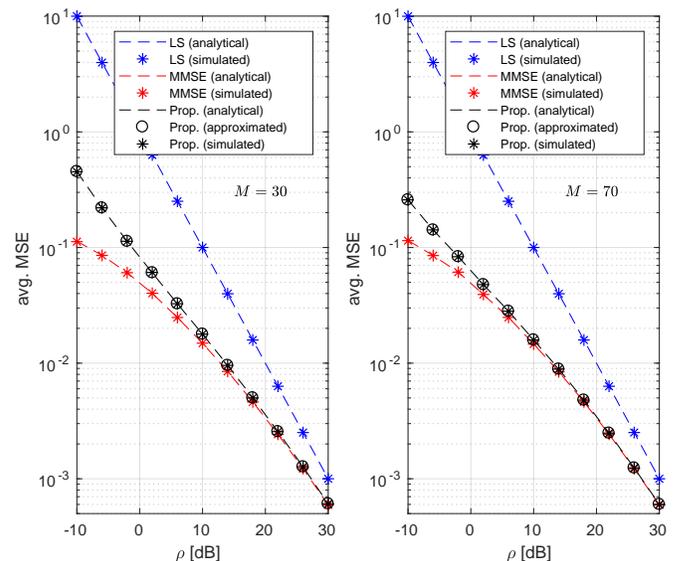


Fig. 9. Average channel estimation MSE under random  $d_k$  versus uplink pilot power,  $\rho$ .

TABLE II  
COMPLEXITIES INVOLVED IN THE STUDIED LINEAR CHANNEL ESTIMATORS.

Estimator	Operation	Flops	$\mathcal{O}(\cdot)$
LS	$\mathbf{Y}\Phi$	$MK(2P-1)$ or $2MKP$ if $P$ is large	$MKP$
MMSE/Prop.	$\mathbf{Y}\Phi\mathbf{D}$	$MK(2P-1) + MK(2K-1)$ or $2MK(P+K)$ if $P$ and $K$ are large	$MK(P+K)$
	Calculation of the elements in $\mathbf{D}$	MMSE: Elements are considered perfectly known Prop.: calculation of $\ \mathbf{y}_k\ ^2 \forall k : K(2M-1)$ flops or $2KM$ if $M$ is large	- $MK$

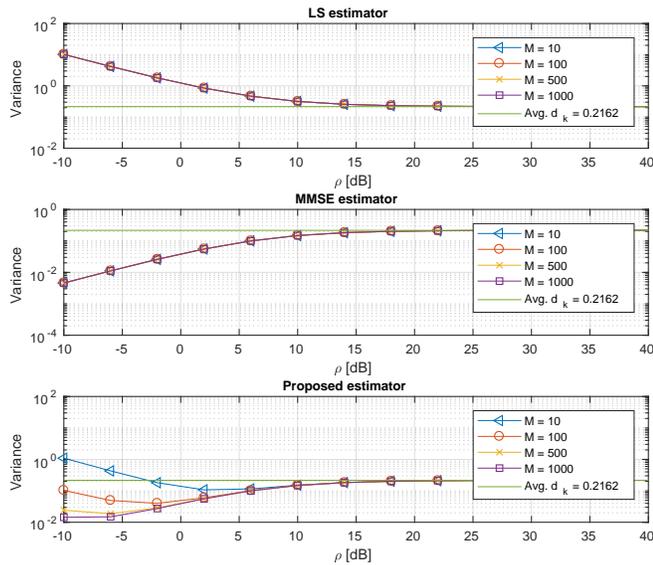


Fig. 10. Comparison of the averaged variances of the channel estimators.

estimator approaches that of the MMSE channel estimator. It is also worth mentioning that the variance of both MMSE and proposed channel estimators converges faster to the average  $d_k$  than the variance of the LS estimator as can be seen in Figure 11. These results are in line with Remarks 1, 2 and 5,

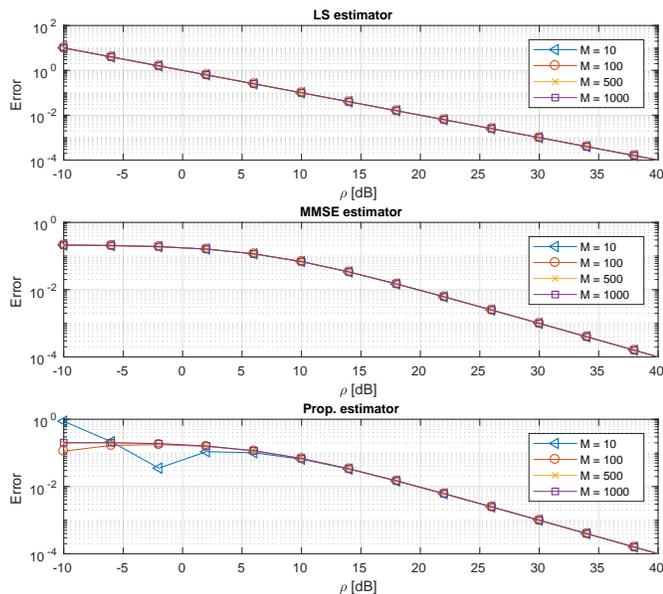


Fig. 11. Average of the absolute error between the large-scale fading,  $d_k$ , and the variance of the studied estimators.

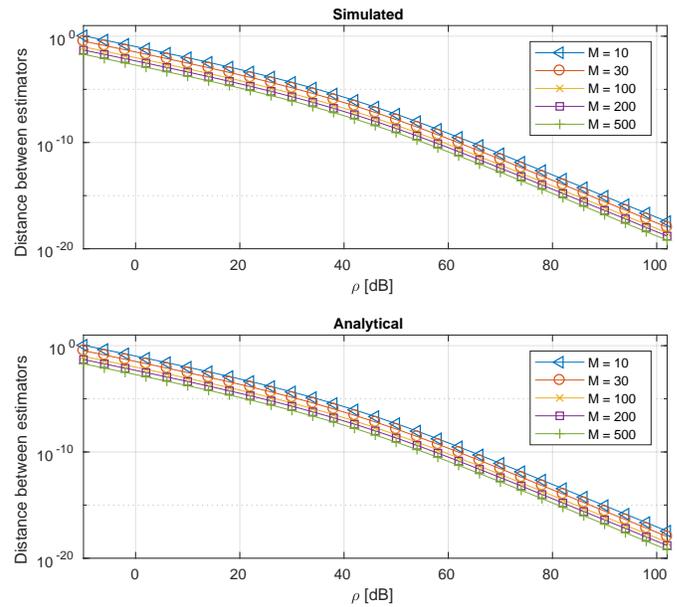


Fig. 12. Averaged distance between proposed and MMSE channel estimators.

showing that the variance of all the studied channel estimators tend to the average  $d_k$ .

The averaged distance between the proposed and MMSE channel estimators for different number of antennas,  $M$ , and

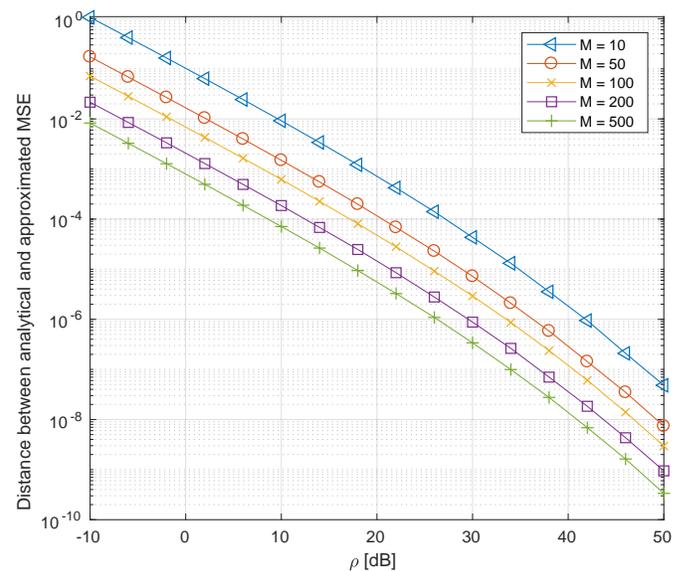


Fig. 13. Averaged absolute distance between closed-form and approximated MSE error expressions,  $|\eta_k^{\text{prop}} - \eta_k^{\text{prop(approx.)}}|$ .

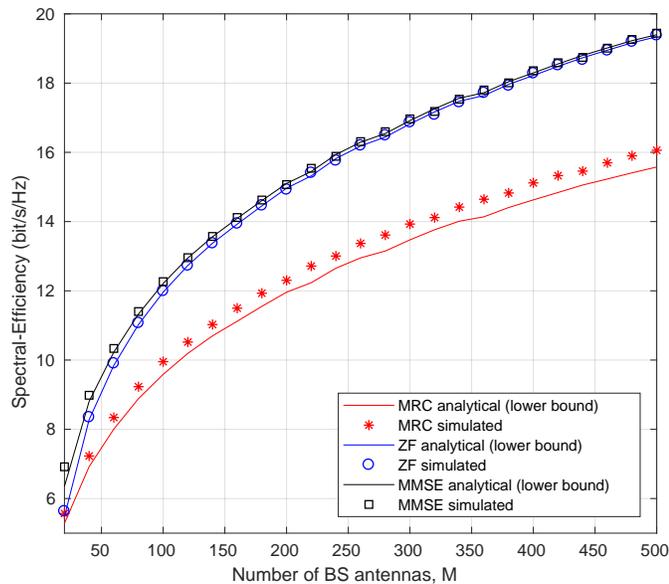


Fig. 14. Lower bounds and numerically evaluated values of the spectral efficiency for different number of BS antennas for MRC, ZF and MMSE detectors.

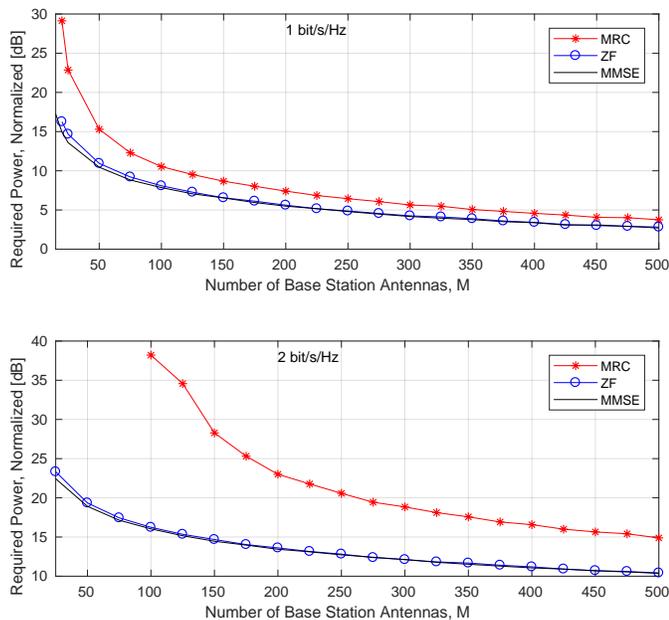


Fig. 15. Transmit power required to achieve 1 and 2 bits/s/Hz per user for MRC, ZF and MMSE linear receivers as a function of the number of antennas  $M$ . The number of users is set to  $K = 10$ , and the propagation parameters are  $\sigma_{\text{shadow}} = 8$  dB and  $v = 3.8$ .

UL Tx power values,  $\rho$ , under random large-scale fading,  $\{d_k\}$ , is depicted in Figure 12. The results are obtained by averaging the MSE values over  $10 \times 10^3$  realizations of  $\{d_k\}$  for the simulated and closed-form distances between the two estimators. As stated in Remark 6, the distance between the estimators decreases as  $M$  and  $\rho$  increase.

In Figure 13, we compare the averaged absolute distance between the approximated MSE expression presented in (45) and the analytical (closed-form) MSE expression presented in (49) under random large-scale fading  $\{d_k\}$  for various UL Tx

power (*i.e.*,  $\rho$ ) and  $M$  values. The results are obtained by averaging the absolute distance between the MSE expressions over  $10 \times 10^3$  realizations of  $\{d_k\}$ . The distance between the MSE expressions is large (around 1.07) at low  $\rho$ , decreasing with  $\rho$  as expected. For  $M = 100$  and  $\rho = -10$  dB the averaged absolute error between the expressions is around 0.072. The results presented in Figure 13 show that the approximated MSE expression can be used instead of the more complex one given by (45) when  $M$  is at least one order of magnitude larger than  $K$  (*i.e.*,  $M \gg K$ ) and/or at high UL Tx power regimes.

Figure 14 shows lower bounds and simulated spectral efficiency for MRC, ZF and MMSE detectors employing MMSE channel estimation under random large-scale fading  $\{d_k\}$ . The spectral efficiency is averaged over  $1 \times 10^6$  realizations of  $\{d_k\}$ . In this simulation  $\rho$  is set to 10 dB and there are  $K = 10$  MTC devices. As can be seen, at this UL Tx power, the spectral efficiency for  $M = 500$  is in the order of 16 – 20 bits/s/Hz, corresponding to a spectral efficiency of 1.6 – 2 bits/s/Hz per MTC device. These values are in line with practical values, for example, 64-QAM with rate 1/2 channel coding corresponds to 3 bits/s/Hz [74]. As can be seen from the figure, the MMSE detector is always better than the MRC and ZF detectors, however, the performance of the ZF detector asymptotically approaches the MMSE with increasing  $M$ . The conclusion here is that even with simple and sub-optimal linear processing such as MRC and ZF it is possible to achieve high spectral efficiency and consequently serve more MTC devices at the same time/frequency resource. It is also worth discussing the achievable rates when the proposed estimator is employed instead of the MMSE one. As it has been shown earlier, the performance of the proposed estimator asymptotically approaches that of the MMSE estimator as  $M$  and/or  $\rho$  increases, therefore, the lower bounds derived here can be thought of as upper bounds for linear detectors using the proposed estimator, *i.e.*, they will never perform better than linear detectors using the MMSE channel estimator once at best, the proposed estimator will be equal to the MMSE one.

In Figure 15 we show the transmit power per user that is needed to reach a fixed spectral efficiency. The figure shows the normalized power,  $\rho$ , required to achieve 1 and 2 bits/s/Hz per user as a function of  $M$ . It can be seen that by doubling the number of antennas,  $M$ , the transmit power can be cut back by approximately 2 dB. When  $M$  is large, the difference in performance between MRC and ZF (or MMSE) is less than 1.5 dB. This difference increases when the target spectral efficiency is increased. For the 2 bits/s/Hz case, the cross-talk interference, *i.e.*, interference from other devices, is more significant (relative to thermal noise) and therefore, ZF and MMSE receivers perform relatively better. This characteristic of massive MIMO is very important to MTC devices, where stringent power constraints have to be respected.

Next, we present some simulations to assess how the BER behaves when a large number of antennas and sub-optimal linear detectors are employed at the BS. For the following simulation results, we assume OFDM parameters similar to LTE: (i) a symbol interval of  $T_s = 500/7 \approx 71.4$  us, (ii) a

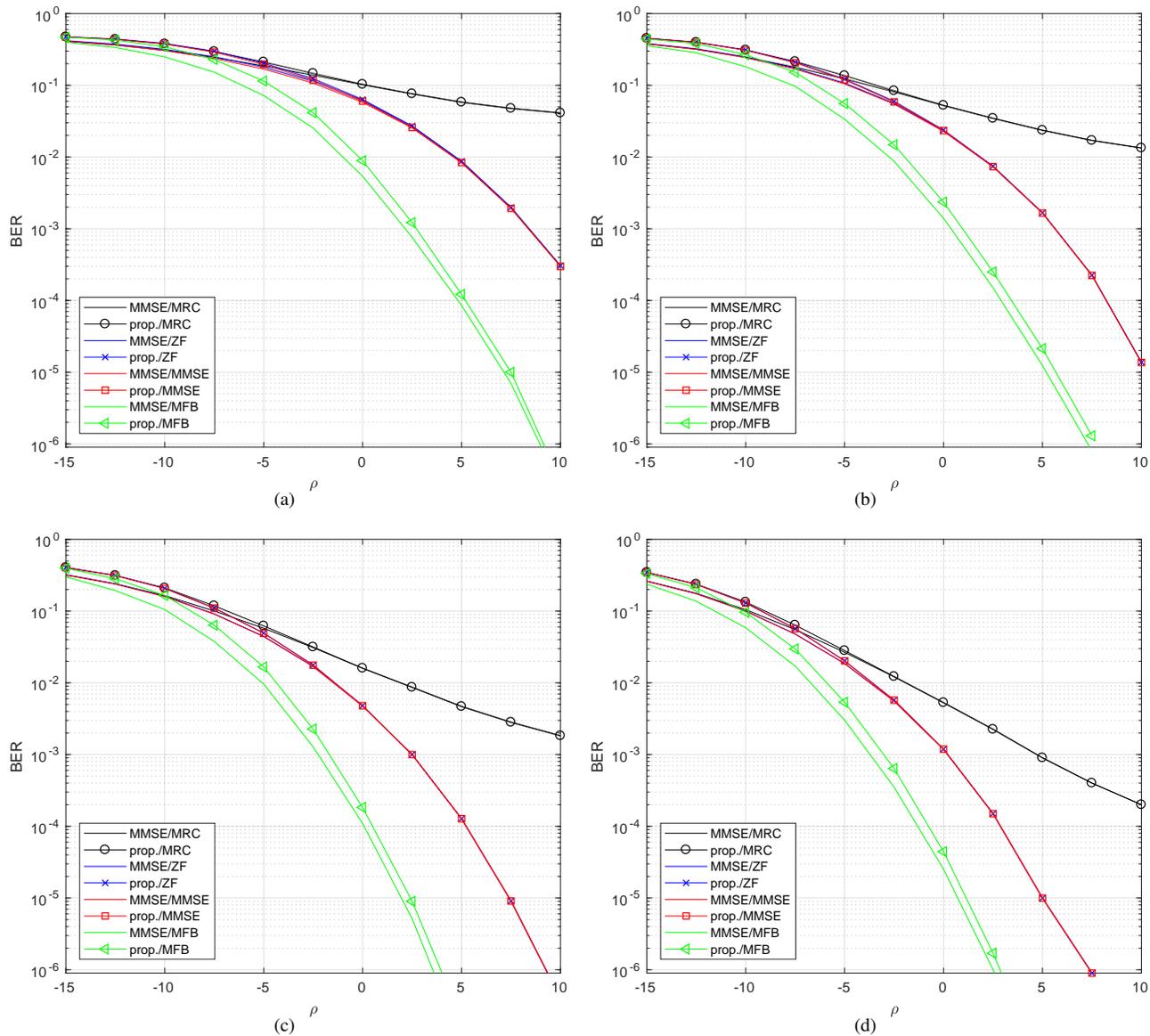


Fig. 16. BER performance of different linear detectors methods for  $K = 10$  single-antenna MTC devices and different array sizes at the BS. MFB is provided as benchmark for comparisons. (a)  $M = 50$  antennas. (b)  $M = 100$  antennas. (c)  $M = 250$  antennas. (d)  $M = 500$  antennas.

subcarrier spacing of  $\Delta f = 15$  KHz, (iii) a useful symbol duration  $T_u = 1/\Delta f \approx 66.7$  us, and (iv) a cyclic prefix interval (guard interval)  $T_g = T_s - T_u \approx 4.76$  us. The frequency smoothness interval is approximately  $N_{\text{smooth}} = 14$  subcarriers. We assume a coherence time of 1 ms (equivalent to 14 OFDM symbols), of which 1 symbol is used to send uplink pilots and the remaining symbols, *i.e.*, 13 symbols, are used to send uplink data. Therefore, in this case, the maximum number of MTC devices,  $K = \tau_p N_{\text{smooth}} = 1 \times 14 = 14$ . Additionally, we consider uncoded QPSK uplink data transmissions.

We use the Matched Filter Bound (MFB) as benchmark for the BER comparisons. The MFB is also known in the literature as the perfect interference-cancellation bound [77]. As the name suggests, MFB performs as the  $k$ -th user of a matched-filter receiver in the absence of other sources of interference [77]. Our motivation for this choice is that for  $M \gg K$  both multi-user interference and small-scale fading

effects tend to disappear (thanks to the processing gains of Massive MIMO), so the performance of the MIMO  $K \times M$  channel, which is assumed to be flat Rayleigh fading inside a coherence block, approaches the MFB as  $M \rightarrow \infty$ . The simulation results discussed in the sequel were averaged over  $10^{10}$  realizations. The simulation type is Monte-Carlo with a bit error counting procedure that compares the transmitted bit vector to the received bit vector.

Figure 16 shows the BER of the linear filtering detectors described in section III-E for a fixed number of  $K = 10$  MTC devices and BS array sizes in the range of  $50 \leq M \leq 500$  antennas. For UL Tx power values smaller than 0 dB, the gap between all the linear detectors (MRC, ZF and MMSE) employing the proposed channel estimator and the ones employing MMSE estimator is noticeable, vanishing as  $\rho$  increases. This gap also decreases as  $M$  increases. In the following discussion, we consider the performance of the

studied linear detectors without differentiating if they employ MMSE or the proposed channel estimator to calculate the equalization matrix  $\mathbf{A}$  once the analysis apply to both cases.

After analyzing the figure, the performance gap inherent to MRC becomes evident, although, as can be seen, it can be dramatically reduced at the expense of larger array sizes at the BS. These results suggest that even low-complexity MRC detectors have the potential to approximate the MFB when  $M$  can be made large enough.

As expected, due to its better balance between interference suppression and noise enhancement, the MMSE detector outperforms both the MRC and ZF detectors in all cases studied. The performance gap between ZF and MMSE, which is small enough to be considered negligible for  $M \leq 50$ , entirely disappears as the BS array size is increased to  $M = 250$  or more antennas. As a matter of fact, the main conclusion that can be drawn from the results presented here is that the MRC, ZF, and MMSE detectors all approach the performance of MFB as  $M \rightarrow \infty$ , however, the gap between the MFB and the ZF/MMSE detectors decreases at a faster pace when compared to that of the MRC detector. It is also important to notice that (4) is an approximation that only becomes an equality when  $M \gg K$ ,  $M \rightarrow \infty$ , which explains why the MRC detection is not optimum for the results presented here.

## VII. CONCLUDING REMARKS AND FUTURE WORK

In this paper we propose the use of a Massive MIMO setup as means to tackle the uplink mixed-service communication problem. By assuming the availability of a physical narrow-band shared channel (PNSCH), the capacity of the network is increased through the creation of clusters of MTC devices that share the same time-frequency physical resource blocks for data transmission.

We present a simple and practical channel estimator that does not need previous knowledge of the large-scale fading coefficients. The proposed channel estimator presents MSE results that asymptotically approach the ones exhibited by the MMSE channel estimator as the number of antennas and/or UL Tx power increases. Additionally, we derive closed-form and approximated MSE expressions for the proposed estimator, which can be useful in system design and performance evaluation.

Moreover, our simulation results suggest that, as the size of the antenna array at the BS is made progressively larger, the performance of sub-optimal linear detectors approaches the perfect interference-cancellation bound.

The findings presented in this paper shed light into and motivate for entirely new research lines towards a better understanding of Massive MIMO for MTC networks.

Next we describe some directions along which the results of this paper can be further extended. First, the results presented in our work assume that there is no inter-cell interference, *i.e.*, coherent and non-coherent interference. Therefore, future work can extend the existing results to account for inter-cell interference. Second, in this paper, we assume independent and identically distributed (*i.i.d.*) Rayleigh fading, however, practical channels are not spatially uncorrelated. Therefore,

future work considering the impact spatially correlated fading (Rician and Rayleigh fading), *i.e.*, spatially correlated channels, has on channel estimation, linear detection and spectral efficiency is of interest.

## APPENDIX A

For the proof of (41), we need the following Lemmas.

*Lemma 1:* Let  $z \sim \mathcal{CN}(0, a)$ , where  $z = u + jv$ , then  $\mathbb{E}[|z|^4] = 2a^2$ .

*Proof:* In order to prove this, we first rewrite  $\mathbb{E}[|z|^4]$  as

$$\mathbb{E}[|z|^4] = \frac{1}{\pi a} \int_{\mathbb{C}} |z|^4 e^{-\frac{|z|^2}{a}} dz. \quad (54)$$

Next, taking the Jacobian determinant of the transformation  $r = \sqrt{u^2 + v^2}$  and  $\theta = \arctan\left(\frac{v}{u}\right)$  results in

$$\begin{aligned} |J| &= \left| \frac{\partial(r, \theta)}{\partial(u, v)} \right| = \left| \begin{array}{cc} \frac{\partial r}{\partial u} & \frac{\partial r}{\partial v} \\ \frac{\partial \theta}{\partial u} & \frac{\partial \theta}{\partial v} \end{array} \right| = \left| \begin{array}{cc} \frac{u}{\sqrt{u^2+v^2}} & \frac{v}{\sqrt{u^2+v^2}} \\ -\frac{v}{u^2+v^2} & \frac{u}{u^2+v^2} \end{array} \right| \\ &= \frac{1}{\sqrt{u^2+v^2}} = \frac{1}{r}. \end{aligned} \quad (55)$$

Therefore, the fourth moment of  $z$  expressed in terms of  $r$  and  $\theta$  is defined by

$$\begin{aligned} \mathbb{E}[r^4] &= \mathbb{E}[|z|^4] \left| \frac{\partial(r, \theta)}{\partial(u, v)} \right| \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} d\theta \frac{1}{a} \int_0^{\infty} r^5 e^{-\frac{r^2}{a}} dr \\ &= \frac{2}{a} \int_0^{\infty} r^5 e^{-\frac{r^2}{a}} dr \\ &= 2a^2. \end{aligned} \quad (56)$$

*Lemma 2:* Let  $\mathbf{z} = [z_1, \dots, z_M]^T \in \mathbb{C}^{M \times 1}$  be a complex random vector with distribution  $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}_M, a\mathbf{I}_M)$ . Then  $\mathbb{E}[\|\mathbf{z}\|^4] = a^2 M(M+1)$ .

*Proof:* We start the proof by expanding  $\mathbb{E}[\|\mathbf{z}\|^4]$  as

$$\begin{aligned} \mathbb{E}[\|\mathbf{z}\|^4] &= \mathbb{E}[(\mathbf{z}^H \mathbf{z})^2] = \mathbb{E} \left[ \left( \sum_{m=1}^M |z_m|^2 \right)^2 \right] \\ &= \sum_{m=1}^M \sum_{m'=1}^M \mathbb{E}[|z_m|^2 |z_{m'}|^2] \\ &= \sum_{m=1}^M \mathbb{E}[|z_m|^4] + \sum_{m=1}^M \sum_{m'=1, m' \neq m}^M \mathbb{E}[|z_m|^2] \mathbb{E}[|z_{m'}|^2] \\ &= a^2 M(M+1), \end{aligned} \quad (57)$$

where in the last equality we used Lemma 1. Lemma 3 can be used to find moments of any order.

*Lemma 3:* Let  $\mathbf{z} = [z_1, \dots, z_M]^T \in \mathbb{C}^{M \times 1}$  be a complex random vector with distribution  $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}_M, a\mathbf{I}_M)$ . Then  $\mathbb{E}[\|\mathbf{z}\|^{2k}] = a^k \prod_{i=0}^{k-1} (M+i)$ , where  $\{k \in \mathbb{Z} \mid k > 0\}$ .

*Proof:* Given the pdf of  $z$

$$f(\mathbf{z}) = \frac{1}{(\pi a)^M} \exp\left(-\frac{1}{a} \|\mathbf{z}\|^2\right). \quad (58)$$

By using the following identity

$$\frac{1}{(\pi a)^M} \int_{\mathbb{C}} \exp\left(-\frac{1}{a}\|\mathbf{z}\|^2\right) d\mathbf{z} = 1, \quad (59)$$

which can be rewritten as

$$\int_{\mathbb{C}} \exp\left(-\frac{1}{a}\|\mathbf{z}\|^2\right) d\mathbf{z} = (\pi a)^M, \quad (60)$$

and next deriving both sides of the identity w.r.t.  $a$ , we find

$$\int_{\mathbb{C}} \frac{d^k}{d^k a} \left[ \exp\left(-\frac{1}{a}\|\mathbf{z}\|^2\right) \right] d\mathbf{z} = \frac{d^k}{d^k a} [(\pi a)^M]. \quad (61)$$

After applying the derivation to (61), it can be rewritten as

$$\int_{\mathbb{C}} \|\mathbf{z}\|^{2k} \exp\left(-\frac{1}{a}\|\mathbf{z}\|^2\right) d\mathbf{z} = \pi^M (M)^{(k)} a^{M+k}, \quad (62)$$

where  $(M)^{(k)} = \prod_{i=0}^{k-1} (M+i)$ . Next, by reorganizing (62), we find

$$\underbrace{\frac{1}{(\pi a)^M} \int_{\mathbb{C}} \|\mathbf{z}\|^{2k} \exp\left(-\frac{1}{a}\|\mathbf{z}\|^2\right) d\mathbf{z}}_{\mathbb{E}[\|\mathbf{z}\|^{2k}]}, \quad (63)$$

which concludes the proof. ■

*Lemma 4:* Let  $\mathbf{z} = [z_1, \dots, z_M]^T \in \mathbb{C}^{M \times 1}$  be a complex random vector with distribution  $\mathcal{CN}(\mathbf{0}_M, a\mathbf{I}_M)$ . Then

$$\lim_{M \rightarrow \infty} \frac{\|\mathbf{z}\|^2}{M} \stackrel{\text{a.s.}}{=} a. \quad (64)$$

*Proof:* As  $M$  increases,  $\frac{\mathbf{z}^H \mathbf{z}}{M}$  becomes more and more deterministic, and consequently, its variance must be zero when  $M \rightarrow \infty$ . Therefore, one way to prove the convergence in (64) is to show that

$$\begin{aligned} & \lim_{M \rightarrow \infty} \text{var} \left[ \frac{\|\mathbf{z}\|^2}{M} \right] \\ &= \lim_{M \rightarrow \infty} \frac{1}{M^2} \text{var} [\|\mathbf{z}\|^2] \\ &= \lim_{M \rightarrow \infty} \frac{1}{M^2} \left\{ \mathbb{E} [\|\mathbf{z}\|^4] - (\mathbb{E} [\|\mathbf{z}\|^2])^2 \right\}. \end{aligned} \quad (65)$$

Applying Lemma 2 to (65) and knowing that  $\mathbb{E} [\|\mathbf{z}\|^2] = aM$ , we find that

$$\lim_{M \rightarrow \infty} \text{var} \left[ \frac{\|\mathbf{z}\|^2}{M} \right] = a^2 \lim_{M \rightarrow \infty} \frac{1}{M} \stackrel{\text{a.s.}}{=} 0. \quad (66)$$

This Lemma is accordance with the Law of large numbers [90], [91].

For the proof of (41) we take into account the distribution of  $\mathbf{y}_k$  defined in (7) and using Lemma 4, after simple computations, we complete the proof. ■

## APPENDIX B

For the proof of the mean-square estimation error of the proposed channel estimator we need to define the following Lemmas.

*Lemma 5:* If  $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}_M, \sigma_x^2 \mathbf{I}_M)$  and  $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}_M, \sigma_y^2 \mathbf{I}_M)$  are independent and  $\frac{\mathbf{x}^H \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} \triangleq R e^{j\theta}$ , therefore,  $\theta$  is uniformly distributed in the range  $[-\pi, \pi]$  and the pdf of  $R$  is defined as

$$f_R(r) = 2Mr(1-r^2)^{M-1}, \quad 0 \leq r \leq 1. \quad (67)$$

*Proof:* The circular symmetry of  $\mathbf{x}$  and  $\mathbf{y}$  results in the uniform distribution of  $\theta$ . The random variable  $Z = \left| \frac{\mathbf{x}^H \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} \right|^2$  exhibits a Beta pdf  $f_Z(z) = M(1-z)^{M-1}$ ,  $z \in [0, 1]$  [92]. Finally, the transformation of random variable  $R = \sqrt{Z}$  yields (67). ■

*Lemma 6:* If a random variable  $R$  has pdf defined as in (67),  $\theta$  is uniformly distributed within the range  $[-\pi, \pi]$  and they are independent, therefore the pdf of the random variable  $W \triangleq R \cos(\theta)$  is given by

$$f_W(w) = \frac{M}{\pi} B\left(\frac{1}{2}, M\right) (1-w^2)^{M-\frac{1}{2}}, \quad |w| \leq 1. \quad (68)$$

*Proof:* Starting from the uniform distribution of  $\theta$ , we have that

$$\mathbb{P}\left(\cos(\theta) \leq \frac{w}{r}\right) = \begin{cases} 0, & \frac{w}{r} < -1 \\ 1 - \frac{1}{\pi} \cos^{-1}\left(\frac{w}{r}\right), & -1 \leq \frac{w}{r} \leq 1 \\ 1, & \frac{w}{r} > 1 \end{cases} \quad (69)$$

Next, the cumulative distribution function (cdf) of the random variable  $W$  is defined as  $F_W(w) = \mathbb{P}(\cos(\theta) \leq w) = \int_0^1 \mathbb{P}(\cos(\theta) \leq \frac{w}{r}) f_R(r) dr$

$$F_W(w) = \begin{cases} 0, & w < -1 \\ \xi(w), & -1 \leq w \leq 0 \\ \xi(w) + \int_0^w f_R(r) dr, & 0 \leq w \leq 1 \\ 1, & w > 1 \end{cases} \quad (70)$$

where  $\xi(w) \triangleq \int_{|w|}^1 \left(1 - \frac{1}{\pi} \cos^{-1}\left(\frac{w}{r}\right)\right) f_R(r) dr$ . Therefore, we find that  $f_W(w) = \frac{dF_W(w)}{dw} = \int_{|w|}^1 \frac{2Mr(1-r^2)^{M-1}}{\pi\sqrt{r^2-w^2}} dr$ ,  $|w| < 1$ . Next, by changing the variable  $z$  as  $z = r^2 - w^2$  and using [93], we complete the proof and find (68). ■

*Lemma 7:* If the random variables  $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$  and  $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$  are independent, therefore,  $U = \frac{\|\mathbf{x}\|}{\|\mathbf{y}\|}$  has its pdf defined by

$$f_U(u) = \frac{2\Gamma(2M)}{(\Gamma(M))^2} \frac{u^{2M-1}}{(u^2+1)^{2M}}, \quad u > 0. \quad (71)$$

*Proof:* We start by recalling that  $\|\mathbf{x}\|^2$  and  $\|\mathbf{y}\|^2$  are central Chi-square random variables with the following pdf:  $f_V(v) = \frac{v^{M-1}}{\Gamma(M)} e^{-v}$ . Next, by using the independence of  $\|\mathbf{x}\|^2$  and  $\|\mathbf{y}\|^2$ , the cdf of  $U^2$  is found and then by differentiating it we find the pdf of  $U^2$  as being defined by  $f_Z(z) = \frac{\Gamma(2M)}{(\Gamma(M))^2} \frac{z^{M-1}}{(z+1)^{2M}}$ ,  $z > 0$ . Finally, by applying the square root variable transformation to  $U^2$ , we find (71). ■

*Lemma 8:* If the random variables  $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}_M, \sigma_x^2 \mathbf{I}_M)$  and  $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}_M, \sigma_y^2 \mathbf{I}_M)$  are independent, therefore

$$\mathbb{E} \left\{ \frac{(\mathbf{x} + \mathbf{y})^H \mathbf{x}}{\|\mathbf{x} + \mathbf{y}\|^2} \right\} = \int_0^\infty \int_{-1}^1 \frac{(ku + w)f_U(u)f_W(w)}{ku + \frac{1}{ku} + 2w} dw du, \quad (72)$$

where  $k \triangleq \sqrt{\frac{\sigma_x^2}{\sigma_y^2}}$ , and  $f_W(w)$  and  $f_U(u)$  are defined in (68) and (71), respectively.

*Proof:* We start this proof by expanding and dividing both the numerator and denominator of the left-hand part of (72) by  $\|\mathbf{x}\|\|\mathbf{y}\|$ . This way, that term can be re-written as

$$\mathbb{E} \left\{ \frac{(\mathbf{x} + \mathbf{y})^H \mathbf{x}}{\|\mathbf{x} + \mathbf{y}\|^2} \right\} = \mathbb{E} \left\{ \frac{kU + Re^{-j\theta}}{kU + \frac{1}{kU} + 2W} \right\}, \quad (73)$$

where  $\frac{\|\mathbf{x}\|}{\|\mathbf{y}\|} \triangleq kU$ ,  $\frac{\mathbf{x}^H \mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} \triangleq Re^{j\theta}$ , and  $W \triangleq R \cos(\theta)$ . Note that  $U$ ,  $R$  and  $\theta$  are independent random variables. Initially, we find that the expected value of the imaginary part of the left-hand side of (73) is equal to zero when we condition it on  $U$  and  $R$ , and average it over  $\theta$ . Therefore, (73) becomes  $\mathbb{E} \left\{ \frac{kU + W}{kU + \frac{1}{kU} + 2W} \right\}$ . Next, by applying Lemmas 6 and 7 to (73) and doing a direct calculation of the expectation in (73) over the pdfs,  $f_U(u)$  and  $f_W(w)$ , results in (72), which completes the proof. ■

#### A. Proof of the mean-square estimation error, $\eta_k^{\text{prop}}$

For the proof of the mean-square estimation error, we first expand it as

$$\eta_k^{\text{prop}} = \frac{1}{M} \mathbb{E} \left[ \|\hat{\mathbf{h}}_k^{\text{prop}}\|^2 \right] + \frac{1}{M} \mathbb{E} \left[ \|\mathbf{h}_k\|^2 \right] - \frac{2}{M} \mathbb{E} \left[ \Re \left[ \left( \hat{\mathbf{h}}_k^{\text{prop}} \right)^H \mathbf{h}_k \right] \right], \quad (74)$$

and find these three expectations. The first expectation can be expanded as

$$\begin{aligned} \frac{1}{M} \mathbb{E} \left[ \|\hat{\mathbf{h}}_k^{\text{prop}}\|^2 \right] &= \frac{1}{M} \mathbb{E} \left[ \left( 1 - \frac{M}{\rho} \frac{1}{\|\mathbf{y}_k\|^2} \right)^2 \|\mathbf{y}_k\|^2 \right] \\ &= \frac{1}{M} \mathbb{E} \left[ \|\mathbf{y}_k\|^2 \right] - \frac{2}{\rho} + \frac{M}{\rho^2} \mathbb{E} \left[ \frac{1}{\|\mathbf{y}_k\|^2} \right], \end{aligned} \quad (75)$$

where the first term  $\frac{1}{M} \mathbb{E} \left[ \|\mathbf{y}_k\|^2 \right] = d_k + \frac{1}{\rho}$  and in order to find the last term we use the fact that  $\frac{1}{\|\mathbf{y}_k\|^2}$  has an Inverse Gamma distribution,  $\Gamma^{-1}(M, d_k + \frac{1}{\rho})$ , with mean,  $\mathbb{E} \left[ \frac{1}{\|\mathbf{y}_k\|^2} \right] = \frac{1}{(d_k + \frac{1}{\rho})(M-1)}$ , therefore, we have the first expectation defined as

$$\begin{aligned} \frac{1}{M} \mathbb{E} \left[ \|\hat{\mathbf{h}}_k^{\text{prop}}\|^2 \right] &= \frac{(\rho d_k + 1)(\rho d_k - 1)(M-1) + M}{\rho(\rho d_k + 1)(M-1)} \\ &= \frac{\rho d_k^2}{(\rho d_k + 1)} + \frac{1}{\rho(\rho d_k + 1)(M-1)}. \end{aligned} \quad (76)$$

Next, for the second expectation, we recall that  $\|\mathbf{h}_k\|^2$  has a Gamma distribution,  $\Gamma(M, d_k)$  and therefore,  $\frac{1}{M} \mathbb{E} \left[ \|\mathbf{h}_k\|^2 \right] = d_k$ .

Finally, in order to find the third expectation, we first expand it as

$$\begin{aligned} & -\frac{2}{M} \mathbb{E} \left[ \Re \left[ \left( \hat{\mathbf{h}}_k^{\text{prop}} \right)^H \mathbf{h}_k \right] \right] \\ &= -\frac{2}{M} \mathbb{E} \left[ \Re \left\{ \left( 1 - \frac{M}{\rho} \frac{1}{\|\mathbf{y}_k\|^2} \right) \mathbf{y}_k^H \mathbf{h}_k \right\} \right] \\ &= -\frac{2}{M} \mathbb{E} \left[ \Re \left\{ \mathbf{y}_k^H \mathbf{h}_k \right\} \right] + \frac{2}{\rho} \mathbb{E} \left[ \Re \left\{ \frac{\mathbf{y}_k^H \mathbf{h}_k}{\|\mathbf{y}_k\|^2} \right\} \right]. \end{aligned} \quad (77)$$

After simple calculations we find that the first term in (77) is given by  $-\frac{2}{M} \mathbb{E} \left[ \Re \left\{ \mathbf{y}_k^H \mathbf{h}_k \right\} \right] = -2d_k$ . In order to find the second term in (77), we apply Lemma 8 to it with  $\mathbf{x} = \mathbf{h}_k$  and  $\mathbf{y} = \mathbf{y}_k$ , where the distribution of  $\mathbf{y}_k$  is defined in (7). Next, for the purpose of making the boundary in the integral finite, we apply the following change of variable,  $t = \frac{1}{1+u^2}$ , and then, we define this integral as  $\theta_k$ . After substituting each one of the three expectations back in the expansion of the mean-square estimation error,  $\eta_k^{\text{prop}}$ , in (74), we conclude the proof.

#### APPENDIX C

In this appendix we present proof for Remark 6. We start with the following expansion

$$\begin{aligned} \frac{1}{M} \mathbb{E} \left[ \|\hat{\mathbf{h}}_k^{\text{prop}} - \hat{\mathbf{h}}_k^{\text{MMSE}}\|^2 \right] &= \frac{1}{M} \mathbb{E} \left[ \|\hat{\mathbf{h}}_k^{\text{prop}}\|^2 \right] + \\ & \frac{1}{M} \mathbb{E} \left[ \|\hat{\mathbf{h}}_k^{\text{MMSE}}\|^2 \right] - \frac{2}{M} \mathbb{E} \left[ \Re \left[ \left( \hat{\mathbf{h}}_k^{\text{prop}} \right)^H \hat{\mathbf{h}}_k^{\text{MMSE}} \right] \right]. \end{aligned} \quad (78)$$

Next, we calculate the three different expectations in (78). First, from Appendix B we know that  $\frac{1}{M} \mathbb{E} \left[ \|\hat{\mathbf{h}}_k^{\text{prop}}\|^2 \right] = \gamma_k + \epsilon_k$ . Next, by recalling that the distribution of  $\hat{\mathbf{h}}_k^{\text{MMSE}}$  is defined in (19), and using the fact that  $\|\hat{\mathbf{h}}_k^{\text{MMSE}}\|^2$  has a Gamma distribution,  $\Gamma \left( M, \frac{\rho d_k^2}{1 + \rho d_k} \right)$ , then we have that  $\frac{1}{M} \mathbb{E} \left[ \|\hat{\mathbf{h}}_k^{\text{MMSE}}\|^2 \right] = \frac{\rho d_k^2}{1 + \rho d_k} = \gamma_k$ . For the last expectation, after substituting  $\hat{\mathbf{h}}_k^{\text{prop}}$  and  $\hat{\mathbf{h}}_k^{\text{MMSE}}$  in the last term of (78) we find

$$\begin{aligned} & -\frac{2}{M} \mathbb{E} \left[ \Re \left[ \left( \hat{\mathbf{h}}_k^{\text{prop}} \right)^H \hat{\mathbf{h}}_k^{\text{MMSE}} \right] \right] \\ &= -\frac{2}{M} \frac{d_k}{d_k + \frac{1}{\rho}} \left\{ \mathbb{E} \left[ \|\mathbf{y}_k\|^2 \right] - \frac{M}{\rho} \right\} \\ &= -2 \frac{\rho d_k^2}{1 + \rho d_k} = -2\gamma_k, \end{aligned} \quad (79)$$

where we have used  $\mathbb{E} \left[ \|\mathbf{y}_k\|^2 \right] = M(d_k + \frac{1}{\rho})$  in the last equality. Finally, after substituting the three found expectations back in (78), we conclude the proof.

#### APPENDIX D

For the proof of the approximated mean-square estimation error per antenna of the proposed estimator we need to define the following Lemma.

*Lemma 9:* Let  $\mu_X$  and  $\mu_Y$  be the expectations of  $X$  and  $Y$ ,  $\sigma_Y^2$  be the variance of  $Y$ , and  $\sigma_{XY}$  be their covariance. Then the expectation,  $\mathbb{E}\{X/Y\}$ , can be approximated by

$$\mathbb{E} \left\{ \frac{X}{Y} \right\} \approx \frac{\mu_X}{\mu_Y} - \frac{\sigma_{XY}}{\mu_Y^2} + \frac{\mu_X}{\mu_Y^3} \sigma_Y^2. \quad (80)$$

*Proof:* For a function that depends on two variables,  $x$  and  $y$ , the second order Taylor expansion series about the point  $(a, b)$  is given by

$$g(x, y) = g(a, b) + g_x(a, b)(x - a) + g_y(a, b)(y - b) + \frac{1}{2!}(g_{xx}(a, b)(x - a)^2 + 2g_{xy}(a, b)(x - a)(y - b) + g_{yy}(a, b)(y - b)^2), \quad (81)$$

where the subscripts denote the respective partial derivatives. The partial derivatives are defined by  $g_y = -X/Y^2$ ,  $g_{yy} = 2X/Y^3$ ,  $g_x = 1/Y$ ,  $g_{xx} = 0$ , and  $g_{xy} = -1/Y^2$ . Applying the derivatives into (81), the second order Taylor expansion of  $g(X, Y) = X/Y$  around the mean point  $(\mu_X, \mu_Y)$ , the following is obtained

$$\frac{X}{Y} \approx \frac{\mu_x}{\mu_y} - \frac{\mu_x}{\mu_y^2}(Y - \mu_y) + \frac{1}{\mu_y}(X - \mu_x) + \frac{1}{2!} \left( \frac{2\mu_x}{\mu_y^3}(Y - \mu_y)^2 - \frac{2}{\mu_y^2}(Y - \mu_y)(X - \mu_x) \right). \quad (82)$$

Finally, applying the expectation operator,  $\mathbb{E}\{\cdot\}$ , to (82) concludes the proof. ■

With the purpose of finding a more tractable expression for the mean-square error per antenna of the proposed channel estimator we derive an approximation to the expectation of the ratio of random variables in the last part of (77), *i.e.*, the term define as  $\theta_k$ .

It is possible to approximate the moments of a function  $g(X, Y)$  using Taylor series expansions, provided that  $g$  is sufficiently differentiable and that the moments of  $X$  and  $Y$  are finite. Therefore, applying Lemma 9 to

$$\theta_k = \mathbb{E} \left[ \Re \left\{ \frac{\mathbf{y}_k^H \mathbf{h}_k}{\|\mathbf{y}_k\|^2} \right\} \right] = \mathbb{E} \left[ \Re \left\{ \frac{(\mathbf{h}_k + \mathbf{w}')^H \mathbf{h}_k}{\|\mathbf{h}_k + \mathbf{w}'\|^2} \right\} \right], \quad (83)$$

we are able to find an approximation to  $\theta_k$ , which is defined as

$$\theta_k \approx \frac{\rho d_k}{1 + \rho d_k}. \quad (84)$$

Note that the approximation does not depend on  $M$  but only on  $\rho$ . The proof is concluded by substituting (84) into the expansion of the mean-square error given in (74).

## REFERENCES

- [1] Cisco, *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015-2020*, White paper, February 2016.
- [2] ITU-R, Recommendation ITU-R M.2083, *IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond*, September 2015.
- [3] Keith Mallinson, *The path to 5G: as much evolution as revolution*, [available at] [http://www.3gpp.org/news-events/3gpp-news/1774-5g\\_wiseharbour](http://www.3gpp.org/news-events/3gpp-news/1774-5g_wiseharbour), May 2016.
- [4] 4G Americas White Papers. *4G Americas' Recommendations on 5G Requirements and Solutions*, [available at] <http://www.4gamericas.org/en/resources/white-papers/>, October 2014.
- [5] Oumer Teyeb, Gustav Wikstrom, Magnus Stattin, Thomas Cheng, Sebastian Faxer, and Hieu Do, *Evolving LTE to fit the 5G future*, Ericsson Technology Review. [available at] <https://www.ericsson.com/en/ericsson-technology-review/archive/2017/evolving-lte-to-fit-the-5g-future>, January 2017.
- [6] Fettweis, G. and Alamouti, S., *5G: Personal Mobile Internet beyond What Cellular Did to Telephony* IEEE Communications Magazine, vol. 52, no. 2, pp. 140-145, February 2014.
- [7] Evangelos Vlachos, Aris S. Lalos, Kostas Berberidis, and Christos Tselios, *Autonomous driving in 5G: Mitigating interference in OFDM-based vehicular communications*, IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), June 2017.
- [8] ITU-R Report M.2370-0, *IMT traffic estimates for the years 2020 to 2030*, July 2015.
- [9] Wunder, G. et al., *5G NOW: Non-Orthogonal Asynchronous Waveforms for Future Mobile Applications*, IEEE Communications Magazine, vol. 52, no. 2, pp. 97-105, February 2014.
- [10] Boswarthick, D., Elloumi, O. and Hersent, O., *M2M Communications: A Systems Approach*, Wiley, March 2012.
- [11] Elmgangoush, A. et al., *Promoting M2M Application Development for Smart City*, Wireless World Research Forum Meeting 29 (WWRF), 2012.
- [12] J. Gubbi et al., *Internet of Things (IoT): A vision, architectural elements, and future directions*, Future Generation Computer Systems, vol. 29, no. 7, pp. 1645-1660, September 2013.
- [13] Corchado, J. M. et al., *Using Heterogeneous Wireless Sensor Networks in a Telemonitoring System for Healthcare*, IEEE Transactions on Information Technology in Biomedicine, vol. 14, no. 2, pp. 234-240, March 2010.
- [14] Papadimitratos, P. et al., *Vehicular Communication Systems: Enabling Technologies, Applications, and Future Outlook on Intelligent Transportation*, IEEE Communications Magazine, vol. 11, no. 1, pp. 84-95, November 2009.
- [15] Moslehi, K. and Kumar, R., *A reliability perspective of the smart grid*, IEEE Transactions on Smart Grid, vol. 1, no. 1, pp. 57-64, June 2010.
- [16] Kyriakides, E. and Polycarpou, M., *Intelligent Monitoring, Control and Security of Critical Infrastructure Systems*, Studies in Computational Intelligence, Springer, 1st edition, ISBN 978-3-662-44159-6, January 2015.
- [17] S. Goel, S. F. Bush and D. E. Bakken, *IEEE Vision for Smart Grid Communications: 2030 and Beyond*, IEEE Press, E-ISBN: 978-0-7381-8460-9, May 2013.
- [18] Pattichis, C. S. et al., *Wireless Telemedicine Systems: An Overview*, IEEE Antennas Propagation Magazine, vol. 44, no. 2, pp. 143-153, April 2002.
- [19] Majerowicz, A. and Tracy, S., *Telemedicine: Bridging Gaps in Healthcare Delivery*, Journal of AHIMA, vol. 81, no. 5, pp. 52-53, May 2010.
- [20] Tadiolo Endeshaw Bogale and Long Bao Le, *Massive MIMO and Millimeter Wave for 5G Wireless HetNet: Potentials and Challenges*, IEEE Vehicular Technology Magazine, vol. 11, no. 1, pp. 64-75, March 2016.
- [21] Araniti, G. et al., *LTE for Vehicular Networking: A Survey*, IEEE Communications Magazine, vol. 51, no. 5, pp. 148-157, May 2013.
- [22] ElBatt, T. et al., *Cooperative Collision Warning using Dedicated Short Range Wireless Communications*, International workshop on Vehicular ad hoc networks (VANET), pp. 1-9, September 2006.
- [23] Ide, C., Kurtz, F. and Wietfeld, C., *Cluster-Based Vehicular Data Collection for Efficient LTE Machine-Type Communication*, IEEE Vehicular Technology Conference (VTC Fall), September 2013.
- [24] Feng Jiang, Jie Chen, A. Lee Swindlehurst, and José A. López-Salcedo, *Massive MIMO for Wireless Sensing With a Coherent Multiple Access Channel*, IEEE Transactions on Signal Processing, vol. 63, no. 12, pp. 3005-3017, June 2015.
- [25] Amirpasha Shirazinia, Subhrakanti Dey, Domenico Ciuonzo, and Pierluigi Salvo Rossi, *Massive MIMO for Decentralized Estimation of a Correlated Source*, IEEE Transactions on Signal Processing, vol. 64, no. 10, pp. 2499-2512, May 2016.
- [26] Domenico Ciuonzo, Pierluigi Salvo Rossi, and Subhrakanti Dey, *Massive MIMO channel-aware decision fusion*, IEEE Transactions on Signal Processing, vol. 63, no. 3, pp. 604-619, February 2015.
- [27] da Costa, F., *Rethinking the Internet of Things: A Scalable Approach to Connecting Everything*, Apress, 1st edition, 2014.
- [28] Wunder, G., Jung, P. and Wang, C., *Compressive Random Access for Post-LTE Systems*, IEEE International Conference on Communications Workshops (ICC), August 2014.
- [29] Villaverde, B. C. et al., *Service Discovery Protocols for Constrained Machine-to-Machine Communications*, IEEE Communications Surveys & Tutorials, vol. 16, no. 1, pp. 41-60, November 2013.
- [30] Wunder, G., Kasparick, M. and Jung, P., *Spline Waveforms and Interference Analysis for 5G Random Access with Short Message Support*, arXiv preprint arXiv:1501.02917, January 2017.
- [31] Yinsheng Liu et al., *Waveform Design for 5G Networks: Analysis and Comparison*, IEEE Access, vol. 5, pp. 19282-19292, February 2017.
- [32] Andrews, J. G. et al., *What Will 5G Be?*, IEEE Journal on Selected Areas in Communications, vol. 32, no. 6, pp. 1065-1082, June 2014.

- [33] Aymen Omri, Mazen O. Hasna and Mohammed Nafie, *Effective area spectral efficiency for wireless communication networks with interference management*, EURASIP Journal on Wireless Communications and Networking, vol. 2015, no. 1, pp. 205-217, August 2015.
- [34] Theodore S. Rappaport, Yunchou Xing, George R. MacCartney, Jr., Andreas F. Molisch, Evangelos Mellios, and Jianhua Zhang, *Overview of Millimeter Wave Communications for Fifth-Generation (5G) Wireless Networks-With a Focus on Propagation Models*, IEEE Transactions on Antennas and Propagation, vol. 65, no. 12, December 2017.
- [35] Theodore S. Rappaport, *Spectrum frontiers: The new world of millimeter-wave mobile communication*, Invited Keynote Presentation, The Federal Communications Commission (FCC) Headquarters, 2016.
- [36] Marja Matinmikko, et al., *Overview and comparison of recent spectrum sharing approaches in regulation and research: From opportunistic unlicensed access towards licensed shared access*, IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN), April 2014.
- [37] Thomas L. Marzetta, Erik G. Larsson, Hong Yang, Hien Quoc Ngo, *Fundamentals of Massive MIMO*, Cambridge U.K., Cambridge University Press, 1st edition, November 2016.
- [38] Larsson, E. G. et al., *Massive MIMO for Next Generation Wireless Systems*, IEEE Communications Magazine, vol. 52, no. 2, pp. 186-195, February 2014.
- [39] Marzetta, T. L., *Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas*, IEEE Transactions on Wireless Communications, vol. 9, no. 11, pp. 3590-3600, November 2010.
- [40] Emil Bjornson, Erik G. Larsson and Thomas L. Marzetta, *Massive MIMO: ten myths and one critical question*, IEEE Communications Magazine, vol. 54, no. 2, pp. 114-123, February 2016.
- [41] Emil Bjornson, Jakob Hoydis and Luca Sanguinetti, *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency*, Foundations and Trends in Signal Processing, vol. 11, no. 3-4, pp. 154-655. DOI: 10.1561/20000000093, November 2017.
- [42] Olakunle Elijah, Chee Yen Leow, Tharek Abdul Rahman, Solomon Nunoo, and Solomon Zakwoi Iliya, *A Comprehensive Survey of Pilot Contamination in Massive MIMO-5G System*, IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 905-923, November 2015.
- [43] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin and R. Zhang, *An Overview of Massive MIMO: Benefits and Challenges*, IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 5, pp. 742-758, October 2014.
- [44] Joao Vieira, Fredrik Rusek, Ove Edfors, Steffen Malkowsky, Liang Liu, and Fredrik Tufvesson, *Reciprocity Calibration for Massive MIMO: Proposal, Modeling, and Validation*, IEEE Transactions on Wireless Communications, vol. 16, no. 5, pp. 3042-3056, May 2017.
- [45] Kamil Senel, and Erik G. Larsson, *Grant-Free Massive MTC-Enabled Massive MIMO: A Compressive Sensing Approach*, arXiv preprint arXiv:1806.10061, June 2018.
- [46] Andres Laya, Luis Alonso, and Jesus Alonso-Zarate, *Is the Random Access Channel of LTE and LTE-A Suitable for M2M Communications? A Survey of Alternatives*, IEEE Communications Surveys & Tutorials, vol. 16, no. 1, pp. 4-16, First Quarter 2014.
- [47] Stefania Sesia, Issam Toufik and Matthew Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*, John Wiley & Sons, ISBN:9780470697160, February 2009.
- [48] Felipe A. P. de Figueiredo, Fabbryccio A. C. M. Cardoso, Renato R. Lopes and Joao Paulo Miranda, *On the Application of Massive MU-MIMO in the Uplink of Machine Type Communication Systems*, International Workshop on Telecommunications (IWT), June 2015.
- [49] Rusek, F. et al., *Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays*, IEEE Signal Processing Magazine, vol. 30, no. 1, pp. 40-60, January 2013.
- [50] B. Hochwald, T. Marzetta, and V. Tarokh, *Multiple-antenna channel hardening and its implications for rate feedback and scheduling*, IEEE Transactions on Information Theory, vol. 50, no. 9, pp. 1893-1909, September 2004.
- [51] Monowar Hasan, Ekram Hossain, and Dusit Niyato, *Random access for machine-to-machine communication in LTE-advanced networks: issues and approaches*, IEEE Communications Magazine, vol. 51, no. 6, pp. 86-93, June 2013.
- [52] Md Shipon Ali, Ekram Hossain, and Dong In Kim, *LTE/LTE-A Random Access for Massive Machine-Type Communications in Smart Cities*, IEEE Communications Magazine, vol. 55, no. 1, pp. 76-83, January 2017.
- [53] Han Seung Jang, Su Min Kim, Kab Seok Ko, Jiyoung Cha, and Dan Keun Sung, *Spatial Group Based Random Access for M2M Communications*, IEEE Communications Letters, vol. 18, no. 6, pp. 961-964, June 2014.
- [54] Shree Krishna Sharma, and Xianbin Wang, *Towards Massive Machine Type Communications in Ultra-Dense Cellular IoT Networks: Current Issues and Machine Learning-Assisted Solutions*, arXiv:1808.02924, August 2018.
- [55] Tarik Taleb, Andreas Kunz, *Machine type communications in 3GPP networks: potential, challenges, and solutions*, IEEE Communications Magazine, vol. 50, no. 3, March 2012.
- [56] Liang Liu, and Wei Yu, *Massive Connectivity With Massive MIMO - Part I: Device Activity Detection and Channel Estimation*, IEEE Transactions on Signal Processing, vol. 66, no. 11, pp. 2933-2946, June 2018.
- [57] Kamil Senel, Emil Bjornson, Erik G Larsson, *Human and Machine Type Communications can Coexist in Uplink Massive MIMO Systems*, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6613-6617, April 2018.
- [58] Rasha Al Khansa, Jean J. Saade, Hassan A. Artail, and Mohamad Asaad, *A small cell approach to optimizing the coverage of MTC systems with massive MIMO and random access using stochastic geometry*, IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), October 2017.
- [59] 3GPP, *TS 22.368 V10.1.0 - Service Requirements for Machine-Type Communications*, June 2010.
- [60] Lien, S. -Y., Chen, K. -C. and Lin, Y., *Toward Ubiquitous Massive Accesses in 3GPP Machine-to-machine Communications*, IEEE Communications Magazine, vol. 49, no. 4, pp. 66-74, April 2011.
- [61] Cheng, M.-Y. et al., *Overload control for machine-type-communications in LTE-advanced system*, IEEE Communications Magazine, vol. 50, no. 6, pp. 38-45, June 2012.
- [62] Gerasimenko, M. et al., *Energy and delay analysis of LTE-advanced RACH performance under MTC overload*, IEEE Globecom Workshops, March 2013.
- [63] Phuyal, U. et al., *Controlling access overload and signaling congestion in M2M networks*, Asilomar Conference on Signals, Systems and Computers (ASILOMAR), March 2013.
- [64] Felipe A. P. de Figueiredo, et al., *Multi-stage Based Cross-Correlation Peak Detection for LTE Random Access Preambles*, Revista Telecomunicações, vol. 15, no. 2, pp. 1-7, 2013.
- [65] Felipe A. P. de Figueiredo, Fabbryccio A. C. M. Cardoso, Ingrid Moerman and Gustavo Fraidenraich, *Channel estimation for massive MIMO TDD systems assuming pilot contamination and flat fading*, EURASIP Journal on Wireless Communications and Networking, vol. 2018, no. 14, pp. 1-10, January 2018.
- [66] R. L. Frank, S. A. Zadoff and R. Heimiller, *Phase shift pulse codes with good periodic correlation properties*, IRE Transactions on Information Theory, vol. 8, no. 6, pp. 381-382, October 1962.
- [67] David C. Chu, *Polyphase codes with good periodic correlation properties*, IEEE Transactions on Information Theory, vol. 18, no. 4, pp. 531-532, July 1972.
- [68] Wallis, J. S., *On the existence of Hadamard matrices*, Journal of Combinatorial Theory, vol. 21, no. 2, pp. 188-195, September 1976.
- [69] Shendi Wang, John S. Thompson, and Peter M. Grant, *Closed-Form Expressions for ICI/ISI in Filtered OFDM Systems for Asynchronous 5G Uplink*, IEEE Transactions on Communications, vol. 65, no. 11, pp. 4886-4898, November 2017.
- [70] Abdoli J., Jia M., and Ma J., *Filtered OFDM: A New Waveform for Future Wireless Systems*, IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), July 2015.
- [71] A. Ashikhmin, T. L. Marzetta, L. Li, *Interference reduction in multi-cell massive MIMO systems I: Large-scale fading precoding and decoding*, arXiv preprint arXiv:1411.4182, November 2014.
- [72] Jinho Choi, *Massive MIMO With Joint Power Control*, IEEE Wireless Communications Letters, vol. 3, no. 4, August 2014.
- [73] Hong Yang, Thomas L. Marzetta, *Massive MIMO With Max-Min Power Control in Line-of-Sight Propagation Environment*, IEEE Transactions on Communications, vol. 65, no. 11, pp. 4685-4693, November 2017.
- [74] Ngo, H. Q., Larsson, E. G. and Marzetta, T. L., *Energy and spectral efficiency of very large multiuser MIMO systems*, IEEE Transactions on Communications, vol. 61, no. 4, pp. 1436-1449, April 2013.
- [75] Hien Quoc Ngo, Erik G. Larsson, and Thomas L. Marzetta, *Aspects of favorable propagation in Massive MIMO*, European Signal Processing Conference (EUSIPCO), pp. 76-80, March 2014.
- [76] Matthaiou, M. et al., *On the condition number distribution of complex wishart matrices*, IEEE Transactions on Communications, vol. 58, no. 6, pp. 1705-1717, June 2010.
- [77] Barry, J. R., Lee, E. A. and Messerschmitt, D. G., *Digital Communication*, Springer, 3rd edition, 2004

- [78] Hien Quoc Ngo, *Massive MIMO: Fundamentals and System Designs*, Linköping Studies in Science and Technology. Dissertations, ISSN 0345-7524, March 2015.
- [79] Sergio Verdu, and Sergio Verd, *Multiuser Detection*, Cambridge University Press, 1st edition, August 1998.
- [80] R. C. de Lamare, *Massive MIMO Systems: Signal Processing Challenges and Research Trends*, arXiv preprint arXiv:1310.7282, October 2013.
- [81] E. G. Larsson, and H. V. Poor, *Joint beamforming and broadcasting in massive MIMO*, IEEE Transactions on Wireless Communications, vol. 15, no. 4, pp. 3058-3070, April 2016.
- [82] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, New York, NY, USA: Pearson Education, 1993.
- [83] Pallaviram Sure, and Chandra Mohan Bhuma, *A survey on OFDM channel estimation techniques based on denoising strategies*, Engineering Science and Technology, an International Journal, vol. 20, no. 2, pp. 629-636, April 2017.
- [84] Edmund Landau, *Handbuch der Lehre von der Verteilung der Primzahlen*, AMS Chelsea Publishing, 1909.
- [85] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, *Pilot contamination and precoding in multi-cell TDD systems*, IEEE Transactions on Wireless Communications, vol. 10, no. 8, pp. 2640-2651, August 2011.
- [86] J. Hoydis, S. ten Brink, and M. Debbah, *Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?*, IEEE Journal on Selected Areas in Communications, vol. 31, no. 2, pp. 160-171, February 2013.
- [87] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, *The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel*, IEEE Transactions on Communications, vol. 61, no. 6, pp. 2350-2361, June 2013.
- [88] Theodore S. Rappaport, *Wireless Communications Principles and Practice*, 2nd edition, Pearson Education Inc, 2002.
- [89] S. R. Saunders, and A. Aragon-Zavala, *Antennas and Propagation for Wireless Communication Systems*, John Wiley & Sons, Chichester, UK, 2nd edition, 2007.
- [90] Zheng Chen and Emil Björnson, *Channel Hardening and Favorable Propagation in Cell-Free Massive MIMO with Stochastic Geometry*, arXiv preprint arXiv:1710.00395, October 2017.
- [91] Hien Quoc Ngo, and Erik G. Larsson, *No Downlink Pilots are Needed in TDD Massive MIMO*, IEEE Transactions on Wireless Communications, vol. 16, no. 5, pp. 2921-2935, May 2017.
- [92] N. Jindal, *MIMO broadcast channels with finite-rate feedback*, IEEE Transactions on Information Theory, vol. 52, no. 11, pp. 5045-5060, November 2006.
- [93] A. Jeffrey and D. Zwillinger, *Table of Integrals, Series, and Products*, New York, NY, USA: Academic, 2007.