Aggregation and Resource Scheduling in Machine-type Communication Networks: A Stochastic Geometry Approach

Onel L. Alcaraz López, Student Member, IEEE, Hirley Alves, Member, IEEE, Pedro H. J. Nardelli, Matti Latva-aho, Senior Member, IEEE

Abstract—Data aggregation is a promising approach to enable massive machine-type communication (mMTC). This paper focuses on the aggregation phase where a massive number of machine-type devices (MTDs) transmit to aggregators. By using non-orthogonal multiple access (NOMA) principles, we allow several MTDs to share the same orthogonal channel in our proposed hybrid access scheme. We develop an analytical framework based on stochastic geometry to investigate the system performance in terms of average success probability and average number of simultaneously served MTDs, under imperfect successive interference cancellation (SIC) at the aggregators, for two scheduling schemes: random resource scheduling (RRS) and channel-aware resource scheduling (CRS). We identify the power constraints on the MTDs sharing the same channel to attain a fair coexistence with purely orthogonal multiple access (OMA) setups. Then, power control coefficients are found, so that these MTDs perform with similar reliability. We show that under high access demand, the hybrid scheme with CRS outperforms the OMA setup by simultaneously serving more MTDs with reduced power consumption.

I. INTRODUCTION

Machine-type Communication (MTC) is typically cited as an integral use case for fifth generation (5G) cellular networks, where fully automatic data generation, exchange, processing and actuation among intelligent machines, are on the agenda. With the rapid penetration of embedded devices, MTC is becoming the dominant communication paradigm for a wide range of emerging smart services including healthcare, manufacturing, utilities, consumer goods and transportation. Specifically, massive MTC (mMTC), as the name suggests, is about massive access by a large number of devices, that is, about providing wireless connectivity to enormous number of often low-complexity low-power machine-type devices (MTDs). Of course, the access is a major concern and a flourishing research area. Different methods have been proposed and addressed to provide more efficient access, e.g., access class barring [1], prioritized random access [2], backoff adjustment scheme [3], delay-estimation based random access [4], distributed queuing [5]. Another promising way to deal with the massive connection problem comes from the concept of data aggregation. Instead of being directly connected to the access network, e.g., by equipping MTDs with own Subscriber Identity Module (SIM) card to have cellular connectivity, the idea is that MTDs may organize themselves locally, creating MTC area networks and exploiting short-range technologies. These MTC area networks may then connect to the core networks through MTC gateways or data aggregators. This is a key solution strategy to collect, process, and communicate data in MTC use cases with static devices, especially if the locations of the devices are known, such as smart utility meters [6] or video surveillance cameras [7]. The traffic from MTDs is first transmitted to the designed data aggregator, while the aggregator then relays the collected packets to the core network [8], e.g., the base station (BS), thus, reducing the congestion and the power consumption at the MTD side [9].

When aggregating such huge number of MTDs, the density of the aggregators, although considerable smaller than the density of the MTDs, obviously will still be large, and the interference coming from the devices sharing the same resource could be significant. There is limited literature that characterizes the interference in mMTC with data aggregation. The analysis in [10] shows how the wireless channel, transmit power, and random deployment of data collectors affect the signal-to-interference-plus-noise ratio (SINR) distribution in random access networks with randomly deployed sensor nodes, and for some special cases, a simple form of signal-to-noise ratio (SNR) or SINR distribution is found. An energy-efficient data aggregation scheme for a hierarchical MTC network is proposed in [11], while authors develop a coverage probability-based optimal data aggregation scheme for MTDs to minimize the energy density of the network. In [12], a theoretical and numerical framework, which aims to assess, model and characterize the network energy consumption profile, is presented by exploiting the stochastic geometry tools. By incorporating some intelligence in the aggregator, the network performance improves, as shown in [9], [13]–[16] for resource scheduling strategies. Among them, only [9] considers a more realistic scenario with a multi-cell network, hence the inter-cell interference, which is a critical issue, is taken into account. The authors introduce a tractable two-phase network model for mMTC, where MTDs first transmit to their serving aggregators (aggregation phase) and then, the aggregated data
is delivered to BSs (relaying phase). Key metrics such as the MTD success probability, average number of successful MTDs and probability of successful channel utilization are investigated for two resource scheduling schemes RRS, where aggregators randomly allocate the limited resources to MTDs, and CRS, where aggregators allocate resources to the MTDs having better channel conditions. Authors show that, compared to the CRS scheme, the RRS scheme can achieve similar performance as long as the resources in the aggregation phase are not very limited.

In its turn, non-orthogonal multiple access (NOMA) has recently attracted a lot of attention as a promising technology for the coming 5G networks to significantly improve the spectral efficiency of mobile communication networks, and/or to meet the demand of massive connectivity. Authors in [17] present NOMA as a promising solution to support a massive number of MTDs in cellular networks, while tackling the main practical challenges and future research directions. Actually, NOMA is considered in [18], [19] when serving a massive number of devices in cellular-based MTC networks. Energy-efficient clustering and medium access control are investigated in [18] to minimize device energy consumption and prolong network battery lifetime, while authors in [19] study the throughput and energy efficiency of the NOMA scenario with a random packet arrival model and derive the stability condition for the system to guarantee the performance. The key idea of NOMA is to exploit the power domain for multiple access, such that multiple users can be multiplexed at different power levels but at the same time/frequency/code. SIC is utilized to separate superimposed messages at the receiver side [20]. In general NOMA can be applied on both, downlink and uplink. A downlink NOMA transmission system is studied in [21], where the authors show that the outage performance of NOMA depends critically on the choices of the targeted data rates and allocated power in scenarios with randomly deployed users; while a dynamic power allocation scheme is proposed in [22] to downlink and uplink NOMA scenarios with two users with various quality of service requirements. A NOMA scheme for uplink that allows more than one user to share the same channel in order to attain a fair coexistence of our scheme with purely OMA setups, while power control coefficients are found too, so these MTDs can perform with similar reliability.

We show that, even when the hybrid scheme could lead to a less reliable system with greater chances of outages per MTD, e.g., due to the additional intra-cluster interference, the number of simultaneous active MTDs could be significantly improved for high access demand scenarios, as long as the success probability does not decrease so much. In that sense, our scheme aims at providing massive connectivity in scenarios with high access demand, which is not covered by traditional OMA setups. Additionally, the CRS scheme requires even a lower average power consumption per orthogonal channel and per MTD, than the OMA setup.

We attain approximated, yet accurate, expressions when analyzing the CRS scheme. Compared to the time-consuming Monte-Carlo simulations, even heavier for our hybrid scheme than for the purely OMA setup, our analytical derivations allow for fast computation. Also, imperfection when implementing SIC is incorporated in our proposed analytical framework.

The remainder of the paper is organized as follows. Section II presents the system model and assumptions. Sections III and IV discuss the RRS and CRS scheduling schemes for our hybrid access protocol, respectively. Section V studies the power consumption of both schemes for OMA and the hybrid protocol. Section VI presents the numerical results, and finally Section VII concludes the paper.
Notation: $E[ \cdot ]$ denotes expectation, while $Pr(A)$ and $Pr(A|B)$ are the probability of event $A$, and $Pr(A)$ conditioned on $B$, respectively. $|S|$ is the cardinality of set $S$, $||x||$ is the Euclidean norm of vector $x$ and $\text{mod}(a,b)$ is the modulo operation. $F_r$ is the generalized hyper-geometric function [28, Eq. (16.2.1)], $\psi(x)$ is the digamma function [29, Eq. (6.3.1)], $\Gamma(x)$ is the gamma function, while $Q(a,x)$ and $\text{Beta}(a,x,b)$ are the regularized incomplete gamma function [28, Eq. (8.2.4)] and incomplete beta function [29, Eq. (6.2.1)], respectively. $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ are the floor and ceiling functions, respectively. $i = \sqrt{-1}$ and $\text{Im}(z)$ is the imaginary part of $z \in \mathbb{C}$. $f_X(x)$ and $F_X(x)$ are the Probability Density Function (PDF) and Cumulative Distribution Function (CDF) of random variable (RV) $X$, respectively. $X \sim \text{Exp}(1)$ is an exponential distributed RV with unit mean, e.g., $f_X(x) = \exp(-x)$ and $F_X(x) = 1 - \exp(-x)$; while $Y \sim \text{Pois}(\bar{m})$ is a Poisson distributed RV with mean $\bar{m}$, e.g., $Pr(Y = y) = \frac{1}{y!} \bar{m}^y \exp(-\bar{m})$ and $F_Y(y) = Q(y+1, \bar{m})$.

II. SYSTEM MODEL AND ASSUMPTIONS

Consider a cellular network overlaid with data aggregators, which are spatially distributed according to a homogeneous Poisson point process (PPP), denoted $\Phi_\text{w}$ with density $\lambda_\text{w}$. Each aggregator, which could function as an ordinary cellular user for certain BS, serves multiple MTDs located nearby. Thus, the result is a cluster point process uniquely defined as:

$$\Phi = \bigcup_{w \in \Phi_\text{w}} w + B_w, \quad (1)$$

where $\Phi_\text{w}$ is the parent PPP and $B_w$ denotes the offspring process where each point at $s \in B_w$ is i.i.d. around the cluster center $w \in \Phi_\text{w}$ with distance distribution $f_{\text{off}}(r_w) = \frac{2\pi r_w}{\pi} e^{-r_w^2}$, e.g., uniformly distributed in the disk region of radius $R_w$. $K \sim \text{Pois}(\bar{m})$ is the instantaneous number of MTDs requiring service in each aggregator, e.g., number of points in $B_w$, thus, the process is a Matérn cluster point process. Notice, that each MTD is associated with a single aggregator even when it could be located within the aggregation areas of several aggregators.

We focus on the uplink where the MTDs across the entire network are served through the same set of orthogonal channels, $N$, available at each aggregator, with $|N| = N$. Differently from [9], here the same orthogonal channel could be used for more than one MTD. Some aggregators could be allocating one MTD per channel because the access demand is not so high, but some other aggregators could be allocating more MTDs per channel to face the increasing access demand, thus we propose and assess a hybrid OMA-NOMA multiple access scenario. Notice that the same aggregator could have channels operating with only one MTD, while others are operating with more. The maximum number of users per orthogonal channel is $L$, where $L = 1$ reduces to the OMA scenario analyzed in [9], which is used here as a benchmark. We focus on the

---

2This is for analytical tractability, but notice that even when some existing results show that NOMA with more devices may provide a better performance gain [26], this may not be practical. The reason is that considering processing complexity for SIC receivers, especially when SIC error propagation is considered, 2-users NOMA is actually more practical in reality [27], [30].

3There are some solutions using NOMA for the random access stage as well, e.g., [17], [32]. In fact, the work in [32] proposes a NOMA scheme where the devices transmit their messages over randomly selected channels, while the random access and data transmissions phases are combined. Readers will realize that our resource scheduling schemes can be easily incorporated to that strategy to improve the overall system performance; however, the details of such implementation are out of the scope of this work.
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2018.2830767, IEEE Transactions on Wireless Communications

III. RRS FOR THE HYBRID ACCESS

Herein we explore the RRS scheduling scheme, for which the CSI is only required at the aggregators when decoding the information and not for resource scheduling. Due to its simplicity, this scheme is used as benchmark when compared to a more evolved strategy discussed in Section IV, where the benefits of the CSI acquisition are exploited also for resource allocation. In Subsection III-A we attain the success probability for each MTD in a given channel, which depends on the Laplace transform of the inter-cluster interference. An accurate approximation of the latter is given, which is fundamental to efficiently evaluate the success probability. In Subsection III-B we characterize the overall system performance.

Under the RRS scheme, \( N \) out of the \( K \) instantaneous MTDs requiring transmissions are independently and randomly, chosen and matched, one-to-one, with the channels in \( \Phi \). If \( K < N \), all MTDs get channel resources, and even \( N - K \) channels will be unused. Otherwise, if \( K > N \), the channel allocation is executed again by allowing the remaining MTDs to share channels with the already served MTDs. This process is executed repeatedly until all the MTDs are allocated or the maximum number of MTDs per channel, \( L \), is reached for all the channels.

**Lemma 1.** The Probability Mass Function (PMF) of the number of MTDs allocated to the same channel is given in (2) at the top of the next page.

**Proof.** See Appendix A.

The point process of the active MTDs on certain channel \( n \in \mathcal{N} \) is obviously a subset of \( \Phi \), defined in (1), and can be defined as

\[
\Phi_n = \bigcup_{w \in \Phi_n} w^N,
\]

where \( \Phi_n^N \) is the offspring point process with instantaneous number of points around the cluster center \( w \) in \( \Phi_n \) obeying (2). Thus, the generating function of the number of active MTDs on certain channel in one cluster is

\[
G_0(z) = \sum_{u=0}^{L} c_u z^u,
\]

where \( c_u = \Pr(U = u) \), while \( c = \sum_{u=1}^{L} u c_u \) is the mean number of active MTDs in each channel of certain cluster.

A. MTD success probability for \( L = 2 \)

The intra-cluster interference coming from the MTDs sharing the same channel is faced with SIC\(^4\). Thus, the SIR\(^5\), of the \( j \)th MTD being decoded on a typical channel of certain cluster, given the number of MTDs \( u \) sharing the same channel and the RRS scheme is

\[
\text{SIR}_{1,1} = \frac{h}{I_r}, \quad \text{SIR}_{1,2} = \max(h', h''), \quad \text{SIR}_{2,2} = \min(h', h''), \quad (5)
\]

where \( I_r = \sum_{x \in \Phi_n^U} g_{w,x} r_{x}^{\frac{r}{\alpha}} \) is the inter-cluster interference for the RRS scheme with \( x \in \Phi_n^U \subset \Phi_n \), denoting both the location and the interfering MTD which occupies certain channel \( n \) in other clusters, and \( r_{x} \) is the distance between the MTD and its serving aggregator. Also, \( \mu \in [0, 1] \) is used to model the impact caused by imperfect SIC \cite{34}, while \( h' \) and \( h'' \) are the channel power coefficients of both MTDs sharing the channel when \( u = 2 \). Notice that we cannot weight the power of the coexistent nodes on the same channel since CSI information is not exploited for resource scheduling when using the RRS scheme. Also, \( \lim_{u \to 0} \text{SIR}_{1,2}^u \) is unbounded, but \( \lim_{u \to 0} \text{SIR}_{2,2}^u \leq \frac{1}{\mu} \), since \( \min(h', h'') \leq \max(h', h'') \), thus the performance of the second MTDs being decoded on the channel is strongly limited by the SIC imperfection parameter.

Assuming a fixed rate coding scheme where the receiver decodes successfully whenever \( \Pr(\text{SIR} \geq \theta) \), where \( \theta \) is the SIR threshold, e.g., information rate of \( \log_2(1 + \theta) \) (bits/symbol), we state the following theorem.

**Theorem 1.** The RRS success probability, \( p_{1,u}^r \), of the \( j \)th MTD sharing a typical channel conditioned on \( u \) MTDs, is given by

\[
p_{1,1}^r = \mathcal{L}_{I_r}(\theta),
\]

\[
p_{1,2}^r = \begin{cases} \frac{1}{1+\theta} \mathcal{L}_{I_r}(\theta), & \text{if } 0 \leq \theta < 1, \\ \frac{2}{1+\theta} \mathcal{L}_{I_r}(\theta), & \text{if } \theta \geq 1, \end{cases}
\]

\[
p_{2,2}^r = \begin{cases} \frac{1}{1+\mu} \mathcal{L}_{I_r}(\theta), & \text{if } 0 \leq \theta \mu < 1, \\ \frac{2}{1+\mu} \mathcal{L}_{I_r}(\theta), & \text{if } \theta \mu \geq 1, \end{cases}
\]

where

\[
\mathcal{L}_{I_r}(s) = \exp \left( -2 \pi \lambda_a \int_0^{\infty} r_{w} \sum_{u=0}^{L} c_u \mathcal{T}(r_{w}, s)^u - 1 \right) dr_{w},
\]

is the Laplace transform of RT \( I_r \) and

\[
\mathcal{T}(r_{w}, s) = \frac{r_{w}^{2s}}{\pi R_0^{2s}} \int_0^{\infty} \int_0^{\infty} \frac{r_{a}^{2s} \exp(-r_{a})}{1 + sr_{w}^{2}(r_{w}^2 + r_{a}^2 + 2r_{w}r_{a} \cos(\omega))} dr_{a} d\omega.
\]

**Proof.** See Appendix B.

**Remark 1.** Although for some pairs \((\theta, \mu)\) it is possible to achieve a similar reliability of both MTDs sharing the same channel, e.g., \( p_{1,2}^r \approx p_{2,2}^r \), this is not attained in general, and it could be a main drawback when using the RRS scheme and certain homogeneity in the Quality of Service (QoS) is expected. Another interesting fact is that \( p_{1,1}^r = p_{1,2}^r = \mathcal{L}_{I_r}(\theta) \) for \( \theta = 1 \), while \( p_{2,2}^r \) would be smaller even when \( \mu \to 0 \), e.g., \( \mathcal{L}_{I_r}(2\theta) \). Additionally, notice that \( \mu < 1/\theta \) is required such
that the second user being decoded on certain channel has chances to succeed. Thus, the greater the SIR threshold, $\theta$, the greater the impact of imperfect SIC, $\mu$.

**Theorem 2.** Expression (11) for $L \geq 2$ is upper (almost surely) and lower bounded by

$$
L_{L_r}(s) \leq \frac{\beta_0 L_{up}(s)}{\beta_1 L_{seq}(s)},
$$

where $\beta = \frac{1}{2} \lambda_s \pi R^2 \Gamma(1 + \frac{2}{\alpha}) \Gamma(1 - \frac{1}{\alpha})$, while

$$
L_{L_r}(s) \approx \beta_0 L_{up}(s) + \beta_1 L_{seq}(s),
$$

provides an approximation with $\beta_0, \beta_1 \in [0,1]$ and $\beta_0 + \beta_1 = 1$.

**Proof.** See Appendix C for derivation of (13) and (14), while (15) is a weighted average of both bounds, thus attaining an approximation that will be more accurate than at least one of the bounds. The weighted average becomes relevant if we know at a priori which of the bounds in (13) and (14) is more accurate for the setup being analyzed and we give it a greater weight, e.g., $\beta_0 \leq \beta_1$, otherwise trivial choice $\beta_0 = \beta_1 = 0.5$ is advisable, as we adopt here.

Fig. 2 shows the bounds and approximation attained in Theorem 2 while comparing them with the exact value given in (11). Specifically, Fig. 2(a) shows the performance for $\alpha \in [3, 5]$, while we set $c_2 = 1$, which conduces to scenarios where the bounds in (13) and (14) are the least tight since $T_1$ in (50) has no impact on (11), e.g., $T_1 = 1$. As noticed, the greater the $\alpha$ the lesser the tightness of the bounds, however the approximation performs very well whatever the setup, which is also appreciated in Fig. 2(b). Notice that for relatively small $N$, the approximation in (15) is accurate even when $\lambda_s$ increases and the bounds are not tight. When $N$ increases, $c_0$ and $c_1$ increase, which increases the contribution of $T_1$ in (50), thus increasing the tightness of the bounds, and the accuracy of the approximation.

**B. Overall Performance for $L = 2$**

The following two lemmas characterize the system performance in terms of average over all MTD success probability, and average number of simultaneously served MTDs since one of the main benefits of NOMA techniques is the possibility of offering service to a great number of devices simultaneously.

**Lemma 2.** Conditioned on being using certain channel, the average over all MTD success probability is

$$
Pr(U = u) = \begin{cases} 
Q(N, \bar{m}) \left(1 - \frac{\bar{m}}{N}\right) + \frac{\exp(-\bar{m})\bar{m}^N}{N!} \sum_{t \in \{0,1\}} 2^{1-t} \left(Q\left(N(u + t), \bar{m}\right)(t + (-1)^t(u - t)) + (-1)^{1-t} \frac{\exp(-\bar{m})\bar{m}^N}{N\left(N(u + t) - 1\right)!}\right), & u = 0; \\
1 + \sum_{t \in \{0,1\}} Q\left(N\left(L - t\right), \bar{m}\right)((-1)^t \frac{(N - L) - t}{N} - t) + (-1)^{1-t} \frac{\exp(-\bar{m})\bar{m}^N}{N\left(N(L - t) - 1\right)!}, & u = L \\
0, & \text{otherwise}
\end{cases}
$$

![Fig. 2. Comparing the exact $L_{L_r}(s)$ (11) as a function of (a) $s$ in dB, for $c_0 = c_1 = 0$, $c_2 = 1$, $\lambda_s = 10^{-4.2}$ (left), and (b) $N$ for $m = 60$, $\alpha = 3.6$, $\lambda_s \in \{10^{-4.2}, 10^{-4.8}\}$ and $s = 0$ dB (right), with the bounds and approximation given in Theorem 2. We also set $R_m = 40m$.](image-url)

**Lemma 3.** The average number of simultaneously served MTDs is given in (17) at the top of the next page, where

$$
A_1 = \bar{m}Q(N + 1, \bar{m}) - \frac{\exp(-\bar{m})\bar{m}^{N+1}}{N!},
$$

$$
A_2 = \exp(-\bar{m})\left(\frac{\bar{m}2^{N-1}}{(2N - 1)!} - \frac{\bar{m}^{N+1}}{N!}\right),
$$

$$
A_3 = N\left(Q\left(N + 1, \bar{m}\right) + 1\right)
$$

**Proof.** Knowing that $c_u = Pr(U = u)$ it is straightforward to attain (16).

**Lemma 2.** Conditioned on being using certain channel, the average over all MTD success probability is

$$
P_{succe} = \frac{c_1}{1 - c_0} p_{1,1}^u + \frac{c_2}{2(1 - c_0)} (p_{1,2}^u + p_{2,1}^u) .
$$

**Proof.** In (17), (a) comes from averaging the number of simultaneously active and successful MTDs, while (b) comes from regrouping terms and using the expressions of the CDF and expected $K$ on one interval (see Appendix A) along with some algebraic transformations.
Remark 2. When \( \bar{m} \) grows above 2\( N \), the last term in equality (a) of (17) becomes the largest contributor for \( \bar{K}_r \), with \( \bar{K}_r \rightarrow N(p_{1,2}^r + p_{2,2}^r) = 2Np_{\text{oma}}^r \) since \( \sum_{k=2N}^{\infty} \text{Pr}(K = k) \approx 1 \) and \( c_2 \approx 1 \) (\( c_0 \approx c_1 \approx 0 \)) in (16). Thus, when comparing with an \( L = 1 \) setup under the same circumstances, e.g., with average number of simultaneously served MTDs being \( Np_{\text{oma}}^r \), the condition required for the hybrid access scheme to overcome the OMA system \( (L = 1) \) is that \( p_{\text{succ}}^r > \frac{1}{2}p_{\text{oma}}^r \). Notice that the hybrid scheme with RRS could lead to a less reliable system with greater chances of outages per MTD, e.g., due to a larger inter-cluster interference and additional intra-cluster interference. However, the number of simultaneous active MTDs could be improved for high access demand scenarios, e.g., \( \bar{m} \sim 2N \), as long as the success probability does not decrease so much.

IV. CRS FOR THE HYBRID ACCESS

Herein we explore the CRS scheduling scheme, which contrary to the RRS scheme, strongly relies on the CSI for resource scheduling. Thus, the CRS scheme is more adjusted to NOMA scenarios, where CSI is a keystone for efficiently decoding multiple user data over the same orthogonal channel with SIC [21], [23], [26]. Similar than in the previous section, we attain first in Subsection IV-A the success probability for each MTD in a given channel along with necessary approximations. Later, in Subsection IV-B, we find the power constraints on the MTDs sharing the same channel in order to attain a fair coexistence of our scheme with purely OMA setups, while power control coefficients are found too, so these MTDs can perform with similar reliability. Finally, in Subsection IV-B we characterize the overall system performance.

Under the CRS scheme, the MTD with better fading (equivalently, better SIR) will be preferentially assigned with the available channel resources. An aggregator with \( K \) instanta- neously MTDs requiring transmission has the knowledge of their fading gains. Let \( \{h_1, \ldots, h_{i-1}, h_i\} \) denote the decreasing ordered channel gains, where \( h_{i-1} > h_i \). If \( K \leq N \) all the MTDs will be chosen, but if \( K > N \) the aggregator will pick the \( N \) MTDs with better channel gains, i.e., \( h_1, \ldots, h_N \), and then will assign the channel set \( \mathcal{N} \) to them [9]. As a continuation, the remaining MTDs can be still allocated sharing those same resources, i.e., users \( N + 1, \ldots, K \) go to the second round for allocation. This process is executed repeatedly until all the MTDs are allocated or the maximum number of MTDs per channel, \( L \), is reached. Both Lemma 1 and the point process of the active MTDs on certain channel \( n \in \mathcal{N} \) characterized through (3) and (4) hold here.

Notice that differently from the RRS scheme, here we can weight the power of coexistent nodes on the same channel through \( a_i \) and \( b_i \) since CSI is used for resource allocation. Of course, some kind of feedback from the aggregators would be required. Once again, the performance of the first MTD being deduced in the channel is somewhat unbounded, e.g., \( \lim_{L \rightarrow 0} \text{SIR}^{(i)}_{2,2} \) is unbounded, while the performance of the second one is not, but this time we can relax that situation by choosing \( a_i < b_i \) since \( \lim_{L \rightarrow 0} \text{SIR}^{(i)}_{2,2} < \frac{b_i}{\mu a_i} \). However, weighting the transmit power of the MTDs sharing the same channel conduces to a marked point process with marks \( m_x = 1 \) when the MTD on \( x \) is alone in the channel, which happens with probability \( c_1 \), and \( m_x \in \{a_i, b_i\} \) for those MTDs sharing the same channel, which happens with probability \( c_2 \). Thus, the inter-cluster interference for the CRS scheme is \( I_c = \sum_{x \in \mathcal{F}_u} g_m x r_m^\alpha \). By letting \( a_i + b_i = \delta \) be a fixed value we impose some kind of total transmission power constraint [22]. This is crucial for NOMA scenarios, and here it is particular important in order to control the interference from the inter/intra-cluster MTDs sharing the same channel. Now, assuming that the receiver can decode successfully (SIR exceeds a threshold \( \theta \)), we state the following theorem.

**Theorem 3.** The CRS success probability, \( p_{\text{succ}}^{(i)} \), of the \( j \)th MTD being decoded on a typical channel, given the first MTD allocated there has the \( i \)th larger channel coefficient, \( h_i \), and there are \( u \) MTDs sharing that same channel, is approximately given by

\[
p_{\text{succ}}^{(i)} \approx \frac{1}{2} - \frac{1}{\pi i} \int_0^\infty \text{Im} \left( \mathcal{L}(\theta - i\varphi) \exp(-i\varphi B/(b_i^j, K)) \right) d\varphi, \tag{24}
\]
where
\[ L_{r}(s) = \exp \left( 2\pi \lambda_a \int_{0}^{\infty} \int_{0}^{\infty} c_0 + c_1 T(r_{w}, s) + c_2 \psi \left( r_{w}, a_i s \right) \left( \delta - a_i s \right) - 1 \right) f_{a_i}(s) \, \mathrm{d}r_{w} \, \mathrm{d}s \],
\[ \approx \sum_{t \in \{1, 2\}} \beta_t - 1 \exp \left( -\left( c_1 + c_2 a_i \right) s^2 \right), \]

is the Laplace transform of \( RV I_c \) for \( L = 2 \). \( \psi(r_{w}, s) \) is defined in (12), and \( \beta_0, \beta_1 \in [0, 1] \), \( \beta_0 + \beta_1 = 1 \), such that (26) with \( \beta_0 = 1 \) and \( \beta_1 = 1 \) provide, almost surely, upper and lower bounds for (25), respectively. Finally,
\[ B_{1,1}^{(i,K)} = \frac{\psi(K + 1) - \psi(i)}{\theta}, \]
\[ B_{1,2}^{(i,K)} = \left( a_i \theta - b_i \right) \psi(K + 1) + b_i \psi(i + N) - \frac{a_i}{\theta} \psi(i), \]
\[ B_{2,2}^{(i,K)} = \left( b_i \theta - \mu a_i \right) \psi(K + 1) + \mu a_i \psi(i) - \frac{b_i}{\theta} \psi(i + N). \]

Proof. See Appendix D.

Notice that even in the case when all MTDs operating on the same channel in the network are using the same \( a_i \) and \( b_i \) coefficients, e.g., \( a_i \) and \( b_i \) with deterministic values, evaluating (24) using (25) is not an easy task. A closer look at that expression makes us suspect on its efficiency. This is because it requires evaluating two inner triple integrals. In fact, several numerical tests we run corroborated that evaluating (24) using (25) is not an easy task. A closer look at (25) and lower bounds for \( I_c \) be required for instance in order to optimize in some way the interference is distributed approximately equal as for the RRS scenario.

**Corollary 1.** An alternative, and easy of evaluating, expression for the CRS success probability, \( p_1^{j,a} \), for a given \( \delta \), is
\[ p_1^{j,a} \approx 1 - 2 \sum_{t \in \{1, 2\}} \beta_t - 1 \exp \left( -1/2 \psi \frac{\pi}{\theta} \right) \sin \left( \frac{\pi}{\theta} \right) \cos \left( \frac{\pi}{\theta} \right), \]

where \( \psi = \psi \left( \frac{\pi}{\theta} \right) \) and \( \psi = \psi \left( \frac{\pi}{\theta} \right) \cos \left( \frac{\pi}{\theta} \right) \) and \( \nu = \chi \left( \frac{c_1}{a_i} + c_2 a_i \right) \).

Proof. See Appendix D.

**Remark 3.** We know that the interference under the CRS scheme with \( a_i = b_i = 1 \) is characterized in the same way as in the RRS case, \( L_{r} = L_{1} \), since no weights to the transmit power, e.g., no marks, are assigned; and notice that (26) captures well this phenomena since with \( \delta = 2 \) we attain

(15). This means that whatever the values of \( a_i, b_i \), as long as \( \delta = 2 \) the interference is distributed approximately equal as for the RRS scenario.

**Lemma 4.** The CRS success probability, \( p_1^{j,a} \), of the jth MTD sharing a typical channel conditioned on \( u \) MTDs, is given in (31) and (32) at the top of the next page, where \( j \in \{1, 2\} \).

Proof. Note that \( a \) comes directly from the total probability theorem, and also averaging \( p_1^{j,a} \) over all possible values of \( i \) given \( k \) and \( N \). On the other hand, \( b \) comes from using the expressions for \( \mathrm{Pr}(K \leq N) \) and \( \mathrm{Pr}(K = k) \) easily obtained.
exist some choice for \( k \) holds, where \( \tau \)

B. Practical issues

we resort to liability for both MTDs sharing the orthogonal channel. Thus,

channel when

\( a \)

it is a major drawback for this method. Alternatively, we state

the knowledge of the inter-cluster interference is required and

over the remaining \( k > k_{\text{max}} \) does not contribute significantly to

the success value in (32). Thus, we could choose \( k_{\text{max}} \) such

\( \sum_{k=N+1}^{\infty} \Pr(K = k) < \tau \rightarrow Q(k_{\text{max}} + 1, \tilde{m}) > 1 - \tau \)

holds, where \( \tau \) is the maximum allowable error we admit when

approximating with (32), e.g., \( \tau = 10^{-5} \) and \( \tilde{m} = 30 \rightarrow k_{\text{max}} = 56 \).

B. Practical issues

If we look closely at (22) and (23) we can notice there must

exist some choice for \( a_i \) and \( b_i \) in order to attain a similar reliability for both MTDs

sharing the orthogonal channel. Thus, we resort to SIR_{i,j}^{(i)} = SIR_{i,j}^{(o)} with \( a_i + b_i = \delta \). However, the knowledge of the inter-cluster interference is required and it is a major drawback for this method. Alternatively, we state the following theorem avoiding that problem.

Theorem 4. A proper approximate choice for \( a_i \) and \( b_i \) in order to

attain a similar reliability for both MTDs sharing the channel when \( \alpha = 2 \) is given by

\[ a_i = \frac{\delta(1 + \frac{1}{\mu})}{(1 + \frac{1}{\mu})\psi(K + 1)} - \psi(i + N) \]  \( \psi(i) \) \( \psi(i + N) \)

\[ b_i = \delta - a_i. \]

\[ (33) \]

\[ (34) \]

Proof. Notice that \( p_{i,1}^{(i)} \) and \( p_{i,2}^{(i)} \) in (24) only differ in the terms \( k_{1,2}^{(i)} \) and \( k_{2,1}^{(i)} \). Thus, solving \( k_{1,2}^{(i)} = B_{2,1}^{(i)} \) with \( a_i + b_i = \delta \) conducts to an approximate choice for these parameters.

Remark 4. The values of \( a_i \) and \( b_i \) aside of the index \( i \), strongly rely on the instantaneous number of MTDs contesting for transmission resources, \( K \), the number of available channels, \( N \), the SIR threshold, \( \theta \), and the SIC imperfection parameter, \( \mu \). These values for \( a_i \) and \( b_i \) do not guarantee the same instantaneous SIR for both MTDs sharing the channel, but when averaging over a long period it guarantees a similar reliability for them. Also, all MTDs have the same chances of occupying any of the channels in \( N \), thus for high loaded systems where \( \tilde{m} > N \), e.g., the probabilities of all the channels being occupied by two MTDs are great, an average similar reliability will be attained.

On the other hand, both the RRS scheme, and CRS scheme with relatively large \( \delta \), do not favor the coexistence with purely OMA setups. The reason is because these NOMA setups would increase the interference, e.g., up to twice and approximately \( \delta \) larger for RRS and CRS schemes, respectively, caused to the OMA setups, and since this is not compensated by multiplexing several users, the performance of the OMA clusters is affected. On the other hand, even when only our hybrid OMA-NOMA scheme is utilized, it is expected that all aggregators will not be under the same average access demand, e.g., same \( \tilde{m} \). Therefore, those with low demand could be operating with OMA almost all the time and the larger interference of previously schemes will be impractical. Thus, limiting the interference is crucial and that could be done by properly selecting a relatively small \( \delta \).

Theorem 5. The required \( \delta, \delta^* \), for a fair coexistence\(^8\) between OMA and our hybrid setup is approximated by the solution of

\[ \xi \delta^{\frac{2}{\alpha}} - 1 + \xi^{2 - \frac{2}{\alpha}} \delta^{\frac{2}{\alpha}} = 2, \]

where \( \xi = \exp(-\sqrt{2}s^{\frac{2}{\alpha}}) \), and it is bounded by

\[ 2^{-\frac{2}{\alpha}} \delta^* \leq \delta^* \leq 1. \]

\[ (35) \]

\[ (36) \]

Proof. See Appendix E.

Remark 5. Interestingly, even though \( \delta^* \) depends completely on the system parameters e.g., \( \lambda_0, R_{\text{th}}, c_\alpha, \theta \) and \( \alpha \), we were able to limit its range only as a function of \( \alpha \). The greater the \( \alpha \) the smaller \( \delta^* \) and consequently the greater the limitation on NOMA setup over nodes. Also, the fact that \( \delta^* < 1 \) is expected, means that the NOMA setup has to operate with an overall consumption power inferior to OMA setup.

C. Overall Performance for \( L = 2 \)

Differently than in the RRS case, where we were able to attain the average over all MTD success probability based only

\( ^8\)We refer to “fair coexistence” when for any aggregator, the interference coming from the outside topology (the inter-cluster interference) remains the same regardless of the alternative (OMA protocol or our hybrid approach) the outside clusters utilize.
The success probability is given in (37) as a function of the ordered channels, \( i \), averaging are different for each \( k \) and dependent on the index of the ordered channels, \( i \). Instead, the average over all MTD success probability is given in (37) as a function of \( p_{j,i}^c \) at the top of the page, where \( (a) \) comes from using the expressions for \( \Pr(K \leq N) \), \( \Pr(K = 0) \) and \( \Pr(K = k) \). On the other hand, the average number of simultaneous served MTDs is given in (38), shown below of (37) on top of the page, where \( \Lambda_1 \) was given in (18) and the approximation is because the same reasons than previously discussed. Notice that Remark 2 holds here for the CRS scheme as well.

### V. POWER CONSUMPTION ANALYSIS

**Lemma 5.** The average transmit power per orthogonal channel for the OMA (\( L = 1 \)) and, RRS and CRS of our hybrid scheme (\( L = 2 \)) is given by

\[
p_t = \begin{cases} 
(1-c_0)\Psi, & \text{for OMA} \\
(c_1+2c_2)\Psi, & \text{for our hybrid scheme with } L = 2 \end{cases},
\]

where \( \Psi = \frac{2\alpha R_a}{\delta c_1 + 2} \).

**Proof.** For OMA, RRS and CRS with \( L = 2 \), the average transmit power is \( \mathbb{E}[\{(1-c_0)\rho]\} = (1-c_0)\rho \mathbb{E}[\rho^\alpha] \), \( \mathbb{E}[c_1\rho^{\alpha} + 2c_2\rho^{\alpha}] = \bar{c}\rho \mathbb{E}[\rho^\alpha] \) and \( \mathbb{E}[c_1\rho^{\alpha} + c_2(a_i+b_i)\rho^{\alpha}] = (c_1 + c_2)\rho \mathbb{E}[\rho^\alpha] \), respectively. For RRS \( \bar{c} = c_1 + \delta c_2 \) since \( \delta = 2 \), and now it is only necessary to compute \( \mathbb{E}[\rho^\alpha] \).

\[
\mathbb{E}[\rho^\alpha] = \int_0^\infty r^\alpha f_r(r)dr = \frac{2}{R_a} \int_0^{\alpha+1} \frac{2R_a^\alpha}{\alpha+2} = \frac{\Psi}{\rho}, \tag{40}
\]

and (39) is attained.

Obviously, the RRS scheme with \( L = 2 \) will always require a greater power consumption than an OMA setup, since whenever two MTDs are transmitting on the same channel of a cluster, the power consumption doubles. On the other hand, the CRS scheme allows to reduce the power consumption by adopting a relatively small \( \delta \), e.g., \( \delta^* \). This behavior is illustrated in Fig. 4 for \( c_0 = 0 \). All the schemes have a common starting point in \( \Psi \) since all of them are equivalent when no MTDs require to share the same orthogonal channel, e.g., \( c_2 = 0 \rightarrow c_1 = 1 \). When \( c_2 \) increases, the power consumption of the RRS and CRS schemes more, as long as \( \delta \neq 2 \), reaching the maximum difference for \( c_2 = 1 \). Notice that choosing \( \delta^* \) we guarantee operating with the same interference than an OMA setup while the power consumption reduces since \( \delta^* < 1 \).

### VI. NUMERICAL RESULTS

Both, simulation and analytical results, are presented in this section in order to investigate the performance of our hybrid scheme as a function of the system parameters while comparing it with an OMA setup. The analytical results for the RRS scheme come from using the exact expressions; while for the CRS scheme we use the approximations. Unless stated otherwise, results are obtained by setting \( \bar{m} = 60 \), \( \lambda_a = 10^{-4.4}/m^2 \), \( 39.81/km^2 \), \( R_a = 40m \), \( \alpha = 3.6 \), \( \mu = 0 \) and \( \theta = 1 \). The value of \( \delta^* \) is found by solving numerically (35) whenever is required. Simulation results are generated using 50000 Monte Carlo runs and a sufficiently large area such 400 aggregators are placed on average.

Fig. 5 shows the success probability of the MTDs sharing the same channel conditioned on \( u = 2 \), e.g., (9) and (10) for RRS, and (32) for CRS. The idea is to show how fair the schemes are when allocating the transmission resources. Notice that for the RRS scheme, the gap between both MTDs performance keeps quite constant, independently of \( N \). This is because that gap relies strongly on the gap between \( \max(h^i, h^o) \) and \( \min(h^i, h^o) \), (see (6) and (7))\(^9\), which only depends on the number of MTDs requiring transmissions and not on the available channels. While for the CRS schemes with fixed \( a_i, b_i \), the gap tends to widen when increasing \( N \) because \( h_i \) becomes larger with respect to \( h_{i+N} \) in (22) and (23). Of course, a given system is projected to work given one value of \( N \), and by properly choosing some fixed \( a_i, b_i \) we can reduce the performance gap between the MTDs sharing the channel, which is not possible for the RRS scheme since weighting the transmit powers is not available. Notice also that by choosing \( a_i \) and \( b_i \) according to (33) and (34) both MTDs reach similar performance, hence the fairest scheme.

\(^9\)Notice the gap also relies strongly on the value of \( \mu \).
such that the implementation of the CRS is almost the same

\[ p_{\text{succ}} = p_{1,1} \frac{\Pr(K \leq N)}{1 - \Pr(K = 0)} + \frac{1}{1 - \Pr(K = 0)} \left( \sum_{k=N+1}^{2N-1} \frac{1}{k} \left( \sum_{i=1}^{\text{mod}(k,N)} c(i) + c(i) \right) + \sum_{i=\text{mod}(k,N)+1}^{N} c(i) \Pr(K = k) \right) \]

\[ \approx p_{1,1} A_1 + \frac{1}{1 - \Pr(K = 0)} \left( \sum_{k=N+1}^{2N-1} \frac{1}{k} \left( \sum_{i=1}^{\text{mod}(k,N)} c(i) + c(i) \right) + \sum_{i=\text{mod}(k,N)+1}^{N} c(i) \Pr(K = k) \right) \]

Fig. 6. (a) Average over all MTD success probability (left) and (b) Average number of simultaneously served MTDs (right), as a function of the number of channels for \( L = \{1, 2\} \).

Fig. 6a shows the average over all MTD success probability for the different schemes and \( L = \{1, 2\} \). Each scheme with \( L = 1 \) performs better than for \( L = 2 \). Also, as expected and in general, the CRS setups outperforms the RRS scheme, except when choosing \( a_i, b_i \) according to (33) and (34), for which a greater fairness is attained instead. Notice that when \( N \) is small, CRS outperforms RRS; while as \( N \) increases, their curves tend to overlap. This is due to the fact that when \( N \) is large, \( i.e., \) greater than \( \bar{m} \) for \( L = 1 \), most of the time the number of MTDs is less than the available resources such that the implementation of the CRS is almost the same

as the RRS; \textit{ii)} \( e.g., \) greater than \( \bar{m}/2 \) for \( L = 2 \), the impact of ordering the channels becomes less significant since most of the time all the MTDs will be allocated. While when \( N \) increases more and more, the system tends to behave as when \( L = 1 \). The change in the CRS curves from decreasing to increasing occurring close to \( \bar{m}/2 \) is somewhat explained by that latter argument. Even when achieving a higher reliability with an \( L = 2 \) setup is not possible, we are able to enhance the average number of simultaneously served MTDs, as shown in Fig. 6b. The success probability does not deteriorate and a significant improvement on \( K \) is attained, fundamentally when \( N \) is not large. Notice that this advantage is evidenced when \( L = 2 \), especially with the CRS scheme, so to cover a high instantaneous access demand \( \bar{m} > N \), which is even more favorable than predicted in Remark 2. By setting \( L = 2 \), the number of efficiently served MTDs could be up to twice the number with a single MTD per orthogonal channel setup. Particularly attractive are the configurations operating with \( \delta^* \) since no change in the perceived interference occurs when switching transmission schemes from \( L = 1 \) and \( L = 2 \) setups and vice versa.

Fig. 7 shows the average number of simultaneously served MTDs as a function of the density of aggregators for \( \mu \in \{0, 10\%\} \). Fig. 7a, and the SIC imperfection coefficient, Fig. 7b. When both, the network is sparse and the SIC imperfection coefficient are not so restrictive, the performance of the \( L = 2 \) setup increases. Since \( N = 30 \), each channel per cluster is operating with two MTDs almost all the time, which are more sensitive to the interference, hence their performance will be affected if either \( \lambda_a \), \( \theta \), or even \( R_a \) are larger. If \( c_2 \) was smaller, \( e.g., \) larger \( N \), this situation becomes less critical, although the gap between the \( L = 1 \) and \( L = 2 \) setups would be smaller as shown previously in Fig. 6. Notice that the imperfect SIC degrades the system performance when \( L = 2 \), but even for a high imperfection such 10\%, the advantage over the OMA setup keeps evident for a wide range of values for the system parameters. For instance, the CRS scheme with \( a_i = b_i = \delta^*/2 \) overcomes the CRS scheme with \( L = 1 \) when \( \lambda_a \lesssim 1.3 \cdot 10^{-4} \) for \( \mu = 0 \), while \( \lambda_a \lesssim 1 \cdot 10^{-4} \) would be required for \( \mu = 10\% \). Since SIC is only related with the
developing a general analytical framework to investigate its performance in terms of average success probability and average number of simultaneously served MTDs for two scheduling schemes RRS and CRS. Also, we found the power constraints on the MTDs sharing the same channel to attain a fair coexistence with purely OMA setups, while power control coefficients are found too, so that both MTDs can perform with similar reliability. Our analytical derivations allow for fast computation compared to the time-consuming Monte-Carlo simulations, which are even heavier for our hybrid scheme than for a purely OMA setup. The numerical results show that

- our hybrid access scheme aims at providing massive connectivity in scenarios with high access demand, which is not covered by traditional OMA setups, and even with lower average power consumption per orthogonal channel and per MTD, the hybrid scheme with CRS outperforms the OMA setup;
- failing to efficiently eliminate the intra-cluster interference could reduce significantly the benefits from NOMA, and can be a challenging issue for implementing NOMA in practice;
- our mathematical derivations, besides being easy to evaluate, are accurate.

Future work could focus on the relaying phase, while finding some strategies to cope with the larger aggregated data. Additionally, we intend to deeply investigate strategies in order to optimally decide when to switch from a purely OMA setup to our hybrid scheme.

**APPENDIX A**

**PROOF OF LEMMA 1**

The PMF of the number of MTDs sharing the same channel, \( u \), conditioned on the number of MTDs requiring transmissions, \( k \), is given by

\[
\Pr(U = u|k) = \begin{cases} 
1, & \text{for } u = L \text{ if } k \geq NL \\
1 - \frac{k}{N} + \left\lfloor\frac{k}{N}\right\rfloor, & \text{for } u = \left\lfloor\frac{k}{N}\right\rfloor \text{ if } k < NL \\
0, & \text{for } u = \left\lfloor\frac{k}{N}\right\rfloor + 1 = \left\lfloor\frac{k}{N}\right\rfloor + 1 \text{ if } k < NL \\
& \text{otherwise}
\end{cases}
\] (41)

Notice that if \( k \) is equal or greater than the number of available resources, \( NL \), there will be for sure \( u = L \) MTDs per channel in the representative cluster. Otherwise, if \( k < NL \), only two consecutive values for \( u \) are possible. For instance, if \( N = 10 \), \( L = 4 \) and \( k = 26 \), then 2 MTDs will be allocated in each of 4 channels \((2 \times 4 = 8 \text{ MTDs})\), and 3 MTDs in the remaining 6 channels \((3 \times 6 = 18 \text{ MTDs})\). Therefore, the probability of one channel being occupied by 2 MTDs is \( 4/10 = 0.4 = 1 - 26/10 + [26/10] \), while the probability of being occupied by 3 is the complement, \( 0.6 = 26/10 - [26/10] \).

VII. CONCLUSION

In this paper, we analyzed the uplink mMTC in a large-scale cellular network system overlaid with data aggregators. We propose a hybrid access scheme, OMA-NOMA, while

10Analytical expressions, even when some of them are numerically evaluated, have an additional value since simulations of these scenarios require a huge amount of computational resources specially for \( L > 1 \).
Now, the required PMF, \( \Pr(U = u) = \Pr(U = u|k) \Pr(K = k) \), can be calculated as follows

\[
\Pr(U = u) = \begin{cases} 
\sum_{k=N_L}^{\infty} \Pr(K = k), & \text{for } u = L \\
\left(1 - \frac{a}{b} + \frac{b}{b} \right) \sum_{k=0}^{N_L-1} \Pr(K = k), & \text{for } u = \left\lfloor \frac{b}{b} \right\rfloor \\
\left(\frac{a}{b} - \frac{b}{b} \right) \sum_{k=0}^{N_L-1} \Pr(K = k), & \text{for } u = \left\lfloor \frac{b}{b} \right\rfloor + 1 = \left\lfloor \frac{b}{b} \right\rfloor \\
0, & \text{otherwise}
\end{cases}
\]

Using the PDF of \( K \), \( \Pr(K = k) \), its CDF, \( \Pr(K \leq k) \), and the expected \( K \) on one interval,

\[
\sum_{k=u}^{b} k \Pr(K = k) = \sum_{k=u}^{b} k \left( \frac{\bar{m} Q(a, \bar{m}) + uQ(1+b, \bar{m})}{b} + \exp(-\bar{m}) \left( \frac{\bar{m}^a}{a!} - \frac{\bar{m}^{a+1}}{b!} \right) \right)
\]

\[
a = \bar{m}Q(1+b, \bar{m}) \exp(-\bar{m}) \left( \frac{\bar{m}^a}{a!} - \frac{\bar{m}^{a+1}}{b!} \right)
\]

\[
b = \bar{m}Q(1+b, \bar{m}) \exp(-\bar{m}) \left( \frac{\bar{m}^a}{a!} - \frac{\bar{m}^{a+1}}{b!} \right)
\]

and regrouping similar terms, we attain (2).

\section*{Appendix B}

\subsection*{Proof of Theorem 1}

We proceed as follows

\[
p^2_{1,1} = \Pr(\text{SIR}^{1,1} > \theta) = \mathbb{E}_T[\Pr(h > \theta I|I)] = \mathbb{E}_T[\exp(\theta I)|I]
\]

\[
p^2_{1,2} = \Pr(\text{SIR}^{1,2} > \theta) = \mathbb{E}_T[\Pr(\text{max}(h', h'') - \theta \min(h', h''), \theta > \theta I|I)] = \mathbb{E}_T[\exp(\theta I)|I]
\]

\[
= \begin{cases} 
\mathbb{E}_T\left[\frac{2\theta}{1+\theta} \exp(-\theta I) - \frac{2\theta}{\theta} \exp\left(-\frac{2\theta}{\theta} I\right) \right], & \text{if } 0 \leq \theta < 1 \\
\mathbb{E}_T\left[\frac{2\theta}{1+\theta} \exp(-\theta I) \right], & \text{if } \theta \geq 1
\end{cases}
\]

\[
p^2_{2,2} = \Pr(\text{SIR}^{2,2} > \theta) = \mathbb{E}_T[\Pr(\text{min}(h', h'') - \theta \mu \max(h', h''), \theta > \theta I|I)] = \mathbb{E}_T[\exp(-\theta I)|I]
\]

\[
= \begin{cases} 
\mathbb{E}_T\left[\frac{1-\theta}{1+\theta} \exp\left(-\frac{2\theta}{\theta} I\right) \right], & \text{if } 0 \leq \theta \mu < 1 \\
0, & \text{if } \theta \mu \geq 1
\end{cases}
\]

where \( v_1 = \max(h', h'') - \theta \min(h', h'') \) and \( v_2 = \min(h', h'') - \theta \mu \max(h', h'') \), while \( (a) \) and \( (b) \) come from using their CDF expressions, which are given by

\[
F_{V_1}(v_1) = \begin{cases} 
1 - \frac{2\theta}{1+\theta} \exp(-\theta I), & \text{if } 0 \leq \theta < 1 \\
1 - \frac{2\theta}{\theta} \exp(-\theta I), & \text{if } \theta \geq 1
\end{cases}
\]

\[
F_{V_2}(v_2) = \begin{cases} 
1 - \frac{1-\theta}{1+\theta} \exp(-\theta I), & \text{if } 0 \leq \theta \mu < 1 \\
1, & \text{if } \theta \mu \geq 1
\end{cases}
\]

for \( v_1, v_2 > 0 \). Notice that the last equalities in (43), (44) and (45) are equivalent to (8), (9) and (10), respectively.

For the derivation of (11) we have to use the fact that the Poisson cluster process defined in (3) is a Neyman-Scott process [35, Definition 3.4] with Probability Generating Functional (PGFL) [35, Corollary 4.13]

\[
G[v] = \exp \left( \lambda_0 \int_{\mathbb{R}^2} \left( \int_{\mathbb{R}^2} v(w+y) f(y) dy \right) - 1 \right) d\text{w}.
\]

Based on \( g \sim \text{Exp}(1) \) which allows to state

\[
v(w+y) = \mathbb{E}_g[\exp(-sg^{\alpha}||w+y||^{\alpha})] = \frac{1}{1 + s\lambda_0^\alpha ||w+y||^{\alpha}}.
\]

and according to the cosine law, e.g., \( ||w+y|| = (r_w^2 + r_a^2 - 2r_w r_a \cos(\omega))^\frac{1}{2} \), while substituting (4) and PDFs expressions \( f_r(r_a) = \frac{2r_a}{r_w^2} \) and \( f_w(\omega) = \frac{1}{2\pi} \) into (48), we reach \( \mathbb{L}_i(s) \) in (11).

\section*{Appendix C}

\subsection*{Proof of Theorem 2}

Performing some algebraic transformations on (11) we have

\[
\mathbb{L}_i(s) = \exp \left( 2\pi \lambda_0 \int_0^\infty \sum_{u=0}^L c_u T(r_w, s)^s - 1 \right) dr_w
\]

\[
\stackrel{(a)}{=} \exp \left( 2\pi \lambda_0 \int_0^\infty \sum_{u=0}^L c_u T(r_w, s)^s - 1 \right) dr_w
\]

\[
\stackrel{(b)}{=} \exp \left( 2\pi \lambda_0 \sum_{u=0}^L c_u \int_0^\infty r_w T(r_w, s)^s - 1 \right) dr_w
\]

\[
\stackrel{(c)}{=} \exp \left( 2\pi c_1 \lambda_0 \int_0^\infty r_w T(r_w, s) - 1 \right) dr_w \cdot \exp \left( 2\pi \lambda_0 \sum_{u=0}^L c_u \int_0^\infty r_w T(r_w, s)^s - 1 \right) dr_w
\]

\[
\stackrel{(d)}{=} \exp \left( 2\pi c_1 \lambda_0 \int_0^\infty r_w T(r_w, s) - 1 \right) dr_w \cdot \exp \left( 2\pi \lambda_0 \sum_{u=0}^L c_u \int_0^\infty r_w T(r_w, s)^s - 1 \right) dr_w.
\]

where (a) comes from \( \sum_{u=0}^L c_u = 1 \), (b) is attained by regrouping terms, and (c) by pulling out the term associated with \( u = 1 \). Notice that \( T_1 \) matches the Laplace transform of an HPPP with density \( c_1 \lambda_0 \) and one active MTD per channel per cluster, which is given by [9, Eq. (23)]

\[
T_1 = \exp(-\chi c_1 s^\frac{1}{2} )
\]
On the other hand, $T_2$ includes the contribution of the clustered MTDs, e.g., $u \geq 2$. For each $u$, the related term in $T_2$ matches the Laplace transform of an HPPP with density $c_u \lambda_0$ and $u$ active MTDs per channel per cluster. It has been observed in [36]–[38] that the SIR complementary CDFs, e.g., Laplace transform of the interference, for different point processes appear to be merely horizontally shifted versions of each other (in dB), as long as their diversity gain is the same. Thus, scaling the threshold $s$ by this SIR gain factor (or shift in dB) $G$, we have

$$\mathcal{L}_I(s) \approx \mathcal{L}_{I,rel}(s/G).$$  \hspace{1cm} (52)

However, $G$ is also a function of $s$ but for many setups it keeps approximately constant and consequently it can be determined by finding its value for an arbitrary value of $s$ [38]. Using the PPP as the reference model, the limit of $G$ as $s \to 0$, $G_0$, is relatively easy to calculate [38]

$$G_0 = \frac{\text{MISR}_{\text{PPP}}}{\text{MISR}},$$  \hspace{1cm} (53)

where the MISR is the mean of the interference-to-(average)-signal ratio $\text{ISR} = \frac{f_{\text{I}}}{f_{\text{S}}}$ (all in dB). Since $S = h$ here and $E[h] = 1$, we have that $\text{MISR} = E[I]$. Now, considering the contribution of each $u$ in $T_2$ separately we have

$$G_{0,u} = \frac{E[I_{\text{PPP}}]}{E[I]} = \frac{1}{u}.$$  \hspace{1cm} (54)

Notice that $E[I_{\text{PPP}}]$ and $E[I]$ are divergent measures for our system model because the resultant point process would not be locally finite since we assumed a path loss model $||x||^{-\alpha}$ [35]. However, the quotient depends merely on the density of both process and since $\lambda_{\text{PPP}} = c_u \lambda_0$ and $\lambda = u c_u \lambda_0$ we attain the last equality in (54). Obviously, $\mathcal{L}_{I_{\text{PPP}}}(us)$ works almost surely\textsuperscript{11} as an upper bound for $\mathcal{L}_{I,s}(s)$ in (11) if $G_0 < 1$ (see [38] for a geometrical perspective), which holds here since $u \geq 2$. Now, using [9, Eq. (23)] as the HPPP of reference with one fixed MTD per orthogonal channel\textsuperscript{12}, we attain

$$T_2 \leq \exp \left(-\chi \sum_{u=2}^{L} c_u(u) \frac{s^2}{4} \right),$$  \hspace{1cm} (55)

with asymptotic equality as $s \to 0$. Substituting (55) and (51) into (50) yields (13).

On the other hand, the lower bound comes from the corresponding HPPP with same intensity that for each $u$ in $T_2$, e.g., $u c_u \lambda_0$. This is because in our system model the performance in terms of success probability of an HPPP with the same intensity that any clustered process will be always worse since we are expecting a greater number of close interfering nodes. Once again using [9, Eq. (23)], we have

$$T_2 \geq \exp \left(-\chi \sum_{u=2}^{L} u c_u s \frac{s^2}{2} \right).$$  \hspace{1cm} (56)

\textsuperscript{11}Except for relatively large $s$ (13) might serve as an upper bound but it becomes an accurate approximation for (11).

\textsuperscript{12}The clustered process with one MTD per orthogonal channel (per cluster) is indeed a HPPP with density $\lambda_0$ because the displacement theorem in stochastic geometry [35].

Now, substituting (56) and (51) into (50) yields (14). For both (13) and (14) the equality fits to the asymptotic cases $\lambda_0 \to \infty$, $0$, $R_a \to \infty$, $0$.

\textbf{APPENDIX D}

\textbf{PROOF OF THEOREM 3}

Let’s assume the case where $K > N$ since for $K \leq N$ the system behaves exactly as in the RRS scheme and the problem is already solved, e.g., $p^c_{j,u}(i) = p^{c}_{j,u}$\textsuperscript{13}. According to high order statistic theory [39], the PDF of the $i$th best channel power gain is

$$f_{H_i,K}(x) = \frac{K!}{(i-1)!(K-i)!} \left[ \frac{1}{\theta} \exp(-\frac{x}{\theta}) - \left(1 - \exp(-\frac{x}{\theta}) \right)^{K-i} \right]^{i-1}.$$  \hspace{1cm} (57)

Now we have $p^{c}_{j,u}(i) = \Pr \left( \text{SNR}^c_{j,u}(i) > \theta \right) = \Pr \left( \Theta^j_{j,u} > \Theta_i \right)$, where $\Theta_{j,1} = h_i$, $\Theta_{j,2} = a_i b_i$, $\Theta_{j,2} = b_i h_{i+N} - \theta a_i h_i$. Unfortunately, the distribution of $\Theta_{j,u}$, even for the simplest case $\Theta_{j,1}$, conduces to very complicated expression for the CDF, preventing us to follow the same path we used for the RRS scheme. Instead we proceed as follows

$$p^{c}_{j,u}(i) = \Pr \left( \Theta_i < \frac{\Theta_{j,u}}{\theta} \right) = \frac{1}{\pi} \int_0^{\infty} \frac{1}{\varphi} \left[ \mathcal{L}_{I_j}(-\varphi) \exp(-\varphi \Theta_{j,u}^{(i)}) \right] d\varphi,$$  \hspace{1cm} (58)

where (a) comes from the Gil-Pelaez inversion theorem [40] and the approximation in (b) comes from the Jensen inequality. Also,

$$E[\Theta_{1,1}] = \int_0^{\infty} x f_{H_1,K}(x) dx = \frac{K!}{(i-1)!(K-i)!} \left[ \left( K^2 x \text{Beta}(x, -K, K - i + 1) + \exp(-Kx) \frac{F_2 \left( -K, -K, -i, -K, 1 - K, 1 - K, \exp(x) \right) \bigg|_{x=\infty} \right) \right] = \frac{\Gamma(i) \Gamma(K - i + 1) \left( \psi(K + 1) - \psi(i) \right) - \Gamma(K - i + 1)}{(i-1)(K-i)!},$$  \hspace{1cm} (59)

while it is straightforward obtaining $E[\Theta_{1,2}]$ and $E[\Theta_{2,2}]$ from (59), yielding (27)-(29). Substituting them into (58) we attain (24).

To attain (25) we require to use the same arguments than previously discussed when deriving the result in (11). However, now the point process has marks $1, a_i, b_i, \cdot$, with $b_i = \delta - a_i$, and we require to include the marks along with their probabilities when evaluating (48) [35, Th. 7.5]. Notice that it is intractable evaluating (25) efficiently when $a_i$, $b_i$ are random since requires an additional integration\textsuperscript{13} in

\textsuperscript{13}For instance, the uniform distribution, which is very simple and probably unrealistic for the scenario discussed, with PDF $f_{\alpha}(\alpha) = \frac{1}{\pi}$, is already cumbersome when evaluating (25).

the exponent, hence we propose using some approximations as follows

\[
\mathcal{L}_{t_i}(s) \approx \exp \left(2\pi c_1 \lambda_s \int_{0}^{\infty} \left(\mathcal{Y}\left(r_w, s\right) - 1\right) dr_w\right)
\]

where \((a)\) comes from using fixed values of \(a_i, b_i\), thus, avoiding the additional integration, and using the relation between the geometric and arithmetic mean, \(\frac{x + y}{2} \geq \sqrt{xy} \rightarrow xy \leq \left(\frac{x + y}{2}\right)^2\), as an approximation, with equality when \(a_i = b_i\); while \((b)\) comes from regrouping terms and \(\sum_{u=0}^{2} c_u = 1\). Now, by using the same procedure than previously discussed in the proof of Theorem 2, e.g., finding upper and lower bounds for \((60)\), and averaging both, we attain

\[
\mathcal{L}_{t_i}(s) \approx \sum_{t \in \{1, 2\}} \beta_{t-1} \exp \left(-\chi \left(c_1 + c_2 \left(\frac{\alpha}{2} \frac{a_i^2 + b_i^2}{2}\right) s^2\right)\right)
\]

where \(\beta_0, \beta_1 \in [0, 1]\) and \(\beta_0 + \beta_1 = 1\). Since \(\frac{2}{\alpha} < 1\), and the relation between the generalized mean (or power mean) with exponents \(\frac{2}{\alpha}\) and 1, the following result holds

\[
\left(\frac{a_i^2 + b_i^2}{2}\right)^{\frac{2}{\alpha}} \leq \frac{a_i + b_i}{2}^{\frac{2}{\alpha}}
\]

By substituting (62) into (61) we attain (26). See Fig. 3 for more insights on the accuracy.

\[\Box\]

**APPENDIX E**

**PROOF OF THEOREM 5**

In order to attain a fair coexistence between the OMA and NOMA setups the interference must be kept the same. Thus, we need to match (26) with the Laplace transform of the interference for an equivalent OMA setup as shown next:

\[
\sum_{t \in \{1, 2\}} \beta_{t-1} \exp \left(-\chi \left(c_1 + c_2 \left(\frac{\alpha}{2} \frac{a_i^2 + b_i^2}{2}\right) s^2\right)\right)
\]

where \((a)\) comes from setting \(\beta_0 = \beta_1 = 0.5\) and evaluating the left-hand sum, \((b)\) comes from dividing both terms by \(\exp(-\chi c_1 s^{2/\alpha})\), and \((c)\) by \(1 - c_0 - c_1 = c_2\) and setting \(\xi = \exp(-\chi c_2 s^{2/\alpha})\). Dividing both terms by \(\xi\) we attain

\[
\exp\left(-\chi \left(c_1 + c_2 \frac{2s^{2/\alpha}}{\delta^2}\right)\right) \geq \exp\left(-\chi (1-c_0) s^{2/\alpha}\right)
\]

\[
c_1 + c_2 \frac{2s^{2/\alpha}}{\delta^2} \leq c_2
\]

\[
\delta \leq 1,
\]

\[
\exp\left(-\chi \left(c_1 + c_2 \frac{2s^{2/\alpha}}{\delta^2}\right)\right) \leq \exp\left(-\chi (1-c_0) s^{2/\alpha}\right)
\]

\[
c_1 + c_2 \frac{2s^{2/\alpha}}{\delta^2} \geq 1 - c_0
\]

\[
2 \frac{s^{2/\alpha}}{\delta^2} \geq 1
\]

\[
\delta \geq 2 \frac{s^{2/\alpha}}{\delta^2}
\]

completing the proof.


