Modeling and Analysis of Tagged Preamble Transmissions in Random Access Procedure for mMTC Scenarios

Salvatore Riolo, *Graduate Student Member, IEEE*, Daniela Panno, *Member, IEEE*, and Luciano Miuccio, *Graduate Student Member, IEEE*

***Abstract*— With the emerging of the massive Machine Type Communication usage scenario, the legacy Random Access (RA) procedures need to be fully renewed to meet the huge number of accesses and the energy-constrained device requirements. In this context, the strategy of transmitting tagged preambles as access requests is gaining importance. In fact, it can reduce the number of signaling transmissions per access attempt, the overall number of phases in the access procedure, and, accordingly, energy consumption per device. However, the effectiveness of detection strategies of tagged preambles by the gNB receiver plays a fundamental role on the RA procedure performance. Up to now, the performance of these advanced detection procedures has been obtained only by running a large number of simulations that are typically highly time-consuming. To the best of our knowledge, this paper presents the first analytical model to analyze the correct detection of both the preamble and the tag transmitted by each device in the presence of interference, due to other preambles and tags, and of noise. The high accuracy of the proposed model is verified through simulations. In addition, we show how our analytical study can be a good tool to investigate and derive innovative detection strategies.**

***Index Terms*— Analytic model, mMTC, performance evaluation, preamble, random access (RA), tag, Zadoff–Chu (ZC) sequence.**

I. Introduction

MOBILE communication applications have shifted from basic voice telephony to empowering a wide range of verticals across various industries, most notably via the rapidly expanding popularity of smartphones, tablets, smartwatches and a massive number of new smart devices representing an integral part of our lives. In this context, three families of usage scenarios have been defined by ITU-R for 5G networks [1]: enhanced Mobile BroadBand (eMBB), that focuses on services that have high requirements for bandwidth [2], [3]; Ultra-Reliable and Low Latency Communications (URLLC), which covers services extremely sensitive to latency [4]; and, finally, massive Machine Type Communications (mMTC), that represents a large number of low-complexity and energy-constrained devices sending very short packets with relaxed delay requirement.

As regards the last scenario, unlike traditional 4G mobile networks that support up to millions connections, the mMTC usage scenario forecasts billions of connected devices [5], including smart grids, environment monitoring, smart metering, industry automation, home automation, and so on. To manage the Random Access (RA) of a massive number of MTC devices, several RA schemes have been proposed. They can be categorized into two types: grant-based and grant-free [6]. In the conventional grant-based RA schemes, the MTC device follows a handshake procedure to receive an uplink grant from the Next Generation NodeB (gNB). If the procedure is successful, the gNB assigns dedicated resources to the MTC in the Physical Uplink Shared Channel (PUSCH), if available, to transmit the data packet. Instead, in grant-free RA schemes, the activated MTC devices directly transmit the data packets with their identity over a randomly selected resource in the PUSCH, without being scheduled by the gNB. As regards the grant-based scheme, the uplink grants prevent the data transmission collisions, at the cost of long latency and high signaling overhead, whereas the grant-free transmissions experience a reduced signaling overhead and low latency but a lower number of devices can be supported [7].

In this paper, we aim to manage the RA procedure of a very large number of MTC devices requiring services with a relaxed delay requirement. So, an improved grant-based scheme, that reduces the signaling overhead and energy consumption, is the most promising solution. Anyway, when a huge number of MTC devices simultaneously initiate their procedure to access to the network, a large number of access requests collide, the MTC devices will re-attempt their access, thus causing a severe congestion problem in the system if the legacy grant-based RA procedure is adopted [8], [9]. At the aim of alleviating this issue, several works [10], [11] consider Access Class Barring (ACB) approaches [12], i.e., congestion control schemes designed for limiting the number of

Manuscript received May 24, 2020; revised October 28, 2020; accepted January 25, 2021. Date of publication February 17, 2021; date of current version July 12, 2021. This work was supported in part by the University of Catania, Fondi di Ateneo "Piaceri 2020-2022", under projects "Linea 2 MANGO" and "Linea 4 OPEN ACCESS"; and in part by the Italian Ministry of University and Research (MIUR) - PNR 2015-2020 under Project ARS01_00353 "MAIA - Monitoraggio attivo dell'infrastruttura" - CUP B56G18000480005. The associate editor coordinating the review of this article and approving it for publication was Y. Chen. *(Corresponding author: Salvatore Riolo.)*

The authors are with the Department of Electrical, Electronic and Computer Engineering, University of Catania, 95125 Catania, Italy, and also with the Consorzio nazionale interuniversitario per le telecomunicazioni (CNIT), 43124 Parma, Italy (e-mail: salvatore.riolo@unict.it; daniela.panno@dieei.unict.it; luciano.miuccio@phd.unict.it).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2021.3057932.

Digital Object Identifier 10.1109/TWC.2021.3057932

simultaneous access attempts, thus reducing the number of RA failed attempts. But in spite of the adoption of an ACB scheme, in a massive scenario the collision probability in each RA attempt remains significantly high, so, the average access attempt number before success is high. This high number of access attempts not only increases delays, but also increases signaling transmissions and, therefore, energy consumption [13] which is a key point for the MTC devices. This means that just applying an ACB scheme is not enough to experience good performance for the RA procedure, i.e., the legacy grant-based RA procedures need to be fully renewed. Let us briefly analyze the LTE RA procedure, that involves the 4-phase message handshake between each device and the Evolved Node B (eNodeB). During the Phase 1, each device attempts the contention-based access by transmitting a preamble sequence in the Physical Random Access Channel (PRACH). The preamble sequence is generated by applying random cyclic shifts to the Zadoff-Chu (ZC) sequence with the same root r. Since the same preamble can be selected by more than one device, the collision probability of the preamble is non-zero. After detecting the received preambles, the eNodeB can only know whether a specific preamble has been transmitted or not, but it cannot recognize how many devices have transmitted it, i.e., if a collision has occurred. During Phase 2, the eNodeB transmits a Random Access Response (RAR) message for each detected preamble. Then, during Phase 3, each device transmits a connection request to set up the Radio Resource Control (RRC) connection with the eNodeB. We note that the eNodeB recognizes eventual preamble collisions occurred in Phase 1 just in this Phase 3. This means that, for any access attempt, the device will still carry out Phases 1 and 3, i.e., two signaling transmissions per access attempt. In addition, if the 4-phase RA procedure was successful, further signaling messages have to be exchanged before starting the data packet transmission [13]. It is clear that the legacy RA procedure is highly inefficient for supporting the very short transmissions of MTC traffic. For this reason, the 3GPP has already introduced, in Release 13, the enhanced Machine Type Communication and the Narrow-Band IoT technologies [14], that are optimized for granting lower complexity, and providing longer battery life. Nevertheless, both proposals continue to adopt the legacy 4-phase RA procedure. Therefore, these technologies, although more suitable to support machine-type communications according to the IMT-Advanced requirements (connection density of 10^5 devices per km^2), are still inefficient for the expected massive scenario where the connection density is at least equal to 10^6 devices per km^2, based on the IMT-2020 requirements. For this reason, an innovative RA procedure is needed. In this context, the academic community is studying and proposing new optimized RA procedures at the aim of avoiding the RRC connection setup overhead, e.g., the 2-phase connectionless RA procedure described in [15], [16]. By adopting this procedure, immediately after the reception of the RAR message (Phase 2), the MTC device transmits its data packet. However, the main disadvantage is that the device sends the data packet regardless of successful or unsuccessful access attempt, wasting energy in the latter case. In parallel, other works [13], [17]–[19] followed a different approach, modifying the first phase of the traditional 4-phase RA procedure to carry additional information to the gNB. In [17] the authors suggest a new access scheme to simultaneously transmit both preambles and a small-sized message in the PRACH as a message embedded preamble sequence by using distinct root numbers. However, we found the innovative approach proposed in [18], [19] very interesting. These works are based on the usage of tagged preamble sequences, i.e., each device sends the sum of a randomly selected preamble sequence with root r and a randomly selected tag sequence with root $k \neq r$. In [18] the authors adopt the tagged preamble sequences to capture multiple Timing Advance (TA) values for a single detected preamble and propose a PUSCH resource allocation method based on the multiple TA values captured. In [19], and also in [13], these tag sequences allow the gNB to detect the preamble collision during the first phase. In this way, the device reduces the number of signaling transmissions per failed access attempt to one. Specifically, in [19] the authors still adopt the conventional 4-phase RA procedure, while in [13] the strategy of transmitting tagged preambles is also used to apply a 2-phase RRC connectionless RA procedure, achieving a further reduction of the energy consumption. However, the expected improved performance for these advanced RA procedures is achieved only if the gNB receiver detector works accurately in the first phase, that is, if all and only the preamble-tag pairs transmitted by the devices are detected.

In this regard, we observe that unlike the conventional preamble transmission in which all the transmitted sequences have the same root r, here the gNB receives the sum of several sequences composed of different roots. This strongly affects the correct detection of both preambles and tags, because preambles and tags generated from different root sequences are not orthogonal. The higher the number of attempting devices, the higher the average number of selected preambles, the higher the number of different ZC roots. Consequently, in a mMTC scenario the interference can be so high that the performance of the gNB receiver is compromised. The tagged preamble sequence detection is hard to model analytically, so its performance evaluation is often reported by means of simulation results [13], [18], [19]. In [13] we assumed an ideal preamble-tag detector, whereas in [18], [19] the authors formulate the problem of preamble-tag detection by introducing the concept of detection thresholds for both preambles and tags. However, they just derive an analytical expression for the preamble and tag detection probabilities, without providing a solution either in closed form or numerically, and based on *a posteriori* information. In fact, the detection performance is only provided by means of average values over a large number of simulations. As reported in [20], simulation results are typically highly time-consuming in a massive scenario and the obtained results are not easily reproducible. For this reason, an analytical model that provides a closed-form solution is valuable. To our best knowledge, in literature no one has yet dealt with evaluating it. The contributions of our work can be summarized as follows.

1) As the main contribution, we present a rigorous methodology for modeling and analyzing the signal processed by the gNB receiver at the first phase of the RA procedure, when tagged preamble sequences are transmitted in mMTC scenarios. In particular, we believe that this is the first work that derives an *exact* expression for the detection probability distributions of both the preamble and tag at the receiver, in presence of Noise and Interference due to other preambles and tags.
2) A strength of our modeling, in addition to its demonstrated accuracy, is that the expressions obtained are in closed form, and we use them to analytically derive the threshold values to detect preambles and tags with a given probability.
3) The proposed models are powerful and computationally simple tools that can be used for determining issues and limits of tagged preamble detection strategies in order to investigate and suggest new ones. For example, we limit the working zones in which a simple detection strategy can work well; we also evaluate the impact of failed tagged preamble detection rates on the signaling overhead and energy consumption for the MTC devices.

The rest of the paper is organized as follows. We describe the conventional preamble and the tagged preamble transmission in Section II. In Section III we present the System Model and give a detailed description of the main issues on the preamble and tag detection procedure. Then, we provide accurate models for preambles and tags detection in Section IV and V, respectively. In Section VI we carry out a performance analysis. Finally, Section VII summarizes the conclusions.

II. Background

A. Conventional Preamble

In LTE, the RA procedure is done by means of ZC sequences [21]. They are defined as:

$$z_r[n] = \exp\left[-j\frac{\pi r n(n+1)}{N_{ZC}}\right], \tag{1}$$

for $n = 0, \ldots, (N_{ZC}-1)$, where N_{ZC} denotes the ZC sequence length, and r is the root index. This latter is an integer value less than N_{ZC} which is chosen to be relatively prime respect to N_{ZC}. Let us note that if N_{ZC} is an odd prime value, then $r \in \{1, \ldots, (N_{ZC}-1)\}$. For this reason, typically N_{ZC} is chosen to be an odd prime value, as we assume in the following.

A preamble sequence is generated cyclically shifting the reference ZC sequence (1) by pN_{CS}. It is denoted as:

$$z_r^p[n] = z_r[(n + pN_{CS})_{N_{ZC}}], \tag{2}$$

where p is the preamble index randomly selected in the set $\{1, \ldots, N_{P_T}\}$, $N_{P_T} = \lfloor N_{ZC}/N_{CS} \rfloor$, N_{CS} is the preamble cyclic shift, and $(\cdot)_{N_{ZC}}$ denotes the modulo-N_{ZC} operation. We note that N_{CS} and r are broadcast by the eNodeB as part of the system information. Among all N_{P_T} available preambles, $N_P \leq N_{P_T}$ are reserved for the contention-based RA procedure, while the remaining ones are reserved for the collision-free Access procedure. The ZC sequences have the following properties:

(i) Ideal cyclic auto-correlation. The correlation function, denoted as $C_{z_r^p, z_r}[\tau]$, between a sequence $z_r^p[n]$ and the reference sequence (1) with the same root r and cyclic shift equal to 0, is non-zero only in $\tau = pN_{CS}$, in which it assumes a positive real value.

$$\begin{aligned} C_{z_r^p, z_r}[\tau] &= \frac{1}{\sqrt{N_{ZC}}} \sum_{n=0}^{N_{ZC}-1} z_r^p[n] z_r^*[n+\tau] \\ &= \sqrt{N_{ZC}} \;\; \delta[\tau - pN_{CS}], \end{aligned} \tag{3}$$

where $(\cdot)^*$ denotes the complex conjugation and $\delta[\tau]$ is the discrete-time unit impulse function. We exploit this property to derive how much a received sequence $z_r^p[n]$ is shifted (i.e., which preamble index p has been selected), by calculating the real part of the correlation function $C_{z_r^p, z_r}[\tau]$, denoted as $\Re\left\{C_{z_r^p, z_r}[\tau]\right\}$.

(ii) Cyclic cross-correlation. The absolute value of the cyclic cross-correlation between two ZC sequences with different root numbers, denoted as $C_{z_r^p, z_k}$, is equal to 1 for each τ.

$$|C_{z_r^p, z_k}[\tau]| = \frac{1}{\sqrt{N_{ZC}}} \left| \sum_{n=0}^{N_{ZC}-1} z_r^p z_k^*[n+\tau] \right| = 1, \tag{4}$$

where $k \in \{1, \ldots, (N_{ZC}-1)\}$, $k \neq r$. So, let us underline that $\Re\left\{C_{z_r^p, z_r}[\tau]\right\}$, for each τ, assumes a generally different value belonging to the interval $[-1, 1]$.

B. Tagged Preambles

The tagged preamble sequence consists of both a preamble and a tag ZC sequence, which are transmitted together [18], [19]. In particular, the tagged preamble sequence is defined as:

$$x_{r,k_p}^{p,t}[n] = P_T(z_r^p[n] + z_{k_p}^t[n]), \tag{5}$$

where P_T denotes the transmit power, and $z_{k_p}^t[n]$ the tag sequence. It is expressed as:

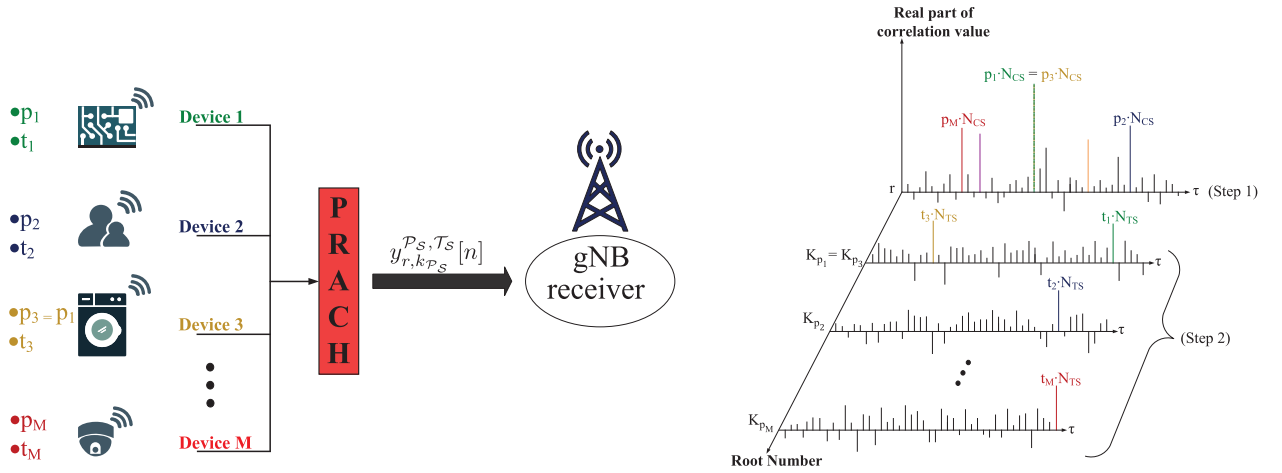
$$z_{k_p}^t = z_{k_p}[(n + tN_{TS})_{N_{ZC}}], \tag{6}$$

where k_p indicates the tag root number related to the pth preamble index, N_{TS} the tag cyclic shift value, and t denotes the tag randomly selected in $\mathcal{T} = \{1, \ldots, (N_T)\}$, in which $N_T = \lfloor N_{ZC}/N_{TS} \rfloor$ is the overall number of tags. The root k_p is determined by a fixed mapping function of the selected preamble index p, i.e., each preamble is mapped into a specific tag root number $k_p \neq r$.

III. System Model and Issues

We are considering the contention-based RA procedure in a PRACH in a given instant time, as shown in Fig. 1a. Let $\mathcal{M} = \{1, \ldots, M\}$ denote the set of attempting devices[1] using the same preamble root number r on the same PRACH, and

[1] $\mathcal{M}$ includes the devices that perform the first access attempt and the re-attempting ones that collided in the past RA cycles.



(a) First phase of the Random Access procedure.

(b) Representation of the real part of the correlations calculated at the receiver.

Fig. 1. An example of the actions done for detecting preambles and tags.

$\mathcal{P} = \{1, \ldots, N_P\}$ denotes the set of total available preambles for the contention-based procedure.

Each device i, with $i \in \mathcal{M}$, selects a preamble $p_i \in \mathcal{P}$ and a tag $t_i \in \mathcal{T}$. Also, let us define:

- **A** as the preamble association matrix of size $M \times N_P$, where each element $a_{i,j}$ is a boolean equal to 1 if and only if device i has selected preamble j,
- **B** as the tag association matrix of size $M \times N_T$, where each element $b_{i,j}$ is a boolean value equal to 1 if and only if device i has selected tag j,
- $\mathcal{P}_\mathcal{S} \subseteq \mathcal{P}$ as the subset of preambles that have been selected by the M devices to perform the RA procedure, i.e., $\mathcal{P}_\mathcal{S} = \{j \mid j \in \mathcal{P} \text{ and } \sum_{i=1}^{M} a_{i,j} > 0\}$,
- $\mathcal{T}_\mathcal{S} \subseteq \mathcal{T}$ as the subset of tags that have been selected by the M devices, i.e., $\mathcal{T}_\mathcal{S} = \{j \mid j \in \mathcal{T} \text{ and } \sum_{i=1}^{M} b_{i,j} > 0\}$.

For each device i, $\sum_{j=1}^{N_P} a_{i,j} = 1$, and $\sum_{j=1}^{N_T} b_{i,j} = 1$. In addition, for each preamble $p \in \mathcal{P}$,

- $\mathcal{M}_p \subseteq \mathcal{M}$ denotes the subset of attempting devices that have chosen the same preamble index p, i.e., $\mathcal{M}_p = \{i \mid i \in \mathcal{M} \text{ and } a_{i,p} = 1\}$.
- $|\mathcal{M}_p|$ as the cardinality of preamble index p, that is, the number of devices that have chosen preamble index p. Let us underline that $\mathcal{M}_p$ can be empty, so $|\mathcal{M}_p|$ can be equal to 0.

For each pair (p, t), with $p \in \mathcal{P}$ and $t \in \mathcal{T}$, $\mathcal{M}_{(p,t)} \subseteq \mathcal{M}$ denotes the subset of attempting devices that have chosen the same pair (p, t), i.e., $\mathcal{M}_{(p,t)} = \{i \mid i \in \mathcal{M}, a_{i,p} = 1 \text{ and } b_{i,t} = 1\}$.

We define $\mathcal{P}_h \subseteq \mathcal{P}$, with $h = 0, 1, \ldots, M$, as the subset of preambles that have been chosen by an amount of devices equal to h, i.e., $\mathcal{P}_h = \{p \mid p \in \mathcal{P} \text{ and } |\mathcal{M}_p| = h\}$. Also in this case, $\mathcal{P}_h$ can be empty. It follows that

$$\mathcal{P} = \bigcup_{h=0,1,\ldots,M} \mathcal{P}_h. \tag{7}$$

Finally, for each selected preamble $p_s \in \mathcal{P}_\mathcal{S}$, we define $\mathcal{M}_{p_s} \subseteq \mathcal{M}$ as the subset of attempting devices that have chosen the preamble p_s, and $\mathcal{T}_{p_s} \subseteq \mathcal{T}$ as the subset of selected tags that are associated to the preamble p_s, i.e., $\mathcal{T}_{p_s} = \{j \mid j \in \mathcal{T}_\mathcal{S} \text{ and } b_{i,j} = 1, \text{ with } i \in \mathcal{M}_{p_s}\}$.

A. Issues on the Preamble and Tag Detection Procedure

In order to identify the issues for a correct preamble and tag detection, we provide here an overview of how the gNB receiver works. As shown in Fig. 1a, the attempting device i choses in a random way the pair (p_i, t_i) and transmits in the PRACH the following tagged preamble:

$$x_{r,k_{p_i}}^{p_i,t_i}[n] = P_{T_i}(z_r^{p_i}[n] + z_{k_{p_i}}^{t_i}[n]). \tag{8}$$

We denote the total received sequence at the gNB with $y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}}[n]$, since it is a function of the set of selected preambles $\mathcal{P}_\mathcal{S}$ and tags $\mathcal{T}_\mathcal{S}$.

$$y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}}[n] = \sum_{i=1}^{M} \sum_{g=1}^{G_i} h_{i,g} x_{r,k_{p_i}}^{p_i,t_i}[(n + d_{i,g})_{N_{ZC}}] + N[n], \tag{9}$$

where G_i denotes the number of multi-paths for each device i, $h_{i,g}$ denotes the channel gain of the gth path related to device i, and $d_{i,g}$ the propagation delay of the gth path related to device i. $N[n] = \Re\{N[n]\} + j\Im\{N[n]\}$ represents the circular symmetry complex Gaussian noise, where $\Re\{N[n]\} \sim \mathcal{N}(0, \sigma_N^2)$ and $\Im\{N[n]\} \sim \mathcal{N}(0, \sigma_N^2)$.

For the sake of simplicity, in the following we assume $P_{T_i} = 1$, $G_i = 1$, $h_{i,g} = 1$, and $d_{i,g} = 0, \forall i \in \mathcal{M}$. The impact of the multi-path propagation, the non-ideal channel conditions, and the propagation delay on the preamble and tag detection are analyzed in Appendix E.

We denote as SNR, the following Signal-to-Noise Ratio measured at the gNB receiver:

$$SNR = \frac{\sum_{n=0}^{N_{ZC}-1} \left| \sum_{i=1}^{M} x_{r,k_{p_i}}^{p_i,t_i}[n] \right|^2}{2\sigma_N^2}. \tag{10}$$

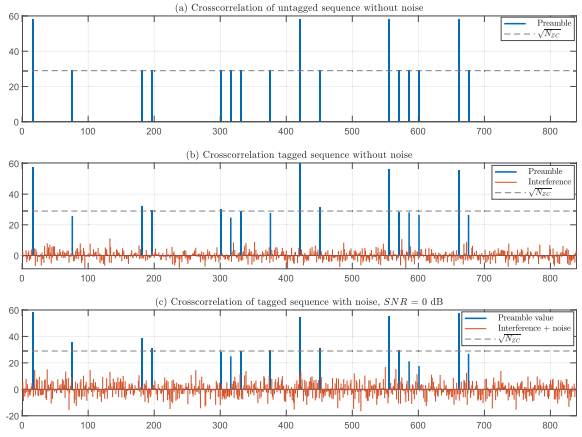


Fig. 2. Example of the real part of the correlation $C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau]$ when $M = 15$ and $N_{ZC} = 839$.

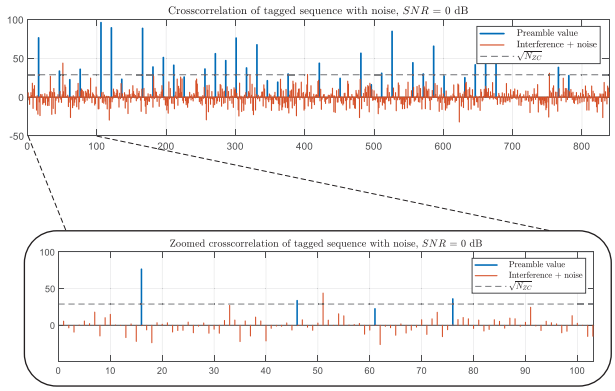


Fig. 3. Example of the real part of the correlation $C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau]$ when $M = 50$ devices attempt the access with tagged preambles and in the presence of noise.

Once the sequence (9) has been received by the gNB, it should detect the transmitted M preamble-tag pairs by performing some actions that can be divided into 2 steps.

Step 1. Since each device $i \in \mathcal{M}$ uses the same root number r to transmit its preamble sequence, the gNB computes the correlation value $C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau]$ between the total received sequence $y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}}[n]$ and the reference sequence (1), in order to detect all the transmitted preambles $p_s \in \mathcal{P}_\mathcal{S}$. In particular, by exploiting the property (i), we consider the real part of the correlation (3), denoted as $\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau]\right\}$.

At the aim of clarifying this step, we show in Fig. 2 an example of the correlation values assumed when $M = 15$. Fig. 2a shows $\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau]\right\}$ in the event that the sequences $x_{r,k_{p_i}}^{p_i,t_i}[n]$ were transmitted without tags, i.e., $z_{k_{p_i}}^{t_i}[n] = 0$, $\forall i \in \mathcal{M}$, and in the absence of noise, i.e., $N[n] = 0$. As expected, in each point $\tau = p_i N_{CS}$ the correlation value is exactly equal to $|\mathcal{M}_{p_i}|\sqrt{N_{ZC}}$, while 0 in all the other points. In Fig.2b each device transmits a tagged sequence, i.e., $z_{k_{p_i}}^{t_i}[n] \neq 0$, $\forall i \in \mathcal{M}$, in the absence of noise. Because of the property (ii), the sum of M cross-correlations produces non-zero values at each point n of the sequence with values that range, in general, in $[-M, M]$. We underline that these values are the interference that affects the ability to correctly detect the preambles. Consequently, not only we obtain non-zero values at the points where no preamble has been selected, but also at $\tau = p_i N_{CS}$ the cross-correlation value can be higher or lower than $|\mathcal{M}_{p_i}|\sqrt{N_{ZC}}$. Finally, in Fig.2c we consider the case in which also $SNR = 0$ dB. In this case, the interference plus noise causes the correlation $\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau]\right\}$ to assume values that could be very different from the ideal ones shown at the top of the figure.

The latter effect is even more evident as the attempting device number increases, as shown in Fig. 3, when $M = 50$. In particular, the zoom at the bottom shows that the correlation value in some points without any selected preamble could exceed the value $\sqrt{N_{ZC}}$ due to the interference, while the correlation value could be lower than $\sqrt{N_{ZC}}$ in some points related to a selected preamble. It can be easily understood that this effect affects the number of preambles correctly detected, both in terms of undetected transmitted preambles (i.e., false negative preambles) and in terms of detected non-transmitted preambles (i.e., false positive preambles). This phenomenon is the focus of next Section in order to obtain adequate probabilistic tools for the correct detection of the transmitted preambles.

Once the preamble detection procedure has been completed, the gNB obtains a new set of detected preambles denoted as $\mathcal{P}_\mathcal{D} \subseteq \mathcal{P}$, and in the ideal condition $\mathcal{P}_\mathcal{D} = \mathcal{P}_\mathcal{S}$.

Step 2. For each preamble $p_s \in \mathcal{P}_\mathcal{D}$ that has been detected by the gNB, a new correlation between the received sequence $y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}}[n]$ and the reference ZC sequence with root k_{p_s}, as function of p_s, should be calculated to detect all tags in $\mathcal{T}_{p_s}$. The aim is to check whether the preamble p_s has been selected by more than one device, i.e., to detect immediately whether it has collided. Also in this case, the noise and the interference due to the cross-correlations cause that the correlation values oscillate respect to the ideal values. In Section V, we study in detail this phenomenon and propose an analytical model.

B. An Example of the Detection Procedure

In Fig. 1a we illustrated an example of the first phase of the RA procedure. Each MTC device selects in a random way one preamble-tag pair (p_i, t_i). We note that the MTC device 1 selects the same preamble of MTC device 3, but a different tag. In Fig. 1b we show step 1 and 2 of the detection procedure in successful conditions, i.e., all preambles and all tags sent are correctly detected. During step 1, the gNB calculates the real part of the correlation between the received sequence $y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}}[n]$ and the reference sequence $z_r[n]$. In the point $p_1 N_{CS} = p_3 N_{CS}$ the correlation value is higher than the one assumed by the other preambles that have been selected once. After the gNB has detected all transmitted preambles, step 2 begins. It calculates, for each $p_s \in \mathcal{P}_\mathcal{D}$, the real part of the correlation between the received sequence and the sequence $z_{k_{p_s}}[n]$. We note that, since $p_1 = p_3$, two different tags

are found in the correlation with the sequence $z_{k_{p_1}} = z_{k_{p_3}}$, then the collision is detected. Instead, in the correlations with $z_{k_{p_2}}[n]$ and $z_{k_{p_M}}[n]$ only one tag is found, i.e., preambles p_2 and p_M were not collided.

IV. Modeling for the Preamble Detection

In this section, we derive the analytical model useful for the detection of the preambles transmitted in the presence of noise and interference due to the tagged sequences.

To detect whether a preamble $p \in \mathcal{P}$ has been selected by at least one device (i.e., if $p \in \mathcal{P}_\mathcal{S}$), the gNB calculates the correlation value between the received sequence (9) and the reference sequence (1). The correlation value is:

$$\begin{aligned} & C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau] \\ &= \frac{1}{\sqrt{N_{ZC}}} \sum_{n=0}^{N_{ZC}-1} y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}}[n] z_r^*[n+\tau] \\ &= \frac{1}{\sqrt{N_{ZC}}} \sum_{n=0}^{N_{ZC}-1} \left(\sum_{i=1}^{M} x_{r,k_{p_i}}^{p_i,t_i}[n] + N[n] \right) z_r^*[n+\tau] \\ &= \frac{1}{\sqrt{N_{ZC}}} \sum_{m=1}^{M} \sum_{n=0}^{N_{ZC}-1} x_{r,k_{p_i}}^{p_i,t_i}[n] z_r^*[n+\tau] + N[\tau]. \end{aligned} \quad (11)$$

Let us put

$$C_{z_r^{p_i},z_r}[\tau] = \frac{1}{\sqrt{N_{ZC}}} \sum_{n=0}^{N_{ZC}-1} z_r^{p_i}[n] z_r^*[n+\tau], \quad (12)$$

and

$$C_{z_{k_{p_i}}^{t_i},z_r}[\tau] = \frac{1}{\sqrt{N_{ZC}}} \sum_{n=0}^{N_{ZC}-1} z_{k_{p_i}}^{t_i}[n] z_r^*[n+\tau]. \quad (13)$$

Then:

$$C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau] = \sum_{i=1}^{M} (C_{z_r^{p_i},z_r}[\tau] + C_{z_{k_{p_i}}^{t_i},z_r}[\tau]) + N[\tau]. \quad (14)$$

Let us analyze the real part of (14):

$$\begin{aligned} \Re\left\{ C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau] \right\} &= \sum_{i=1}^{M} \Re\left\{ C_{z_r^{p_i},z_r}[\tau] \right\} \\ &\quad + \sum_{i=1}^{M} \Re\left\{ C_{z_{k_{p_i}}^{t_i},z_r}[\tau] \right\} + \Re\{N[\tau]\}. \end{aligned} \quad (15)$$

By applying property (3), it follows

$$\begin{aligned} \Re\left\{ C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau] \right\} &= \sum_{p_s \in \mathcal{P}_\mathcal{S}} |\mathcal{M}_{p_s}| \sqrt{N_{ZC}} \;\; \delta[\tau - p_s N_{CS}] \\ &\quad + \sum_{i=1}^{M} \Re\left\{ C_{z_{k_{p_i}}^{t_i},z_r}[\tau] \right\} + \Re\{N[n]\}. \end{aligned} \quad (16)$$

The term $\Re\left\{ C_{z_{k_{p_i}}^{t_i},z_r}[\tau] \right\}$ can be written as:

$$\Re\left\{ C_{z_{k_{p_i}}^{t_i},z_r}[\tau] \right\} = \frac{1}{\sqrt{N_{ZC}}} \sum_{n=0}^{N_{ZC}-1} \cos\left(\theta_{k_{p_i},r}^{t_i}[n,\tau] \right), \quad (17)$$

where $\theta_{k_{p_i},r}^{t_i}[n,\tau] = \frac{\pi}{N_{ZC}} \{ k_{p_i} (n + t_i N_{TS})_{N_{ZC}} ((n + t_i N_{TS})_{N_{ZC}} + 1)) - r(n+\tau)_{N_{ZC}} ((n+\tau)_{N_{ZC}} + 1)) \}$. At a general time instant τ, $\Re\left\{ C_{z_{k_{p_i}}^{t_i},z_r}[\tau] \right\}$ becomes a random variable, denoted in the following as H_i, for $i = 1, \ldots, M$ with expected value $\mu_{H_i} = 0$, and variance $\sigma_{H_i}^2 = \frac{1}{2}$.

Let us note that each random variable H_i depends on the pair (p_i, t_i) that has been selected by the device i. Since the probability that at least two devices select the same pair (p, t) can be neglected, as proved in Appendix A, the sequence of random variables $\{H_1, \ldots, H_M\}$ is a sequence of independent and identically distributed (i.i.d.) variables drawn from the same distribution with finite variance. Then, by exploiting the central limit theorem, the distribution of $H = \sum_{i=1}^{M} H_i$ approaches a Gaussian distribution with $\mu_H = M\mu_{H_i} = 0$ and $\sigma_H^2 = M\sigma_{H_i}^2 = \frac{M}{2}$, as M gets larger. Let us note that the above analysis would not work properly when the number of devices M is low; however, in this case, there would be no significant interference and collision issues. Instead, the target of this paper is to support an mMTC scenario, so the M value is enough large that the hypothesis of convergence of the central limit theorem is valid.

Assessing the relationship (16) as a whole, at each time $\tau' \neq p_s N_{CS}, \forall p_s \in \mathcal{P}_\mathcal{S}$, we note that the first addend of the second member of (16) is equal to 0, the second addend (which corresponds to H) follows the distribution $\mathcal{N}\left(0, \frac{M}{2}\right)$ and the third addend follows the distribution $\mathcal{N}\left(0, \sigma_N^2\right)$. Thus, the Cumulative Distribution Function (CDF) of the random variable $\Phi = \Re\left\{ C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau'] \right\}$, for $\tau' \neq p_s N_{CS}, \forall p_s \in \mathcal{P}_\mathcal{S}$, is equal to:

$$F_\Phi(\phi) = Pr\{\Phi \leq \phi\} = \frac{1}{2} \left[1 + \mathrm{erf}\left(\frac{\phi}{\sqrt{M + 2\sigma_N^2}} \right) \right], \quad (18)$$

where $\mathrm{erf}(\cdot)$ is the Gauss error function.

Now, let us analyze (16), when $\tau'' = p_s N_{CS}, \forall p_s \in \mathcal{P}_\mathcal{S}$. The first addend of the second member of (16) is equal to $|\mathcal{M}_{p_s}|\sqrt{N_{ZC}}$, while the second and the third one follow the distribution $\mathcal{N}\left(0, \frac{M}{2} + \sigma_N^2\right)$. Thus, the CDF of the random variable $\Phi_P = \Re\left\{ C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau''] \right\}$, for $\tau'' = p_s N_{CS}, \forall p_s \in \mathcal{P}_\mathcal{S}$, is equal to:

$$F_{\Phi_p}(\phi) = \frac{1}{2} \left[1 + \mathrm{erf}\left(\frac{\phi - |\mathcal{M}_{p_s}|\sqrt{N_{ZC}}}{\sqrt{M + 2\sigma_N^2}} \right) \right]. \quad (19)$$

V. Modeling for the Tag Detection

In this section, we derive the analytical model for the detection of the tags transmitted in the presence of noise and interference due to the other preambles and related tags. For each preamble $p_s \in \mathcal{P}_\mathcal{D}$ that has been correctly detected by the gNB, a new correlation between the received sequence $y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}}[n]$ and the reference ZC sequence with root k_{p_s}, as function of p_s, is calculated to detect tags in $\mathcal{T}_{p_s}$. The aim is to check whether the preamble p_s has been selected by more than one device, i.e., it has collided. The

cross-correlation is:

$$C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_{k_{p_s}}}[\tau]=\frac{1}{\sqrt{N_{ZC}}}\sum_{n=0}^{N_{ZC}-1} y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}}[n]z_{k_{p_s}}^*[n+\tau]. \quad (20)$$

Let us put

$$C_{z_{k_{p_i}}^{t_i},z_{k_{p_s}}}[\tau]=\frac{1}{\sqrt{N_{ZC}}}\sum_{n=0}^{N_{ZC}-1} z_{k_{p_i}}^{t_i}[n]z_{k_{p_s}}^*[n+\tau], \quad (21)$$

and

$$C_{z_r^{p_i},z_{k_{p_s}}}[\tau]=\frac{1}{\sqrt{N_{ZC}}}\sum_{n=0}^{N_{ZC}-1} z_r^{p_i}[n]z_{k_{p_s}}^*[n+\tau]. \quad (22)$$

Then

$$C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_{k_{p_s}}}[\tau]=\sum_{i=1}^{M} C_{z_{k_{p_i}}^{t_i},z_{k_{p_s}}}[\tau]+\sum_{i=1}^{M} C_{z_r^{p_i},z_{k_{p_s}}}[\tau]+N[\tau]. \quad (23)$$

Now, let us analyze the real part of (23), which can be rewritten as:

$$\begin{aligned}\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_{k_{p_s}}}[\tau]\right\} &= \sum_{t_i\in\mathcal{T}_{p_s}}\sqrt{N_{ZC}}\ \ \delta[\tau-t_iN_{TS}]\\ &+\sum_{i\in\mathcal{M}-\mathcal{M}_{p_s}}\Re\left\{C_{z_{k_{p_i}}^{t_i},z_{k_{p_s}}}[\tau]\right\}\\ &+\sum_{i=1}^{M}\Re\left\{C_{z_r^{p_i},z_{k_{p_s}}}[\tau]\right\}+\Re\{N[\tau]\}.\end{aligned} \quad (24)$$

Let us underline that in the first addend of the second member we have neglected the possibility that more than one device have selected the same preamble-tag pair, see Appendix A.

The term $\Re\left\{C_{z_{k_{p_i}}^{t_i},z_{k_{p_s}}}[\tau]\right\}$ can be written as:

$$\Re\left\{C_{z_{k_{p_i}}^{t_i},z_{k_{p_s}}}[\tau]\right\}=\frac{1}{\sqrt{N_{ZC}}}\sum_{n=0}^{N_{ZC}-1}\cos\left(\theta_{k_{p_i},k_{p_s}}^{t_i}[n,\tau]\right), \quad (25)$$

where $\theta_{k_{p_i},k_{p_s}}^{t_i}[n,\tau]=\frac{\pi}{N_{ZC}}\{k_{p_i}(n+t_iN_{TS})_{N_{ZC}}((n+t_iN_{TS})_{N_{ZC}}+1))-k_{p_s}(n+\tau)_{N_{ZC}}((n+\tau)_{N_{ZC}}+1))\}$. At a general time instant τ, $\Re\left\{C_{z_{k_{p_i}}^{t_i},z_{k_{p_s}}}[\tau]\right\}$ becomes a random variable denoted as I_i with expected value $\mu_{I_i}=0$, and variance $\sigma_{I_i}^2=\frac{1}{2}$. By exploiting the central limit theorem, and considering that no device has selected the same pair (p,t) of another device, the distribution $I=\sum_{i\in\mathcal{M}-|\mathcal{M}_{p_s}|} I_i$ approaches a Gaussian distribution with $\mu_I=0$ and $\sigma_I^2=\frac{1}{2}(M-|\mathcal{M}_{p_s}|)$, as M gets larger. The main problem of this result is that $|\mathcal{M}_{p_s}|$ is not known a priori. To overcome this issue, we replaced the variance σ_I^2 with its expected value $\mathbb{E}\{\sigma_I^2\}=\frac{M}{2}-\frac{\mathbb{E}\{|\mathcal{M}_{p_s}|\}}{2}$, where $\mathbb{E}\{|\mathcal{M}_{p_s}|\}$ becomes a known value given the number of devices M and the number of available preambles N_P, as proved in Appendix B. Under the above approximation, it follows that

$$I\sim\mathcal{N}\left(0,\frac{M}{2}-\frac{M-1}{2N_P}-\frac{1}{2}\right). \quad (26)$$

As regards the term $\Re\left\{C_{z_r^{p_i},z_{k_{p_s}}}[\tau]\right\}$ related to the third addend of (24), it can be written as:

$$\Re\left\{C_{z_r^{p_i},z_{k_{p_s}}}[\tau]\right\}=\frac{1}{\sqrt{N_{ZC}}}\sum_{n=0}^{N_{ZC}-1}\cos\left(\theta_{r,k_{p_s}}^{p_i}[n,\tau]\right), \quad (27)$$

where $\theta_{r,k_{p_s}}^{p_i}[n,\tau']=\frac{\pi}{N_{ZC}}\{r(n+p_iN_{CS})_{N_{ZC}}((n+p_iN_{CS})_{N_{ZC}}+1))-k_{p_s}(n+\tau)_{N_{ZC}}((n+\tau)_{N_{ZC}}+1))\}$. At a general time instant τ, $\Re\left\{C_{z_r^{p_i},k_{p_s}}[\tau]\right\}$ becomes a random variable denoted as L_{p_i}, with expected value $\mu_{L_{p_i}}=0$, and variance $\sigma_{L_{p_i}}^2=\frac{1}{2}$.

However, for the random variable $L=\sum_{i\in\mathcal{M}}L_{p_i}$ the hypotheses of the central limit theorem are no longer valid. In fact, in this case the value assumed by L_{p_i} depends only on the preamble p_i selected, and the probability that at least two devices have chosen the same preamble p_i cannot be neglected. Consequently, we can no longer consider all the variables L_{p_i}, with $i=1,\ldots,M$, as independent variables. To overcome this problem, we rewrite the third addend of the second member of (24) as follows:

$$\begin{aligned}L&=\sum_{i=1}^{M}\Re\left\{C_{z_r^{p_i},z_{k_{p_s}}}[\tau]\right\}=\sum_{p\in\mathcal{P}}|\mathcal{M}_p|\Re\left\{C_{z_r^{p},z_{k_{p_s}}}[\tau]\right\}\\ &=\frac{1}{\sqrt{N_{ZC}}}\sum_{p\in\mathcal{P}}\sum_{n=0}^{N_{ZC}-1}|\mathcal{M}_p|\cos\left(\theta_{r,k_{p_s}}^{p}[n,\tau]\right)\\ &=\sum_{p\in\mathcal{P}}|\mathcal{M}_p|L_p.\end{aligned} \quad (28)$$

where we imposed $L_p=\Re\left\{C_{z_r^{p},z_{k_{p_s}}}[\tau]\right\}$, with $p\in\mathcal{P}$. By applying (7) in (28), it follows:

$$L=\sum_{h=0}^{M}\sum_{p\in\mathcal{P}_h}hL_p=0+\sum_{p\in\mathcal{P}_1}L_p+\sum_{p\in\mathcal{P}_2}2L_p+\cdots\sum_{p\in\mathcal{P}_M}ML_p. \quad (29)$$

Each random variable hL_p is characterized by the expected value $\mu_{hL_p}=0$, and variance $\sigma_{hL_p}^2=\frac{h^2}{2}$. So, if each set $\mathcal{P}_h$ is either empty or with great cardinality, then the central limit theorem can be applied for each addend of (29), i.e., $\sum_{p\in\mathcal{P}_h}hL_p$ approaches a Gaussian distribution with mean value 0 and variance equal to $|\mathcal{P}_h|\frac{h^2}{2}$. The random variable L can be written as:

$$L=\sum_{h=1}^{M}\mathcal{N}\left(0,|\mathcal{P}_h|\frac{h^2}{2}\right)=\mathcal{N}\left(0,\frac{1}{2}\sum_{h=1}^{M}h^2|\mathcal{P}_h|\right). \quad (30)$$

Otherwise, if at least one set $\mathcal{P}_h$ has a low cardinality, then the central limit theorem cannot be applied for the related addend. In particular, we examine the worst case, that is, when the cardinality of each set is $\mathcal{P}_h=1$, with $h=1,\ldots,N<M$, and we obtain the sum of N independent random variables each one not equally distributed with the other ones. However, in Appendix C we verify that the sequence of independent random variables $\{L_p,2L_p,\ldots,NL_p\}$ satisfies the Lindeberg's condition [22]. Accordingly, the central limit theorem can be

applied, and the random variable $L = \sum_{h=1}^{N} hL_p$ follows $\mathcal{N}\left(0, \frac{1}{2}\sum_{h=1}^{L} h^2\right)$, if N is large enough.

Globally, the random variable $L \sim \mathcal{N}\left(0, \frac{1}{2}\sum_{h=1}^{M} |\mathcal{P}_h| h^2\right)$, whatever the size of the sets $\mathcal{P}_h$. However, $|\mathcal{P}_h|$ is not known a priori. To overcome this issue, we replaced the variance σ_L^2 with its expected value $\mathbb{E}\{\sigma_L^2\}$, that is a fixed and known value given the number of devices M and the number of available preambles N_P, as proved in Appendix D. So, by applying (59), the random variable L can be approximated as follows:

$$L \sim \mathcal{N}\left(0, \frac{1}{2}\left[\frac{M(M-1)}{N_P} + M\right]\right). \tag{31}$$

Assessing the relationship (24) as a whole, at each time $\tau' \neq t_i N_{TS}$, $\forall t_i \in \mathcal{T}_{p_s}$, we obtain a random variable denoted as $\Omega = \Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}}, z_{k_{p_s}}}[\tau']\right\}$, which assumes the values related to the interference and noise in the tag detection procedure. We note that in $\tau = \tau'$ the first addend of the second member is equal to 0, while the remaining addends are three independent Gaussian variables: I, L, and $\mathcal{N}(0, \sigma_N^2)$. Therefore, it follows that the random variable Ω is:

$$\Omega \sim \mathcal{N}\left(0, \frac{1}{2}\left[\frac{(M-1)^2}{N_P} + 2M - 1\right] + \sigma_N^2\right). \tag{32}$$

The CDF of Ω is equal to:

$$F_\Omega(\omega) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\omega}{\sqrt{\frac{(M-1)^2}{N_P} + 2M - 1 + 2\sigma_N^2}}\right)\right]. \tag{33}$$

Now, let us analyze (24), when $\tau'' = t_i N_{TS}$, $\forall t_i \in \mathcal{T}_{p_s}$. We obtain a random variable, denoted as Ω_T, which assumes the values related to the transmitted tags. It follows $\Omega_T = \Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}}, z_{k_{p_s}}}[\tau'']\right\} = \sqrt{N_{ZC}} + \Omega$. So, the CDF of Ω_T is equal to:

$$F_{\Omega_T}(\omega) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\omega - \sqrt{N_{ZC}}}{\sqrt{\frac{(M-1)^2}{N_P} + 2M - 1 + 2\sigma_N^2}}\right)\right]. \tag{34}$$

VI. Performance Analysis

In this Section, we show the accuracy of proposed analytical approach by comparing our results with those obtained by simulation in MATLAB Environment. The parameters adopted are enlisted in Table I. Simulations were run N_S times until all the results averaged up to the N_Sth simulation differ from those averaged up to the $(N_S - 1)$th simulation by less than 0.01%. Furthermore, the proposed model allows us to estimate the application limits (i.e., the working zone) of an efficient tag-preamble detector, in terms of maximum number of attempting devices M_{max} and SNR requirement at the gNB receiver.

TABLE I
Simulation Parameters

Parameter	Symbol	Value
Zandoff-Chu sequence length	N_{ZC}	839
Preamble Cyclic Shift	N_{CS}	13
Number of preambles available for the contention-based procedure	N_P	54
Tag Cyclic Shift	N_{TS}	13
Preamble root index	r	1
Tag root index	k_p	$N_{CS} \cdot p$
Total number of attempting MTC devices in one RA cycle	M	$1 : 100$[a]
Signal-Noise Ratio at the gNB receiver	SNR	$0 : 20$ dB
Detection probability	ϵ	$[0, 1]$

[a] These values are consistent with the massive MTC scenario adopted in [13].

A. Preamble Detection Analysis and Thresholds

Let us remember that the random variable Φ_p represents the correlation $\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}}, z_r}[\tau'']\right\}$ in a point $\tau'' = p_s N_{CS}$, $\forall p_s \in \mathcal{P}_\mathcal{S}$ (i.e., a point in which the preamble has been transmitted by at least one device), while $F_\Phi(\phi)$ represents the correlation $\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}}, z_r}[\tau']\right\}$ in a point $\tau' \neq p_s N_{CS}$, $\forall p_s \in \mathcal{P}_\mathcal{S}$ (i.e., a point in which no preamble has been transmitted).

Fig. 4a shows the comparison between the analytical CDF $F_\Phi(\phi)$, derived in (18), with the numerical results obtained by simulations, for $M = 25, 50, 75, 100$ and $SNR = 20$ dB; while Fig. 4b shows an analogous comparison for the CDF of Φ_P, derived in (19), when $|\mathcal{M}_{p_s}| = 1$, i.e., the worst case for the successful preamble detection. In both figures, the curves have the same trend, with mean value equal to 0 for Fig. 4a and to $\sqrt{N_{ZC}}$ for Fig. 4b, that are the ideal value assumed in the absence of noise and interference by Φ and Φ_P, respectively. It can be clearly observed that the results obtained by the model and by simulations are extremely similar for most of the ϕ values. However, in order to provide with an in-depth look at the accuracy of our model, in Table II the $F_\Phi(\phi)$ values obtained by simulation, by our analytic model and the error values are reported as an example for $M = 50$.

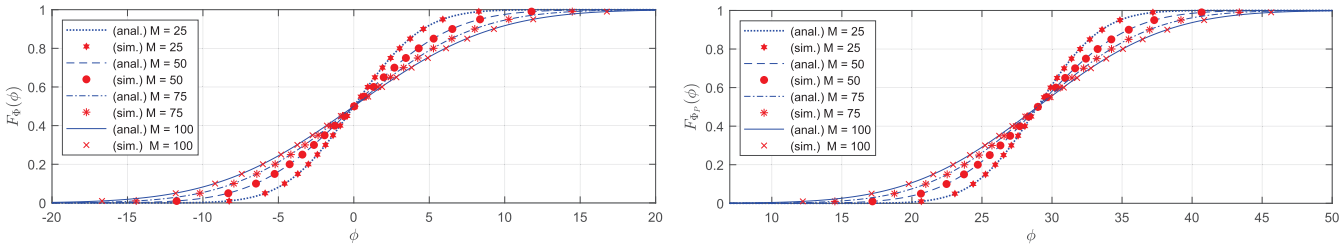
Once we have proven the accuracy of the proposed analytic model, we can use it to evaluate the effectiveness of preamble detection strategies as M and the SNR values change. Starting from the above results, in order to detect with probability ϵ the preambles $p_s \in \mathcal{P}_\mathcal{S}$, we define a proper threshold, T_P^ϵ, such that $Pr\left[\Phi_p \geq T_P^\epsilon\right] = \epsilon$ and, by using (19), we obtain:

$$1 - \frac{1}{2}\left[1 + \text{erf}\left(\frac{T_P^\epsilon - |\mathcal{M}_{p_s}|\sqrt{N_{ZC}}}{\sqrt{M + 2\sigma_N^2}}\right)\right] = \epsilon. \tag{35}$$

Given the probability ϵ, since the preamble detection probability should be respected even in the worst case, i.e., $|\mathcal{M}_{p_s}| = 1$, the threshold T_P^ϵ is computed in this case, as:

$$T_P^\epsilon = \sqrt{N_{ZC}} - \sqrt{M + 2\sigma_N^2}\ \ \text{erf}^{-1}(2\epsilon - 1). \tag{36}$$

Let us note that in a generic point $\tau' \neq p_s N_{CS}$, $\forall p_s \in \mathcal{P}_\mathcal{S}$, a false positive preamble could be detected if the interference



(a) CDF of Φ.

(b) CDF of Φ_P (when $|\mathcal{M}_{p_s}| = 1$).

Fig. 4. CDFs of Φ_P (when $|\mathcal{M}_{p_s}| = 1$) and Φ obtained by simulation and our model, for several M values, with $SNR = 20$dB.

TABLE II

SIMULATED VS ANALYTICAL RESULTS OF $F_\Phi(\phi)$, WHEN $M = 50$ AND $SNR = 20$ dB

Percentile	Simulation	Analytical	Error
0.99	11.7772	11.8003	-0.0231
0.95	8.3655	8.3435	0.0220
0.90	6.5271	6.5006	0.0265
0.85	5.2885	5.2573	0.0312
0.80	4.3001	4.2691	0.0310
0.75	3.4509	3.4213	0.0295
0.70	2.6894	2.6600	0.0294
0.65	1.9822	1.9545	0.0277
0.60	1.3101	1.2851	0.0250
0.55	0.6628	0.6374	0.0254
0.50	0.0200	0	0.0200
0.45	-0.6144	-0.6374	0.0230
0.40	-1.2660	-1.2851	0.0191
0.35	-1.9378	-1.9545	0.0167
0.30	-2.6432	-2.6600	0.0168
0.25	-3.4084	-3.4213	0.0130
0.20	-4.2533	-4.2691	0.0158
0.15	-5.2476	-5.2573	0.0097
0.10	-6.4842	-6.5006	0.0164
0.05	-8.3190	-8.3435	0.0245
0.01	-11.7423	-11.8003	0.0580

TABLE III

SIMULATED VS ANALYTICAL RESULTS OF $F_\Omega(\omega)$, WHEN $M = 50$ AND $SNR = 20$ dB

Percentile	Simulation	Analytical	Error
0.99	19.7529	19.8029	-0.0500
0.95	14.0663	14.0017	0.0646
0.90	10.9918	10.9091	0.0827
0.85	8.9039	8.8226	0.0814
0.80	7.2377	7.1642	0.0734
0.75	5.7996	5.7415	0.0581
0.70	4.5086	4.4639	0.0447
0.65	3.3115	3.2800	0.0315
0.60	2.1704	2.1566	0.0138
0.55	1.0690	1.0697	-0.0007
0.50	-0.0177	0	-0.0177
0.45	-1.1000	-1.0697	-0.0303
0.40	-2.2035	-2.1566	-0.0469
0.35	-3.3401	-3.2800	-0.0601
0.30	-4.5350	-4.4639	-0.0711
0.25	-5.8231	-5.7415	-0.0816
0.20	-7.2520	-7.1642	-0.0878
0.15	-8.9152	-8.8226	-0.0927
0.10	-10.9853	-10.9091	-0.0761
0.05	-14.0311	-14.0017	-0.0294
0.01	-19.6485	-19.8029	0.1544

assumes high values. For this reason, in order to non-detect false positive preambles with probability ϵ, we also define for the other variable Φ a threshold, denoted as T_{IP}^{ϵ}, that is, $Pr\left[\Phi \leq T_{IP}^{\epsilon}\right] = \epsilon$. Then, given the probability ϵ, the threshold T_{IP}^{ϵ} can be calculated, by using (18), as $T_{IP}^{\epsilon} = \sqrt{N_{ZC}} - T_P^{\epsilon}$.

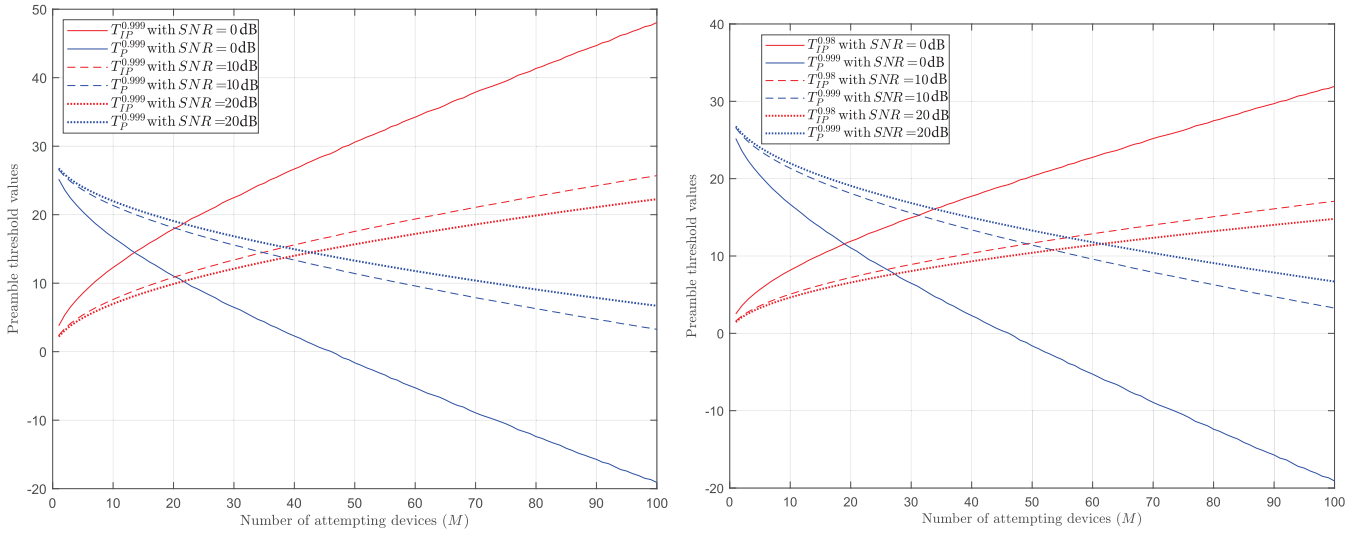
In Figs. 5a we show the thresholds T_P^{ϵ} and T_{IP}^{ϵ} vs the number of devices M, with different SNR values, when $\epsilon = 0.999$. Let us analyze Fig. 5a. $T_{IP}^{0.999}$ is the minimum threshold which guarantees a true negative preamble probability at least equal to 0.999 in a point $\tau' \neq p_s N_{CS}$, $\forall p_s \in \mathcal{P}_{\mathcal{S}}$, while $T_P^{0.999}$ is the maximum threshold which guarantees a true positive preamble probability at least equal to 0.999 in a point $\tau'' = p_s N_{CS}$, $\forall p_s \in \mathcal{P}_{\mathcal{S}}$ with $|\mathcal{M}_{p_s}| = 1$.

As expected, for guaranteeing a given true positive preamble probability, the $T_P^{0.999}$ threshold is a monotonically increasing function as the number of devices (and, consequently, the interference) increases. An opposite behavior occurs for the threshold $T_{IP}^{0.999}$ to guarantee a given true negative preamble probability.

In order to ensure a very low detection error probability, the detection strategy should ensure that both true positive preamble and true negative preamble probabilities are high, that is, in a generic instant time τ, it should be $T_{IP}^{\epsilon} \leq C_{y_{r,k_{\mathcal{P}_{\mathcal{S}}}}^{\mathcal{P}_{\mathcal{S}},\mathcal{T}_{\mathcal{S}}},z_r}[\tau] \leq T_P^{\epsilon}$, with ϵ approaching 1. Accordingly, it is possible to derive a valid working zone as long as $T_P^{\epsilon} \geq T_{IP}^{\epsilon}$, that is, up to the point of intersection between the two curves. This working zone corresponds to a maximum number of attempting devices M_{max} which can be effectively managed by the detector. As depicted in Fig. 5a, it corresponds to $M_{max} = 13$ when $SNR = 0$ dB, $M_{max} = 35$ when $SNR = 10$ dB, and $M_{max} = 43$ when $SNR = 20$ dB. We note that, for high SNR values, the contribution of interference dominates with respect to the noise.

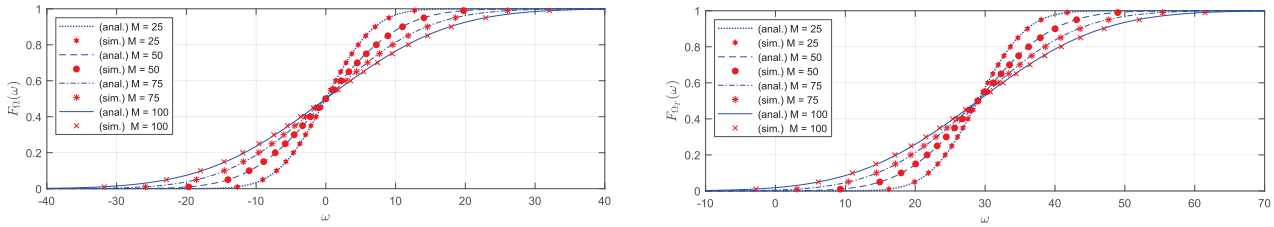
Clearly, this approach is also valid for thresholds with different probability requirements. In fact, on basis of a given RA procedure strategy, the effect of a false positive preamble could be less severe than a false negative preamble. For instance, in Fig. 5b we show the curves related to a probability of true positive preamble equal to 0.999 and a probability of true negative preamble equal to 0.98. It results $M_{max} = 19$ when $SNR = 0$ dB, $M_{max} = 49$ when $SNR = 10$ dB, and $M_{max} = 61$ when $SNR = 20$ dB.

The above analysis shows that the effective working zones could be small and, generally, depend on the pair of values (M, SNR). This result suggests that a adaptive detection strategy may be needed (e.g., based on a dynamic threshold value that adapts to the current traffic load M and SNR conditions). However, if we consider high SNR values (e.g.



(a) $T_P^{0.999}$ and $T_{IP}^{0.999}$ (b) $T_P^{0.999}$ and $T_{IP}^{0.98}$

Fig. 5. Preamble thresholds vs the number of attempting devices M, with different SNR values.



(a) CDF of Ω. (b) CDF of Ω_T.

Fig. 6. CDFs of Ω and Ω_T obtained by simulation and by our model, for several M values, with $SNR = 20$ dB.

$SNR > 10$ dB) the threshold values depend substantially only on the value of M (i.e., the contribution of interference dominates with respect to the noise). In this case, a fixed threshold value can be considered as a function of M_{max}, and an appropriate ACB factor can be adopted to limit the maximum number of attempting devices to M_{max}. Therefore, in a scenario where a massive number of access attempts are expected, it may be necessary to carry out adequate cellular planning, and/or transmission power control to maintain a high SNR value on the receiver.

B. Tag Detection Analysis and Thresholds

The analysis carried out here is similar to that of the previous subsection. In fact, we begin by comparing the analytical CDFs $F_\Omega(\omega)$, derived in (33), and $F_{\Omega_T}(\omega)$, derived in (34), with the numerical results obtained by simulation. Let us remember that the random variable Ω_T represents the correlation $\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}},\mathcal{T}_\mathcal{S},z_{k_{p_s}}}[\tau'']\right\}$ in a point $\tau'' = t_i N_{TS}$, in which the preamble $t_i \in \mathcal{T}_{p_s}$ has been transmitted by at least one device, while $F_\Omega(\omega)$ represents the correlation $\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}},\mathcal{T}_\mathcal{S},z_{k_{p_s}}}[\tau']\right\}$ in a point $\tau' \neq t_i N_{CS}$, $\forall t_i \in \mathcal{T}_{p_s}$, in which no tag has been transmitted.

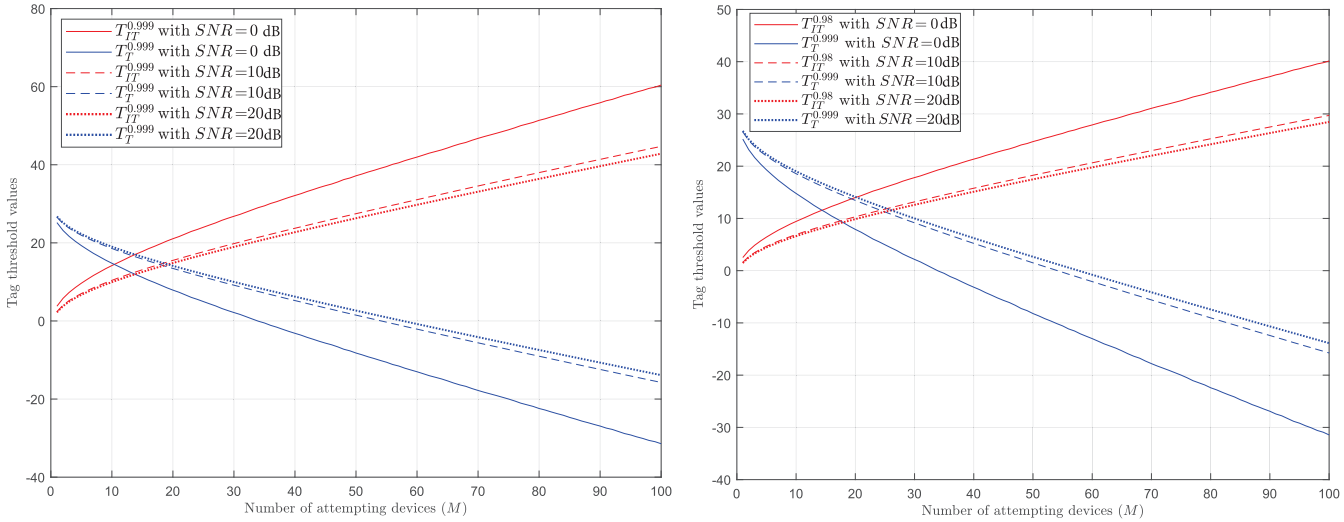
In Fig. 6a we illustrate the comparison between the analytical and the simulation results of $F_\Omega(\omega)$ with $M = 25, 50, 75, 100$ with $SNR = 20$ dB, whereas in Fig. 6b we show $F_{\Omega_T}(\omega)$. Similarly to the previously analyzed preamble CDFs, both curves have the same trend, with mean value equal to 0 for Fig. 6a and to $\sqrt{N_{ZC}}$ for Fig. 6b, that are the ideal values assumed by Ω and Ω_T, respectively. Also, since the results obtained by our analytical model and by simulation are very similar, in Table III we show, as example, the values of $F_\Omega(\omega)$ for $M = 50$.

Having proved the accuracy of the proposed analytic model related to the tag detection, we use it to evaluate the effectiveness of tag detection strategies by deriving proper thresholds at the aim of discerning the tags from the interference. Therefore, we define the threshold T_T^ϵ, with the aim of detecting a tag $t_i \in \mathcal{T}_{p_s}$ with a probability equal to ϵ, and a threshold, T_{IT}^ϵ, with the aim of non-detecting an interference point as a tag with a probability equal to ϵ. In a similar way to what described for the preambles, given a probability ϵ, the threshold T_{IT}^ϵ can be calculated as:

$$T_{IT}^\epsilon = \sqrt{\frac{(M-1)^2}{N_P} + 2M - 1 + 2\sigma_N^2}\ \ \mathrm{erf}^{-1}(2\epsilon - 1), \quad (37)$$

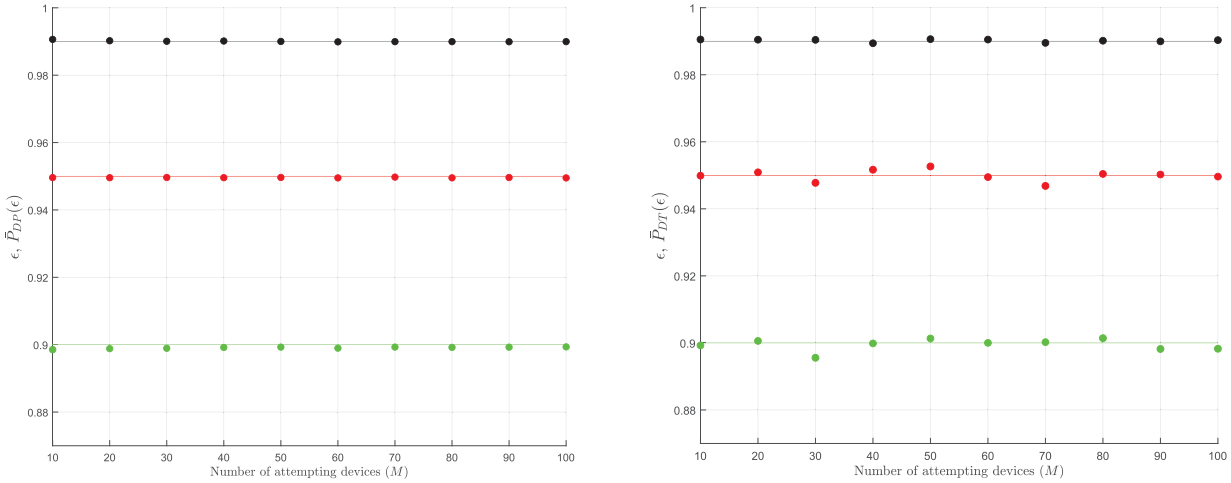
whereas the threshold T_T^ϵ can be calculated as $T_T^\epsilon = \sqrt{N_{ZC}} - T_{IT}^\epsilon$.

In Figs. 7a we show the thresholds T_T^ϵ and T_{IT}^ϵ, when $\epsilon = 0.999$ vs the number of devices M, with different SNR values. In order that $T_{IT}^\epsilon \leq C_{y_{r,k_{\mathcal{P}_\mathcal{S}}},\mathcal{T}_\mathcal{S},z_r}[\tau] \leq T_T^\epsilon$ (i.e., $T_{IT}^\epsilon \leq T_T^\epsilon$), we get $M_{max} = 10$ when $SNR = 0$ dB, $M_{max} = 18$



(a) $T_T^{0.999}$ and $T_{IT}^{0.999}$ (b) $T_T^{0.999}$ and $T_{IT}^{0.98}$

Fig. 7. Tag thresholds vs the number of attempting devices M, with different SNR values.



(a) $\bar{P}_{DP}(\epsilon)$ (b) $\bar{P}_{DT}(\epsilon)$

Fig. 8. Simulation results for $\epsilon \in \{0.90; 0.95; 0.99\}$, with $SNR = 10$ dB.

when $SNR = 10$ dB, and $M_{max} = 19$ when $SNR = 20$ dB. Also for the tag detector, on basis of the access procedure strategy, the effect of a false positive tag could be less severe than a false negative tag. For instance, in Fig. 7b we show the curves related to a probability of true positive tag equal to 0.999 and a probability of true negative tag equal to 0.98. It results $M_{max} = 14$ when $SNR = 0$ dB, $M_{max} = 24$ when $SNR = 10$ dB, and $M_{max} = 26$ when $SNR = 20$ dB.

Compared to preamble detection, the tag detection has a more limited working zone, due to more severe interference. This means that, with the same preamble and tag detection probability, the working zone is constrained by the tag detection. So, a possibility to increase the working zone is to adopt a less stringent tag detection probability than the preamble probability. All the considerations made at the end of the previous subsection are valid also here for the tag detection. However, we highlight that the working areas valid for tag detection are unfortunately very small, and in general a straightforward strategy is not suitable for a massive MTC scenario. The overall working zones analysis suggests to investigate new optimized strategies for tagged preamble sequence detection.

C. Assessment of the Analytical Thresholds Accuracy

For the sake of completeness, in this subsection, we evaluate by simulation the accuracy of the thresholds derived from the analytical model in subsection VI-A and VI-B. In this way, the accuracy of working zones is also verified.

We denote with $\bar{P}_{PD}(\epsilon)$ the average percentage of detected preambles above the threshold T_P^{ϵ}, and with $\bar{P}_{DT}(\epsilon)$ the average percentage of detected tags above the threshold T_T^{ϵ}. In Fig. 8 we show $\bar{P}_{DP}(\epsilon)$ and $\bar{P}_{DT}(\epsilon)$, for $\epsilon \in \{0.90, 0.95, 0.99\}$ with $SNR = 10$ dB. As regards $\bar{P}_{DP}(\epsilon)$, the simulation results demonstrate that the average percentage of detected preambles above the T_P^{ϵ} threshold value derived by the analytical model is very close to the expected ϵ value. Similar results were obtained with other SNR values. As for $\bar{P}_{DT}(\epsilon)$, we can see that the accuracy is slightly lower compared to the model

used for the preambles. This result was expected because the model adopted for the tags introduces some approximations. In particular, the variances σ_I^2 and σ_L^2 have been approximated with their average value. However, the variation of the simulation results with respect to the theoretical expected values is very small and, consequently, the rightness of the working zones identified in the previous subsections is confirmed.

Similar considerations are valid for the percentage of interference/noise points below the thresholds T_{IP}^{ϵ} and T_{IT}^{ϵ}, but the related graphs are not reported for space reasons.

D. Impact of Failed Preamble-Tag Detections on the Signaling Overhead and Energy Consumption

In this subsection, we analyze the performance of the detector inside the gNB receiver, when a simple preamble-tag detection strategy and the two-step RA procedure [13] are applied. In particular, we evaluate how much the erroneous detections impact on the signaling overhead and on the extra energy consumption experienced by the MTC devices, under different preamble and tag detection probabilities, ϵ_P and ϵ_T, respectively.

Given a RA cycle, for each preamble $p \in \mathcal{P}$, the detector may either detect the preamble (i.e., $p \in \mathcal{P}_{\mathcal{D}}$), or not. In the first case, thanks to the tags, the preamble can be classified either as successfully transmitted (i.e., $p \in \mathcal{P}_{\mathcal{SD}}$), or as collided (i.e., $p \in \mathcal{P}_{\mathcal{CD}}$). Obviously, $\mathcal{P}_{\mathcal{SD}} \cup \mathcal{P}_{\mathcal{CD}} = \mathcal{P}_{\mathcal{D}}$ and $\mathcal{P}_{\mathcal{SD}} \cap \mathcal{P}_{\mathcal{CD}} = \emptyset$. Let us define the following disjoint failure events for the detector.

Event A: The MTC device has selected a preamble $p_s \in \mathcal{P}_{\mathcal{S}}$ that was successfully transmitted (i.e., $|\mathcal{M}_{p_s}| = 1$), but it was either not detected or classified as collided (i.e., $p_s \in (\mathcal{P} - \mathcal{P}_{\mathcal{D}}) \cup \mathcal{P}_{\mathcal{CD}}$).

Event B: The MTC device has selected a preamble $p_s \in \mathcal{P}_{\mathcal{S}}$ that was collided (i.e., $|\mathcal{M}_{p_s}| \geq 2$), and it was classified by the detector as successfully transmitted (i.e., $p_s \in \mathcal{P}_{\mathcal{SD}}$).

Event C: The MTC device has selected a preamble $p_s \in \mathcal{P}_{\mathcal{S}}$ that was a collided preamble and it was not detected by the detector (i.e., $p_s \notin \mathcal{P}_{\mathcal{D}}$).

Event D: No MTC device has selected the preamble p (i.e., $p \in \mathcal{P} - \mathcal{P}_{\mathcal{S}}$), but it was detected as sent by the detector, (i.e. $p \in \mathcal{P}_{\mathcal{D}}$).

As regards the MTC device experiencing event A, it is not affected by any additional signaling overhead. However, it is affected by an additional energy consumption, denoted as E_A, due to the preamble transmission and the permanence in the RX active state waiting for the RAR message. As for the MTC device experiencing Event B, it is subject to a signaling overhead of O_B bytes related to the transmission of the data packet including the device ID, and an additional energy consumption, E_B, due to the data packet transmission and the waiting for the ACK message. As regards Events C and D, they do not involve any increase in both signaling overhead and energy consumption from the point of view of the MTC device.

Let us denote with $Pr\{A\}$ and $Pr\{B\}$ the probability that an attemping MTC device is experiencing Event A and Event B, respectively. The additional average energy consumption per access attempt is $E_{add} = Pr\{A\}E_A + Pr\{B\}E_B$, whereas the additional signaling overhead is $O_{add} = Pr\{B\}O_B$ bytes. The values of $Pr\{A\}$ and $Pr\{B\}$ depend on the detection strategy adopted and on ϵ_P and ϵ_T values.

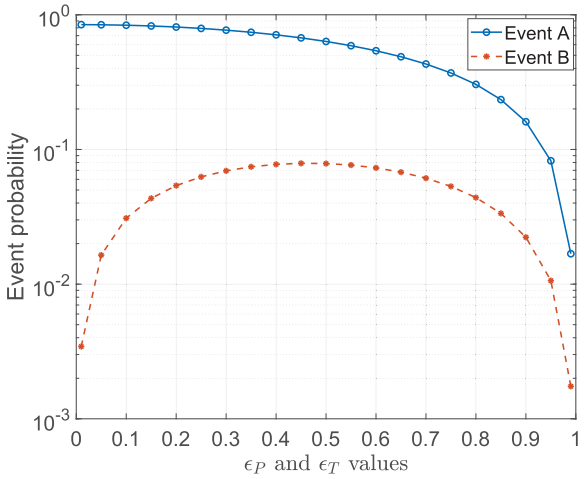


Fig. 9. Variation of $Pr\{A\}$ and $Pr\{B\}$ with respect to ϵ_P and ϵ_T, with $SNR = 20$ dB, for $M = 10$.

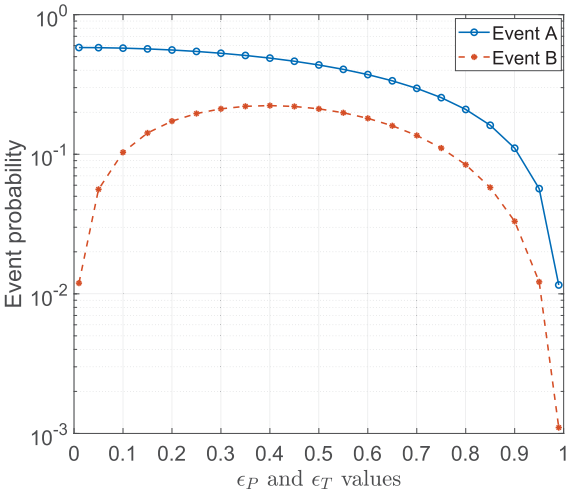


Fig. 10. Variation of $Pr\{A\}$ and $Pr\{B\}$ with respect to ϵ_P and ϵ_T, with $SNR = 20$ dB, for $M = 30$.

As regards the preamble detection strategy, we note that the detector does not know a priori whether a point τ of $\Re\left\{C_{y_{r,k_{\mathcal{P}_{\mathcal{S}}}},z_r}[\tau]\right\}$ is an interference point (i.e., $\tau \neq p_s N_{CS}$) or not. So, for any point τ, a simple strategy is to apply a unique threshold $T_P^{\epsilon_P}$ in order to detect the transmitted preambles. The same consideration applies to the tag detection, and we assume the unique threshold $T_T^{\epsilon_T}$ for each point of $\Re\left\{C_{y_{r,k_{\mathcal{P}_{\mathcal{S}}}},z_{k_{p_s}}}[\tau]\right\}$.

Then, in Figs. 9 and 10 we show the variation of $Pr\{A\}$ and $Pr\{B\}$ with respect to ϵ_P and ϵ_T, with $SNR = 20$ dB, for $M = 10$ and $M = 30$, respectively. For low values of ϵ_P and ϵ_T, the $Pr\{A\}$ value is high and $Pr\{B\}$ is low because there is a high probability that the preamble is not detected. As ϵ_P and ϵ_T increase to about 0.5, $Pr\{B\}$ increases because

the probability of detecting a preamble increases, but the probability of correctly detecting the tags remains low (less than 0.5). Obviously, for $\epsilon_P = \epsilon_T > 0.5$, $Pr\{A\}$ and $Pr\{B\}$ decrease. Comparing Fig. 9 with 10, it can be seen that for $M = 10$ the $Pr\{A\}$ curve is higher than the one for $M = 30$, whereas the reverse holds true for $Pr\{B\}$, because for $M = 10$ the probability that a MTC device transmits a preamble with success is higher and the collision probability is lower.

Finally, as a case study, to quantify the signaling overhead and energy consumption of the MTC device, we consider the two-step RA procedure and the energy model proposed in [13]. The MTC device that experiences Event A will consume 3 mJ for transmitting the preamble, and 1.277 mJ for permaning in the RX active state during the RAR message window, i.e., $E_A = 4.277$ mJ. As regards the MTC device experiencing Event B, it consumes the same amount of energy to transmit the data packet and to wait for the relative ACK reception, i.e., $E_B = E_A$. In addition, there will be a signaling overhead of $O_B = 9$ bytes, related to the additional information piggybacked with the data. As example, if $\epsilon_P = \epsilon_T = 0.95$ and $M = 10$, it results $E_{add} = 0.3978$ mJ and $O_{add} = 0.095$ bytes, whereas with $M = 30$, it follows $E_{add} = 0.2947$ mJ and $O_{add} = 0.11$ bytes.

VII. Conclusion

In this paper, we have presented a rigorous methodology for modeling and analyzing the signal processed by the gNB receiver, when tagged preamble sequences are transmitted at the first phase of the RA procedure. More specifically, we have derived the expression of the probability distributions of the related variables, in closed form, in presence of additive white Gaussian noise and interference due to other preambles and tags. Under several traffic conditions, we have assessed the accuracy of proposed models compared to simulation results. We used our analytical approach to derive the threshold values to detect both the preamble and the tag with a given probability, and to determine the working zones where a basic preamble detection strategy can well operate. The high accuracy obtained allows the researcher, for example, to investigate and elaborate optimized detection strategies or to evaluate the implications of different ACB factor values without carrying out a large number of simulations, which are typically highly time-consuming in a massive scenario.

We emphasize that the detection strategy for the tag-preamble pairs adopted by the gNB receiver plays a fundamental role on the performance of the contention-based access procedures. For example, by considering a simple detection strategy and the 2-phase RA procedure, our models allowed us to estimate the relationship between the tag-preamble pair detection probabilities and both the signaling overhead and the energy consumption per access attempt from the MTC device's perspective, under different traffic loads.

In conclusion, our analytical approach is a powerful tool that can be easily adopted to investigate and to propose new and effective strategies of tagged preamble detection and, accordingly, to evaluate innovative and efficient RA strategies tailored for the mMTC scenario.

Appendix A
Probability That at Least Two Devices Select the Same Pair (p, t)

Each device i, with $i = 1, \ldots, M$, extracts independently its own preamble p_i and its own tag t_i. The probability that a device selects the pair (p, t), with $p \in \mathcal{P}$ and $t \in \mathcal{T}$ is equal to $(N_P N_T)^{-1}$. The probability that the pair (p, t) has been selected by a number of devices equal to h, with $h = 0, 1 \ldots, M$ follows the binomial distribution:

$$Pr\left(|\mathcal{M}_{(p,t)}| = h\right) = \binom{M}{h} \left(\frac{1}{N_P N_T}\right)^h \left(1 - \frac{1}{N_P N_T}\right)^{M-h}. \quad (38)$$

So, the probability that at least two devices select the same pair (p, t) is:

$$Pr\left(|\mathcal{M}_{(p,t)}| \geq 2\right) = 1 - \left(1 - \frac{1}{N_P N_T}\right)^M + - M \left(\frac{1}{N_P N_T}\right) \left(1 - \frac{1}{N_P N_T}\right)^{M-1}. \quad (39)$$

This probability increases with M, therefore considering the values in Table I, the maximum value that (39) could assume is $4.07 \cdot 10^{-4}$ for $M = 100$. For this reason, the occurrence of this event has been neglected in our analysis model.

Appendix B
Calculation of $\mathbb{E}\{|\mathcal{M}_{p_s}|\}$

$|\mathcal{M}_{p_s}|$ is the number of devices that have selected the preamble p_s, given that the preamble p_s has been selected by at least one device. Then, the probability that $|\mathcal{M}_{p_s}| = h$ results:

$$Pr\{|\mathcal{M}_{p_s}| = h\} = \binom{M-1}{h-1} \left(\frac{1}{N_P}\right)^{h-1} \left(1 - \frac{1}{N_P}\right)^{M-h}, \quad (40)$$

with $h = \{1, \ldots, M\}$. The expected value of $|\mathcal{M}_{p_s}|$ is defined as:

$$\mathbb{E}\{|\mathcal{M}_{p_s}|\} = \sum_{h=1}^{M} h Pr\left\{|\mathcal{M}_{p_s}| = h\right\}. \quad (41)$$

Let us put $j = h - 1$, $n = M - 1$, $p = \frac{1}{N_P}$, and $q = 1 - p$. Then:

$$\mathbb{E}\{|\mathcal{M}_{p_s}|\} = \sum_{j=0}^{n} (j+1) \binom{n}{j} p^j q^{n-j} = \sum_{j=0}^{n} j \binom{n}{j} p^j q^{n-j} + \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j}. \quad (42)$$

By exploiting the binomial formula, it follows:

$$\mathbb{E}\{|\mathcal{M}_{p_s}|\} = \sum_{j=0}^{n} j \binom{n}{j} p^j q^{n-j} + (p+q)^n. \quad (43)$$

Since the first addend of the second member is the average value of a binomial distribution with parameters n and p, and the second added is equal to 1, it follow:

$$\mathbb{E}\{|\mathcal{M}_{p_s}|\} = np + 1 = \frac{M-1}{N_P} + 1. \tag{44}$$

APPENDIX C
LINDEBERG'S CONDITION

In this section, we verify that the sequence of N independent random variables $\{L_p, 2L_p, \ldots, NL_p\}$ satisfies the Lindeberg's condition.

Let $s_N^2 = \sum_{h=1}^{N} \sigma_{hL_p}^2 = \sum_{h=1}^{N} \frac{h^2}{2}$. The Lindeberg's condition is the following:

$$\lim_{N\to\infty} \frac{1}{s_N^2} \sum_{h=1}^{N} \mathbb{E}[(hL_p - \mu_{hL_p})^2 \cdot \mathbf{1}\{|hL_p - \mu_{hL_p}| > \varepsilon s_N\}] = 0 \tag{45}$$

for all $\varepsilon > 0$, where $\mathbf{1}\{\cdot\}$ is the indicator function.

Let us analyze the term $\mathbb{E}[(hL_p - \mu_{hL_p})^2 \cdot \mathbf{1}\{|hL_p - \mu_{hL_p}| > \varepsilon s_N\}]$. In our case, it becomes:

$$\mathbb{E}\left[(hL_p)^2 \cdot \mathbf{1}\{|hL_p| > \varepsilon s_N\}\right] h^2 \mathbb{E}\left[L_p^2 \cdot \mathbf{1}\{h|L_p| > \varepsilon s_N\}\right], \tag{46}$$

where

$$\mathbf{1}\{h|L_p| > \varepsilon s_N\} = \begin{cases} 1 & \text{if } h|L_p| > \varepsilon s_N \\ 0 & \text{otherwise.} \end{cases} \tag{47}$$

Let us analyze the indicator function:

$$h|L_p| > \varepsilon s_N \Rightarrow h|L_p| > \varepsilon \sqrt{\sum_{h=1}^{N} \frac{h^2}{2}} \Rightarrow h^2 L_p^2 > \frac{\varepsilon^2}{2} \sum_{h=1}^{N} h^2. \tag{48}$$

By exploiting the Faulhaber's formula [23], it follows:

$$h^2 L_p^2 > \frac{\varepsilon^2}{2} \frac{N(N+1)(2N+1)}{6}. \tag{49}$$

Since $0 \le L_p^2 \le 1$ and $1 \le h \le N$, when N approaches infinity, the second member goes to infinity faster than the first one. Therefore, for N large enough, it follows $\mathbb{E}\left[L_p^2 \cdot \mathbf{1}\{h|L_p| > \varepsilon s_N\}\right] = 0$, and the Lindeberg's condition (45) is verified.

APPENDIX D
CALCULATION OF THE AVERAGE VALUE OF σ_L^2

$$\mathbb{E}\{\sigma_L^2\} = \mathbb{E}\left\{\frac{1}{2}\sum_{h=1}^{M} h^2 |\mathcal{P}_h|\right\} = \frac{1}{2}\sum_{h=1}^{M} h^2 \mathbb{E}\left\{|\mathcal{P}_h|\right\}, \tag{50}$$

where $\mathbb{E}\{|\mathcal{P}_h|\}$ is the average number of preambles that have been selected by h out of M devices. We define the random variables X_h^j, with $h \in \{1, \ldots, M\}$, and $j \in \{1, \ldots, N_P\}$ as:

$$X_h^j = \begin{cases} 1 & \text{if preamble } j \text{ has been selected by } h \text{ devices} \\ 0 & \text{otherwise.} \end{cases} \tag{51}$$

So, the number of preambles selected by h devices can be written as:

$$|\mathcal{P}_h| = \sum_{j=1}^{N_P} X_h^j, \tag{52}$$

and the mean value results:

$$\mathbb{E}\{|\mathcal{P}_h|\} = \mathbb{E}\left\{\sum_{j=1}^{N_P} X_h^j\right\} = \sum_{h=1}^{N_P} \mathbb{E}\{X_h^j\}. \tag{53}$$

Since X_h^j are N_P random variables identically distributed, the mean value of X_h^j, for each j, is equal to:

$$\begin{aligned} \mathbb{E}\{X_h^j\} &= \Pr(X_h^j = h) \\ &= \binom{M}{h} \left(\frac{1}{N_P}\right)^h \left[1 - \left(\frac{1}{N_P}\right)\right]^{M-h}. \end{aligned} \tag{54}$$

So:

$$\begin{aligned} \mathbb{E}\{|\mathcal{P}_h|\} &= N_P \mathbb{E}\{X_h^j\} \\ &= N_P \binom{M}{h} \left(\frac{1}{N_P}\right)^h \left[1 - \left(\frac{1}{N_P}\right)\right]^{M-h}. \end{aligned} \tag{55}$$

Equation (50) becomes:

$$\mathbb{E}\{\sigma_L^2\} = \frac{1}{2} N_P \sum_{h=1}^{M} h^2 \binom{M}{h} \left(\frac{1}{N_P}\right)^h \left[1 - \left(\frac{1}{N_P}\right)\right]^{M-h}. \tag{56}$$

Let us put $p = \frac{1}{N_P}$, $q = 1 - p$, and substitute h^2 with $h(h-1) + h$. Then:

$$\begin{aligned} \mathbb{E}\{\sigma_L^2\} &= \left[\sum_{h=2}^{M} h(h-1) \binom{M}{h} p^h q^{M-h} + \sum_{h=0}^{M} h \binom{M}{h} p^h q^{M-h}\right] \frac{1}{2p} \\ &= \left[M(M-1)p^2 \sum_{h=2}^{M} \frac{(M-2)!}{(h-2)!(M-h)!} p^{h-2} q^{M-h} + Mp\right] \frac{1}{2p}. \end{aligned} \tag{57}$$

Let us put $k = h - 2$ and $N = M - 2$ in the summation, it follows:

$$\begin{aligned} \mathbb{E}\{\sigma_L^2\} &= \left[M(M-1)p^2 \sum_{k=0}^{N} \binom{N}{k} p^k q^{N-k} + Mp\right] \frac{1}{2p} \\ &= \left[M(M-1)p^2 (p+q)^N + Mp\right] \frac{1}{2p}. \end{aligned} \tag{58}$$

Since $(p+q)^N = 1$, it results:

$$\begin{aligned} \mathbb{E}\{\sigma_L^2\} &= \left[M(M-1)\left(\frac{1}{N_P}\right)^2 + \frac{M}{N_P}\right] \frac{N_P}{2} \\ &= \frac{1}{2}\left[\frac{M(M-1)}{N_P} + M\right]. \end{aligned} \tag{59}$$

APPENDIX E
MULTI-PATH FADING, PROPAGATION DELAY, AND DOPPLER SPREAD ANALYSIS

In the development of the analytical model, we assumed a single path, ideal channel conditions, and the propagation delay was neglected. In this section, we add some discussions on the implications of the delay spread due to multi-path fading, the propagation delay, and the Doppler spread, due to the relative motion between the MTC device and the gNB. Clearly, this signal degradation reduces the orthogonality of the tagged preamble sequences. To overcome this issue, we introduced, as reported in Section II, one cyclic shift both for the preambles, N_{CS}, and for the tags, N_{TS}, aiming to guarantee the orthogonality of the sequences regardless of all the above phenomena. Particularly, both the cyclic shifts should be properly dimensioned based on the considered mMTC scenario. In the following we focus on N_{CS}, but similar considerations are valid for N_{TS}. As reported by 3GPP in [24], the recommended deployment scenario for mMTC is denoted as "Urban coverage for massive connection", and it is characterized by the carrier frequency (f) equal to 700 MHz, and the movement speed (v) fixed to 3 km/h, identical for all the MTC devices. First of all, we observe that this test scenario implies a negligible frequency offset due to Doppler spread, calculated as $\frac{vf}{c}$, where c is the speed of light. In fact, it is equal to 1.944 Hz and results very lower than the PRACH sub-carrier spacing, Δf =1.25 kHz. So, it is known that:

$$N_{CS} \geq \left\lceil (2\delta_{p,max} + \tau_{DS,max}) \frac{N_{ZC}}{T_{SEQ}} \right\rceil, \tag{60}$$

where $\delta_{p,max} = R/c$ is the maximum propagation delay, R is the cell radius, $\tau_{DS,max}$ is the maximum delay spread, $T_{SEQ} = 1/\Delta f$ is the sequence duration.

Let us start neglecting the propagation delay, i.e., $\delta_{p,max} = 0$. Then, Eq. (60) depends only on $\tau_{DS,max}$, that is, a measure of the multi-path richness of the communication channel. At this aim of estimating $\tau_{DS,max}$, we consider the Tapped Delay Line (TDL) models for wireless channels provided by the 3GPP in [25]; these models include all fading phenomena considered in (9). In particular, we utilize the TDL-C model for Non-Line-Of-Sight (NLOS), and the TDL-E model for Line-Of-Sight (LOS). Each TDL model is scaled in delay so that the model achieves a desired Root Mean Square (RMS) delay spread, $\tau_{DS_{RMS}}$, equal to 93 ns for the short-delay profile, and 363 ns for the normal-delay profile [25]. From the TDL-C model, the maximum delay spread is $\tau_{DS,max} = 8.6523 \cdot \tau_{DS_{RMS}}$, so by using (60), the minimum N_{CS} value is 1 for the short-delay profile, and 4 for the normal-delay profile. Conversely, in the TDL-E model, $\tau_{DS,max} = 20.6519 \cdot \tau_{DS_{RMS}}$, then N_{CS} should be at least equal to 3 and 8, for the short-delay profile and the normal-delay profile, respectively. Since the minimum value provided by the standard is $N_{CS} = 13$, we can estimate the maximum cell size, by taking into account the propagation delay $\delta_{p,max}$. In the TDL-C model, the maximum cell radius R is 1.78 km for the short-delay profile, and 1.38 km for the normal-delay profile. Conversely, in the TDL-E model, R is 1.57 km and 0.73 km, for the short-delay profile and the normal-delay profile, respectively. Clearly, in the case of larger cells, a greater value of N_{CS} needs to be set. also, we note that our analytical model is derived for a generic N_{CS}, so it is still valid. As regards N_{TS}, the same dimensioning is applied, but there is no a minimum value set by the standard.

Now, we want to evaluate the overall effects of the multi-path transmissions, non-ideal channel gains ($h_{i,g} \neq 1$) and $d_{i,g} \neq 0$ on the preamble detection. Firstly, for each device $i \in \mathcal{M}$ transmitting an untagged preamble sequence $x_r^{p_i}[n]$, we calculated the output sequence across an independent TDL-E and TDL-C modeled channel, with normal-delay profile, $N_{CS} = 13$, and $N[n] = 0$. Then, we calculated the sum of the above sequences and evaluate the cross-correlation $\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau]\right\}$, reported in Figs.11a and 11b. As regards the LOS (see Fig. 11a), due to the multiple paths, in each point $\tau = p_i N_{CS}$, $\forall p_i \in \mathcal{P}_\mathcal{S}$, the correlation value is slightly lower than the expected value $|\mathcal{M}_{p_i}|\sqrt{N_{ZC}}$, and different from 0 in the other points. Specifically, significant secondary peaks occur in the ranges $\{p_i N_{CS}+1, \ldots, p_i N_{CS}+N_{CS}-1\}$, $\forall p_i \in \mathcal{P}_\mathcal{S}$. Therefore, given $p_i \in \mathcal{P}_\mathcal{S}$, we introduce the "Detection Zone for preamble p_i", $DZ_{p_i} = \{p_i N_{CS}, \ldots, p_i N_{CS} + N_{CS} - 1\}$, as the interval of values containing the main and the secondary peaks of the correlation $\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau]\right\}$. Outside these ranges, i.e., $\tau \notin DZ_{p_i}, \forall p_i \in \mathcal{P}_\mathcal{S}$, the values assumed by $\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau]\right\}$ approaches zero. Conversely, as regards the NLOS (see Fig. 11b), the sign of the main peaks of the correlation $\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau]\right\}$ can be either positive or negative. In addition, $\left|\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau]\right\}\right|$ can be significantly lower than $|\mathcal{M}_{p_i}|\sqrt{N_{ZC}}$. Also in this case, $\forall p_i \in \mathcal{P}_\mathcal{S}$, several peaks occur in the range DZ_{p_i}, while outside these detection zones, the values assumed approaches zero. Also, we note that the main peak of $\left|\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau]\right\}\right|$ may be in a point $\tau \neq p_i N_{CS}$, but $\tau \in DZ_{p_i}$; this is due to the absence of a direct path which dominates in terms of power compared to the other secondary paths and arrives with an additional delay equal to 0.

Having introduced the detection zones, we are now able to appropriately adapt the Φ_P and Φ random variables in our model. At this aim, the range $\{0, \ldots, (N_{ZC}-1)\}$ needs to be divided into N_P detection zones $DZ_p, \forall p \in \mathcal{P}$, and for each zone, the scheduler should calculate:

$$\tau'' = \underset{\forall \tau \in DZ_p}{\arg\max} \left|\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau]\right\}\right|. \tag{61}$$

The random variable Φ_P should be re-defined as $\Phi_P = \left|\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau'']\right\}\right|$, where τ'' is calculated in (61), $\forall p \in \mathcal{P}_\mathcal{S}$. We also need to re-define the random variable Φ as $\Phi = \left|\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}},z_r}[\tau'']\right\}\right|$, where τ'' is calculated in (61), $\forall p \in \mathcal{P} - \mathcal{P}_\mathcal{S}$. Then, the T_P^ϵ and T_{IP}^ϵ values and the related working zones need to be recalculated.

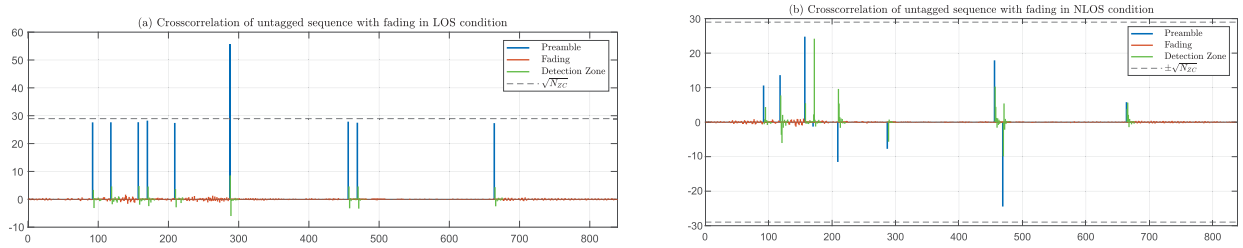


Fig. 11. Example of $\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}},z_r}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}}[\tau]\right\}$ when $M = 10$ with fading in LOS (a) and NLOS (b) conditions.

As already pointed out, in LOS the effect of fast fading does not severely change the values assumed by both the adapted Φ_P and Φ variables compared to the ideal values (i.e., with the absence of noise, fading, and interference). We therefore expect a very slight variation in the threshold values and the related working zone. Instead, in NLOS, only the variation of the Φ values is negligible compared to the ideal ones, while Φ_P is subject to large variations. For this reason, we expect T_P^ϵ to assume much lower values compared to the ones of the analysis previously made and, therefore, the working zones will be reduced accordingly.

Now, we take into account the propagation delays effects, in absence of multi-path propagation and noise. Typically, each device $i \in \mathcal{M}$ experiences a propagation delay different from the other ones. Since the N_{CS} value in (60) is properly dimensioned taking into account the $\delta_{p,max}$ value, the Detection Zones strategy is valid for the preamble transmissions of any device. In addition, we note that, given a preamble $p_s \in \mathcal{P}_\mathcal{S}$, the gNB receiver could detect in $\Re\left\{C_{y_{r,k_{\mathcal{P}_\mathcal{S}}},z_r}^{\mathcal{P}_\mathcal{S},\mathcal{T}_\mathcal{S}}[\tau]\right\}$, for $\tau \in DZ_{p_s}$, $|\mathcal{M}_{p_s}|$ different peaks with amplitude $\sqrt{N_{ZC}}$ belonging to the same detection zone DZ_{p_s}, instead of a single peak of amplitude $|\mathcal{M}_{p_s}|\sqrt{N_{ZC}}$ in $\tau = p_s N_{CS}$. Consequently, the value assumed by the adapted random variable Φ_P would be equal to $\sqrt{N_{ZC}}$. However, in our model the preamble threshold T_P^ϵ has been derived in the worst case (i.e., $|\mathcal{M}_{p_s}| = 1$), therefore, the preamble p_s will still be detected with the related ϵ probability and the eventual collision will be detect by the tag analysis. Finally, we also add the multi-path effects. Since each gth path of the MTC device i will suffer the same additional propagation delay, the main and secondary peaks of the cross-correlation will be rigidly shifted forward. Nevertheless, since the N_{CS} value in (60) is properly dimensioned taking into account both $\delta_{p,max}$ and $\tau_{DS,max}$, the peaks remain within the Detection Zone, allowing the detector to properly work.

REFERENCES

[1] *IMT Vision—Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond*, document M.2083, ITU-T, 2015.

[2] D. Panno and S. Riolo, "An enhanced joint scheduling scheme for GBR and non-GBR services in 5G RAN," *Wireless Netw.*, vol. 26, no. 4, pp. 3033–3052, May 2020.

[3] D. Panno and S. Riolo, "A new centralized access control scheme for D2D-enabled mmWave networks," *IEEE Access*, vol. 7, pp. 80697–80716, 2019.

[4] P. Popovski *et al.*, "Wireless access for ultra-reliable low-latency communication: Principles and building blocks," *IEEE Netw.*, vol. 32, no. 2, pp. 16–23, Mar. 2018.

[5] M. Blockstrand, T. Holm, L.-Ö. Kling, R. Skog, and B. Wallin, "More than 50 billion connected devices," Ericsson, Stockholm, Sweden, White Paper 284 23-3149 Uen, 2011, vol. 14, p. 124. [Online]. Available: https://vdna.be/publications/Wp-50-Billions.Pdf

[6] Z. Zhang, Y. Li, C. Huang, Q. Guo, C. Yuen, and Y. L. Guan, "DNN-aided block sparse Bayesian learning for user activity detection and channel estimation in grant-free non-orthogonal random access," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 12000–12012, Dec. 2019.

[7] Z. Ding, R. Schober, P. Fan, and H. V. Poor, "Simple semi-grant-free transmission strategies assisted by non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4464–4478, Jun. 2019.

[8] A. Laya, L. Alonso, and J. Alonso-Zarate, "Is the random access channel of LTE and LTE-A suitable for M2M communications? A survey of alternatives," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 4–16, Feb. 2014.

[9] L. Ferdouse, A. Anpalagan, and S. Misra, "Congestion and overload control techniques in massive M2M systems: A survey," *Trans. Emerg. Telecommun. Technol.*, vol. 28, no. 2, p. e2936, Feb. 2017, doi: 10.1002/ett.2936.

[10] L. Tello-Oquendo, J.-R. Vidal, V. Pla, and L. Guijarro, "Dynamic access class barring parameter tuning in LTE—A networks with massive M2M traffic," in *Proc. 17th Annu. Medit. Ad Hoc Netw. Workshop (Med-Hoc-Net)*, Jun. 2018, pp. 1–8.

[11] L. Tello-Oquendo *et al.*, "Reinforcement learning-based ACB in LTE-A networks for handling massive M2M and H2H communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–7.

[12] *Service Accessibility*, document TS 22.011 version 9.4.0, 06, 3GPP, 2010.

[13] L. Miuccio, D. Panno, and S. Riolo, "Joint control of random access and dynamic uplink resource dimensioning for massive MTC in 5G NR based on SCMA," *IEEE Internet Things J.*, vol. 7, no. 6, pp. 5042–5063, Jun. 2020.

[14] B. Fekade, T. Maksymyuk, M. Kyryk, and M. Jo, "Probabilistic recovery of incomplete sensed data in IoT," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2282–2292, Aug. 2018.

[15] M. Tavares, D. Samardzija, H. Viswanathan, H. Huang, and C. Kahn, "A 5G lightweight connectionless protocol for massive cellular Internet of Things," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Mar. 2017, pp. 1–6.

[16] H. S. Dhillon, H. Huang, and H. Viswanathan, "Wide-area wireless communication challenges for the Internet of Things," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 168–174, Feb. 2017.

[17] H. S. Jang, S. M. Kim, H.-S. Park, and D. K. Sung, "Message-embedded random access for cellular M2M communications," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 902–905, May 2016.

[18] H. S. Jang, S. M. Kim, H.-S. Park, and D. K. Sung, "A preamble collision resolution scheme via tagged preambles for cellular IoT/M2M communications," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1825–1829, Feb. 2018.

[19] H. S. Jang, S. M. Kim, H.-S. Park, and D. K. Sung, "An early preamble collision detection scheme based on tagged preambles for cellular M2M random access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5974–5984, Jul. 2017.

[20] I. Leyva-Mayorga, L. Tello-Oquendo, V. Pla, J. Martinez-Bauset, and V. Casares-Giner, "On the accurate performance evaluation of the LTE–A random access procedure and the access class barring scheme," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 7785–7799, Dec. 2017.

[21] *Physical Channels and Modulation*, document TS 36.211, 10, 3GPP, 2020.

[22] J. W. Lindeberg, "Eine neue herleitung des exponentialgesetzes in der wahrscheinlichkeitsrechnung," *Math. Zeitschrift*, vol. 15, no. 1, pp. 211–225, 1922.
[23] K. J. McGown and H. R. Parks, "The generalization of Faulhaber's formula to sums of non-integral powers," *J. Math. Anal. Appl.*, vol. 330, no. 1, pp. 571–575, Jun. 2007.
[24] *Study on Scenarios and Requirements for Next Generation Access Technologies*, document TR 38.913, 07, 3GPP, 2020.
[25] *Study on Channel Model for Frequencies From 0.5 to 100 GHz*, document TR 38.901, 01, 3GPP, 2020.

Salvatore Riolo (Graduate Student Member, IEEE) received the bachelor's degree in electronics engineering and the master's degree *(cum laude)* in telecommunications engineering from the Department of Electrical, Electronics, and Computer Engineering, University of Catania, Italy, in 2012 and 2017, respectively, where he is currently pursuing the Ph.D. degree in systems, energy, computer, and telecommunications engineering. From 2017 to 2018, he was an early-stage Researcher in high-bit rate device-to-device services for 5G mobile networks with the University of Catania. His scientific research interests include radio resource management for 5G mobile networks, green networking, mmWave D2D communications, and access protocols for massive machine type communications.

Daniela Panno (Member, IEEE) received the Laurea degree *(cum laude)* in electrical engineering from the University of Catania, Italy, in 1989, and the Ph.D. degree in electronic engineering and computer science engineering from the University of Palermo, Italy, in 1993. In 1989, she joined the Department of Electrical, Electronics, and Computer Engineering with the University of Catania, where she is currently an Associate Professor of telecommunications. She has attended several international workshops and symposia as an Invited Speaker. Her research interests include radio resource management, green networking, mmWave D2D communication, SDN and NFV technologies for 5G cellular networks, and mobile services in IoT environments. She has served as a Guest Editor for Computer Communications.

Luciano Miuccio (Graduate Student Member, IEEE) received the bachelor's degree in electronics engineering and the master's degree (Hons.) in telecommunications engineering from the Department of Electrical, Electronics, and Computer Engineering, University of Catania, Italy, in 2015 and 2018, respectively, where he is currently pursuing the Ph.D. degree in systems, energy, computer, and telecommunications engineering. From May 2019 to September 2019, he was an early-stage researcher in Study of methodologies for the Multi-objective optimization of Parametric Systems with the University of Catania. His scientific interests include radio resource management for energy-constrained devices in 5G mobile networks, NOMA techniques, green networking, and access protocols for massive machine type communications.