



HAL
open science

Autonomous power decision for grant free access MUSA scheme in mMTC scenario

Wissal Ben Ameer, Philippe Mary, Jean-François H elard, Marion Dumay,
Jean Schwoerer

► **To cite this version:**

Wissal Ben Ameer, Philippe Mary, Jean-Fran ois H elard, Marion Dumay, Jean Schwoerer. Autonomous power decision for grant free access MUSA scheme in mMTC scenario. *Sensors*, 2021, 21 (1), pp.116. 10.3390/s21010116 . hal-03101354

HAL Id: hal-03101354

<https://hal.science/hal-03101354>

Submitted on 7 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin e au d p t et   la diffusion de documents scientifiques de niveau recherche, publi s ou non,  manant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv s.

Article

Autonomous power decision for grant free access MUSA scheme in mMTC scenario

Wissal Ben Ameer ^{1,2*} , Philippe Mary ¹ , Jean-François Hélard ¹ , Marion Dumay ²  and Jean Schwoerer ² 

¹ Univ. Rennes, INSA Rennes, CNRS, IETR, Rennes, France; Philippe.Mary@insa-rennes.fr (P.M.); Jean-François.Helard@insa-rennes.fr (J.F.H.)

² Orange Labs, Meylan, France; marion-dumay@orange.com (M.D.); jean.schwoerer@orange.com (J.S.)

* Correspondence: wissal.benameur@orange.com (W.B.A.)

Version January 5, 2021 submitted to Journal Not Specified

Abstract: Non orthogonal multiple access schemes with a grant free access have been recently highlighted as a prominent solution to meet the stringent requirements of mMTC. In particular, multi user shared access (MUSA) scheme has shown a great potential to allow grant free access to the available resources. For the sake of simplicity, MUSA is generally conducted with successive interference cancellation (SIC) receiver which offers a low decoding complexity. However, this family of receivers requires a sufficiently diversified received user powers in order to ensure the best performance and avoid the error propagation phenomenon. The power allocation has been considered as a complicated issue especially for a decentralized decision with a minimum signaling overhead. In this paper, we propose a novel algorithm for an autonomous power decision with a minimal overhead based on a tight approximation of the bit error probability (BEP) while considering the error propagation phenomenon. We investigate the efficiency of multi-armed bandit (MAB) approaches for this problem in two different reward scenarios: i) in scenario 1, each user reward only informs on its own packet whether it was successfully transmitted or not; ii) in scenario 2, each user reward may carry information about the other interfering users packets. The performances of the proposed algorithm and the MAB techniques are compared in terms of the successful transmission rate. The simulations results prove that the MAB algorithms show a better performance in the second scenario compared to the first one. However, in both scenarios, the proposed algorithm outperforms the MAB techniques with a lower complexity at user equipment.

Keywords: Non orthogonal multiple access (NOMA); multi-user shared access (MUSA); successive interference cancellation (SIC); grant free access; bit error probability (BEP); power allocation; multi-armed bandit (MAB) algorithms.

1. Introduction

The future radio access network of the fifth generation is expected to support a variety of applications with different quality of service (QoS). These services are classified by the international telecommunications union and the third generation partnership project into three main use cases with different stringent requirements, namely enhanced mobile broadband (eMBB), ultra reliable and low latency communications (uRLLC) and massive machine type communications (mMTC). This latter is also known as massive IoT as it is designed to mainly deal with a massive number of connected devices [1], i.e., one million connected devices per km². The mMTC use case is characterized by short packet communications, i.e., on the order of few bytes, low system complexity and low energy consumption which leads to a battery life on the order of ten years. The conventional orthogonal multiple access (OMA) schemes are limited by the restricted number of the available orthogonal resources and thereby

33 they may not be suitable to handle the huge number of devices to be connected in the mMTC scenario.
34 However, the non orthogonal multiple access (NOMA) schemes have been underlined as a prominent
35 solution to address the connectivity issue [2]. In fact, they allow multiple users to simultaneously and
36 non-orthogonally share the same resources, which increases the system overload.

37 In the existing technologies, users used to go through a contention based random access protocol
38 for data transmission. For LTE/LTE-A network, the eNB initially broadcasts information about
39 the available physical random access channel (PRACH) to all users. Then, each user launches a
40 coordination process over the PRACH to ensure its alignment with the eNB. After that, for each
41 transmission attempt, each user should send a grant acquisition request to the eNB to reserve its
42 resource. The coordination random access channel (RACH) process is performed through four
43 handshake steps [3]: 1) the preamble transmission; 2) the random access response; 3) the radio
44 resource control (RRC) connection request and 4) the RRC connection setup. However, RACH and
45 resource allocation processes may be very expensive in terms of signaling overhead, especially for
46 mMTC devices.

47 According to [4], the transmission of 100 bytes of useful data in the uplink while going through
48 the RACH process, security procedures and connection release generates a signaling overhead of 59
49 bytes on the uplink and 136 bytes on the downlink. This induces an excessive waste of resource, a high
50 energy consumption and thus a shorter battery life for the transmission of small packets. Moreover, the
51 very high number of devices may lead to unacceptable high latency for certain mMTC applications. In
52 fact, a large number of simultaneous connections may imply the overuse of the resources and increase
53 the decoding error probability. For instance, under ideal system conditions, the RACH process induces
54 a latency of 9.5 ms, which would increase significantly in case of collision [3]. As a consequence, the
55 random radio resource access strategy may be a bottleneck in some mMTC scenarios.

56 In this context, NOMA with grant free access option has gained a lot of interest and it has been
57 promoted by the scientific community as a promising solution to support mMTC scenarios with a
58 minimum signaling overhead, which ensures a low energy consumption. Authors in [5] has presented
59 the evolution steps towards the uplink NOMA schemes combined with the grant free access. They
60 suggested two possible communication scenarios for grant free access in the uplink. Users can either
61 go with RACH-based with grant free transmission or RACH-less with grant free transmission. In the
62 first scenario, the RACH process allows one to establish a connection with the base station and ensure
63 user synchronisations. Then, each user transmits its data without waiting for the allocated resources
64 from the base station. This option has never been possible for OMA schemes since grant free access
65 may yield to a severe system congestion when users transmit on the same resources. In the second
66 scenario, users transmit their data without any beforehand communication with the base station, which
67 significantly minimizes the signaling overhead but at the cost of non synchronized communications.
68 Therefore, robust multi-user detection (MUD) receivers are required for signal detection.

69 Since the announcement of the advent of 5G, several NOMA schemes have been emerged during
70 the last few years, namely power domain NOMA (PD-NOMA) [6], sparse code multiple access (SCMA)
71 [7], multi-user shared access (MUSA) [8], pattern division multiple access (PDMA) [9], to cite a few.
72 These schemes are different multiplexing techniques based on different keys such as user codebook,
73 power or multiple domains. Authors in [10] aimed at handling the critical transmission latency issue
74 for vehicle-to-vehicle services through a grant free access option with NOMA schemes. Two novel
75 algorithms known as hyper-fraction and genetic algorithms were proposed to respectively reduce the
76 system latency and improve the system throughput while guaranteeing a rate fairness between users.
77 In [11], authors dealt with asynchronous transmissions due to grant free access. In order to improve the
78 decoding process, multiple copies of the same message are transmitted and then used at the receiver
79 with successive interference cancellation (SIC) technique as a kind of users diversity. The authors
80 proposed closed-form expressions of the successful transmission probability, the battery lifetime and
81 the energy efficiency. The proposed approach may be useful for short packet communications, but
82 at the cost of a complex decoding process. In addition, one problem of the grant free access is the

83 estimation of the number of active users. This issue has been addressed in [12] by proposing a deep
84 learning algorithm which uses the recorded user activities at the base station to predict their future
85 behavior. This prediction is given as an input to a modified orthogonal matching pursuit algorithm
86 to improve the multi-user detection and reduce the error probability. In [13], a sinusoidal code is
87 proposed for the signals separation in the context of mMTC scenario with grant free access. The
88 proposed spreading sequences permit to use non-iterative algorithms for multi-user detection without
89 a prior knowledge of the channel state information and the number of active users. Authors in [14]
90 dealt with the problem of packet collisions in a grant free access context without a re-transmission
91 opportunity. A novel grant free access framework was proposed where the non-decoded users consider
92 the occurred collisions as interference. Moreover, the system performance was evaluated analytically
93 and authors provided simplified expressions of the outage probability and the system throughput.

94 SCMA has particularly been studied with grant free access protocols. For instance, in [15],
95 authors studied the application of SCMA with a faster than Nyquist signaling which improves the
96 spectral efficiency, but at the expense of a higher inter-symbol and inter-user interference. Therefore, a
97 novel algorithm based on the expectation propagation was proposed for the channel estimation, the
98 detection of user activities and the signal decoding. The work in [16] have investigated an iterative
99 message passing algorithm for grant free access SCMA, based on the belief propagation. The proposed
100 algorithm permits to jointly estimate the channel coefficients, identify the number of active users and
101 detect the transmitted data while improving the bit error rate compared to the other techniques.

102 Regarding the system design, MUSA has the potential to enable grant free access with minimum
103 signaling overhead in the context of mMTC applications. Unlike the SCMA scheme which requires
104 the assignment of codebook beforehand, in MUSA each user randomly and autonomously selects a
105 spreading sequence within a predefined constellation. In other words, users can transmit their data
106 at any moment without going through a resource allocation process with the base station, which
107 minimizes the amount of signaling overhead. MUSA scheme is typically used with a SIC receiver
108 for multi user detection, which provides a low decoding complexity. However, the SIC technique
109 may suffer from the error propagation phenomenon when the received powers are similar [17]. The
110 power allocation process is usually performed in a centralized manner [18,19] where the base station
111 knows the channel state information of all users. For a grant free access, each user performs a blind
112 transmission with no information about its propagation environment and interfering users, which
113 makes the power determination more complex.

114 Autonomous power decision for NOMA schemes with grant free access strategy has recently been
115 investigated in several works. An interesting solution is to use multi-armed bandit (MAB) algorithms
116 which belong to the global reinforcement learning paradigm [20,21]. MAB techniques can be applied
117 to the problem of dynamic resource allocation by balancing between exploration and exploitation
118 phases. At each time, each agent selects an arm, i.e., representing the physical resource to be shared,
119 among a set according to a predefined policy in order to maximize its cumulative reward and hence
120 minimize its regret. The MAB algorithms have been used in several applications such as marketing,
121 advertising and cellular communications. For instance, authors in [22] applied the MAB algorithms
122 to the autonomous power decision problem in order to maximize the user rates for the PD-NOMA
123 scheme. The user rewards are their rates. However, these may be carried on many bits which increases
124 the signaling overhead and hence it may not be really adapted for mMTC scenarios. MAB have also
125 been merged with NOMA schemes in [23] where authors proposed a distributed NOMA-based MAB
126 approach to handle the channel access problem in cognitive radio networks. Moreover, authors in
127 [24] have performed the MAB algorithms in the LTE cellular network for an autonomous subcarriers
128 allocation in a dense network while taking into consideration the dynamic resource occupation in each
129 surrounding cell.

130 To the best of our knowledge, no work has investigated the problem of autonomous power
131 decision for grant free access with MUSA scheme. The characteristics of spreading sequences and the
132 principle of SIC receiver make the power decision more complex. Therefore, in this paper, we deal

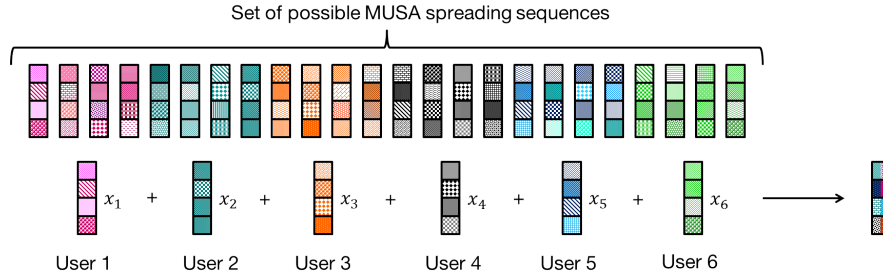


Figure 1. MUSA scheme system for $J = 6$ and $K = 4$.

133 with this issue with minimum signaling overhead to address the mMTC requirements. The goal is to
 134 improve the system performance measured with the successful transmission rate in order to achieve the
 135 performance of an optimal centralized power allocation. This latter is quite difficult to obtain, especially
 136 for SIC receivers with the error propagation problem. To do so, we start by proposing an approximated
 137 expression for the bit error probability (BEP) while considering the inter-user interference and the
 138 effect of error propagation. The optimal power value of users are obtained as the solution of the
 139 minimization of the global average BEP. Based on the derived BEP expression, we propose a novel
 140 algorithm for power selection for MUSA scheme with a reduced signaling overhead. The proposed
 141 algorithm is compared with known index-based MAB algorithms adapted to the power selection by
 142 each user. In this part, we propose to investigate two scenarios for selecting the best arm by each MAB
 143 algorithm. A scenario where the arm index computation by a user is only based on the decoding status
 144 of its own packet, i.e. success or failure, and another scenario where it depends on the decoding status
 145 of the other users' packets in addition to its own packet decoding status.

146 This paper is organized as follows. The system model and the fundamentals of MUSA are
 147 introduced in Section 2. SIC receiver is revisited in Section 3 while a closed-form expression for users'
 148 bit error probability is derived in Section 4. Then, the proposed algorithm for autonomous power
 149 decision is described in Section 5. The multi-armed bandit algorithms and the studied scenarios are
 150 introduced in Section 6. A comparison of all power decision approaches is provided in Section 7.
 151 Numerical results and performance analysis are conducted in Section 8 and conclusions are drawn in
 152 Section 9.

153 **Notations:** Vectors and matrices are denoted in lower and upper cases respectively and in bold font,
 154 while scalars use normal font weight. The complex and real number sets are denoted by \mathbb{C} and \mathbb{R} ,
 155 respectively. Moreover $(\cdot)^T$ and $(\cdot)^H$ stand for transpose and hermitian operations. $\text{diag}(\mathbf{a})$ represents
 156 the diagonal matrix created with the elements of vector \mathbf{a} in the main diagonal.

157 2. System Model

An uplink communication system of J users transmitting over K orthogonal subcarriers is considered. The active users share the available resources using the MUSA scheme with a grant free access. Each user bits are mapped to a series of symbols through a M-ary modulation block. Then, the modulated symbols are multiplied by the users spreading sequences and spread over the available subcarriers, as illustrated in Figure 1. Users sequences $\mathbf{s}_j, \forall j \in \{1, \dots, J\}$ are such that $\mathbf{s}_j \in \{a + jb\}^K$, where $(a, b) \in \{-1, 0, 1\}^2$. The received signal on subcarrier k of each OFDM symbol is:

$$y_k = \sum_{j=1}^J \sqrt{p_j} h_{kj} s_{kj} x_j + n_k \quad (1)$$

where h_{kj} and s_{kj} are the k -th component of the j -th user channel vector and spreading sequence, i.e. \mathbf{h}_j and \mathbf{s}_j , respectively. Moreover x_j, p_j are the transmitted symbol and the transmission power of the j -th user, respectively, and n_k is the additive white Gaussian noise component on the k -th subcarrier

with $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_K)$, where \mathbf{I}_K is the K -by- K identity matrix. The multiplexed received signals on all subcarriers can be written as:

$$\mathbf{y} = \mathbf{G} \mathbf{P}^{\frac{1}{2}} \mathbf{x} + \mathbf{n} \quad (2)$$

where $\mathbf{P} = \text{diag}(p_1, p_2, \dots, p_J) \in \mathbb{R}_+^{J \times J}$ is the transmission power matrix, $\mathbf{x} = [x_1, x_2, \dots, x_J]^T$ is the transmitted users' symbols with $\mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{I}_J$ and \mathbf{G} is the equivalent channel matrix including the spreading sequences such that:

$$\mathbf{G} = \mathbf{H} \odot \mathbf{S} \quad (3)$$

158 where $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_J]$, $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_J]$ and \odot is the Hadamard product, i.e., $g_{kj} = h_{kj}s_{kj}$.

159 3. Multi-user detection

The SIC receiver offers a low decoding complexity compared to other MUD algorithms, namely message passing algorithm or maximum a posteriori algorithm [25]. However, the SIC performance depends on the user received powers and the receiver performs better when the received powers are sufficiently different. MUSA is typically used with ordered-SIC jointly with a linear detection receiver such as the minimum mean square error (MMSE). The MMSE matrix is calculated as in [26]:

$$\mathbf{W}^H = (\mathbf{P}^{\frac{1}{2}} \mathbf{G}^H \mathbf{G} \mathbf{P}^{\frac{1}{2}} + \sigma^2 \mathbf{I})^{-1} \mathbf{P}^{\frac{1}{2}} \mathbf{G}^H. \quad (4)$$

The main principle of the ordered-SIC technique is to successively estimate the user symbol, reconstruct the generated interference and then subtract it from the received signal. Users symbols are decoded in a descending order of their SINRs. Assuming that the received signal at the j -th iteration is:

$$\mathbf{y}^j = \sqrt{p_j} \mathbf{g}_j x_j + \sum_{i=j+1}^J \sqrt{p_i} \mathbf{g}_i x_i + \mathbf{n}^j, \quad (5)$$

where \mathbf{g}_j is the j -th column of the matrix \mathbf{G} . Then, the SINR of the picked user j to be decoded is

$$\beta_j(\mathbf{p}) = \frac{p_j |\mathbf{w}_j^H \mathbf{g}_j|^2}{\sum_{i=j+1}^J p_i |\mathbf{w}_j^H \mathbf{g}_i|^2 + \sigma^2 \|\mathbf{w}_j^H\|^2}, \quad (6)$$

where \mathbf{w}_j is the j -th column of the MMSE matrix \mathbf{W} . After that, the user symbol is estimated by multiplying the row vector \mathbf{w}_j^H by the received column signal as follows:

$$\hat{x}_j = \mathbf{w}_j^H \mathbf{y}. \quad (7)$$

The interference generated by the j -th user is reconstructed and then subtracted from the received signal which is updated as follows:

$$\mathbf{y} = \mathbf{y} - \mathbf{g}_j \hat{x}_j. \quad (8)$$

160 After each iteration, the j -th column of the matrix \mathbf{G} , corresponding to the decoded user j , is removed
161 and the MMSE matrix is recalculated as in (4). This process is repeated until all users are decoded.

162 4. BEP analysis

The error propagation is one of the critical issue of SIC receivers, which significantly deteriorates the system performance and makes the derivation of the BEP expression more complicated. For a Gray mapping, two adjacent symbols are different in only one single bit. Hence, assuming the inter-user interference as noise, the erroneous detection often leads to the detection of an adjacent symbol with

only one wrong bit compared to the correct symbol [27]. Therefore, the average system BEP is well approximated as:

$$P_{b, \text{MMSE-SIC}} \approx \frac{1}{J \log_2(M)} \sum_{j=1}^J P_{ej} \quad (9)$$

163 where P_{ej} is the symbol error probability (SEP) of the j -th user. In the following, we investigate the
 164 BEP of the MMSE-SIC receiver with two different hypotheses; i) Perfect SIC without error propagation;
 165 ii) Imperfect SIC with error propagation.

166 4.1. Perfect SIC without error propagation

In this case, since there is no error propagation in the receiver, the BEP is calculated similarly as for the MMSE receiver while updating the MMSE matrix at each iteration and the SINRs are calculated as in (6). For a QPSK modulation and assuming the inter-user interference as noise [28], the j -th user SEP is approximated as [27]:

$$P_{ej} \approx 2Q\left(\sqrt{\beta_j^{\text{NEP}}(\mathbf{p})}\right) \left(1 - 0.5Q\left(\sqrt{\beta_j^{\text{NEP}}(\mathbf{p})}\right)\right). \quad (10)$$

167 4.2. Imperfect SIC with error propagation

In that case, the BEP of each user depends on the previously decoded users. In this paper, we are inspired by the proposed approach in [28] and thereby the SEP of the j -th user is calculated as:

$$P_{\varepsilon_j} = \sum_{i=0}^{N_j-1} P\{\varepsilon_j | \mathbf{b}_i^j\} P\{\mathbf{b}_i^j\}, \quad (11)$$

where $N_j = 2^{j-1}$ is the number of possible $(j-1)$ -dimensional binary sequences and $\mathbf{b}_i^j = (b_{i,1}^j, b_{i,2}^j, \dots, b_{i,j-1}^j) \forall i \in \{0, \dots, N_j-1\}$ and $j \in \{1, \dots, J\}$, with $b_{i,k}^j = 0$ if the symbol of the k -th decoded user is correctly detected and 1 otherwise. Each sequence refers to the state, correctly decoded or not, of all the previously $(j-1)$ decoded users. The event ε_j indicates an erroneous detection of the j -th user symbol. Hence, $P\{\varepsilon_j | \mathbf{b}_i^j\}$ is the error probability of the j -th user symbol conditioned on the sequence \mathbf{b}_i^j . Considering an eventual error propagation occurrence, the received signal at the j -th SIC iteration is represented as:

$$\mathbf{y}^j = \sqrt{p_j} \mathbf{g}_j x_j + \sum_{i=j+1}^J \sqrt{p_i} \mathbf{g}_i x_i + \sum_{k=1}^{j-1} \sqrt{p_k} \mathbf{g}_k (x_k - \hat{x}_k) + \mathbf{n}^j, \quad (12)$$

where \hat{x}_k is the faulty estimation of x_k . The additional term compared to (5), is generated by the erroneous detection of the previous users. This may significantly affect the system performance. Therefore, the experienced noise and the new interference term can be combined in $\mathbf{n}_{eq} = \sum_{k=1}^{j-1} \sqrt{p_k} \mathbf{g}_k (x_k - \hat{x}_k) + \mathbf{n}^j$. The resulting term is approximated as a centered Gaussian random variable, where $\mathbb{E}\{\mathbf{n}_{eq}\} = \mathbf{0}$ and $\mathbb{E}\{\mathbf{n}_{eq} \mathbf{n}_{eq}^H\} = (\sum_{k=1}^{j-1} p_k \|\mathbf{g}_k\|^2 \mathbb{E}\{\|x_k - \hat{x}_k\|^2\} + \sigma^2) \mathbf{I} = (\sum_{k=1}^{j-1} p_k \|\mathbf{g}_k\|^2 \delta_k d + \sigma^2) \mathbf{I}$. We define d as the square of the euclidean distance between the neighboring symbols and $\delta_k = 1$ if $x_k \neq \hat{x}_k$ and 0 otherwise. As a consequence, the SINR of the j -th user, corresponding to the detection combination \mathbf{b}_i^j , is calculated as follows:

$$\beta_{j,i}^{\text{EP}}(\mathbf{p}) = \frac{p_j |\mathbf{w}_j^H \mathbf{g}_j|^2}{\sum_{i=j+1}^J p_i |\mathbf{w}_j^H \mathbf{g}_i|^2 + (\sum_{k=1}^{j-1} p_k \|\mathbf{g}_k\|^2 \delta_k d + \sigma^2) \|\mathbf{w}_j^H\|^2} \quad (13)$$

Two main terms should be calculated to obtain the user SEP. Starting by the conditional probability which is calculated according to (10) and (13), we have:

$$P \{ \varepsilon_j | b_i^j \} = 2Q \left(\sqrt{\beta_{j,i}^{EP}(\mathbf{p})} \right) \left(1 - 0.5Q \left(\sqrt{\beta_{j,i}^{EP}(\mathbf{p})} \right) \right). \quad (14)$$

However, the probability of the combination \mathbf{b}_i^j is readily calculated as:

$$P \{ \mathbf{b}_i^j \} = P \{ \cap_{n=1}^{j-1} b_{i,n}^j \} = \prod_{n=1}^{j-1} P \{ b_{i,n}^j | \cap_{m=1}^{n-1} b_{i,m}^j \}, \quad (15)$$

where $P \{ b_{i,n}^j | \cap_{m=1}^{n-1} b_{i,m}^j \}$ is the probability that the n -th symbol of user j is correctly decoded or not, i.e., $b_{i,n}^j = 0$ or $b_{i,n}^j = 1$, conditioned on the estimation of the previously decoded $(n - 1)$ symbols. It is calculated as:

$$P \{ b_{i,n}^j | \cap_{m=1}^{n-1} b_{i,m}^j \} = \quad (16)$$

$$\begin{cases} 1 - 2Q \left(\sqrt{\beta_{n,i}^{EP}(\mathbf{p})} \right) \left(1 - 0.5Q \left(\sqrt{\beta_{n,i}^{EP}(\mathbf{p})} \right) \right) & \text{if } b_{i,n}^j = 0 \\ 2Q \left(\sqrt{\beta_{n,i}^{EP}(\mathbf{p})} \right) \left(1 - 0.5Q \left(\sqrt{\beta_{n,i}^{EP}(\mathbf{p})} \right) \right) & \text{otherwise.} \end{cases} \quad (17)$$

For an uplink transmission, devices are restricted by a maximum transmission power, p^U , imposed by the regulation authorities and the equipment design restrictions. Therefore, an optimal centralized power allocation \mathbf{p}_{opt} , that minimizes the global average error probability, can be obtained by solving the following problem:

$$OP_1 \begin{cases} \min_{\mathbf{p}} & \frac{1}{J \log_2(M)} \sum_{j=1}^J \sum_{i=0}^{N_j-1} P \{ \varepsilon_j | \mathbf{b}_i^j \} P \{ \mathbf{b}_i^j \} \\ & p_j \leq p^U \quad \forall j \in \mathcal{J} \end{cases} \quad (18a)$$

$$p_j \leq p^U \quad \forall j \in \mathcal{J} \quad (18b)$$

168 where $\mathcal{J} = \{1, 2, \dots, J\}$ is the set of active users. The derived expression of user SEP is quite
 169 complicated to be analysed theoretically with the Karush–Kuhn–Tucker (KKT) conditions. Therefore,
 170 we use an advanced optimization algorithm, i.e. particle swarm optimization [29], to solve the power
 171 allocation problem above. This algorithm is known to be efficient for complex problem [30].

172 5. Proposed autonomous power decision algorithm

173 Each user has to decide its transmission power autonomously with no information about the
 174 propagation environment and the interference. In this section, we aim at proposing an autonomous
 175 power decision algorithm for uplink communication. It allows each user to select an adequate power
 176 value close to the optimal one, \mathbf{p}_{opt} , obtained by solving OP_1 .

177 The key idea is to perform an iterative algorithm that takes advantage from the natural base
 178 station acknowledgement (ACK). Each user gradually updates its transmitted power from the received
 179 ACK in order to converge toward the nearest power level from \mathbf{p}^{opt} . For example, the j -th user initially
 180 transmits its data with a randomly selected power p^j within the interval $[p_{\min}^j, p_{\max}^j]$, where p_{\min}^j and
 181 p_{\max}^j are respectively the initial minimum and maximum power values memorized in the j -th user
 182 equipment (UE). Then, the base station detects the user signal and compares its transmission power
 183 with $p^{j,\text{opt}}$, that base station has computed on its own. An acknowledgment will be sent back to each
 184 user to adjust its power. In order to minimize the signaling overhead, the acknowledgment is carried
 185 on two bits and can hence encode four possible states; 1) ACK = 3 if user should simply transmit with
 186 its maximum authorized power p^U . This case may be gainful for the cell edge users that experience
 187 bad propagation conditions. 2) ACK = 2 if $p^j > p^{j,\text{opt}}$; 3) ACK = 1 if $p^j < p^{j,\text{opt}}$ and 4) ACK = 0 if

188 $p^j = p^{j,\text{opt}}$. Each user updates its interval by shifting p_{\min}^j and p_{\max}^j values. After that, it picks up
 189 another random value in the new power interval for the next packet transmission until it arrives at
 190 the appropriate power value. However, the channel conditions may change along the way. Hence,
 191 the algorithm must take this into consideration in order to ensure its convergence and assure the best
 192 performance. For that reason, the base station may, sometimes, send another extra bit "Stat" to notify
 193 user by this occurrence. In this case, UE will try to initialize its power interval while taking advantage
 194 from the previous sent packets. This process is described in details in Algorithm 1.

Algorithm 1: Autonomous power decision

Require: $p_{\max}^j = p^U, p_{\min}^j > 0 \forall j = 1, 2, \dots, J$;

Ensure: \mathbf{p}

- 1: Each user picks up its spreading sequences.
- 2: Each user selects a random power level $p^j \in [p_{\min}^j, p_{\max}^j]$.
- 3: The BS detects users signals.
- 4: The BS calculates the optimal power $p^{j,\text{opt}}$.
- 5: The BS compares each user power p^j with the nearest power level from $p^{j,\text{opt}}$.
- 6: BS send an acknowledgement to each user:
 - a) If $p^{j,\text{opt}} = p_{\max} \Rightarrow \text{ACK} = 3$
 - b) If $p^j > p^{j,\text{opt}} \Rightarrow \text{ACK} = 2$
 - c) If $p^j < p^{j,\text{opt}} \Rightarrow \text{ACK} = 1$
 - d) If $p^j = p^{j,\text{opt}} \Rightarrow \text{ACK} = 0$
- 7: If the propagation environment is changed, the BS sends one-bit ACK: Stat = 1.
- 8: Each user updates his p_{\min}^j or p_{\max}^j :
 - a) If $\text{ACK} = 3 \Rightarrow p^j = p^U > 0$
 - b) If $\text{ACK} = 2 \Rightarrow p_{\max}^j = p^j \leq p^U$
 - If Stat = 1 $\Rightarrow p_{\min}^j = 0$
 - c) If $\text{ACK} = 1 \Rightarrow p_{\min}^j = p^j > 0$
 - If Stat = 1 $\Rightarrow p_{\max}^j = p^U$
 - d) If $\text{ACK} = 0 \Rightarrow$ no update
- 9: Return to step 2

195 The channel should not change too fast in order to allow the convergence of the algorithm.
 196 However, as it will be seen in simulation results, the proposed algorithm converges to the near-optimal
 197 power value quite quickly. In addition, users transmission powers must be known at the BS to perform
 198 the proposed algorithm. However, these powers values are obviously needed in order to apply the
 199 SIC receiver properly. Therefore, a calibration phase between the BS and the UE should always be
 200 established.

201 **6. Power allocation with multi-armed bandits**

202 In this section, we revisit three known MAB algorithms, i.e. ϵ -greedy, upper confidence bound
 203 (UCB1) and Thompson sampling (THS), that we apply to our autonomous power selection problem. A
 204 MAB is a model with N resources, called arms, each of them being associated to a reward following
 205 a specific probability distribution. At each time slot t , each agent j plays an arm a_j according to its
 206 policy. Then, it receives the corresponding reward $r_j^t(a_j)$. Based on this and the number of time each
 207 arm has been played so far, $n_j^t(a_j)$, each agent chooses the appropriate arm for the next time slot $t + 1$,

208 according to the calculated index that depends on each algorithm policy. Over time, these techniques
209 will prioritize the arms showing the best performance and exclude the worst ones.

All MAB algorithms search for the maximization of the cumulative rewards of each agent over the time horizon T , i.e., $\sum_{t=1}^T r_j^t(a_j)$ and thereby the minimization of its regret R_j defined as the difference between the rewards obtained using the chosen policy and the expected reward we would obtain if the best arm would always be played i.e. r_j^* . The j -th user regret during a maximum period of T slots is calculated as follows:

$$R_j = Tr_j^* - \sum_{t=1}^T \mathbb{E}\{r_j^t(a_j)\} \quad (19)$$

210 In our case, we consider a multi-agent system where the agent refers to the UE and the arms represent
211 the power levels. At the t -th iteration, the successful transmission rate of the j -th user is defined as
212 the ratio between the cumulative number of its correctly received packets during t time slots and the
213 total number of plays so far. The MAB algorithms are investigated in two different scenarios detailed
214 hereafter.

215 a) Scenario 1:

216 The base station acknowledgement at the t -th iteration is carried on 1 bit representing the
217 corresponding user reward, i.e., $r_j^t \in \{0, 1\}$. At each time slot t , $r_j^t(a_j) = 1$ if the packet of the
218 j -th user is successfully decoded and $r_j^t(a_j) = 0$ otherwise. Therefore, the successful transmission rate
219 of the j -th user at the t -th iteration is calculated as $Q_j^t = \frac{\sum_{i=1}^t r_j^i(a_j)}{t}$. In this scenario, the reward of each
220 user only depends on the decoding status of its own packet without any consideration to the other
221 users. However, the successful decoding event of one packet depends on the successful decoding of
222 the others, because of the SIC receiver. Hence every user has interest on good power selection for the
223 other users and not only for itself. The scenario 2, we propose hereafter, takes into account this fact.

224 b) Scenario 2:

225 The base station acknowledgement at the t -th iteration is now carried on two bits $\{b_{2,j}^t, b_{1,j}^t\}$. The
226 first bit informs whether all users are correctly decoded, $b_{1,j}^t = 1$, or, at least, one packet is erroneously
227 detected, $b_{1,j}^t = 0$. The second bit notifies each user whether its own packet is correctly received,
228 $b_{2,j}^t = 1$, or not, $b_{2,j}^t = 0$. For a picked power p_j by user j , there are three possible states for the j -th user
229 acknowledgement $\{b_{2,j}^t, b_{1,j}^t\} \in \{11, 10, 00\} = \{3, 2, 0\}$. The case where $\{b_{2,j}^t, b_{1,j}^t\} = 01$ is not possible
230 because $b_{1,j}^t = 1$ means that all packets have been correctly decoded, including the j -th user packet,
231 and hence $b_{2,j}^t$ is automatically equal to 1. In order to meet the conditions of convergence theorems
232 derived in [31], the rewards should be supported in $[0, 1]$. Therefore, users rewards are defined as a
233 normalization of the associated acknowledgements, i.e., $r_j^t \in \{1, \frac{2}{3}, 0\}$. The successful transmission rate,
234 at the t -th iteration, of the j -th user is then calculated based only on the second bit $b_{2,j}^t$, i.e., $Q_j^t = \frac{\sum_{i=1}^t b_{2,j}^i}{t}$.
235 In this scenario, the inter-user dependence is involved in the associated rewards.

236 6.1. UCB1

UCB1 has been inspired by the Agrawal's index-based policy [31]. This algorithm has an uniformly logarithmic regret over time. Generally, the UCB family algorithms rely to a confidence interval on the average reward of each arm [32]. UCB1 index gathers two functions; the average reward and the exploration term. This index refers to an estimation of the upper bound of the true expectation of the

arm reward. It is an upper bound because the square root term is an estimation of the variance of the expected return when playing the arm a_j and is defined as follows, at time slot t :

$$\frac{1}{n_j^t(a_j)} \sum_{i=1}^t r_j^i(a_j) + \sqrt{\frac{\theta \log(t)}{n_j^t(a_j)}} \quad (20)$$

237 where $\theta > 0$ is the exploration parameter. Originally, UCB1 was proposed with $\theta = 2$, however,
 238 authors in [32] have mentioned that $\theta = 0.5$ performs better empirically although $\theta > 0.5$ is strongly
 239 recommended for the theoretical analysis.

240 At the initialization phase, UCB1 explores each arm once in order to have an estimation of the
 241 reward of each arm. Then, at each iteration, each user selects the arm with the highest index, as
 242 illustrated in Algorithm 2. The calculated index (20) ensures the balance between the exploration of
 243 the most uncertain arms and the exploitation of the best arm so far. UCB1 prescribes the principle of
 244 "optimism face uncertainty" which means that the less visited arm seems more uncertain and thereby
 245 it may optimistically be the best arm to play.

Algorithm 2: UCB1 algorithm

Require: θ and N
 Each user plays all the arms once during N plays:
for $t = N + 1 : T$ **do**
 for $j = 1 : J$ **do**
 Select the arm: $\operatorname{argmax}_{a_j} m_j^{t-1}(a_j) + \sqrt{\frac{\theta \log(t-1)}{n_j^{t-1}(a_j)}}$
 Update the following variables;
 a) $n_j^t(a_j) = n_j^{t-1}(a_j) + 1$
 b) $m_j^t(a_j) = \frac{1}{n_j^t(a_j)} \sum_{i=1}^t r_j^i(a_j)$
 end
end

246 6.2. ϵ -greedy

This algorithm deals with the exploration and the exploitation dilemma randomly. At each iteration, each user either explores arbitrarily a new arm with probability ϵ or it plays the best arm corresponding to the highest average reward so far with a probability of $1 - \epsilon$. However, for a constant exploration parameter ϵ , the system regret evolves linearly overtime instead of being logarithmic. On the one hand, for a high ϵ value, i.e., $\epsilon \approx 1$, user will continue to only explore random arms even if it came out with the best arm, and on the other hand, for a low ϵ value, i.e., $\epsilon \ll 1$, the algorithm will tend to exploit all the time even if it has not sufficiently explored the other arms. In both cases, an important performance loss will be experienced. Therefore, the ϵ value is a critical parameter. A revised version called ϵ -decreasing greedy has been proposed, where the exploration probability is decreasing toward zero over time with a rate of $\frac{1}{t}$. This allows one to essentially explore at the beginning of the learning and mostly to exploit the best arm found so far after a certain amount of time. The new exploration probability is defined as [22,31]:

$$\epsilon(t) = \min \left\{ 1, \frac{CN}{d^2 t} \right\} \triangleq \min \left\{ 1, \frac{LN}{t} \right\}. \quad (21)$$

247 Where $L > 0$ is the exploration parameter. However, the main challenge of this policy is how to
 248 properly set the value of L . The ϵ -decreasing greedy algorithm is described in details in Algorithm 3.

<p>Algorithm 3: ϵ-decreasing greedy algorithm</p> <p>Require: L and N</p> <p>for $t = 1 : T$ do</p> <p style="padding-left: 20px;">for $j = 1 : J$ do</p> <p style="padding-left: 40px;">Select a random arm with probability $\epsilon(t) = \min \left\{ 1, \frac{LN}{t} \right\}$</p> <p style="padding-left: 40px;">Select with probability $1 - \epsilon(t)$ the best arm: $\operatorname{argmax}_{a_j} m_j^{t-1}(a_j)$</p> <p style="padding-left: 40px;">Update the following variables:</p> <p style="padding-left: 60px;">a) $n_j^t(a_j) = n_j^{t-1}(a_j) + 1$</p> <p style="padding-left: 60px;">b) $m_j^t(a_j) = \frac{1}{n_j^t(a_j)} \sum_{i=1}^t r_j^i(a_j)$</p> <p style="padding-left: 20px;">end</p> <p>end</p>

249 6.3. Thompson sampling algorithm

250 This approach shows a robust performance for stochastic problems and sometimes outperforms
 251 other MAB algorithms. THS algorithm belongs to the Bayesian MAB family. The j -th user starts by a
 252 uniform prior beta distribution $\beta(\alpha_{j,k}, \gamma_{j,k})$ for all arms with initial values $\alpha_{j,k} = \gamma_{j,k} = 2 \forall j \in \{1, \dots, J\}$
 253 and $\forall k \in \{1, \dots, N\}$, where k refers to the arm index among N power levels. Then, inspired by the case
 254 where rewards follow a Binomial distribution [33] and based on the observed reward, the parameters
 255 of the posterior beta distribution are updated such that $\alpha_{j,k} = \alpha_{j,k} + 3r_j^t$ and $\gamma_{j,k} = \gamma_{j,k} + 3(1 - r_j^t)$.
 256 At the next time slot, each user draws a sampled index from the updated beta distribution for each
 257 arm, i.e., $i_{j,k} \sim \beta(\alpha_{j,k}, \gamma_{j,k}) \forall k \in 1, \dots, N$ and $\forall j \in 1, \dots, J$. The arm with the highest index, i.e.,
 258 $\hat{i}_{j,k} = \max_{k \in \mathcal{N}}(i_{j,k}) \forall j \in 1, \dots, J$, is hence elected for this transmission attempt. Through time,
 259 Thompson sampling prioritizes the arm with the highest probability of being the optimal one and
 260 avoids other arms that have demonstrated poor performance so far.

<p>Algorithm 4: Thompson sampling algorithm</p> <p>Require: N and $\alpha_{j,k} = \gamma_{j,k} = 2 \forall k = 1 \dots N$ and $\forall j = 1 \dots J$</p> <p>for $t = 1 : T$ do</p> <p style="padding-left: 20px;">for $j = 1 : J$ do</p> <p style="padding-left: 40px;">Select a sampled index from the beta distribution of each arm $i_{j,k} \sim \beta(\alpha_{j,k}, \gamma_{j,k})$ $\forall k = 1, \dots, N$</p> <p style="padding-left: 40px;">Play the arm a_j with the highest index $\hat{i}_{j,q} = \max_{k \in \mathcal{N}}(i_{j,k})$</p> <p style="padding-left: 40px;">Update the following variables:</p> <p style="padding-left: 60px;">a) $n_j^t(a_j) = n_j^{t-1}(a_j) + 1$</p> <p style="padding-left: 60px;">b) $m_j^t(a_j) = \frac{1}{n_j^t(a_j)} \sum_{i=1}^t r_j^i(a_j)$</p> <p style="padding-left: 60px;">c) $\alpha_{j,k} = \alpha_{j,k} + 3r_j^t(a_j)$</p> <p style="padding-left: 60px;">d) $\gamma_{j,k} = \gamma_{j,k} + 3(1 - r_j^t(a_j))$</p> <p style="padding-left: 20px;">end</p> <p>end</p>

261 7. Complexity and overhead analysis

262 A quantitative comparison of all the examined techniques in the context of mMTC scenario is
 263 summarized in Table 1. The random power selection and the centralized allocation are taken as

264 reference scenarios. The centralized allocation is the reference in terms of performance and the random
 265 selection is the simplest one.

	Signaling overhead	Complexity at UE	Power decision
Centralized allocation	$O(J \cdot k)$ if k bits (Depend on DCI)	$O(1)$	Attributed by BS
Random selection	1 bit	$O(1)$	Random
Proposed algorithm	2 or 3 bits	$O(1)$	Iterative decision
ϵ -decreasing greedy	Scenario1: 1 bit	$O(N)$	Random with ϵ probability
	Scenario2: 2 bits		
UCB1	Scenario1: 1 bit	$O(N)$	Index-based
	Scenario2: 2 bits		
Thompson sampling	Scenario1: 1 bit	$O(N)$	Bayesian distribution
	Scenario2: 2 bits		

Table 1. Quantitative comparison of the signaling overhead and the complexity at user equipment in each iteration for all algorithms

266 The centralized power allocation algorithm computes, at the base station, the power to allocate
 267 to the users at each transmission attempt, based on the users received SINRs. All the complexity
 268 is located at the base station and users have to set their transmitting power at the values sent back
 269 from the BS, hence the algorithm complexity at the user side is $O(1)$. The signaling overhead of this
 270 scheme cannot be assessed precisely since it strongly depends on the downlink control information
 271 (DCI) format. However, the power computed is quantized over k bits, which would likely be much
 272 larger than 1 or 2 bits, for each user. Hence, for a large number of users the signaling would be at least
 273 in $O(J \cdot k)$. Thus, it may be very expensive in terms of energy consumption leading to a significant
 274 reduction of battery lifetime.

275 The random power selection does not manifest any algorithmic complexity since the power
 276 selection is realized randomly. Therefore, the generated signaling overhead is minimal, i.e. 1 bit, as it
 277 only relies on the acknowledgment sent by the BS for each user's packet, whether it is successfully
 278 received or not.

279 The proposed autonomous power decision algorithm is based on four acknowledgment levels,
 280 used to update the power at the user side, which can be carried with two bits. Moreover, one may add
 281 one additional bit if the BS detects a channel variation in order to notify the corresponding user of this
 282 event. The generated complexity is on the order of $O(1)$ as no computation is required at UE during
 283 this process.

284 All the MAB techniques have the same signaling overhead and algorithmic complexity for each
 285 transmission attempt. UCB1, ϵ -decreasing greedy and Thompson sampling can be seen as index-based
 286 policies. Hence, the algorithmic complexity consists in sorting N indexes, representing the rating of
 287 the arms w.r.t. the objective of the agent, and taking the arm that corresponds to the highest index.
 288 Therefore, their complexity is on the order of $O(N)$. Furthermore, the generated signaling overhead
 289 depends particularly on the applied learning scenario. In scenario 1, the indexes update by an agent
 290 is only based on the processing output of its own packet using a given power, i.e. either the packet
 291 is successfully received or not, and hence it takes 1 bit. In scenario 2, the update of an agent index is
 292 made by taking into account the decoding status of the other users' transmissions, in addition to that
 293 of its own packet, which is carried out with 2 bits. It is worth noting that the computational complexity
 294 is not considered here. Moreover, the complexity of calculating a sampled index from the beta function
 295 for each arm with the Thompson sampling algorithm is higher than that of UCB1 and ϵ -decreasing
 296 greedy indexes.

297 8. Numerical results and analysis

298 We consider an uplink system with 150% of overload, where $J = 12$ and $K = 8$. Users are
 299 uniformly scattered in the cell while experiencing an AWGN channel with different path-losses. Each
 300 user can pick its transmission power over a set of $N = 10$ possible power levels in the interest

301 of selecting the appropriate value ensuring the best performance in both scenarios 1 and 2. Users
 302 spreading sequences are normalized to unitary energy. The algorithms are investigated in term of
 303 the successful transmission rate, i.e. the total number of correctly decoded packets over the total
 304 number of sent packets. Simulations are averaged over 150 network realizations, i.e. the successful
 305 transmission rate is averaged over the path losses and the spreading sequences. Regarding UCB1
 306 algorithm, the exploration of new power values is conducted by the parameter θ . As mentioned above,
 307 this parameter is originally set to 2, but in the literature $\theta = 0.5$ is admitted empirically as it provides
 308 better performance. In order to choose the optimal value of θ , the average transmission rate achieved
 309 by UCB1 has been investigated w.r.t. θ and the value $\theta = 0.5$ is the one that allows to achieve the best
 310 transmission rate. The figure is not reported here not to clutter the exposure. The other simulation
 311 parameters are reminded in Table 2.

Channel	AWGN with path losses
Users	$J = 12$
Subcarriers	$K = 8$
Maximum individual power	20 dBm
N	10 levels
Noise power	$\sigma^2 = -14$ dBm
T	1000 slots
θ	0.5

Table 2. Simulation settings

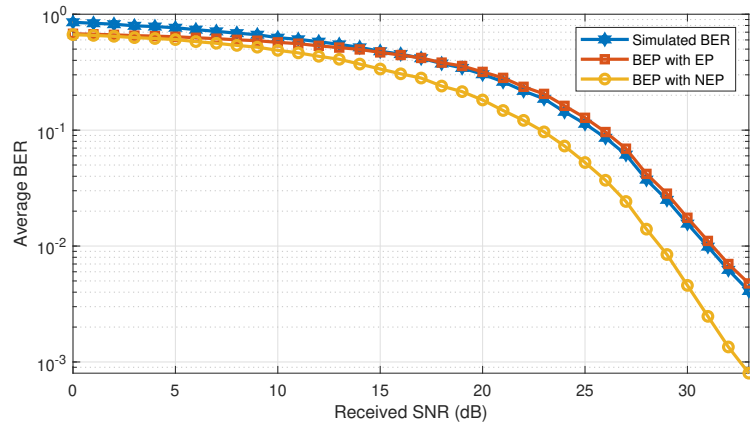


Figure 2. Performance comparison of the simulated BER and the analytical BEP for an AWGN channel with different users path-loss and equal transmission powers.

312 Figure 2 compares the simulated average BER, i.e., averaged over the spreading sequences and
 313 positions, and the analytical average BEP obtained by the proposed expression in (9) for an AWGN
 314 channel and uniformly distributed users over the cell w.r.t the global received SNR. We remark that
 315 the expression that takes into account the error propagation phenomenon almost matches with the
 316 simulated BER. However, removing the error propagation effect induces a wide gap in the performance
 317 because it is too optimistic. In addition, we notice that, for high SNR values, the BEP with EP gets
 318 closer to the simulated BER. This can be explained by the fact that the QPSK approximation in (10) is
 319 more robust for high SNR.

320 The performance of the ϵ -decreasing greedy algorithm depends on the ϵ value which in turn
 321 depends on the coefficient L . It is important to choose the coefficient that allows the algorithm to
 322 achieve its best performance. Therefore, the main challenge of the ϵ -decreasing greedy approach is to
 323 handle the exploration and the exploitation dilemma by properly set the value of L in (21). Figure 3
 324 investigates the performance of this algorithm for different L in scenario 1 after $T = 1000$ iterations.
 325 We note that $L = 0.1$ gives the best performance in term of average transmission rate and hence it is

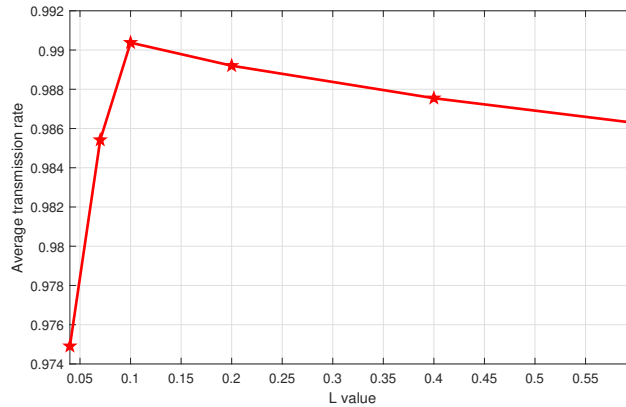


Figure 3. Performance comparison of ϵ -decreasing greedy for different L values after $T = 1000$ iterations in scenario 1.

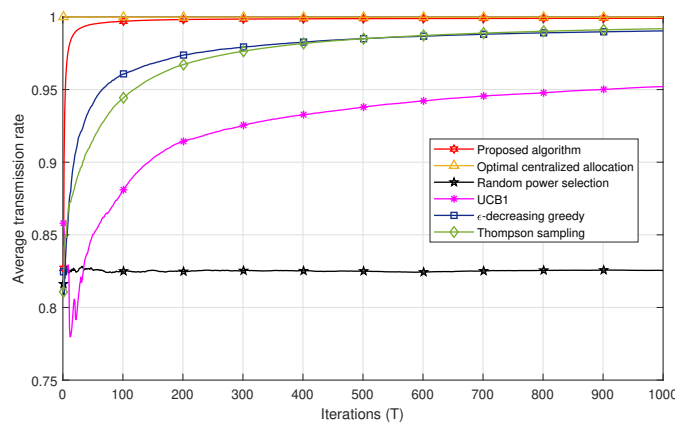


Figure 4. Successful transmission rate comparison for all algorithms in scenario 1.

326 kept for the rest of the simulations. The same behaviour is observed in scenario 2 but not reported
 327 here to limit the redundancy.

328 Figures 4 and 5 compare the successful transmission rate of the algorithms under study, i.e. the
 329 centralized power allocation, the proposed algorithm, the MAB algorithms (ϵ -decreasing greedy, UCB1
 330 and THS) and the random power selection in scenarios 1 and 2, respectively. The proposed algorithm
 331 outperforms all the MAB techniques with a faster convergence to the optimal power in both scenarios.
 332 We also remark in Fig. 4 that the ϵ -decreasing greedy algorithm converges faster than THS and UCB1
 333 algorithms. This can be explained by the optimal selection of L value that ensures a trade-off between
 334 the exploration and the exploitation phases in order to achieve the best performance. The ϵ -decreasing
 335 greedy and THS algorithms converge to the same successful transmission rate after 400 iterations.
 336 However, the gap between ϵ -decreasing greedy and THS is less important in scenario 2 in Fig. 5.
 337 In fact, after $T = 100$ iterations, THS is slightly better than ϵ -decreasing greedy. THS seems to take
 338 advantage of the additional information carried by the feedback whether there is a decoding error
 339 among the users or not. However, both algorithms, i.e. ϵ -decreasing greedy and THS, are far better
 340 than UCB1 in both scenarios. UCB1 takes more time to explore suboptimal powers which slows down
 341 its convergence to the optimal power values and thereby induces more packet losses. The random
 342 power allocation presents the lowest performance bound in both scenarios since no strategy is applied
 343 for an adequate power selection which induces error propagation and hence packet losses.

344 For a given number of iterations T , the figures represent the average successful transmission
 345 rate achieved after averaging over the network realizations and the spreading sequences, i.e. 150
 346 realizations, and T being the number of packets sent, a.k.a. the number of iterations in each algorithm.
 347 The performance achieved by the algorithms under fast variations of the propagation environment
 348 is directly obtained from Figures 4 and 5 by shortening them to the desired value of T . In other

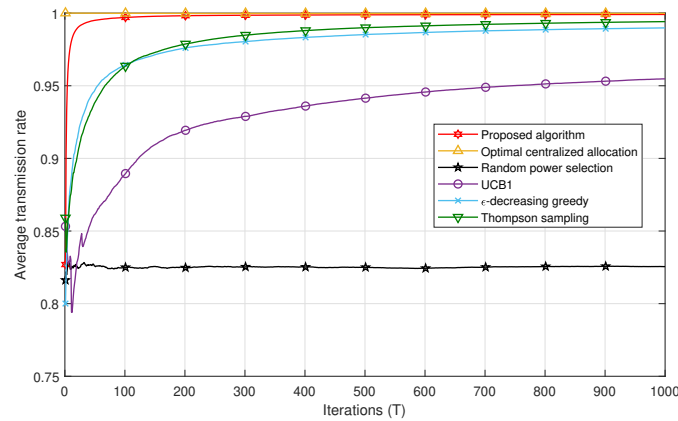


Figure 5. Successful transmission rate comparison for all algorithms in scenario 2.

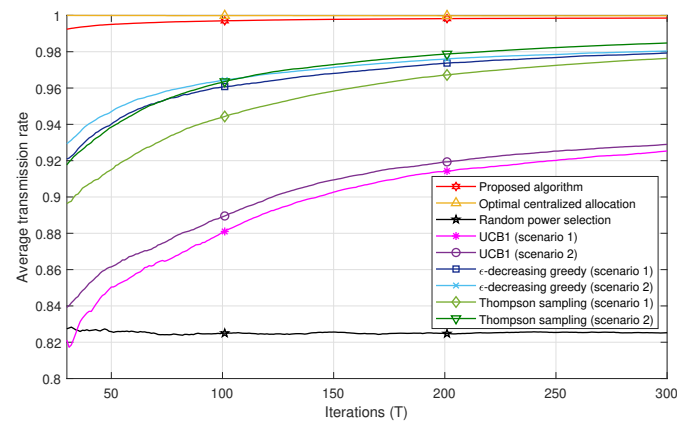


Figure 6. Successful transmission rate comparison for all algorithms in scenarios 1 and 2.

349 words, if one would want to obtain the achievable successful rate of the different algorithms when the
 350 environment changes every 100 packets, then one should collect the points at $T = 100$ in each figure
 351 above. Moreover, a fading channel could have been considered also, however, this would only affect
 352 the absolute performance, as the statistic of the rewards would have been changed, but not the relative
 353 behaviors of the algorithms. Therefore, in this paper and for the sake of simplicity, we consider only an
 354 AWGN channel with different path losses among users and we show the behavior of the investigated
 355 techniques as the number of iterations increases averaged over several network realizations.

356 Figure 6 shows a performance comparison of all algorithms in scenarios 1 and 2 for $30 \leq T \leq 300$.
 357 One can remark that all MAB techniques achieve better performances in scenario 2 compared to
 358 scenario 1. For instance, after $T = 50$ iterations, the Thompson sampling algorithm achieves a
 359 successful transmission rate of ≈ 0.94 in scenario 2, whereas, in scenario 1, it attains the value of 0.91.
 360 This may be explained by the fact that scenario 2 conveys more information compared to scenario
 361 1 to select the best set of powers. In other words, the reward a user gets in scenario 2 is not only a
 362 function of the successful decoding of its own packet, but also whether all other users succeeded in
 363 their transmissions or not. This strategy allows each user to take into account a kind of *global interest*
 364 in the selection of its power. In addition, the successful transmission rate achieved with the proposed
 365 algorithm converges to the one obtained with the optimal centralized solution after a few number
 366 of iterations compared to the MAB techniques. For example, after $T = 30$ iterations, the proposed
 367 algorithm achieves a rate of 0.99 of correctly received packets whereas the ϵ -decreasing greedy has
 368 a rate of 0.93. It should be noted that, after a large number of iterations, the performances of MAB
 369 algorithms in scenario 1 converge to those in scenario 2.

370 9. Conclusion

371 The autonomous power decision for NOMA schemes with a grant free access strategy has been
372 an issue to satisfy the mMTC requirements. To the best of our knowledge, no work has been granted
373 on this problem for MUSA scheme in order to enhance users performance with a minimal signaling
374 overhead. In this paper, we addressed this issue by proposing a novel algorithm for autonomous
375 power decision based on the proposed BEP approximation and the base station acknowledgements.
376 Moreover, we studied the efficiency of some MAB algorithms for the power allocation with two
377 different implementation scenarios, i.e. one where the rewards of a user are only dependent on the
378 decoding output status of its own packet and another one where they depend also whether all users
379 have successfully transmitted their packets or not. The proposed algorithm converges very fast to the
380 obtained solution with a centralized resource allocation that is considered as a baseline. Moreover, the
381 MAB algorithms have an acceptable performance but at the cost of a larger convergence time and a
382 higher UE complexity compared to the proposed algorithm. This latter shows the best performance
383 with a faster convergence rate but also with a slightly higher signaling overhead compared to the MAB
384 algorithms, particularly for a variant propagation environment.

385 **Author Contributions:** Conceptualization, W.B.A., P.M., J.F.H., M.D. and J.S.; methodology, W.B.A.; software,
386 W.B.A.; validation, W.B.A., P.M., J.F.H. and M.D.; formal analysis, W.B.A.; investigation, W.B.A.; writing–original
387 draft preparation, W.B.A. and P.M.; visualization, W.B.A.; supervision, P.M., J.F.H., M.D. and J.S. All authors have
388 read and agreed to this version of the manuscript.

389 **Funding:** This research received no external funding.

390 **Conflicts of Interest:** The authors declare no conflict of interest.

391 References

- 392 1. Popovski, P.; Trillingsgaard, K.F.; Simeone, O.; Durisi, G. 5G Wireless Network Slicing for
393 eMBB, URLLC, and mMTC: A Communication-Theoretic View. *IEEE Access* **2018**, *6*, 55765–55779.
394 doi:10.1109/ACCESS.2018.2872781.
- 395 2. Shirvanimoghadam, M.; Dohler, M.; Johnson, S.J. Massive Non-Orthogonal Multiple Access for Cellular
396 IoT: Potentials and Limitations. *IEEE Communications Magazine* **2017**, *55*, 55–61.
- 397 3. Shahab, M.B.; Abbas, R.; Shirvanimoghadam, M.; Johnson, S.J. Grant-free non-orthogonal multiple access
398 for IoT: A survey. *IEEE Communications Surveys & Tutorials* **2020**.
- 399 4. Shariatmadari, H.; Ratasuk, R.; Iraji, S.; Laya, A.; Taleb, T.; Jäntti, R.; Ghosh, A. Machine-type
400 communications: current status and future perspectives toward 5G systems. *IEEE Communications*
401 *Magazine* **2015**, *53*, 10–17.
- 402 5. Docomo, N.; others. Uplink Multiple Access Schemes for NR. *3GPP Draft* **2016**.
- 403 6. Saito, Y.; Kishiyama, Y.; Benjebbour, A.; Nakamura, T.; Li, A.; Higuchi, K. Non-Orthogonal Multiple Access
404 (NOMA) for Cellular Future Radio Access. 2013 IEEE 77th Vehicular Technology Conference (VTC Spring),
405 2013, pp. 1–5. doi:10.1109/VTCSpring.2013.6692652.
- 406 7. Nikopour, H.; Baligh, H. Sparse code multiple access. 2013 IEEE 24th Annual International Symposium on
407 Personal, Indoor, and Mobile Radio Communications (PIMRC), 2013, pp. 332–336.
- 408 8. Yuan, Z.; Yu, G.; Li, W.; Yuan, Y.; Wang, X.; Xu, J. Multi-User Shared Access for Internet of Things. 2016
409 IEEE 83rd Vehicular Technology Conference (VTC Spring), 2016, pp. 1–5.
- 410 9. Chen, S.; Ren, B.; Gao, Q.; Kang, S.; Sun, S.; Niu, K. Pattern Division Multiple Access - A Novel
411 Nonorthogonal Multiple Access for Fifth-Generation Radio Networks. *IEEE Transactions on Vehicular*
412 *Technology* **2017**, *66*, 3185–3196.
- 413 10. Lee, S.; Kim, J.; Park, J.; Cho, S. Grant-Free Resource Allocation for NOMA V2X Uplink Systems Using a
414 Genetic Algorithm Approach. *Electronics* **2020**, *9*, 1111.
- 415 11. Azari, A.; Popovski, P.; Miao, G.; Stefanovic, C. Grant-Free Radio Access for Short-Packet Communications
416 over 5G Networks. GLOBECOM 2017 - 2017 IEEE Global Communications Conference, 2017, pp. 1–7.
417 doi:10.1109/GLOCOM.2017.8255054.
- 418 12. Miao, X.; Guo, D.; Li, X. Grant-Free NOMA with Device Activity Learning Using Long Short-Term Memory.
419 *IEEE Wireless Communications Letters* **2020**.

- 420 13. Hasan, S.M.; Mahata, K.; Hyder, M.M. Fast Uplink Grant-Free NOMA with Sinusoidal Spreading Sequences. *arXiv preprint arXiv:2010.00199* **2020**.
- 421
- 422 14. Abbas, R.; Shirvanimoghaddam, M.; Li, Y.; Vucetic, B. A novel analytical framework for massive grant-free
- 423 NOMA. *IEEE Transactions on Communications* **2018**, *67*, 2436–2449.
- 424 15. Yuan, W.; Wu, N.; Zhang, A.; Huang, X.; Li, Y.; Hanzo, L. Iterative receiver design for FTN signaling aided
- 425 sparse code multiple access. *IEEE Transactions on Wireless Communications* **2019**, *19*, 915–928.
- 426 16. Wei, F.; Chen, W.; Wu, Y.; Ma, J.; Tsiftsis, T.A. Message-passing receiver design for joint channel estimation
- 427 and data decoding in uplink grant-free SCMA systems. *IEEE Transactions on Wireless Communications* **2018**,
- 428 *18*, 167–181.
- 429 17. Seung Hoon Nam.; Kwang Bok Lee. Transmit power allocation for an extended V-BLAST system. The
- 430 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2002, Vol. 2,
- 431 pp. 843–848 vol.2. doi:10.1109/PIMRC.2002.1047341.
- 432 18. Evangelista, J.V.; Sattar, Z.; Kaddoum, G.; Chaaban, A. Fairness and sum-rate maximization via joint
- 433 subcarrier and power allocation in uplink SCMA transmission. *IEEE Transactions on Wireless Communications*
- 434 **2019**, *18*, 5855–5867.
- 435 19. Ali, M.S.; Tabassum, H.; Hossain, E. Dynamic User Clustering and Power Allocation for Uplink
- 436 and Downlink Non-Orthogonal Multiple Access (NOMA) Systems. *IEEE Access* **2016**, *4*, 6325–6343.
- 437 doi:10.1109/ACCESS.2016.2604821.
- 438 20. Kaufmann, E. Analyse de stratégies Bayésiennes et fréquentistes pour l'allocation séquentielle de ressources.
- 439 PhD thesis, Paris, ENST, 2014.
- 440 21. Slivkins, A. Introduction to multi-armed bandits. *arXiv preprint arXiv:1904.07272* **2019**.
- 441 22. Adjif, M.A.; Habachi, O.; Cances, J. Joint Channel Selection and Power Control for NOMA: A Multi-Armed
- 442 Bandit Approach. 2019 IEEE Wireless Communications and Networking Conference Workshop (WCNCW),
- 443 2019, pp. 1–6.
- 444 23. Tian, Z.; Wang, J.; Wang, J.; Song, J. Distributed NOMA-Based Multi-Armed Bandit Approach for Channel
- 445 Access in Cognitive Radio Networks. *IEEE Wireless Communications Letters* **2019**, *8*, 1112–1115.
- 446 24. Feki, A.; Capdevielle, V. Autonomous resource allocation for dense LTE networks: A Multi Armed
- 447 Bandit formulation. 2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio
- 448 Communications, 2011, pp. 66–70.
- 449 25. Ameer, W.B.; Mary, P.; Dumay, M.; H elard, J.; Schwoerer, J. Performance study of MPA, Log-MPA and
- 450 MAX-Log-MPA for an uplink SCMA scenario. 2019 26th International Conference on Telecommunications
- 451 (ICT), 2019, pp. 411–416. doi:10.1109/ICT.2019.8798841.
- 452 26. Cho, Y.S.; Kim, J.; Yang, W.Y.; Kang, C.G. *MIMO-OFDM wireless communications with MATLAB*; John Wiley
- 453 & Sons, 2010.
- 454 27. Proakis, J.G.; Salehi, M. *Digital communications*; Vol. 4, McGraw-hill New York, 2001.
- 455 28. Zarella, A.; Chiani, M.; Win, M.Z. MMSE reception and successive interference cancellation for MIMO
- 456 systems with high spectral efficiency. *IEEE Transactions on Wireless Communications* **2005**, *4*, 1244–1253.
- 457 doi:10.1109/TWC.2005.847103.
- 458 29. Kennedy, J.; Eberhart, R. Particle swarm optimization. Proceedings of ICNN'95-International Conference
- 459 on Neural Networks. IEEE, 1995, Vol. 4, pp. 1942–1948.
- 460 30. Sahab, M.G.; Toropov, V.V.; Gandomi, A.H. A review on traditional and modern structural optimization:
- 461 problems and techniques. *Metaheuristic applications in structures and infrastructures* **2013**, pp. 25–47.
- 462 31. Auer, P.; Cesa-Bianchi, N.; Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Machine*
- 463 *learning* **2002**, *47*, 235–256.
- 464 32. Bonnefoi, R.; Besson, L.; Moy, C.; Kaufmann, E.; Palicot, J. Multi-Armed Bandit Learning in IoT Networks:
- 465 Learning helps even in non-stationary settings. International Conference on Cognitive Radio Oriented
- 466 Wireless Networks. Springer, 2017, pp. 173–185.
- 467 33. Gupta, N.; Granmo, O.; Agrawala, A. Thompson Sampling for Dynamic Multi-armed Bandits. 2011 10th
- 468 International Conference on Machine Learning and Applications and Workshops, 2011, Vol. 1, pp. 484–489.
- 469 doi:10.1109/ICMLA.2011.144.

470 © 2021 by the authors. Submitted to *Journal Not Specified* for possible open access publication
471 under the terms and conditions of the Creative Commons Attribution (CC BY) license
472 (<http://creativecommons.org/licenses/by/4.0/>).