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

Wissal Ben Ameur, Philippe Mary, Jean-François Hélard, Marion Dumay, Jean Schwoerer

MDPI

Article

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

Wissal Ben Ameur [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 brodband (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 km^2. 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

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

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

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

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

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

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

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

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

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

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

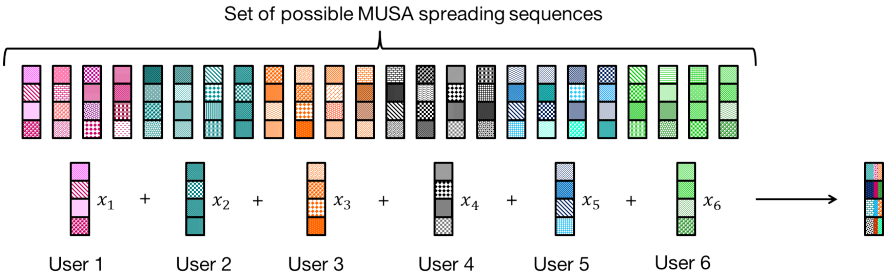

Figure 1. MUSA scheme system for $J = 6$ and $K = 4$.

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

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

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

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 $\boldsymbol{s}_j, \forall j \in \{1, \cdots, J\}$ are such that $\boldsymbol{s}_j \in \{a + jb\}^K$, where $(a, b) \in \{-1, 0, 1\}^2$. The received signal on subcarrier k of each OFDM symbol is:

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

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:

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

where $\mathbf{P} = \text{diag}(p_1, p_2, ..., p_J) \in \mathbb{R}_+^{J\times J}$ is the transmission power matrix, $\mathbf{x} = \left[x_1, x_2, ..., x_J\right]^T$ 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:

$$\mathbf{G} = \mathbf{H} \odot \mathbf{S} \tag{3}$$

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

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]:

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

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:

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

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

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

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:

$$\hat{x}_j = \mathbf{w}_j^H\mathbf{y}. \tag{7}$$

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

$$\mathbf{y} = \mathbf{y} - \mathbf{g}_j\hat{x}_j. \tag{8}$$

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

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:

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

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

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]:

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

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:

$$P_{\varepsilon_j} = \sum_{i=0}^{N_j-1} P\left\{\varepsilon_j|\mathbf{b_i^j}\right\} P\left\{\mathbf{b_i^j}\right\}, \tag{11}$$

where $N_j = 2^{j-1}$ is the number of possible $(j-1)$-dimensional binary sequences and $\mathbf{b_i^j} = \left(b_{i,1}^j, b_{i,2}^j, \cdots, b_{i,j-1}^j\right)$ $\forall i \in \{0, \cdots, N_j - 1\}$ and $j \in \{1, \cdots, J\}$, 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, $P\left\{\varepsilon_j|\mathbf{b_i^j}\right\}$ 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:

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

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 $\boldsymbol{n}_{eq} = \sum_{k=1}^{j-1} \sqrt{p_k}\mathbf{g}_k(x_k - \hat{x}_k) + \mathbf{n}^j$. The resulting term is approximated as a centered Gaussian random variable, where $\mathbb{E}\left\{\boldsymbol{n}_{eq}\right\} = \mathbf{0}$ and $\mathbb{E}\left\{\boldsymbol{n}_{eq}\boldsymbol{n}_{eq}^H\right\} = (\sum_{k=1}^{j-1} p_k \left\|g_k\right\|^2 \mathbb{E}\left\{\left\|x_k - \hat{x}_k\right\|^2\right\} + \sigma^2)\mathbf{I} = (\sum_{k=1}^{j-1} p_k \left\|g_k\right\|^2 \delta_k d + \sigma^2)\mathbf{I}$. 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:

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

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:

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

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

$$P\left\{\mathbf{b}_i^j\right\} = P\left\{\cap_{n=1}^{j-1} b_{i,n}^j\right\} = \prod_{n=1}^{j-1} P\left\{b_{i,n}^j \mid \cap_{m=1}^{n-1} b_{i,m}^j\right\}, \tag{15}$$

where $P\left\{b_{i,n}^j \mid \cap_{m=1}^{n-1} b_{i,m}^j\right\}$ 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:

$$P\left\{b_{i,n}^j \mid \cap_{m=1}^{n-1} b_{i,m}^j\right\} = \tag{16}$$

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

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 $\mathbf{p}_{\mathrm{opt}}$, that minimizes the global average error probability, can be obtained by solving the following problem:

$$OP_1 \begin{cases} \min\limits_{\mathbf{p}} & \dfrac{1}{J\log_2(M)} \displaystyle\sum_{j=1}^{J} \sum_{i=0}^{N_j-1} P\left\{\varepsilon_j | \mathbf{b_i^j}\right\} P\left\{\mathbf{b_i^j}\right\} & \text{(18a)} \\ & p_j \leq p^U \quad \forall j \in \mathcal{J} & \text{(18b)} \end{cases}$$

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

5. Proposed autonomous power decision algorithm

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

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

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

Algorithm 1: Autonomous power decision

Require: $\mathbf{p}^j_{\max} = \boldsymbol{p}^U$, $\mathbf{p}^j_{\min} > \mathbf{0}\ \forall j = 1, 2, \cdots, J$;
Ensure: $\mathbf{p}$

1: Each user picks up its spreading sequences.
2: Each user selects a random power level $p^j \in [p^j_{\min}, p^j_{\max}]$.
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 $p^j_{\min}$ or $p^j_{\max}$:

a) If $\text{ACK} = 3 \Rightarrow p^j = p^u > 0$
b) If $\text{ACK} = 2 \Rightarrow p^j_{\max} = p^j \leq p^U$
   - If $\text{Stat} = 1 \Rightarrow p^j_{\min} = 0$
c) If $\text{ACK} = 1 \Rightarrow p^j_{\min} = p^j > 0$
   - If $\text{Stat} = 1 \Rightarrow p^j_{\max} = p^U$
d) If $\text{ACK} = 0 \Rightarrow$ no update

9: Return to step 2

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

6. Power allocation with multi-armed bandits

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

according to the calculated index that depends on each algorithm policy. Over time, these techniques 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:

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

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

a) Scenario 1:

The base station acknowledgement at the t-th iteration is carried on 1 bit representing the 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 j-th user is successfully decoded and $r_j^t(a_j) = 0$ otherwise. Therefore, the successful transmission rate 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 user only depends on the decoding status of its own packet without any consideration to the other users. However, the successful decoding event of one packet depends on the successful decoding of the others, because of the SIC receiver. Hence every user has interest on good power selection for the other users and not only for itself. The scenario 2, we propose hereafter, takes into account this fact.

b) Scenario 2:

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

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:

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

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

At the initialization phase, UCB1 explores each arm once in order to have an estimation of the reward of each arm. Then, at each iteration, each user selects the arm with the highest index, as illustrated in Algorithm 2. The calculated index (20) ensures the balance between the exploration of the most uncertain arms and the exploitation of the best arm so far. UCB1 prescribes the principle of "optimism face uncertainty" which means that the less visited arm seems more uncertain and thereby 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: $\underset{a_j}{\text{argmax}}\ m_j^{t-1}(a_j) + \sqrt{\frac{\theta \log(t-1)}{n_j^{t-1}(a_j)}}$

    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

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., $\epsilon << 1$, 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]:

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

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

Algorithm 3: ϵ-decreasing greedy algorithm

Require: L and N
for $t = 1 : T$ **do**
for $j = 1 : J$ **do**
Select a random arm with probability $\epsilon(t) = \min\left\{1, \frac{LN}{t}\right\}$
Select with probability $1 - \epsilon(t)$ the best arm: $\underset{a_j}{\text{argmax}}\ m_j^{t-1}(a_j)$
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

6.3. Thompson sampling algorithm

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

Algorithm 4: Thompson sampling algorithm

Require: N and $\alpha_{j,k} = \gamma_{j,k} = 2\ \forall k = 1...N$ and $\forall j = 1...J$
for $t = 1 : T$ **do**
for $j = 1 : J$ **do**
Select a sampled index from the beta distribution of each arm $i_{j,k} \sim \beta(\alpha_{j,k}, \gamma_{j,k})$ $\forall k = 1, \cdots, N$
Play the arm a_j with the highest index $\hat{i}_{j,q} = \max_{k \in \mathcal{N}}(i_{j,k})$
Update the following variables:

a) $n_j^t(a_j) = n_j^{t-1}(a_j) + 1$
b) $m_j^t(a_j) = \frac{1}{n_j^t(a_j)} \sum_{i=1}^{t} r_j^i(a_j)$
c) $\alpha_{j,k} = \alpha_{j,k} + 3r_j^t(a_j)$
d) $\gamma_{j,k} = \gamma_{j,k} + 3(1 - r_j^t(a_j))$

end
end

7. Complexity and overhead analysis

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

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

	Signaling overhead	Complexity at UE	Power decision
Centralized allocation	O(J · 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

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

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

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

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

8. Numerical results and analysis

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

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

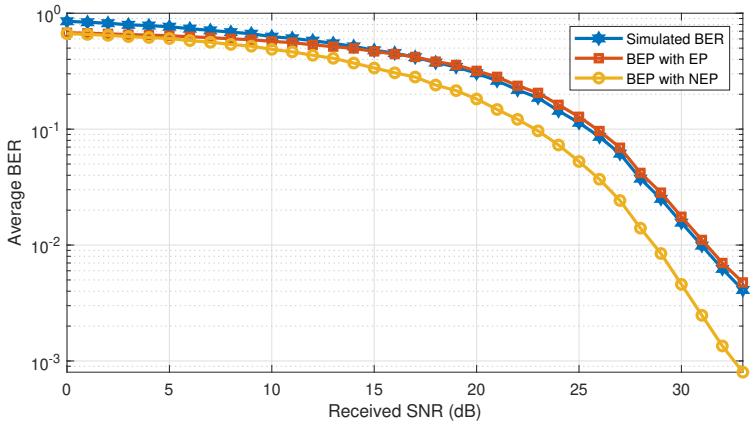

Figure 2. Performance comparison of the simulated BER and the analytical BEP for an AWGN channel with different users path-loss and equal transmission powers.

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

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

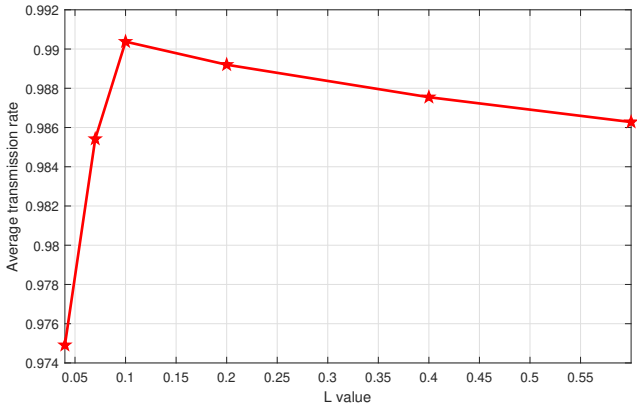

Figure 3. Performance comparison of ϵ-decreasing greedy for different L values after $T = 1000$ iterations in scenario 1.

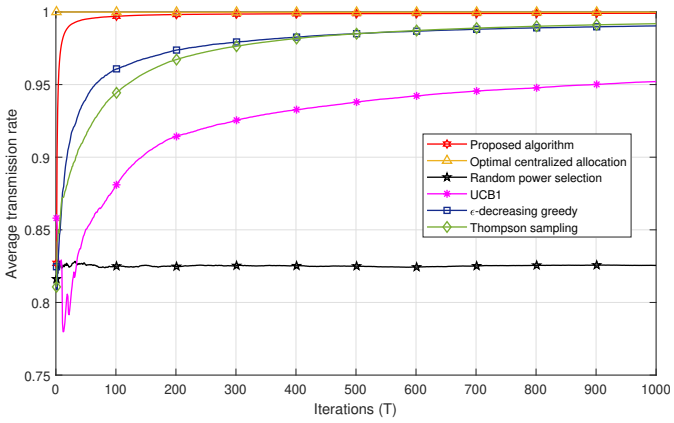

Figure 4. Successful transmission rate comparison for all algorithms in scenario 1.

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

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

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

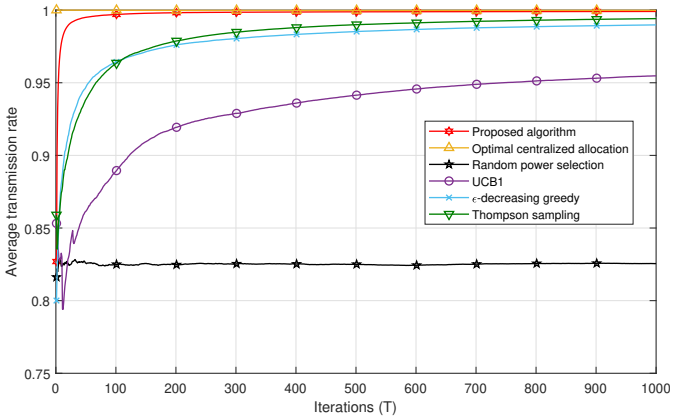

Figure 5. Successful transmission rate comparison for all algorithms in scenario 2.

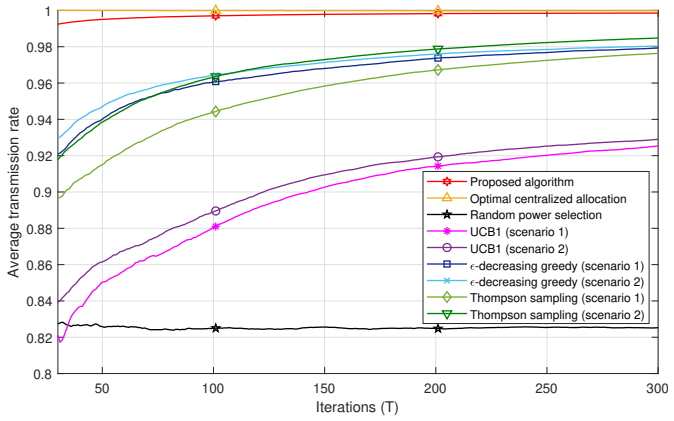

Figure 6. Successful transmission rate comparison for all algorithms in scenarios 1 and 2.

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

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

9. Conclusion

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

Author Contributions: Conceptualization, W.B.A., P.M., J.F.H., M.D. and J.S.; methodology, W.B.A.; software, 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 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 read and agreed to this version of the manuscript.

Funding: This research received no external funding.

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

References

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

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