



**HAL**  
open science

# Study of non-orthogonal multiple access schemes for the massive deployment of IoT in future cellular networks

Wissal Ben Ameer

## ► To cite this version:

Wissal Ben Ameer. Study of non-orthogonal multiple access schemes for the massive deployment of IoT in future cellular networks. Networking and Internet Architecture [cs.NI]. INSA de Rennes, 2021. English. NNT : 2021ISAR0011 . tel-03640601

**HAL Id: tel-03640601**

**<https://theses.hal.science/tel-03640601>**

Submitted on 13 Apr 2022

**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 destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THESE DE DOCTORAT DE

L'INSTITUT NATIONAL DES SCIENCES  
APPLIQUEES RENNES

ECOLE DOCTORALE N° 601  
*Mathématiques et Sciences et Technologies  
de l'Information et de la Communication*  
Spécialité : *Télécommunications*

Par

**Wissal BEN AMEUR**

**Etude des mécanismes d'accès multiples non orthogonaux pour le  
déploiement massif de l'IoT dans les futurs réseaux cellulaires**

Thèse présentée et soutenue à « Grenoble » le « 25/05/2021 »

Unité de recherche : Institut d'Électronique et de Télécommunications de Rennes (IETR), INSA Rennes  
TGI/OLS/BIZZ/MIS/CITY, Orange Labs Meylan

Thèse N° : 21ISAR 11 / D21 - 11

## Rapporteurs avant soutenance :

Laurent ROS                      Professeur à l'INP Grenoble  
Claire GOURSAUD              Maître de Conférences HDR à l'INSA Lyon

## Composition du Jury :

Président :	Jean-Pierre CANCES	Professeur à l'ENSIL de Limoges
Examineurs :	Laurent ROS	Professeur à l'INP Grenoble
	Claire GOURSAUD	Maître de Conférences HDR à l'INSA Lyon
	Jean-Pierre CANCES	Professeur à l'ENSIL de Limoges
	Laurent CLAVIER	Professeur à Télécom Lille
	Marion DUMAY	Docteur-Ingénieur à Orange Labs Meylan
	Jean SCHWOERER	Docteur-Ingénieur à Orange Labs Meylan
Dir. de thèse :	Jean-François HELARD	Professeur à l'INSA Rennes
Co-dir. de thèse :	Philippe MARY	Maître de Conférences HDR à l'INSA Rennes

*This thesis is dedicated to my family  
for their love, unwavering support  
and belief in me.*



# *Acknowledgments*

The last three years of my thesis have been intense and challenging with a lot of ups and downs. However, this experience was rich and constructive both on a human and professional level.

First and foremost, I would like to express my sincere gratitude to my esteemed supervisors Jean-François H elard, Philippe Mary, Marion Dumay and Jean Schworer for their patience, their availability and especially their sound advice which contributed to my reflection. Their professionalism, thoroughness and invaluable attention greatly contributed to the success of this thesis. In addition, their human qualities during these three years touched me deeply.

I would like to express my warmest thanks to all the members of the jury who have done me the honor of carefully studying my work : Laurent Ros and Claire Goursaud for having accepted to be reviewers of this thesis ; Jean-Pierre Cances and Laurent Clavier for agreeing to examine this thesis. I am deeply grateful to them for their very careful reading, their valuable remarks and the interest shown in my work.

From the bottom of my heart I would like to offer my special thanks to my parents, my brothers, my sisters, my sister in law and my little nephew Talid, whom without this would have not been possible. I would like to thank them for their unwavering encouragement and treasured support all through my studies, especially during my difficult times.

Finally, my appreciation also goes out to my friends, and my colleagues in Orange Labs for a cherished time spent together.



---

# Résumé en français

## Introduction

Depuis quelques années, le concept d'Internet of Things (IoT) prend de plus en plus d'ampleur avec des dizaines de milliards d'objets connectés, environ 100 milliards d'objets connectés d'ici 2030 selon Cisco [1]. En fait, l'apparition de nouvelles applications avec différentes qualités de service conduit à une évolution significative des futurs réseaux. Les nouveaux services sont classés par l'union internationale des télécommunications (UIT) et le projet de partenariat de troisième génération (3GPP) en trois catégories principales avec des exigences différentes, à savoir le haut débit mobile amélioré (eMBB), les communications ultra fiables à faible latence (URLLC) et les communications massives de type machine (mMTC). Ces scénarios visent à surmonter les limitations des technologies précédentes concernant, respectivement, la bande passante, la latence et la capacité de connectivité :

- ▶ eMBB est le premier service commercialisé au grand public de la 5G. Il est considéré comme une évolution considérable des applications haut débit par rapport au réseau 4G garantissant une meilleure expérience utilisateur [2]. Les principaux défis consistent à offrir une efficacité spectrale plus élevée avec un débit de données important, 10 Gbps en liaison montante et 20 Gbps en liaison descendante, et une fiabilité de  $10^{-3}$  de taux d'erreur paquet [3]. Les exigences d'eMBB sont nécessaires pour prendre en charge certains services tels que les communications dans les transports avec une mobilité à grande vitesse et les grands événements nécessitant un débit élevé de données pour desservir tous les utilisateurs dans la même zone. Selon Ericsson, environ 1,9 milliard d'abonnés aux réseaux 5G devraient bénéficier du service eMBB d'ici 2024 [4].
- ▶ URLLC est l'un des scénarios les plus révolutionnaires des réseaux 5G [5]. Il a été révélé lors de la première phase d'étude de la 5G. URLLC est particulièrement dédié aux applications critiques en termes de latence, de sécurité et de fiabilité telles que les véhicules autonomes et les applications e-santé qui sont considérés comme des services innovants pour les futurs réseaux. Cependant, les technologies existantes souffrent de valeurs de latence élevées. Par exemple, le réseau LTE-Advanced offre une latence d'environ 10 ms. Cette valeur ne convient pas aux applications d'URLLC qui nécessitent une interaction rapide

avec la station de base. Par exemple, pour les services de télé-chirurgie, la vie du patient est en jeu et donc aucun retard ne peut être toléré. Par conséquent, selon les spécifications de l'UIT et du 3GPP, les services d'URLLC nécessitent des transmissions efficaces avec une fiabilité de bout en bout de  $10^{-5}$  de taux d'erreur paquet [3] et une faible latence de 1 ms [6].

- Le scénario mMTC est considéré comme l'un des aspects clés pour la création de futures villes intelligentes et ultra-connectées [7]. En fait, le scénario mMTC est conçu pour traiter un grand nombre d'objets connectés. Cette exigence de connectivité sera satisfaite par des cellules denses avec 1 million d'objets connectés par  $\text{km}^2$  [6], [8]. Les objets pris en charge sont conçus avec des matériaux ultra-économiques, ce qui réduit leur prix à l'ordre de quelques dollars [7], [9]. Cela permettra la mise en place de plusieurs applications mMTC, telles que les compteurs intelligents, les applications de suivi et les maisons connectées, avec un faible coût de déploiement. Les scénarios mMTC sont caractérisés par des transmissions intermittentes de paquets courts, de l'ordre de quelques octets, avec une faible complexité du système. En fait, l'aspect MTC est basé sur des communications de machine à machine avec une intervention humaine minimale. Pour cette raison, les appareils connectés doivent fonctionner de manière autonome pendant une longue période de 10 à 15 ans [8]. En outre, étant donné que les scénarios du mMTC doivent couvrir un grand nombre d'objets qui peuvent être dispersés sur toute la cellule et subir de très mauvaises conditions de canal, ces objets devraient donc bénéficier d'une large couverture tout en tolérant une perte de canal maximale de 164 dB [10].

Depuis l'introduction du concept MTC [11], des efforts énormes ont été fait pour répondre aux demandes pressantes de ses applications tout en introduisant plusieurs technologies, soit par le 3GPP [12] ou d'autres sociétés propriétaires [13], qui offrent un débit de données relativement faible, une faible complexité du système et une consommation d'énergie efficace pour des objets à faible coût. Cependant, les solutions proposées ne peuvent être convenables que pour certains cas d'utilisation. D'une part, les technologies non 3GPP, telles que LoRa [14] et SigFox [15], sont des solutions propriétaires qui sont déployées dans des bandes de fréquences sans licence avec des normes privées et sans aucune garantie de performance du fait de l'utilisation partagée de la bande. Par conséquent, pour un grand nombre de connexions simultanées, le niveau d'interférence augmente considérablement, entraînant une dégradation significative des performances [9]. De plus, leurs infrastructures ne sont disponibles que dans certaines régions du monde, ce qui limite leur disponibilité et donc leur adaptabilité aux services à grande échelle, tels que les applications de localisation et les voitures connectées, nécessitant un service de roaming lors de leurs déplacements. D'autre part, pour les solutions 3GPP standardisées telles que eMTC [16], NB-IoT [17] et EC-GSM-IoT [18] avec une infrastructure globale et des bandes de fréquences sous licence, le nombre maximum d'objets servis est très limité par rapport à l'objectif du mMTC. Par exemple, NB-IoT et eMTC sont capables de servir un nombre maximum de 50,000 objets par cellule [19], ce qui est loin du défi du mMTC de 1 million d'objets connectés par  $\text{km}^2$  [19]. En outre, les solutions 3GPP sont dérivées des réseaux LTE et GSM. Par conséquent, elles génèrent une

quantité excessive de données de signalisation qui peut être plus importante que la taille des données transmises dans les scénarios mMTC. Cela peut entraîner un gaspillage de ressources, une consommation d'énergie élevée et une latence élevée pour la transmission de quelques octets.

Pour les mécanismes d'accès multiples orthogonaux (OMA) classiques, le nombre d'utilisateurs pouvant communiquer simultanément dépend notamment de la quantité de ressources disponibles et de la granularité de l'ordonnancement, ce qui conduit à un niveau limité d'interférence inter-utilisateur avec des techniques simples de détection de signaux. Cependant, avec la croissance rapide du nombre d'objets connectés, on peut s'interroger sur la capacité de ces mécanismes d'accès orthogonaux à satisfaire toutes les demandes de connexions simultanées du mMTC. Cette limite peut être dépassée en renonçant à l'orthogonalité inter-utilisateur. Dans ce contexte, les mécanismes d'accès multiple non orthogonaux (NOMA) sont considérés comme une solution prometteuse pour relever les défis du mMTC [20]. Ils permettent à plusieurs utilisateurs de partager simultanément et non-orthogonalement la même ressource en utilisant différentes techniques de multiplexage des utilisateurs telles que des codes clairsemés, des séquences d'étalement ou des distributions de puissance différentes, ce qui augmente le nombre d'utilisateurs pris en charge. Mais, ils introduisent également des interférences inter-utilisateur importantes, nécessitant des techniques robustes de détection de signaux avec une complexité de décodage potentiellement élevée.

Dans les systèmes cellulaires, la transmission de données nécessite toujours l'établissement d'un processus d'allocation de ressources. Cette opération peut être très coûteuse en termes de latence et de données de signalisation échangées, entraînant une consommation d'énergie élevée qui peut réduire considérablement l'autonomie de la batterie. Plus précisément, en considérant un scénario mMTC, la quantité de données de signalisation générée peut être plus importante que la taille des données utiles transmises, ce qui induit un gaspillage de ressources pour la transmission de quelques octets. Dans ce contexte, les futurs réseaux se doivent de proposer de nouvelles solutions pour relever les défis critiques du mMTC. Parmi les technologies candidates, nous citons notamment l'accès libre sans allocation de ressources dédiées [21]. En fait, chaque utilisateur peut transmettre ses données selon ses besoins sans aucune communication préalable avec la station de base pour l'allocation de ressources, ce qui réduit la quantité de données de signalisation ainsi que la latence des communications. Cette stratégie n'a jamais été possible avec les schémas OMA conventionnels où chaque ressource doit être strictement attribuée à un seul utilisateur. Ainsi, en utilisant un accès aléatoire sans allocation de ressources avec les schémas OMA, plusieurs utilisateurs peuvent choisir simultanément les mêmes ressources, entraînant une collision entre leurs signaux. La probabilité d'occurrence de ce problème augmente considérablement pour un grand nombre d'objets connectés. Dans ce cas, la solution typique avec des tentatives de retransmission n'est pas efficace, et cela compliquerait encore le problème en termes de gaspillage de ressources, de délai de transmission et de signalisation générée. Les discussions du 3GPP ont abouti à deux options possibles pour réaliser des schémas NOMA avec un accès libre sans allocation de ressources [22]. Pour la première option, chaque utilisateur établit un processus d'accès aléatoire avec la station de base pour assurer sa synchronisa-

tion. Ensuite, une fois que l'utilisateur est connecté à la station de base, il transmet ses paquets directement sans demander de ressources dédiées. Pour la deuxième option, chaque utilisateur envoie ses paquets sans aucune communication préalable avec la station de base, mais également sans aucune garantie de synchronisation avec la station de base, notamment en liaison montante.

Étant donné le grand potentiel des schémas NOMA pour surmonter les restrictions des schémas OMA conventionnels concernant les défis du mMTC, cette thèse fournit une analyse approfondie des schémas NOMA dans le contexte du mMTC en liaison montante. L'idée est de répondre aux limites de NOMA en termes de complexité d'implémentation des émetteurs et des récepteurs, et de performance en proposant de nouvelles optimisations afin de répondre au mieux aux défis du mMTC. De plus, nous cherchons à aborder le problème d'accès au canal, avec ou sans allocation de ressources. Notre objectif est d'assurer un compromis entre la performance souhaitée et la satisfaction des exigences du mMTC en termes de nombre de connexions simultanées, de consommation d'énergie et de complexité du système.

Dans le chapitre 1, nous présentons l'état de l'art sur les enjeux des futurs réseaux 5G. Nous présentons également les technologies les plus prometteuses pour répondre aux exigences de l'aspect MTC. De plus, nous introduisons les principales limites et inconvénients de ces solutions face aux défis du MTC massif. Dans le chapitre 2, nous fournissons une étude qualitative des différents mécanismes NOMA. Ensuite, une comparaison de performance de ces schémas est présentée dans le chapitre 3. En outre, dans le chapitre 4, nous abordons le problème d'allocation de puissance centralisée par la station de base afin d'améliorer les performances du système et de traiter le problème des interférences entre les utilisateurs. Dans le chapitre 5, nous nous intéressons au cas d'accès libre sans allocation de ressources en proposant un nouvel algorithme pour une décision de puissance décentralisée au niveau de chaque utilisateur.

Cette thèse s'est déroulée au sein de l'équipe CITY d'Orange Labs Meylan en collaboration avec le laboratoire IETR (Institut d'Electronique et des Technologies du numéRique) de l'Institut National des Sciences Appliquées de Rennes (INSA Rennes).

## **Chapitre 1 : Contexte de la 5G et des communications de type machine**

Le chapitre 1 s'intéresse à l'état de l'art. Il présente les principaux services des futurs réseaux 5G, à savoir eMBB, URLLC et mMTC. Plus précisément, il s'est concentré sur le concept MTC tout en présentant les principaux enjeux de la communication machine à machine et les étapes d'évolution vers le scénario mMTC. Ce chapitre présente les solutions candidates pour répondre aux exigences du MTC. Ces solutions se répartissent en deux grandes catégories : i) Les solutions 3GPP, telles que eMTC, EC-GSM-IoT et NB-IoT, qui sont dérivées des réseaux LTE et GSM existants et bénéficient d'une infrastructure présente dans le monde entier, conduisant à un faible coût de déploiement ; ii) Les solutions non 3GPP telles que LoRa et SigFox qui disposent d'infrastructures propriétaires et de bandes de fréquences sans licence. En

outre, un aperçu des principales limites de ces technologies concernant le scénario mMTC est fourni. Ce chapitre introduit également les limitations des schémas OMA conventionnels et les principaux enjeux de la stratégie d'accès aléatoire aux canaux avec une allocation de ressources pour répondre aux exigences du mMTC en termes de capacité de connectivité et de consommation d'énergie.

## Chapitre 2 : Schémas d'accès multiples non orthogonaux

Au cours des dernières années, plusieurs schémas NOMA ont été introduits avec différentes techniques de multiplexage telles qu'une distribution de puissance non uniforme, différents codes ou différents entrelaceurs. Parmi ces mécanismes, certains ont particulièrement attiré l'attention de la communauté scientifique à savoir, le schéma NOMA dans le domaine de puissance (PD-NOMA), l'accès multiple par code clairsemé (SCMA) [23], l'accès partagé multi-utilisateurs (MUSA) [24], l'accès multiple par division de motif (PDMA) [25] et l'accès multiple par division d'entrelaceur (IDMA) [26]. Ce chapitre présente les caractéristiques clés de ces schémas. De plus, il fournit une comparaison qualitative entre eux par rapport aux exigences strictes du mMTC telles que la capacité de connectivité, la complexité du décodage et leurs scénarios les plus appropriés. Ce chapitre résume également les principaux avantages et inconvénients de chaque schéma. En outre, il présente un aperçu des stratégies existantes pour l'accès aux ressources ainsi que l'évolution progressive vers les options d'accès libre sans allocation de ressources.

## Chapitre 3 : Évaluation des schémas NOMA

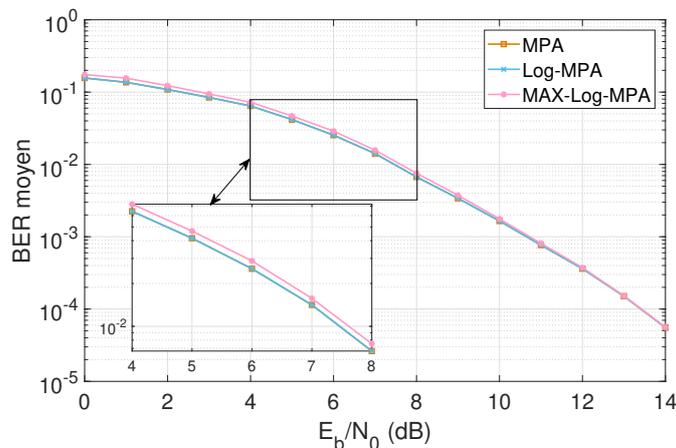
Le chapitre 3 fournit une évaluation quantitative des schémas étudiés. Ce chapitre vise à renforcer l'étude qualitative rapportée dans le chapitre précédent. Les principaux schémas NOMA sont comparés dans les mêmes conditions du système, en considérant un canal de Rayleigh avec des utilisateurs équidistants de la station de base ou uniformément dispersés dans la cellule avec différentes pertes de trajet. L'objectif est de choisir la technique la plus adaptée à notre cas d'étude qui offre le meilleur compromis entre les performances souhaitées et la satisfaction des exigences du service mMTC. Les résultats obtenus prouvent que les deux schémas SCMA et MUSA sont des solutions prometteuses, mais elles ont des compromis différents entre les performances et la complexité du système. La technique SCMA offre les meilleures performances avec une forte résistance à la variation des environnements de propagation et l'augmentation du nombre d'utilisateurs par rapport aux schémas MUSA, PDMA et PD-NOMA, mais au prix d'une complexité du système élevée. Alors que MUSA a une complexité du système inférieure, mais avec une dégradation des performances. Étant donné que la complexité du système est un aspect important pour mMTC, dans ce qui suit, nous étudions les performances des récepteurs candidats du schéma SCMA.

Généralement, SCMA est implémenté avec le récepteur MPA (message passing

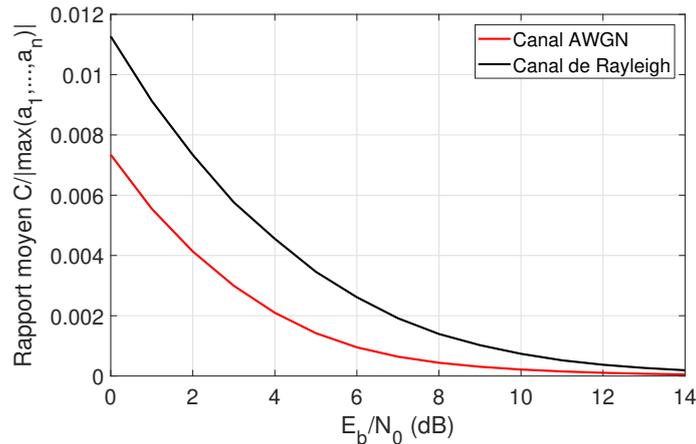
algorithm) [27], qui est connu pour ses bonnes performances de détection multi-utilisateurs par rapport à d'autres algorithmes tels que l'algorithme d'annulation successive des interférences (SIC). Cependant, l'algorithme MPA présente une complexité élevée due au grand nombre de termes exponentiels utilisés. De ce fait, deux versions simplifiées du MPA, nommées MAX-Log-MPA [28] et Log-MPA [29], inspirées des simplifications sous-optimales de l'algorithme de maximum a posteriori (MAP), Log-MAP et MAX-Log-MAP [30], ont été étudiées.

L'idée principale de Log-MPA et MAX-Log-MPA est d'appliquer le processus de décodage itératif du MPA dans le domaine logarithmique afin de réduire sa complexité. Cependant, le principal défi dans le domaine logarithmique est de calculer précisément la fonction  $\log(\exp(a_1) + \dots + \exp(a_n))$ . Cette complexité peut être réduite dans l'algorithme MAX-Log-MPA en utilisant la fonction approximative de maximisation suivante  $\log(\exp(a_1) + \dots + \exp(a_n)) \approx \max(a_1, \dots, a_n)$ . Par conséquent, MAX-Log-MPA est considéré comme un algorithme sous-optimal en raison de cette approximation de maximisation. Log-MPA vise à atteindre les mêmes performances que MPA mais avec une complexité réduite. L'idée principale est d'ajouter un terme de correction  $C$  à l'approximation de l'algorithme MAX-Log-MPA pour compenser sa dégradation des performances. Grâce au terme de correction, la perte de performance due à l'utilisation de la fonction de maximisation devrait être réduite. Cependant, cela se fait au prix d'une complexité supplémentaire, par rapport à l'algorithme MAX-Log-MPA, en raison de l'utilisation d'une somme de termes exponentiels.

À notre connaissance, aucun travail existant ne regroupe, dans une étude complète, les performances des algorithmes MPA, Log-MPA et MAX-Log-MPA dans un scénario SCMA de liaison montante pour un canal gaussien (AWGN) et un canal de Rayleigh. De plus, la comparaison de ces algorithmes ne peut pas être déduite de plusieurs articles de la littérature, dû notamment au manque de conditions communes de simulation ou dans la conception des codes [27]. Pour cette raison, nous présentons une comparaison approfondie des performances entre ces algorithmes via des simulations numériques.



**Fig. 1.** BER moyen pour un canal AWGN,  $J = 6$  utilisateurs et  $K = 4$  sous-porteuses.



**Fig. 2.** Comparaison du rapport moyen pour les canaux AWGN et de Rayleigh.

La Figure 1 montre le BER obtenu pour les trois algorithmes étudiés dans un canal AWGN. Nous observons que Log-MPA ne peut pas être distingué du MPA grâce à l'utilisation du terme de correction, ce qui permet de compenser la perte de performance par rapport à l'algorithme MAX-Log-MPA. En effet, une perte de performance peut être enregistrée entre MAX-Log-MPA et MPA (et donc Log-MPA) de 0.26 dB pour un BER de  $3 \times 10^{-2}$ . Cet écart provient de l'approximation utilisée. De plus, le BER obtenu avec MAX-Log-MPA converge vers ceux obtenus avec MPA/Log-MPA lorsque le  $E_b/N_0$  augmente. Ce comportement est dû au fait que le biais introduit par l'approximation diminue fortement pour les valeurs élevées de  $E_b/N_0$  comme illustré dans la Figure 2. Les algorithmes se comportent de la même manière dans un canal de Rayleigh.

Nous étudions dans la Figure 2 l'évolution du rapport moyen, sur tous les mots de code transmis possibles, entre le terme de correction  $C$  et la valeur absolue de la fonction de maximisation, c-à-d  $|\max(a_1, \dots, a_n)|$ , utilisés dans l'algorithme Log-MPA. Nous remarquons que, pour un canal AWGN et un canal à évanouissement de Rayleigh, ce rapport a des valeurs très basses et il décroît asymptotiquement vers zéro lorsque  $E_b/N_0$  augmente. Cela signifie que le terme de correction devient négligeable par rapport au terme de maximisation lorsque le SNR augmente, ce qui explique le faible écart entre les courbes des algorithmes MPA/Log-MPA et celle du MAX-Log-MPA. De plus, nous remarquons que le rapport pour le canal de Rayleigh est visiblement plus important que celui du canal Gaussien pour un rapport de  $E_b/N_0$  donné, ce qui justifie le fait que la dégradation de la performance de ce dernier est légèrement inférieure à celle du canal de Rayleigh.

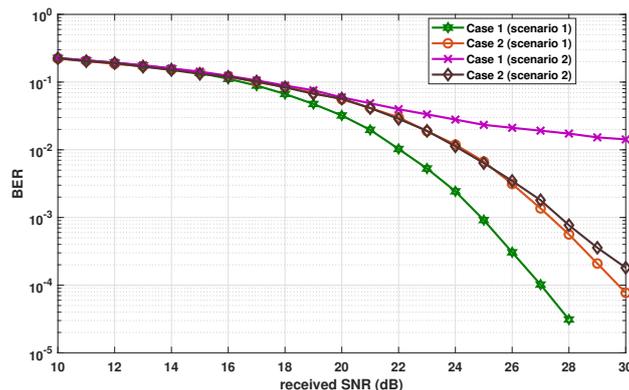
Cependant, malgré les versions simplifiées du MPA, la complexité du décodage du schéma SCMA est toujours élevée. De plus, la technique SCMA nécessite une attribution de codes au préalable, ce qui induit une grande quantité de signalisation entre les utilisateurs et la station de base, et peut donc être une contrainte importante dans le scénario mMTC avec une faible consommation d'énergie. La complexité de la conception des mots de code demeure également un problème important. En outre, le nombre maximal de codes possibles pour SCMA, avec leurs fortes contraintes de conception, peut être insuffisant pour atteindre un grand nombre de connexion simultanées tel qu'envisagé pour le service mMTC.

## Chapitre 4 : Allocation de puissance pour la minimisation du BER du schéma MUSA

Selon les résultats présentés dans le chapitre 3, la technique MUSA a également montré de bonnes performances par rapport aux autres schémas, notamment en cas de puissances reçues différentes. En fait, le schéma MUSA est généralement mis en oeuvre avec un récepteur SIC pour la séparation des signaux, qui a une faible complexité. Cependant, contrairement à certains récepteurs tels que MPA et MAP qui fonctionnent mieux avec des puissances reçues égales, la technique SIC atteint ses meilleures performances pour des puissances reçues suffisamment différentes. Par conséquent, l'allocation de puissance est considérée comme un axe majeur d'amélioration pour le schéma MUSA. Une allocation inappropriée de puissance peut conduire à une détérioration des performances à cause d'un problème de propagation d'erreur. Pour cette raison, dans le chapitre 4, nous proposons une approche d'allocation de puissance afin de palier à ce problème. Nous formulons un problème d'optimisation pour minimiser le BER moyen prenant en compte soit la contrainte de la puissance totale du système ou soit la contrainte de la puissance d'émission individuelle maximale, et une autre contrainte sur la différence de puissance reçue entre l'utilisateur actuellement décodé et ceux qui interfèrent. Puisque les termes de puissances des utilisateurs sont présents à la fois dans le numérateur et le dénominateur du SINR de chaque utilisateur, le problème d'optimisation n'est pas convexe sur le vecteur de puissance. Dans ce cas, les conditions de Karush–Kuhn–Tucker ne sont pas suffisantes pour garantir une solution globale optimale. Par conséquent, une approximation simplifiée du SINR a été proposée dans [31], où seule la puissance de l'utilisateur actuellement décodé est impliquée dans l'expression du SINR. Le problème d'optimisation est divisé en deux sous-problèmes avec deux scénarios différents considérant une puissance totale du système dans le premier scénario ou une puissance individuelle maximale dans le deuxième. Ensuite, chaque scénario est étudié dans deux cas différents, avec ou sans la contrainte de différence de puissance comme expliqué dans la Table 1 pour un système à  $J$  utilisateurs.

TABLE 1: Les cas et les scénarios étudiés

	Cas 1	Cas 2
Scénario 1	$\sum_{j=1}^J p_j \leq P_{tot}$	$\sum_{j=1}^J p_j \leq P_{tot}$ $p_j \ \mathbf{g}_j\ ^2 - \sum_{i=j+1}^J p_i \ \mathbf{g}_i\ ^2 \geq P_{SIC} \forall j \in \{1, 2, \dots, J\} \setminus j$
Scénario 2	$p_j \leq P_{max} \forall j \in \{1, 2, \dots, J\}$	$p_j \leq P_{max} \forall j \in \{1, 2, \dots, J\}$ $p_j \ \mathbf{g}_j\ ^2 - \sum_{i=j+1}^J p_i \ \mathbf{g}_i\ ^2 \geq P_{SIC} \forall j \in \{1, 2, \dots, J\} \setminus j$



**Fig. 3.** Comparaison des performances pour les différents cas et scénarios considérés pour un canal AWGN,  $J = 6$  et  $K = 4$ .

La Figure 3 montre la comparaison du BER pour les différents cas et scénarios étudiés avec un canal AWGN en fonction du SNR reçu. Nous observons que le premier cas du scénario 1 avec seulement la contrainte de puissance totale atteint les meilleures performances. En effet, en se basant uniquement sur les différents SINRs des utilisateurs, qui sont générés par la mise à jour de la matrice d'erreur quadratique moyenne minimale (MMSE) à chaque itération du récepteur MMSE-SIC, la contrainte de puissance totale offre un degré de liberté maximal pour l'allocation de puissance, ce qui conduit à une amélioration significative du BER moyen. Considérant le cas réaliste d'une contrainte de puissance individuelle (cas 1 du scénario 2), nous observons que les performances obtenues sont médiocres. En effet, dans ce cas, tous les utilisateurs transmettent avec leur puissance maximale autorisée, et donc tous les signaux sont reçus avec la même puissance puisque tous les utilisateurs ont le même canal. En fait, les SINRs des utilisateurs ne se distinguent que par leurs séquences d'étalement sélectionnées, ce qui n'est pas suffisant pour assurer un décodage correct. Par conséquent, la séparation des signaux subit une propagation d'erreur sévère, ce qui dégrade considérablement le BER du système avec un plancher d'erreur d'environ  $10^{-2}$  à 30 dB. Ensuite, l'ajout de la contrainte SIC de différence minimale entre les puissances reçues, dans le second cas des deux scénarios, conduit à des résultats similaires pour les deux scénarios, mais pas aussi bons que le cas 1 du scénario 1. En revanche, il permet d'améliorer significativement les performances par rapport au cas 1 du scénario 2. Cette contrainte assure une distribution adéquate de la puissance reçue, conduisant à une réduction du phénomène de propagation d'erreur et donc à une meilleure détection des signaux. En outre, il faut noter que, dans le cas 2 du scénario 2, la combinaison de contraintes ne permet pas de consommer entièrement le budget de puissance disponible, alors les résultats seraient encore meilleurs en considérant une puissance totale reçue égale à celle du scénario 1. Nous remarquons aussi que, pour  $\text{SNR} \leq 20$  dB, il n'y a pas de différence significative entre tous les cas étudiés. En fait, à cause du bruit élevé, le processus de décodage itératif est perturbé par un problème de propagation d'erreur, ce qui induit un BER important.

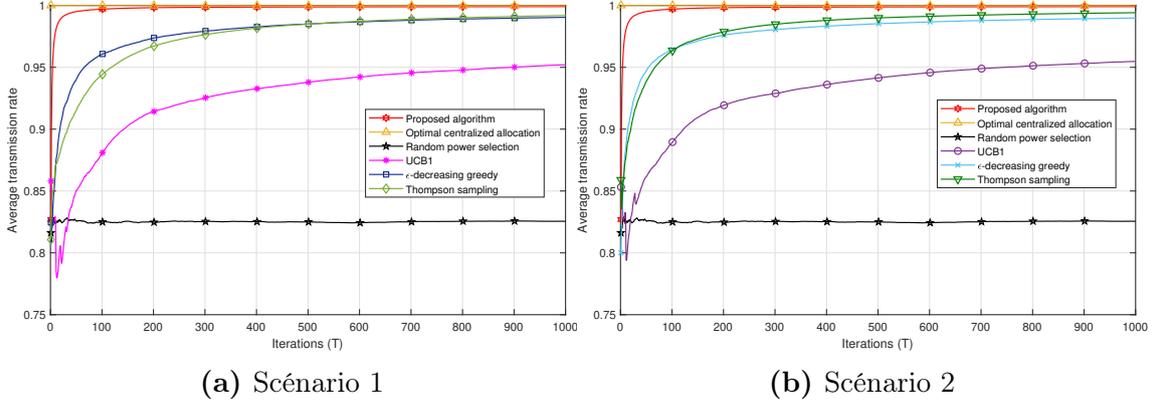
## Chapitre 5 : Décision autonome de puissance pour un accès sans allocation de ressources

Le dernier chapitre est dédié à l'étude d'un scénario d'accès libre sans allocation de ressources. L'objectif principal est de permettre à chaque utilisateur de transmettre directement ses données sans attendre une autorisation ou une allocation de ressources de la station de base afin de réduire les données de signalisation. Le schéma MUSA est présenté comme l'un des schémas NOMA les plus prometteurs avec un grand potentiel pour permettre un accès libre sans allocation de ressources puisque les séquences d'étalement peuvent être conçues facilement et de manière autonome par les utilisateurs. Cependant, la technique MUSA avec le récepteur SIC souffre principalement du problème de propagation d'erreur qui peut être traité avec une distribution de puissance adéquate comme étudié au chapitre précédent. Le processus d'allocation de puissance est généralement effectué de manière centralisée où la station de base connaît les informations du canal de tous les utilisateurs alors que, pour un accès libre sans allocation de ressources, chaque utilisateur effectue une transmission aveugle sans aucune information sur son environnement de propagation et les utilisateurs interférents, ce qui rend le calcul de la puissance plus complexe.

Dans ce chapitre, nous abordons ce problème de décision autonome de puissance en considérant une signalisation minimale pour répondre aux exigences du mMTC. L'objectif est de trouver un compromis entre le temps de convergence vers une allocation performante similaire à celle de l'allocation de puissance centralisée optimale et le coût de la signalisation. Pour cela, nous étudions les performances du système en termes de taux de transmissions réussies. La valeur de la puissance optimale est assez difficile à calculer, en particulier pour les récepteurs SIC avec le problème de la propagation d'erreur. Pour ce faire, nous commençons par proposer une expression approximative de la probabilité d'erreur binaire (PEB) en considérant l'interférence inter-utilisateur. La valeur de puissance optimale est obtenue en minimisant la PEB globale moyenne. Ensuite, en exploitant l'expression dérivée de la PEB, nous proposons un nouvel algorithme pour une sélection de puissance autonome avec une signalisation réduite. L'idée clé est de faire exécuter par chaque utilisateur un algorithme itératif qui profite de l'acquiescement naturel, envoyé par la station de base lors du décodage du paquet de chaque utilisateur. Chaque utilisateur met à jour progressivement sa puissance d'émission en fonction de la valeur reçue d'acquiescement pour converger vers les puissances les plus proches des valeurs optimales qui assurent un processus de décodage correct pour tous les utilisateurs tout en générant une quantité minimale des données de signalisation.

L'algorithme proposé est comparé à certains algorithmes de bandits (MAB) qui sont adaptés pour ce problème. Les algorithmes MAB sont appliqués avec un modèle de  $N$  puissances discrétisées, appelées bras. Chaque puissance est associée à une récompense suivant une distribution de probabilité spécifique. A chaque intervalle de temps, chaque utilisateur joue un bras selon la stratégie adoptée. Ensuite, il reçoit la récompense correspondante à ce bras. Nous considérons deux scénarios pour la sélection du meilleur bras : i) Dans le premier scénario, la récompense de chaque utilisateur est portée sur un seul bit. Elle ne dépend que de l'état de décodage de son

propre paquet, c-à-d s'il a été correctement reçu ou non, sans aucune considération des autres utilisateurs, ii) Dans le second scénario, la récompense de chaque utilisateur est portée sur deux bits. Elle porte des informations sur l'état de décodage de son propre paquet ainsi que sur l'état des paquets des autres utilisateurs, qu'ils soient reçus avec succès ou non.



**Fig. 4.** Comparaison du taux de transmissions réussies pour tous les algorithmes dans les scénarios 1 et 2.

La Figure 4 compare le taux de transmissions réussies des algorithmes étudiés, c-à-d l'allocation centralisée des puissances, l'algorithme proposé, les algorithmes MAB considérés (" $\epsilon$ -decreasing greedy", "UCB1" et "Thompson sampling") et la sélection aléatoire des puissances, pour un canal AWGN avec des utilisateurs uniformément dispersés dans la cellule. Nous notons que l'algorithme proposé obtient de meilleures performances que les techniques MAB avec une convergence plus rapide vers la puissance optimale obtenue par une allocation centralisée. Nous observons également que l'algorithme " $\epsilon$ -decreasing greedy" converge plus vite que les algorithmes "Thompson sampling" et "UCB1". Cependant, pour un nombre d'itération élevé, l'algorithme "Thompson sampling" surpasse légèrement l'algorithme " $\epsilon$ -decreasing greedy", ce qui est plus clair dans le scénario 2. L'algorithme "UCB1" atteint des performances plus faibles que les algorithmes " $\epsilon$ -decreasing greedy" et "Thompson sampling" dans les deux scénarios car il prend plus de temps pour explorer les puissances sous-optimales, ce qui ralentit sa convergence vers les puissance optimales et induit plus de perte de paquets. De plus, l'allocation de puissance aléatoire présente les pires performances dans les deux scénarios puisqu'aucune stratégie n'est utilisée pour une sélection de puissance adéquate, ce qui induit une propagation d'erreur et donc des pertes de paquets. Nous remarquons que les performances obtenues dans le scénario 2 sont meilleures que celles du scénario 1, mais avec l'inconvénient de données de signalisations plus élevées.

## Conclusion et perspectives

Dans cette thèse, nous avons proposé une étude quantitative et qualitative des différents mécanismes NOMA tels que SCMA, MUSA, PDMA et PD-NOMA. Nous avons abordé également le problème de la complexité du décodage pour le schéma SCMA en comparant les performances du récepteur MPA, avec ses deux simplifications dans le domaine logarithmique, Log-MPA et MAX-Log-MPA. Les comportements de ces algorithmes ont été analysés via une étude exhaustive des informations extrinsèques échangées. De plus, nous avons abordé le problème d'allocation de puissance pour le schéma MUSA afin d'atténuer le problème de propagation d'erreur. Pour ce faire, nous avons proposé un nouvel algorithme d'allocation de puissance centralisée. Cet algorithme a été étudié dans différents scénarios avec différentes combinaisons de contraintes. Enfin, nous avons étudié le problème d'accès sans allocation de ressources en traitant le problème de décision de puissance pour le schéma MUSA dans ce contexte. Nous avons proposé un nouvel algorithme qui permet à chaque utilisateur de choisir de manière autonome, sans aucune information sur l'environnement de propagation ainsi que les autres utilisateurs, sa puissance optimale permettant de réduire le problème de propagation d'erreur avec un décodage correct des paquets de tous les utilisateurs. Cet algorithme a été comparé aux algorithmes MAB et à l'allocation de puissance optimale de manière centralisée pour évaluer son efficacité pour ce problème.

Malgré les problèmes abordés pour les mécanismes NOMA dans cette thèse, d'autres problèmes restent non résolus et pourraient être étudiés plus en détails dans les travaux futurs. On note principalement le problème de la synchronisation. En effet, tous les schémas ont été proposés avec une hypothèse de synchronisation parfaite. Cependant, pour un accès sans allocation de ressources, cette hypothèse ne peut pas être garantie, en particulier pour les transmissions montantes. De plus, la conception et la distribution des codes d'utilisateurs, pour les schémas basés sur des codes, sont considérées comme un défi majeur et peuvent être optimisées afin d'améliorer les performances du système et réduire sa complexité, en particulier pour le schéma SCMA.

# Liste des publications

## Revue internationale

1. Wissal Ben Ameer, Philippe Mary, Jean-François Hélard, Marion Dumay, Jean Schwoerer, "Autonomous Power Decision for the Grant Free Access MUSA Scheme in the mMTC Scenario." *Sensors* 21.1 (2021) : 116.

## Communications internationales

1. Wissal Ben Ameer, , Philippe Mary, Marion Dumay, Jean-François Hélard, Jean Schwoerer, "Performance study of MPA, Log-MPA and MAX-Log-MPA for an uplink SCMA scenario." 2019 26th IEEE International Conference on Telecommunications (ICT 2019).
2. Wissal Ben Ameer, Philippe Mary, Marion Dumay, Jean-François Hélard, Jean Schwoerer, "Power allocation for BER minimization in an uplink MUSA scenario." 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring).



---

# List of Acronyms

1G	First Generation
2G	Second Generation
3G	Third Generation
3GPP	3rd Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
AWGN	Additive White Gaussian Noise
BEP	Bit Error Probability
BER	Bit Error Rate
BLER	Block Error Rate
BP	Belief Propagation
BPSK	Binary Phase Shift Keying
BSR	Buffer Status Report
CDMA	Code Division Multiple Access
CM	Cubic Metric
C-RNTI	Cell Radio Network Temporary Identity
CSI	Channel State Information
CS-MPA	Compressive Sensing MPA
DFT	Discrete Fourier Transform
EC-GSM-IoT	Extended Coverage GSM Internet of Things
eDRX	Extended Discontinuous Reception
eGPRS	Enhanced GPRS
eNB	Evolved Node B
eMBB	Enhanced Mobile BroadBand
eMTC	Enhanced Machine Type Communication
EPA	Expectation Propagation Algorithm
ESE	Elementary Signal Estimator
FDMA	Frequency Division Multiple Access
FN	Function Node
IDMA	Interleave Division Multiple Access
IMT-2020	International Mobile Telecommunications-2020
IoT	Internet of Things
ISM	Industrial, Scientific and Medical
ITU	International Telecommunication Union
GA	Grant Acquisition
GPRS	General Packet Radio Service

GSM	Global System for Mobile Communications
KKT	Karush–Kuhn–Tucker
LDPC	Low Density Parity Check
LDS-OFDM	Low Density Spreading Orthogonal Frequency Division Multiplexing
LDS-CDMA	Low Density Spreading Code Division Multiple Access
LLR	Log Likelihood Ratio
LPWAN	Low Power Wide Area Network
LTE	Long Term Evolution
LTE-A	LTE Advanced
LTE-Cat 0	LTE Category 0
LTE-Cat M1	LTE Category M1
LTE-Cat NB1	LTE Category NarrowBand
LTE-M	LTE Machine Type Communication
MAB	Multi Armed Bandit
MAC	Medium Access Control
MAP	Maximum A Posteriori
MC-CDMA	Multi-Carrier Code Division Multiple Access
MCL	Maximum Coupling Loss
MIMO	Multiple Input Multiple Output
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
mMTC	Massive Machine Type Communications
MPA	Message Passing Algorithm
MPDCCH	Machine Type Communication Physical Downlink Control Channel
MTC	Machine Type Communication
MU-MIMO	Multi-User Multiple Input Multiple Output
MUSA	Multi User Shared Access
MUST	Multi User Superposition Transmission
NB-IoT	NarrowBand Internet of Things
NOMA	Non Orthogonal Multiple Access
NR	New Radio
OFDMA	Orthogonal Frequency Division Multiple Access
OFDM	Orthogonal Frequency Division Multiplexing
OMA	Orthogonal Multiple Access
OSIC	Ordered SIC
PAPR	Peak to Average Power Ratio
PDCCH	Physical Downlink Control Channel
PDMA	Pattern Division Multiple Access
PD-NOMA	Power Domain NOMA
PDSCH	Physical Downlink Shared Channel
PER	Packet Error Rate
PN	Pseudo Noise
PRACH	Physical Random Access Channel
PRB	Physical Resource Block
PSM	Power Saving Mode
PUCCH	Physical Uplink Control Channel

PUSCH	Physical Uplink Shared Channel
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RA	Random Access
RACH	Random Access Channel
RAR	Random Access Response
RA-RNTI	Random Access Radio Network Temporary Identity
RRC	Radio Resource Control
SEP	Symbol Error Probability
SIC	Successive Interference Cancellation
SC-FDMA	Single-Carrier FDMA
SCMA	Sparse Code Multiple Access
SINR	Signal to Interference plus Noise Ratio
SMARTER	New Services and Markets Technology Enablers
SNR	Signal to Noise Ratio
TDMA	Time Division Multiple Access
THS	Thompson Sampling
TRxP	Transmission Reception Point
UCB	Upper Confidence Bound
UE	User Equipment
UNB	Ultra NarrowBand
URLLC	Ultra Reliable and Low Latency Communications
VN	Variable Node
WLAN	Wireless Local Area Network



---

# Mathematical Notations

$x$	Scalar
$\mathbf{x}$	Vector
$\mathbf{X}$	Matrix
$\mathbf{X}^{-1}$	Inverse of matrix $\mathbf{X}$
$\mathbf{X}^T$	Transpose of matrix $\mathbf{X}$
$\mathbf{X}^H$	Conjugate transpose of matrix $\mathbf{X}$
$\ \mathbf{x}\ $	L-2 norm of vector $\mathbf{x}$
$\mathbb{E}\{x\}$	Expectation of random variable $x$
$\mathbb{C}$	Complex field
$\mathbb{R}$	Real field
$\mathcal{P}$	A set
$ \mathcal{P} $	Cardinality of the set $\mathcal{P}$
$\odot$	Hadamard product
$\setminus$	Set difference
$\mathcal{O}(\cdot)$	The order of computation
$e^x$	Exponential function of $x$
$\log(x)$	Natural logarithm
$\mathcal{CN}(\mu, \sigma^2)$	Complex normal distribution with mean $\mu$ and variance $\sigma^2$
$\prod$	Product of sequence
$\sum$	Summation of sequence
$\operatorname{argmax}$	Argument of the maximum
$\operatorname{diag}(\mathbf{x})$	Diagonal matrix with diagonal entries $\mathbf{x}$
$p(x)$	Probability of $x$
$Q(x)$	The error function $Q$ of $x$



# Contents

## Acknowledgements

Résumé i

List of Acronyms xv

Mathematical Notations xix

Introduction 5

**1 Context on 5G and machine-type communications 11**

1.1 The 5G standard . . . . . 11

1.1.1 Why do we need 5G? . . . . . 11

1.1.2 5G services and their requirements . . . . . 12

1.1.3 5G standardization timeline . . . . . 17

1.1.4 Conclusion . . . . . 18

1.2 Existing MTC technologies . . . . . 19

1.2.1 Non-3GPP technologies . . . . . 19

1.2.2 3GPP technologies . . . . . 21

1.3 Conclusion . . . . . 25

1.3.1 Limitations of existing MTC technologies . . . . . 25

1.3.2 mMTC enablers . . . . . 26

**2 Non orthogonal multiple access schemes 29**

2.1 Introduction . . . . . 29

2.2 Details of NOMA schemes . . . . . 32

2.2.1 Power domain NOMA . . . . . 32

2.2.2 Low density spreading CDMA . . . . . 36

2.2.3 Low density spreading OFDM . . . . . 37

2.2.4 Sparse code multiple access . . . . . 39

2.2.5 Multi user shared access . . . . . 43

2.2.6 Pattern division multiple access . . . . . 48

2.2.7 Interleave division multiple access . . . . . 50

2.3 Qualitative comparison . . . . . 51

2.3.1	Receivers complexity . . . . .	52
2.3.2	Design complexity . . . . .	53
2.3.3	Overload capacity . . . . .	54
2.3.4	Adapted scenarios . . . . .	54
2.3.5	Conclusion . . . . .	55
2.4	Grant free access . . . . .	55
2.4.1	Random access . . . . .	55
2.4.2	Limitation of random access process . . . . .	59
2.4.3	Grant free access evolution . . . . .	60
2.5	NOMA standardization . . . . .	62
2.6	Conclusion . . . . .	64
<b>3</b>	<b>NOMA schemes evaluation</b>	<b>67</b>
3.1	Introduction . . . . .	67
3.2	Quantitative comparison . . . . .	68
3.3	Performance study of SCMA receivers . . . . .	73
3.3.1	MAX-Log-MPA algorithm . . . . .	74
3.3.2	Log-MPA algorithm . . . . .	75
3.3.3	Algorithmic complexity . . . . .	77
3.3.4	Numerical results and analysis . . . . .	79
3.4	Conclusion . . . . .	82
<b>4</b>	<b>Power allocation for BER minimization in an uplink MUSA scenario</b>	<b>83</b>
4.1	Introduction . . . . .	83
4.2	State of the art on resource allocation . . . . .	83
4.3	Average BER minimization . . . . .	85
4.3.1	Problem formulation . . . . .	85
4.3.2	Problem resolution . . . . .	87
4.4	Numerical results and analysis . . . . .	91
4.4.1	Perfect AWGN channel . . . . .	91
4.4.2	AWGN channel with different path losses . . . . .	93
4.5	Conclusion . . . . .	95
<b>5</b>	<b>Autonomous power decision for grant free access</b>	<b>97</b>
5.1	Introduction . . . . .	97
5.2	BEP Analysis . . . . .	100
5.2.1	Perfect SIC without Error Propagation . . . . .	100
5.2.2	Imperfect SIC with Error Propagation . . . . .	101
5.3	Autonomous Power Decision . . . . .	103
5.4	Power Allocation with Multi-Armed Bandits . . . . .	105
5.4.1	Upper confidence bound . . . . .	106
5.4.2	$\epsilon$ -Greedy . . . . .	107
5.4.3	Thompson Sampling Algorithm . . . . .	108
5.5	Complexity and Overhead Analysis . . . . .	109
5.6	Numerical Results and Analysis . . . . .	110
5.7	Conclusion . . . . .	113

## CONTENTS

---

<b>Conclusion and future perspectives</b>	<b>115</b>
<b>Bibliography</b>	<b>119</b>



# List of Figures

1	BER moyen pour un canal AWGN, $J = 6$ utilisateurs et $K = 4$ sous-porteuses. . . . .	vi
2	Comparaison du rapport moyen pour les canaux AWGN et de Rayleigh. . . . .	vii
3	Comparaison des performances pour les différents cas et scénarios considérés pour un canal AWGN, $J = 6$ et $K = 4$ . . . . .	ix
4	Comparaison du taux de transmissions réussies pour tous les algorithmes dans les scénarios 1 et 2. . . . .	xi
5	The main 5G use cases [32] . . . . .	6
1.1	5G deployment modes [42]. . . . .	12
1.2	Overview of 5G requirements [43] . . . . .	12
1.3	5G standardization timeline . . . . .	18
1.4	The new NR-Light interface in Release 17 [49] . . . . .	19
1.5	The operation mode of NB-IoT technology [15]. . . . .	23
2.1	Illustration of OMA and NOMA schemes [61] . . . . .	30
2.2	Illustration of NOMA power domain for $J=2$ [69]. . . . .	34
2.3	Uplink LDS-OFDM system [89] . . . . .	38
2.4	MPA factor graph for $J = 6$ and $K = 4$ . . . . .	41
2.5	MUSA BLER for different overload factors for $K = 4$ [110] . . . . .	45
2.6	MUSA scheme transmitter for $J = 6$ and $K = 4$ . . . . .	46
2.7	Iterative process of MMSE-SIC receiver. . . . .	47
2.8	MPA factor graph for PDMA scheme with $J = 6$ and $K = 4$ . . . . .	50
2.9	IDMA transmitter for $J$ users. . . . .	51
2.10	RACH process for LTE/LTE-A networks. . . . .	57
2.11	Preamble collision resolution . . . . .	65
2.12	Collision in step 3 of the random access process. . . . .	65
2.13	The sources of delay in LTE system [21] . . . . .	66
2.14	2-step RACH. . . . .	66
3.1	Users transmission scenario for $J = 6$ users. . . . .	68
3.2	Performance comparison of MUSA, SCMA, PDMA and OFDMA for 100% of resource occupation. . . . .	70

3.3	Performance comparison of MUSA, SCMA and PDMA for 150% of resource overload. . . . .	71
3.4	Performance comparison of MUSA, SCMA and PDMA for 150% of resource overload in scenario 1 and scenario 2. . . . .	71
3.5	Performance comparison of MUSA, SCMA, PD-NOMA and PDMA for 200% of resource overload in scenario 2. . . . .	72
3.6	A comparison of the decoding complexity. . . . .	78
3.7	Performance comparison for different number of iterations, AWGN channel, $J = 6$ and $K = 4$ . . . . .	79
3.8	MPA, Log-MPA and MAX-Log-MPA performance comparison for AWGN channel, $J = 6$ , $K = 4$ and $T = 5$ . . . . .	80
3.9	MPA, Log-MPA and MAX-Log-MPA performance comparison for Rayleigh fading channel, $J = 6$ , $K = 4$ and $T = 5$ . . . . .	80
3.10	A comparison of the average ratio $\frac{C(x)}{\max(a_1, \dots, a_n)}$ for AWGN and Rayleigh fading channels. . . . .	81
3.11	A comparison of the average distance $ a_j - a_i $ for AWGN channel. . . . .	81
4.1	BER w.r.t. $P_{SIC}$ for $SNR_1 = 20\text{dB}$ , $SNR_2 = 23\text{ dB}$ and AWGN channel. . . . .	92
4.2	Performance comparison for the different cases and scenarios for AWGN channel, $J = 6$ and $K = 4$ . . . . .	92
4.3	BER w.r.t. $P_{SIC}$ for $SNR_1 = 20\text{dB}$ , $SNR_2 = 25\text{ dB}$ and AWGN channel with different path loss. . . . .	93
4.4	Performance comparison for the different cases and scenarios for AWGN channel with different path losses, $J = 6$ and $K = 4$ . . . . .	94
5.1	Performance comparison of the simulated BER and the analytical BEP for an AWGN channel with different users' path losses and equal transmission powers. . . . .	103
5.2	Average transmission rate of $\epsilon$ -decreasing greedy for different $L$ values after $T = 1000$ iterations in Scenario 1. . . . .	111
5.3	Successful transmission rate comparison for all algorithms in Scenario 1. . . . .	112
5.4	Successful transmission rate comparison for all algorithms in Scenario 2. . . . .	112
5.5	Convergence comparison of all algorithms in Scenarios 1 and 2. . . . .	113
6.1	The main contributions . . . . .	115

# List of Tables

1	Les cas et les scénarios étudiés . . . . .	viii
1.1	Average spectral efficiency for different test environments [6]. . . . .	13
1.2	5-th percentile user spectral efficiency for different test environments [6]. . . . .	13
1.3	Differences between LTE and NR [45] . . . . .	14
1.4	Performance requirements for low latency and high reliability scenarios [47]. . . . .	15
2.1	Main characteristics of different NOMA schemes . . . . .	52
2.2	The applicable receivers for the studied NOMA schemes . . . . .	52
2.3	Decoding complexity of typical NOMA receivers. . . . .	53
2.4	Performances comparison of NOMA schemes . . . . .	54
2.5	Advantages, drawbacks and adapted scenarios of NOMA schemes . . . . .	56
3.1	Simulation settings. . . . .	69
3.2	Computational complexity of the studied algorithms. . . . .	77
5.1	Quantitative comparison of the signaling overhead and the complexity at user equipment in each iteration for all algorithms. . . . .	109
5.2	Simulation settings . . . . .	110



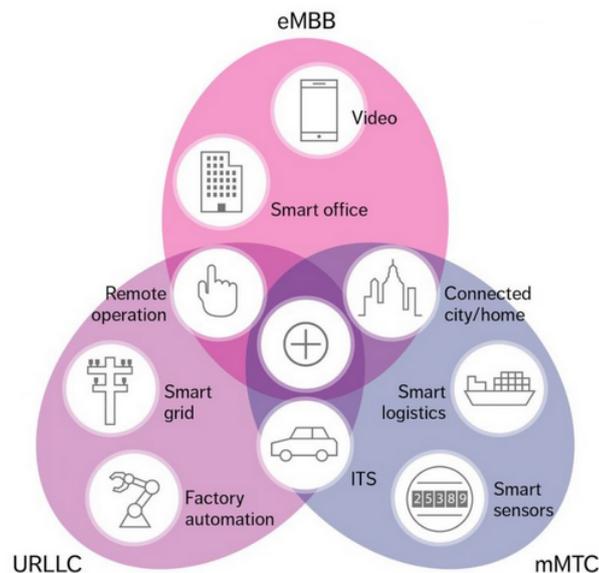
# Introduction

## Context and motivations

The capacity of the future radio access networks are expected to skyrocket within the next few years with billions of connected devices, i.e., 100 billions connected devices by 2030, due to the appearance of variety of new applications with different qualities of service (QoS) [1]. These services are classified by the international telecommunications union (ITU) and the third generation partnership project (3GPP) into three main broad capabilities with different stringent requirements namely, enhanced mobile broadband (**eMBB**), ultra reliable and low latency communications (**URLLC**) and massive machine type communications (**mMTC**) [3], [6], as illustrated in Fig. 5. The identified **5G** scenarios aim to satisfy, respectively, the bandwidth, the latency and the connectivity issues unsolved by previous technologies. eMBB is the first revealed service to the 5G public. It is considered as an evolution of 4G broadband applications while ensuring a better user experience [2]. The main challenges of eMBB consist of providing an increased spectral efficiency with an important data rate, on the order of Gigabits per second, and a moderate reliability, i.e., about  $10^{-3}$  of packet error rate (PER) [3]. The eMBB requirements will be crucial to support certain services such as the enhanced multimedia, the public transportation with a high mobility issue and the large scale events where a high data rate is needed to serve densely populated areas.

However, the URLLC [5] use case is drawn to particularly deal with the critical applications such as remote surgery, autonomous cars, smart grids, etc. These services require especially high reliable transmissions with  $10^{-5}$  of PER and low latency communications, i.e., 1 ms in uplink transmission [3], [33]. It is expected that the advent of this use case will be revolutionary for the upcoming technologies exceptionally for the vital sectors.

Finally, the mMTC use case is considered as one of the key feature to build the future smart cities while connecting everything and anything [7]. The mMTC scenario is also known as massive IoT since it is designed to mainly deal with a huge number of connected devices, i.e., about one million connected devices per  $\text{km}^2$  [10], [33]. In addition, the mMTC devices are characterized by their intermittent transmissions of short packets, i.e., on the order of few bytes. However, they are restricted by low system complexity and low energy consumption in order to offer a long battery life, i.e., on the order of ten years [8]. In contrast to the URLLC



**Fig. 5.** The main 5G use cases [32]

and eMBB capabilities, the mMTC use case is latency-tolerant and it is designed to particularly support low cost devices with low data rate transmissions. Over the last years, many efforts have been made to enable the MTC aspect. Therefore, many low power wide area (LPWA) cellular solutions have been highlighted by the 3GPP standards in Releases 13 and 14 namely, enhanced machine type communications (eMTC), narrow-band internet of things (NB-IoT) and extended coverage GSM for the IoT (EC-GSM-IoT) [9], [34], [35]. These solutions are derived from the existing networks of LTE and GSM. They guarantee a robust QoS, transmission reliability, relatively low latency and long coverage range. Despite their efforts to increase the connectivity capacity and reduce the power consumption, these techniques, in their current versions [19], may not be really adaptable to the strict requirements of the massive MTC scenario and should be further improved in the future 3GPP releases [9].

In addition, regarding the huge number of devices to be connected, the conventional orthogonal multiple access (**OMA**) schemes such as time/frequency division multiple access (TDMA/FDMA) may not be suitable to meet the connectivity issue. Indeed, each resource is attributed to a single user, which determines the maximum number of users that can be served. Therefore, these schemes are restricted by the limited number of the orthogonal resources and the fine synchronization tuning to guarantee interference-free communications, which induces high energy consumption and thereby short battery life. In this context, the non orthogonal multiple access (**NOMA**) [8] schemes have been highlighted by the scientific community as a prominent solution to enable the mMTC challenges. The main idea is to allow multiple users to share simultaneously and non-orthogonally the same resources, leading to network overload. Therefore, NOMA schemes show great potential to overcome the limitations of the OMA schemes regarding the mMTC requirements. Since the introduction of the IoT paradigm, several NOMA schemes have emerged. Most of them can be classified into two main categories with different multiplexing techniques : i)

NOMA power domain [36] which applies a non-uniform power distribution between users according to their SNRs; ii) NOMA code domain techniques namely, sparse code multiple access (SCMA) [23], multi-user shared access (MUSA) [24], pattern division multiple access (PDMA) [37], etc, where users signals are distinguished using different codes. Each scheme has its own principle for codes design and their attribution among users. However, users overlapping over the same radio resource may induce a high interference level at the receiver. This issue may be handled using robust multi user detection techniques, but at the cost of significant decoding complexity.

Although NOMA schemes can deal with the massive connectivity issue by increasing the resources overload, the way each user should access to the channel remains a challenging problem. In the existing technologies, for each transmission attempt, each user establishes a connection with the base station through a contention-based random access (RA) process. Then, it goes through the resource allocation procedures for data transmission. This operation may be very expensive in terms of signaling overhead, inducing high energy consumption, significant delay and often a network bottleneck. Furthermore, for the mMTC scenarios with small packets, the amount of signaling overhead may be higher than the transmitted data. Therefore, excessive resources will be exploited for the transmission of few bytes, which yields to resource wasting. In this context, NOMA schemes with a **grant free access** option [21] have gained a lot of interest and have been underlined as a promising solution to reduce the signaling overhead and the energy consumption. The grant free access strategy allows each user to transmit its data as per its need without waiting for resource reservation from the base station. Two main options have been proposed under the umbrella of grant free access while ensuring or not the uplink synchronization. For each transmission attempt, each device can apply contention based protocols, i.e., ALOHA [38] or slotted-ALOHA [39], to access to the radio resources [21], [40].

In this thesis, we mainly focus on the uplink mMTC scenario as the traffic model of mMTC applications is mainly in the uplink direction (tracking, metering, etc.) and because it is more complex than the downlink. Indeed, for a downlink system, all transmissions are managed by the base station with sufficient information on the active users and their propagation environments. More precisely, the base station is able to adjust the resource allocation and manage the inter-user interference. However, for uplink communications, the system performance is restricted by the capacity of power-limited user equipment (UE) without any coordination between users, which makes the implementation of NOMA schemes and grant free access option more challenging.

The standardization activities on the 5G fundamentals (Release 16) were completed in June 2020. Despite the work done on NOMA schemes for the 5G networks, it was decided not to include NOMA in 5G since the benefit over OMA scheme was limited and given the additional complexity, in particular at the UE side. Indeed, these schemes induce a high inter-user interference requiring robust multi-user detection techniques with a significant complexity. In addition, they need more processing power and some hardware updates compared to the conventional OMA schemes. NOMA schemes are considered quite complex regarding the low performance en-

hancement compared to the orthogonal schemes. As a consequence, the 5G services are deployed with OFDMA technique, as for 4G networks. However, it should be noted that the 3GPP roadmap only considers the eMBB and URLLC services. The standardization of the mMTC service has been postponed as it was stated that this service would be addressed by evolving eMTC and NB-IoT. As a consequence, NOMA techniques are not totally abandoned. They could be considered in future networks for mMTC scenarios if an acceptable trade-off could be established between the achieved performance, the system complexity and the massive number of supported users.

## Objectives and contributions

Motivated by the importance of the mMTC use cases as well as the great potential of NOMA schemes to overcome the restrictions of the conventional OMA schemes regarding the mMTC challenges, this thesis provides a qualitative and a quantitative analysis of NOMA schemes in the context of an uplink mMTC scenario. The idea is to address the challenges of NOMA in terms of implementation, complexity and performance by proposing a novel enhancement in order to meet the stringent constraints of mMTC devices. In addition, we seek to tackle the channel access issue, i.e., grant-based or grant free access. Our objective is to guarantee a trade-off between the desired system QoS and the satisfaction of the mMTC requirements in terms of number of connected devices, energy consumption and system complexity. Our main contributions to address the mMTC challenges by deploying NOMA schemes are summarized as follows :

- ▶ A performance assessment of the most promising NOMA schemes in the context of mMTC namely, PD-NOMA, SCMA and MUSA is provided in order to select the appropriate schemes meeting the defined requirements.
- ▶ In order to address the system complexity problem, the performance of SCMA scheme, with different multi-user detection techniques, i.e., message passing algorithm (MPA), Log-MPA and MAX-Log-MPA, is deeply investigated while studying the exchanged extrinsic information behavior.
- ▶ A novel centralized power allocation algorithm for the bit error rate (BER) minimization in an uplink MUSA scheme is proposed with a simplified expression of the bit error probability (BEP) and based on special constraints on received powers.
- ▶ A closed-form approximation of the BEP for MUSA scheme with the successive interference cancellation (SIC) technique is introduced while taking into consideration the inter-user interference and the error propagation phenomenon.
- ▶ A novel algorithm for an autonomous power decision is proposed in the context of a grant free access with minimum signaling overhead. This algorithm exploits the proposed BEP expression to determine the appropriate power for each user.

- The efficiency of different multi armed bandit (MAB) algorithms is investigated for the autonomous power decision problem with a minimum signaling overhead. Then, their performances are compared with that of the proposed algorithm.

## Dissertation outline and organization

This manuscript includes 5 chapters. The theoretical background and the literature review are introduced in chapter 1. It highlights the evolution steps towards the introduction of 5G use cases and beyond. It presents as well the main limitations of the existing technologies regarding the stringent mMTC challenges.

Chapter 2 introduces the main principles of NOMA schemes which are highlighted as promising for the future radio access network. Moreover, it provides a qualitative comparison of these schemes w.r.t the mMTC requirements. Furthermore, an overview of the existing resource access strategies as well as the step-wise evolution towards grant free access options are represented.

Chapter 3 represents a performance assessment of the most promising NOMA schemes to meet the mMTC challenges, e.g., PD-NOMA, SCMA, MUSA and PDMA. Then, in order to address the decoding complexity problem of SCMA schemes, an exhaustive performance comparison is provided between MPA and its two log-domain simplifications known as Log-MPA and MAX-Log-MPA for Gaussian and fading channels. The algorithm performances are deeply evaluated through a study of the extrinsic information behavior.

Chapter 4 investigates the problem of the centralized power allocation for MUSA schemes in order to alleviate the error propagation problem of SIC receivers. Therefore, an optimization problem is formulated in order to minimize the average system BER while guaranteeing a sufficient power difference between the concurrent received signals. In order to ensure a global optimal solution and simplify the problem resolution, a simplified expression of user BER is used. Then, the optimization problem is investigated in two different scenarios with different constraints combinations. The optimization problem behaviour is analyzed theoretically. Afterwards, a novel iterative algorithm is proposed for its resolution. At the end of this chapter, a performance evaluation is shown in order to reveal the impact of the power allocation on the system BER as well as the efficiency of each power constraint.

Chapter 5 is dedicated to the autonomous power decision issue in the context of a grant free access strategy with minimum signaling overhead. In the first section a closed-form approximation of the BEP for SIC receivers is proposed while involving the inter-user interference and the effect of the error propagation. This expression is optimized using an advanced algorithm to obtain the optimal centralized power allocation. Then, in the second section, the BEP expression is exploited to propose a novel algorithm for an autonomous power decision for MUSA schemes with reduced signaling overhead. In the third section, we stress the efficiency of one of the reinforcement learning techniques known as MAB algorithms for an autonomous power selection. Several MAB algorithms namely, Thompson sampling, UCB1 and  $\epsilon$ -decreasing greedy are investigated in different implementation scenarios while adapting users rewards to meet the reduced signaling overhead. At the last section,

the obtained performance using the MAB algorithms is compared with that of the proposed algorithm.

Finally, at the last part of this thesis, conclusions are drawn while underlining the ongoing research and the future perspectives.

# Context on 5G and machine-type communications

## 1.1 The 5G standard

### 1.1.1 Why do we need 5G ?

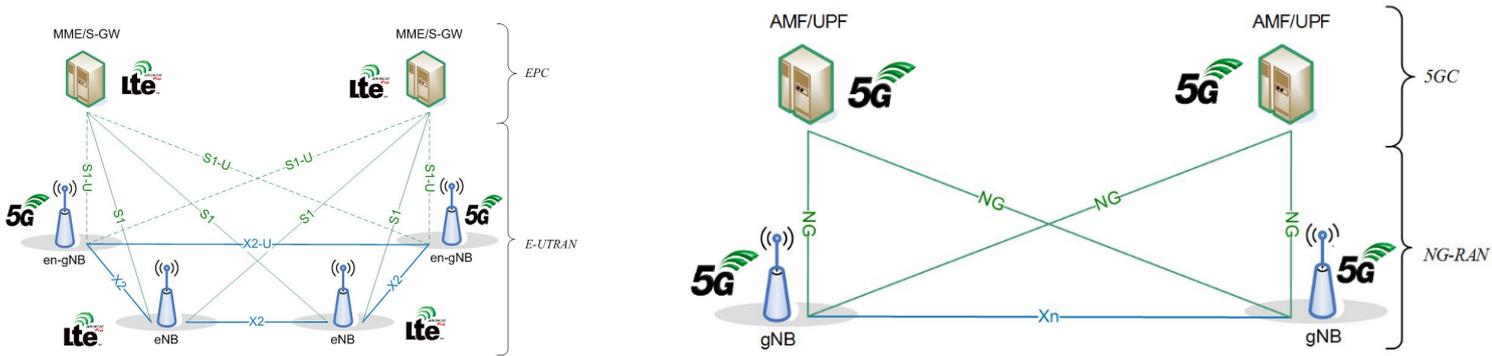
The radio access network of the fifth generation is experiencing a revolutionary upgrade in order to handle the new revealed services requiring different QoS. As a consequence, the managed data traffic is rising significantly with billions of connected devices. The 3GPP SMARTER project has identified about 70 new use cases of 5G services such as smart cities, ultra-reliable communications, connected vehicles, expanded industrial IoT, improved healthcare network, etc [41].

The main challenges of 5G networks are to meet the stringent requirements of the heterogeneous services in terms of connection density, data rate, transmissions latency and reliability while ensuring a trade-off with energy consumption, system complexity and network deployment cost. Therefore, according to the ITU specifications and the 3GPP standards, known as IMT-2020 standard, the identified uses cases are classified under the umbrella of three main scenarios with different exigencies : i) eMBB for a better user experience with a peak of data rate exceeding 10 Gbps ; ii) URLLC for high reliable transmissions with a minimum latency of 1 ms ; iii) mMTC for an extreme connection density with 1 million connected devices per km<sup>2</sup>.

These scenarios are addressed in two 3GPP Releases : i) Release 15 studied the first deployment phase of 5G networks and mainly focuses on the eMBB scenario standardization ; ii) Release 16 is interested on the second phase of 5G including URLLC and mMTC scenarios. Regarding the 5G networks infrastructure, this later can be rolled out with two deployment modes :

- ▶ Non-Stand-Alone mode [42] : is the first commercialized version of 5G new radio (NR) interface. It exploits the existing 4G core network with 5G base stations, i.e., known as en-gNB. The architecture of the Non-Stand-Alone mode is illustrated in Figure 1.1a. It permits to reduce the 5G deployment cost and accelerate its commercialization.

# CHAPITRE 1. CONTEXT ON 5G AND MACHINE-TYPE COMMUNICATIONS



(a) The Non-Stand-Alone architecture.

(b) The Stand-Alone architecture.

Fig. 1.1. 5G deployment modes [42].

- Stand-Alone mode [42] : is based on a full deployment of 5G networks independently from 4G infrastructure, as seen in Figure 1.1b. This mode ensures the support of all 5G features with a full benefit from its new functionalities.

## 1.1.2 5G services and their requirements

The design goals of eMBB, URLLC and mMTC are illustrated in Figure 1.2. They are reviewed in the following subsections.

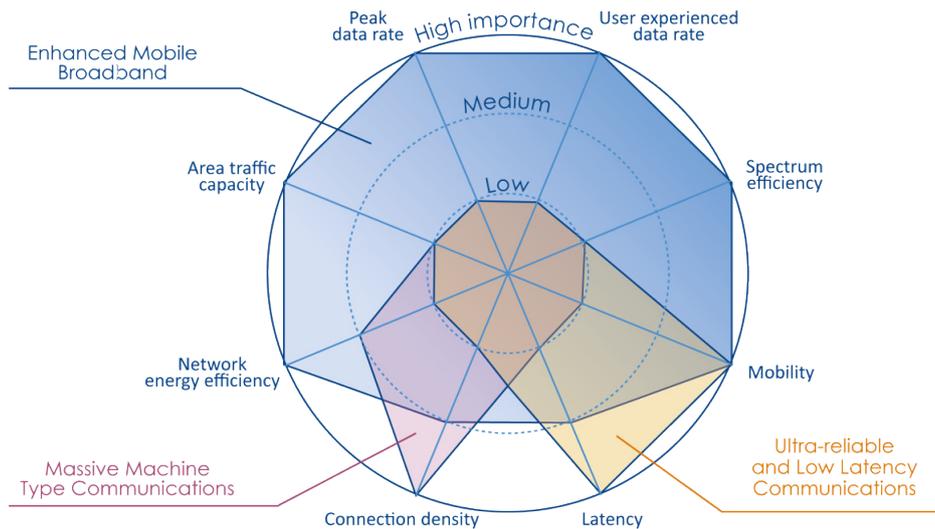


Fig. 1.2. Overview of 5G requirements [43]

### 1.1.2.1 Enhanced mobile broadband

The eMBB capability [2] can be seen as a radical overhaul of 4G services with an improved data rate and an enhanced user experience compared to the existing mobile broadband services. The eMBB capability is considered as the first phase of 5G networks. It is among the early delivered services in 2020. According to Ericsson, about 1.9 billion 5G subscribers are expected for eMBB services by 2024 [4]. The ITU report and the 3GPP specification activities [6], [33] defined the main eMBB requirements as follows :

- **High data rate** : eMBB is an evolution of the existing LTE networks in terms of the achieved data rate which permits to handle the congestion problem and enable new applications, such as 4K-8K video streaming, augmented reality and virtual reality. For the ideal system conditions with no transmission errors, the peak data rate would jump significantly from 1 Gbps in LTE to 20 Gbps for downlink transmission and 10 Gbps for uplink in 5G networks. As a consequence, the user experienced data rate, which is calculated as the 5-th percentile of the cumulative distribution function of the user rate, should attain the values of 100 Mbps on the downlink and 50 Mbps on the uplink.
- **Peak and average spectral efficiency** : The peak spectral efficiency is calculated by dividing the peak data rate by the channel bandwidth. The eMBB capability aims to ensure an enhanced spectral efficiency with a peak value of 30 bit/s/Hz in downlink transmission and 15 bit/s/Hz in uplink. However, the average spectral efficiency is calculated by dividing the average system spectral efficiency by the number of transmission-reception points (TRxP), e.g., the number of available antenna by cell. The minimum required values are defined according to different test environments as summarized in Table. 1.1.

Test environment	Downlink (bit/s/Hz/(TRxP))	Uplink (bit/s/Hz/(TRxP))
Indoor Hotspot	9	6.75
Dense urban	7.8	5.4
Rural	3.3	1.6

TABLE 1.1: Average spectral efficiency for different test environments [6].

- **5-th percentile user spectral efficiency** : is calculated as the 5-th percentile of the cumulative distribution function of the spectral efficiency. The minimum required value depends on the propagation environment where three main scenarios are identified as summarized in Table. 1.2.

Test environment	Downlink (bit/s/Hz)	Uplink (bit/s/Hz)
Indoor hotspot	0.3	0.21
Dense urban	0.225	0.15
Rural	0.12	0.045

TABLE 1.2: 5-th percentile user spectral efficiency for different test environments [6].

- ▶ **Area traffic capacity** : One of the new advantages of the eMBB capability is the increase of the amount of supported data traffic per  $m^2$ , which would improve the user experience in crowded areas such as sport events, shopping malls or down street. Therefore, eMBB aims to provide an area traffic capacity on the order of  $10 \text{ Mbit/s/m}^2$  in the case of indoor hotspot.
- ▶ **User plane latency** : is the needed time between the packet transmission and its reception. The eMBB services require a moderate user plane latency of 4 ms in the ideal system conditions.
- ▶ **Control plane latency** : is the time spent by each device to transit from the idle to the active state. Its value for eMBB scenario is set to 20 ms. However a lower value of 10 ms is recommended.
- ▶ **Energy efficiency** : is calculated as the ratio between the spectral efficiency and the consumed power. The energy efficiency should be guaranteed by an optimal transmission power for high data traffic, and a minimum energy consumption when no data is available, which can be ensured by supporting high ratio and long duration of sleep cycles.
- ▶ **Mobility** : The eMBB capability seeks to support a fast devices mobility, i.e., up to 500 km/h while ensuring a spectral efficiency of 0.45 bit/s/Hz.
- ▶ **Mobility interruption time** : is the minimum time for handover establishment. In LTE networks, this duration is typically between 30 and 60 ms [44]. However, according to IMT-2020 standard, the mobility interruption time for the eMBB services should achieve the lowest technically possible value, i.e., as close as possible to zero.
- ▶ **Reliability** : The eMBB services can tolerate a PER value of  $10^{-3}$  which is lower than the required value in LTE networks of  $10^{-1}$  [3].

	LTE	NR
Frequency of operation	Up to 6 GHz	Up to 6 GHz, 28 GHz, 39 GHz, other mmWave bands (up to 52 GHz)
Carrier bandwidth	Max : 20 MHz	Max : 100MHz (at <6 GHz) Max : 1 GHz (at >6 GHz)
Carrier aggregation	Up to 32	Up to 16
Analog beamforming (dynamic)	Not supported	Supported
Digital beamforming	Up to 8 Layers	Up to 12 Layers
Channel coding	Data : Turbo Coding Control : Conventional Coding	Data : LDPC coding Control : Polar coding
Subcarrier spacing	15 KHz	15 KHz, 30 KHz, 60 KHz, 120 KHz, 240 KHz
Self-contained subframe	Not supported	Can be implemented
Spectrum occupancy	90 % of channel bandwidth	Up to 98 % of channel bandwidth

TABLE 1.3: Differences between LTE and NR [45]

Similarly to the LTE networks, the first phase of 5G, i.e., eMBB services, is standardized with the orthogonal frequency division multiplexing (OFDM) techniques. However, several differences in terms of frequency bands, carrier bandwidth, carrier aggregation, channel coding, etc, can be identified as listed in Table. 1.3.

### 1.1.2.2 Ultra reliable and low latency communications

The URLLC service was introduced in the initial phase of the 5G study in Release 15 [42]. Then, it was enhanced with new features in the second 5G phase in Release 16 [46]. The URLLC scenario is dedicated particularly for the emergent applications with critical requirements in terms of latency, safety and reliability such as autonomous vehicles, robotics, drones, eHealth applications, etc [5]. The existing technologies suffer from high latency values. For instance, the LTE-Advanced (LTE-A) networks offer a control plane latency on the order of  $\sim 50$  ms and a user plane latency of  $\sim 10$  ms. These values can not be suitable for URLLC applications that require fast interaction with the base station. Table. 1.4 provides an overview of certain application scenarios requiring low latency and high reliable transmissions.

Scenario	End-to-end latency	Availability	Reliability	User experienced data rate	Payload size	Traffic density	Connection density	service area dimension
Discrete automation-motion control	1 ms	99.9999%	99.9999%	1 Mbps up to 10 Mbps	Small	1 Tbps/km <sup>2</sup>	100 000/km <sup>2</sup>	100 × 100 × 30 m
Discrete automation	10 ms	99.99%	99.99%	10 Mbps	Small to big	1 Tbps/km <sup>2</sup>	100 000/km <sup>2</sup>	1000 × 1000 × 30 m
Process automation-remote control	50 ms	99.9999%	99.9999%	1 Mbps up to 100 Mbps	Small to big	100 Gbps/km <sup>2</sup>	1000/km <sup>2</sup>	300 × 300 × 50 m
Process automation-monitoring	50 ms	99.9%	99.9%	1 Mbps	Small	10 Gbps/km <sup>2</sup>	10 000/km <sup>2</sup>	300 × 300 × 50 m
Electricity distribution-medium voltage	25 ms	99.9%	99.9%	10 Mbps	Small to big	100 Gbps/km <sup>2</sup>	1000/km <sup>2</sup>	100 km along power line
Electricity distribution-highvoltage	5 ms	99.9999%	99.9999%	10 Mbps	Small	100 Gbps/km <sup>2</sup>	1000/km <sup>2</sup>	200 km along power line
Intelligent transport systems-infrastructure backhaul	10 ms	99.9999%	99.9999%	10 Mbps	Small to big	10 Gbps/km <sup>2</sup>	1000/km <sup>2</sup>	2 km along a road
Tactile interaction	0.5 ms	99.999%	99.999%	Low	Small	Low	Low	TBC
Remote control	5 ms	99.999%	99.999%	From Low to 10 Mbps	Small to big	Low	Low	TBC

TABLE 1.4: Performance requirements for low latency and high reliability scenarios [47].

The IMT-2020 standard identified a set of special requirements for URLLC applications [6] :

- ▶ **User plane latency** : The URLLC services are generally critical and require a fast interaction between the UE and the base station, which is translated to low latency transmission. For instance, for tele-surgery services with vital risks, the patient life is at stake. Therefore, any delay can not be tolerated. The IMT-2020 standard precises a very low user plane latency on the order of 1 ms.
- ▶ **Control plane latency** : The URLLC capability is not more stringent than eMBB in terms of control plane latency which was already set at 20 ms with a recommendation for the value of 10 ms.
- ▶ **Reliability** : is one of the serious requirements of the critical services. In fact, URLLC use case aims to ensure high reliable transmissions with a minimum packet error rate of  $10^{-5}$  [3].

- ▶ **Mobility interruption time** : The URLLC capability is able to support fast handover between base stations. Similarly to the eMBB scenarios, the required mobility interruption time is set to be as close as possible to 0 ms.

### 1.1.2.3 Massive machine type communications

The future cellular network is expected to evolve considerably during the next decades due to the multiplicity of demanded services such as smart-grid, traffic and parking management, smart city, health monitoring, etc. More precisely, the number of connected devices will grow significantly up to 100 billions connected devices worldwide by 2030 [1]. In fact, the mMTC capability, known also as massive IoT, is drawn to particularly deal with a large number of low cost devices with stringent requirements in terms of packet size, energy consumption and transmission frequency [7]. The specific requirements to support the mMTC services are determined by the IMT-2020 standard and the 3GPP contributors as follows :

- ▶ **Massive connectivity** : The mMTC capability is mainly introduced to involve a variety of services and hence a massive number of connected devices. According to Statistica, around 80 billions devices will be connected to cellular networks worldwide by 2025 [48]. Therefore, this connectivity will be satisfied by overloading the network cells and serving around 1 million connected devices per km<sup>2</sup> [6], [8].
- ▶ **Low cost device** : Regarding the massive connectivity target and in order to reduce the deployment cost, the mMTC devices are designed with an ultra low-cost hardware compared to mobile broadband subscribers, on the order of few dollars [7], [9]. This criteria will enable the implementation of several mMTC applications such as smart water meters, tracking applications, connected home, smart watches, etc.
- ▶ **Long battery life** : The mMTC devices are mostly battery-powered. Therefore, the energy efficiency is one of the key feature that should be satisfied. In addition, the mMTC aspect is based on machine-to-machine communications with a minimum human intervention. For that reason, the connected devices should operate autonomously for a long time with a minimum battery life of 10 – 15 years [8] depending on the mMTC applications.
- ▶ **Low system complexity** : In order to respect the stringent requirement of energy efficiency and hardware design, and according to 3GPP specifications, the connected devices should experience a low system complexity which ensures a low energy consumption and hence a long battery life.
- ▶ **Low data rate** : The mMTC capability is designed to mainly deal with sporadic transmissions of small packets and low data rate, on the order of 10-20 bytes per second. Therefore, the mMTC terminals may not be active continuously which permits to reduces the consumed energy by going through sleep cycles.

- **Extended coverage :** The mMTC services should cover a large number of devices that may be scattered over the entire cell and experience very bad channel conditions. For instance, certain indoor applications, as smart metering use case, require a deeper coverage in order to guarantee a minimum QoS. Therefore, an extended network coverage for wide area is inevitable in order to support the expected massive-IoT services. The mMTC service should be able to support a maximum coupling loss (MCL), i.e., representing the maximum channel loss between the UE and the base station, of 164 dB [10].

During the 3GPP standardization activities, the first phase of 5G networks has given priority to eMBB services to accelerate its commercialization. The URLLC and mMTC scenarios were part of the second 5G phase, i.e. Release 16 [7].

According to 5G stakeholders, regarding the today's connection density, one can focus only on satisfying the MTC requirements namely, short packets, low cost devices, i.e., order of a few dollars, low power consumption, long battery life of more than 10 years and wide coverage area. In this context, the existing MTC solutions, namely LTE-M and NB-IoT, are able to meet some of these requirements. As a result, Release 16 focused on improving the existing solutions instead of defining a new mMTC service based on 5G NR. However, given the rapid growth in the number of connected devices in the coming years, some emergent issues must be addressed in the future mMTC service, such as the massive connection density and the minimum signaling overhead, requiring advanced techniques and innovative approaches.

### 1.1.3 5G standardization timeline

The fifth wireless network generation is designed to support heterogeneous services with different performance requirements that are defined by the international telecommunication union in 2015. The 3GPP standardization of 5G network is carried in two phases. First, the 3GPP Release 15, started in late 2017 and ended in June 2019, introduced the phase 1 of the 5G NR while focusing on the specifications of eMBB scenarios. More precisely, the stringent eMBB services with a direct impact on the hardware design namely, waveform, numerology, frame structure and channel coding were given a special attention. The Release 15 provided the basic axes for the future standardization activities. It defined the two 5G deployment modes, known as Non-Standalone and Standalone modes, with a full 5G infrastructure. This release included also some improvements of the highly reliable low latency communications, the machine type communications, Vehicle-to-everything communications, critical mission and some other features concerning WLAN, unlicensed spectrum and LTE networks. The commercialization of the initial 5G phase was launched in early 2020.

The second phase of the 5G NR with a full 5G system has been explored in Release 16, which ended in June 2020, as shown in Figure 1.3. The main features of this phase are to improve both the URLLC and enhanced MTC scenarios. Specifically, when it comes to connection density requirements in current systems, it is sufficient to build on some of the existing 3GPP solutions for MTC networks, which were already included in previous releases 13-15 such as eMTC and NB-IoT, while introducing further enhancements in terms of power consumption, battery life, and device cost. However, the huge connection density issue of future mMTC services is left for

# CHAPITRE 1. CONTEXT ON 5G AND MACHINE-TYPE COMMUNICATIONS

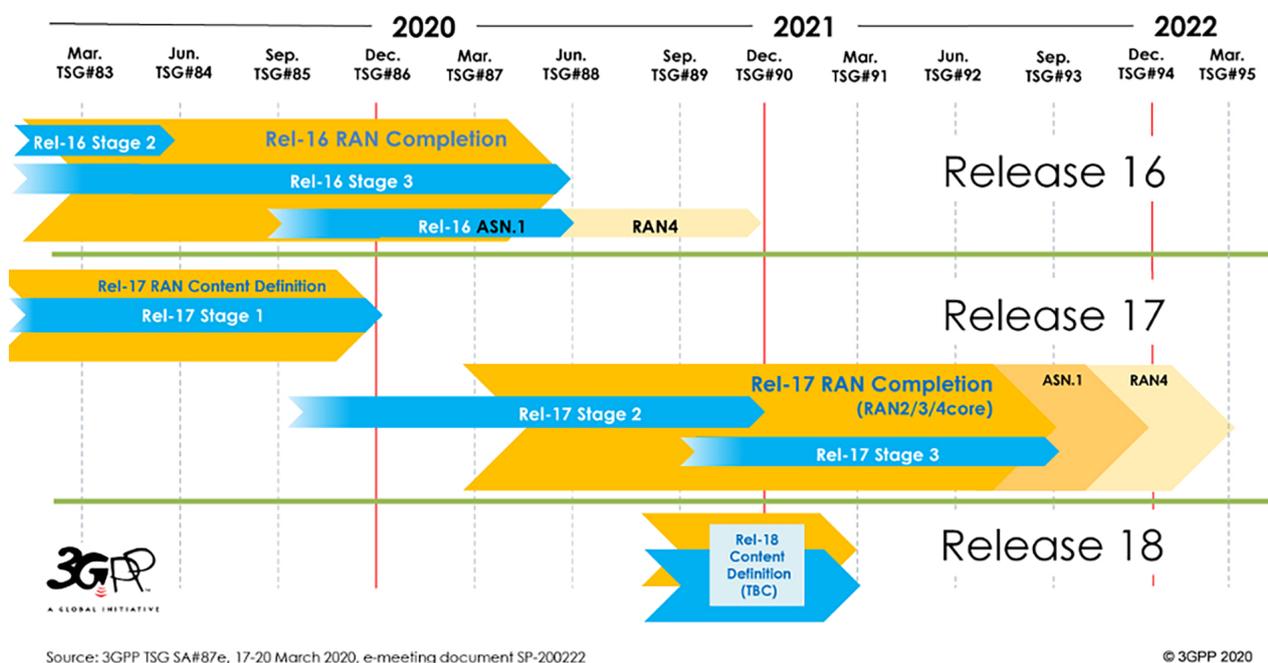


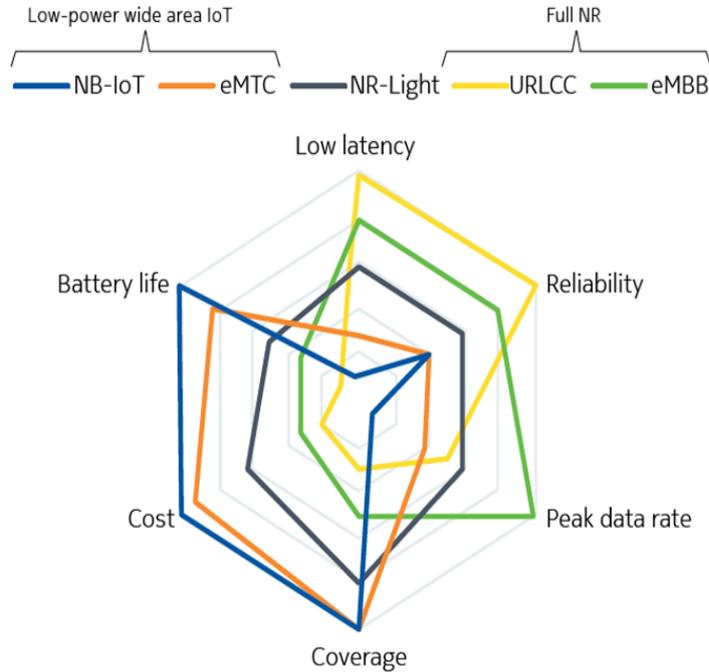
Fig. 1.3. 5G standardization timeline

future work. Release 16 also included improvements to the functionalities defined in the previous Release 15 such as localization, multicast, multi-antenna transmissions and power control. It involved new use cases for vehicle-to-everything communications. In addition, it provided improvements to industrial IoT and URLLC services as well as, new operations for 5G NR in unlicensed bands.

The 3GPP Release 17 is still ongoing. It is expected to be completed by mid-2022, as shown in Figure 1.3. It aims to make improvements to 5G systems. The main feature of this Release is the definition of the new 5G radio interface called NR-Light. The latter seeks to ensure a compromise between the different requirements of the three 5G scenarios, eMBB, URLLC and mMTC as illustrated in Figure 1.4. Moreover, it is also expected that Release 17 will include other features such as broadcast and multicast services, MIMO system enhancements, NR operations on high frequency bands and, NB-IoT and eMTC enhancements.

## 1.1.4 Conclusion

The 5G standardization in both Releases 15 and 16 was completed. Most of the previously defined performance requirements of the main 5G scenarios, eMBB, URLLC and mMTC, have been satisfied. However, as already said, the today's approved 5G features for the support of mMTC only consisted in improving the existing MTC solutions to get closer to the mMTC requirements. As a consequence, some of the



**Fig. 1.4.** The new NR-Light interface in Release 17 [49]

mMTC challenges remain not being supported, especially the massive connection density and the signaling overhead. This issue should be resolved in the short term in order to meet all connection requests. In this view, many 3GPP study items have proposed some promising solutions in the interest of addressing the density issue, but no final decision has been approved until now [50], [51]. Therefore, future 3GPP releases are expected to endorse final solutions in order to meet all mMTC requirements. The following sections provide an overview of the most promising MTC solutions, 3GPP and non-3GPP, in order to explain their main advantages and limitations regarding mMTC services. They introduce also the main candidate techniques to enable the mMTC service.

## 1.2 Existing MTC technologies

The buzzword IoT has motivated the telecommunications stakeholders from academia and industry to suggest their contributions. In fact, since the introduction of IoT paradigm, several technologies have been involved to satisfy the specific MTC requirements. They may be classified into two main categories namely, non-3GPP and 3GPP solutions [7], [52], according to their own principles.

### 1.2.1 Non-3GPP technologies

The non-3GPP technologies are carried out with proprietary standards that are proposed by non-governmental organizations [13]. Most of them are tagged as low power and wide area network (LPWAN) techniques that exploit free unlicensed spectrum which reduces the deployment cost. The LPWAN solutions have gained

a lot of interest from research communities thanks to their low data rate, low cost communications, i.e., on the order of few dollars for operating and radio chipset cost, and long range communications. In addition, one of the key features of these technologies is their power-efficient behavior which ensures long battery life, i.e., more than 10 years. Therefore, they may be suitable to serve several IoT applications. During last years, several LPWAN technologies have been developed with different system specifications namely, LoRa [15], SigFox [15], INGENU [9], etc. It should be noted that other low power techniques with short coverage range have been also proposed such as ZigBee, low power WiFi, Bluetooth, etc. However, due to their limited coverage, they may not be adaptable to all IoT applications. In this section, we only focus on LPWAN technologies, particularly LoRa and Sigfox, which seem more suitable for MTC concept.

#### 1.2.1.1 Sigfox

Sigfox solution was initially proposed in 2010 by the French LPWAN operator Sigfox [9], [15]. This technology is now commercialized in 45 countries worldwide [53]. It allows an efficient transmission of small packets based on low system bandwidth. In addition, Sigfox is realized with proprietary base stations based on cognitive software-defined radios which permit to attach them to the back-end servers through an IP network. Packets transmission is performed with binary phase shift keying (BPSK) modulation and ultra-narrow band (UNB) of 100 Hz within sub-GHz industrial, scientific and medical (ISM) band carriers. The exploited ISM bands are unlicensed spectrum, which reduces the communications cost and provides a long coverage range, i.e., up to 30-50 km in rural zones and 3-10 km in urban areas with a line of sight transmission. For instance, the exploited ISM band in North America is 915 MHz, 868 MHz in Europe and 433 MHz in Asia.

UNB criteria allow high spectral efficiency with low system noise, which ensures robust decoding performance while using a low cost antenna and consuming a small amount of power, but with a limited data rate of 100 bps [15]. The Sigfox solution was originally proposed for uplink transmission and then, it was enhanced to involve downlink communications. However, a downlink message can be transmitted only after an uplink transmission. The network capacity is set to a maximum of 140 messages of 12 bytes per day in uplink and 4 messages of 8 bytes in downlink. Therefore, it may not be possible to acquit all the uplink messages. For that reason, the message replication technique with time and frequency diversity is used to enhance the transmission reliability. Indeed, each user transmits multiple copies of its packet, i.e., generally three copies, using various frequency bands in order to ensure a proper decoding process. For the sake of simplicity, each user can easily select its random frequency band which decreases the system complexity, minimizes the cost of devices and optimizes the hardware implementation.

#### 1.2.1.2 LoRa

LoRa is considered as a physical layer solution which was innovated by the French startup Cycleo in Grenoble in 2009 [14], [15]. Then, it was bought to the American

company Semtech in 2012. Three years later, the LoRa alliance standardized this solution.

Now, LoRa technology is commercialized in more than 42 countries and aims to invade others. It provides low power consumption, low cost devices and wide coverage area. Users data are transmitted over unlicensed ISM frequency bands while applying the spread-spectrum technique [54] with six different proprietary spreading factors. In fact, the robust spread-spectrum technique [55] permits to propagate a generated signal in narrow frequency band over a large bandwidth for uplink and downlink communications. Therefore, the experienced system noise decreases significantly leading to a mitigated interference impact, and thus a better system performance. The base stations of LoRa networks can simultaneously receive all the transmitted signals that are modulated with different spreading factors.

The selected factor has a direct impact on the achieved data rate, i.e., between 300 bps and 50 kbps, and the reached coverage, up to 15 km. More precisely, a low spreading factor provides a high data rate, but at the cost of a short coverage range, and vice versa. For LoRa networks, the payload size is limited to 243 bytes, which seems suitable for most of IoT applications. LoRa technology uses a key communication protocol known as LoRaWAN which applies ALOHA access strategy [9]. All base stations located in the device coverage area are able to receive its message, which yields to multiple copies of the same message with the advantage of a better decoding performance. The redundant copies are treated by the back-end server which orients each message to its corresponding destination and sends back an acknowledgment to the sender. The fact that the user's message is received by multiple base stations offers a better device localization. In addition, it allows to support users mobility while eliminating the handover process. However, several neighboring base stations are needed, implying a high deployment cost.

### 1.2.2 3GPP technologies

The existing 3GPP cellular technologies namely 2G, 3G and LTE provide a lot of satisfactions such as long coverage range, robust performance with low system interference, high security level, unique allocated resources as well as optimized and stable network infrastructure. However, these solutions have been developed to mainly deal with mobile broadband scenarios requiring high data rate and high spectral efficiency, which induces high system complexity, high power consumption and thereby short battery life. As a consequence, they are not able to support the MTC applications with their particular requirements.

In this context, three promising LPWAN solutions for MTC scenarios have been proposed during the last decade namely, eMTC, EC-GSM-IoT and NB-IoT [12]. In order to take advantage from the existing network infrastructures and reduce the deployment cost, the proposed MTC technologies are considered as enhancements of previous cellular networks. For instance, eMTC and NB-IoT are derived from LTE network, whereas EC-GSM-IoT is derived from GSM network. These technologies provide wider coverage range, lower power consumption, lower system complexity and thus, longer battery life compared to LTE and GSM networks. The eMTC and NB-IoT enhancements and extensions have continued throughout the successive

3GPP releases, up to Release 17, in order to guarantee the basic requirements of mMTC applications.

### 1.2.2.1 Enhanced machine type communications (eMTC)

The first introduction of MTC feature with LTE network was in Release 10 [16]. Then, this aspect was expanded in Release 11 by proposing an extended access barring class in order to avoid overloading the access and the core networks. After that, Release 12 [16] involved the LTE-Cat 0 technology with the first low complexity devices and the power saving mode (PSM) technique. Release 13 had taken the major step toward satisfying the stringent requirements of IoT-like devices by introducing an enhanced LPWAN version of LTE networks, known as LTE-Cat M1, LTE-M or eMTC [9], [56], [57].

eMTC offers an extension of the coverage range, compared to LTE networks, that allows to support devices in challenging locations with MCL value up to 164 dB. The power efficiency of connected devices was also improved in Release 13 using the extended discontinuous reception (eDRX) and the PSM features. From 75% to 80% of the system complexity is optimized compared to the based LTE network. This can be explained by the reduction of system bandwidth from 20 MHz, in LTE system, to 1.4 MHz which is divided into eight physical resource blocks of 140 KHz. Each user can only exploit a maximum of six resources while the other two are used as guard bands in order to reduce the system interference. Reducing the channel bandwidth induces some modifications in LTE air interface, e.g., the substitution of the wide PDCCH band by a narrower MTC band, called MPDCCH, as well as the reduction of system throughput to 1 Mbps. The maximum uplink transmission power of eMTC devices is set at 20 dBm in order to reach a battery life of more than 10 years. As a consequence, eMTC solution can support low cost devices with low system complexity. eMTC can coexist with the current 3GPP technologies on the same network infrastructure which optimizes the deployment cost.

This technology has been improved in the subsequent 3GPP releases in order to further improve its performance in terms of MTC requirements. Release 14 introduced an enhancement of eMTC technology by increasing the data rate up to 4 Mbps in the downlink and 7 Mbps in the uplink. This may be beneficial for certain IoT applications requiring more data traffic. The improved eMTC version offers a more accurate device localization based on the time difference observed between the received reference signals from all base stations. In addition, it allows the message multicast feature which permits to optimize the resources exploitation and to reduce the power consumption leading to a longer battery lifetime.

Release 15 included the first data transmission feature that minimizes the transmission overhead and latency, leading to higher battery life. Added to this, it introduced the wake-up signal concept for the optimization of the activation and the power saving duration of connected device, which reduces the power consumption. The sub-PRB resource allocation was also one of the new features of Release 15 in the interest of enhancing the quality of uplink communications for users in bad locations.

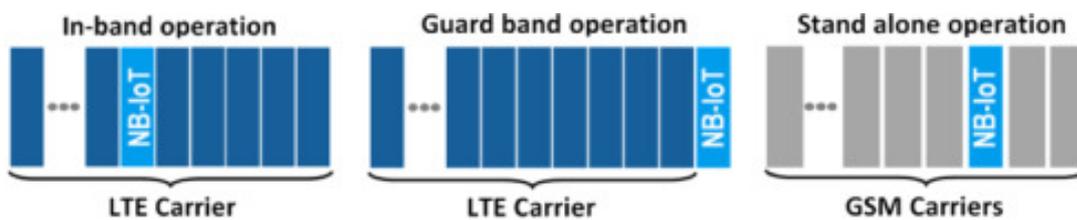
Release 16 concerned the evolution of LTE solutions towards 5G requirements, particularly mMTC scenario. It aimed to further increase the spectral efficiency and

reduce the power consumption using an enhanced early data transmission, wake-up signaling for a group of users, pre-configured resource allocation, scheduling of multiple transport blocks and mobility improvement for quasi-stationary users.

### 1.2.2.2 NB-IoT

The narrowband IoT technology, known as NB-IoT or LTE-Cat NB1, is a new LP-WAN and cost-efficient solution with great potential to address the IoT challenges [9], [14], [17]. It was firstly introduced in the 3GPP Release 13 with a maximum transmission power of 20 dBm. NB-IoT and eMTC are very complementary to each other, while addressing different types of use cases. NB-IoT typically targets a massive number of UEs transmitting small amounts of data, while eMTC endeavors to deliver a higher performance to a smaller number of UEs. NB-IoT can be deployed in the same infrastructure as the existing technologies, i.e., GSM and LTE networks, while using a set of licensed frequency bands, e.g., 700 MHz, 800 MHz and 900 MHz, which simplifies its deployment design and reduces its deployment cost. However, for the uplink and the downlink transmissions, NB-IoT is carried out over an ultra-narrow frequency bandwidth of 180 KHz corresponding to a single physical resource block of GSM or LTE networks. Therefore, it may be realized under three possible scenarios as shown in Figure 1.5 [15] :

- ▶ **In-band operation** : NB-IoT operates over the same frequency bands of 180 KHz as the LTE network.
- ▶ **Guard band operation** : NB-IoT application occupies the unexploited guard bands of LTE network.
- ▶ **Stand-alone operation** : consists of reusing the same frequency bands of 200 KHz as the GSM network.



**Fig. 1.5.** The operation mode of NB-IoT technology [15].

However, it is strongly recommended in the 3GPP specifications to perform NB-IoT solution over the existing LTE infrastructure, which may be realized by simple improvements of software using devices with low-cost hardware. NB-IoT is able to support about 50,000 connected devices per cell [19], which offers a large capacity of connectivity. Added to this, as the NB-IoT communications are realized over a single narrow frequency band, the system data-rate is set to 25 kbps in the downlink and 65 kbps in the uplink [17]. Signals are modulated by a quadrature phase shift keying (QPSK) block and then transmitted using the single carrier frequency division multiple access (SC-FDMA) technique for uplink and orthogonal FDMA (OFDMA)

for downlink. In order to reduce the power consumption, the payload size of NB-IoT is limited to 1600 bytes, at maximum, with the use of eDRX and PSM techniques. As a consequence, device battery can operate for 10 years continuously with an average daily transmission of 200 bytes. NB-IoT offers an extended network coverage by 15 dB compared to LTE networks, giving a coverage range of 11 km, and thereby it can support a high MCL value of 164 dB. In short, NB-IoT seems an auspicious LPWAN candidate technology to satisfy the IoT applications.

The 3GPP Release 14 introduced new features and improvements for NB-IoT technology by enabling message multicast feature, supporting devices mobility and involving a new method for devices localization [17]. Moreover, a new reduced power class of 14 dBm was added to that of Release 13 of 20 dBm, which further increases the battery life, but at the cost of lower network coverage with MCL value of 155 dB. In this case, the network coverage can be improved through an excessive resources exploitation. Therefore, a new feature was introduced to control the coverage enhancement process using an authorization aspect. In addition, a new multi-carrier improvement was proposed in order to increase the connection density by managing the resources allocated for the signaling exchange. Release 14 offered an increase in data rate, i.e., up to 106 kbps in uplink and 79 kbps in downlink, in order to expand the set of supported services. The originally proposed NB-IoT in Release 13 did not include the handover process for non-stationary users, which results in a failure of the radio link in case of a cell change. Therefore, Release 14 came out with a new solution allowing the re-establishment of the radio resource connection when users move from one cell to another. In order to further reduce the power consumption, Release 14 offered a new indication-based feature that minimizes the activation time of each user. This technique permits to put the active user in the idle mode, as soon as the transmission or reception process of its data is terminated.

NB-IoT technology was rebraded also in the 3GPP Release 16 in order to satisfy the mMTC requirements while including further performance enhancements. It entailed the same features proposed for eMTC in Release 16 to improve the spectral efficiency and reduce the power consumption. In addition, it benefited from further improvements for mobility and network management.

It must be pointed out that, in December 2016, the first NB-IoT message was sent by Vodafone and Huawei operators to a connected water meter via the Spanish Vodafone network. At the end of 2019, the NB-IoT was deployed in 142 networks worldwide. This number is expected to increase in the next years.

### 1.2.2.3 EC-GSM-IoT

The EC-GSM-IoT was introduced in Release 13 of the 3GPP standardization based on the general packet radio service (GPRS) and enhanced GPRS (eGPRS) technologies [9], [15], [18]. Therefore, the channel bandwidth is limited to 200 KHz as for GSM system. eGPRS/GPRS networks are one of the most widespread cellular technologies over the world while using the GSM frequency bands, i.e., 850, 900, 1800 and 1900 MHz, which provides a common infrastructure between several countries and thus, supports roaming option. This may be very interesting for telecommunications stakeholders in order to simplify the deployment design and accelerate the time to commercialize this technology. The EC-GSM-IoT network is enabled by a

simple software update over the existing GSM infrastructures which improves the coverage range by 20 dB compared to eGPRS and minimizes the deployment time and cost.

EC-GSM-IoT device has a maximum transmission power of 23 dBm or 33 dBm which corresponds, respectively, to MCL value of 154 dB or 164 dB. Furthermore, thanks to the offered low system complexity and the deployed eDRX and PSM features, the devices enjoy low power consumption and thus long battery life of more than 10 years [9]. For instance, Orange operator is currently deploying the EC-GSM-IoT network for the measurement of the environmental conditions such as air pollution and temperature [57]. Two modulation options can be applied with the EC-GSM-IoT networks; eighth phase shift keying which offers a system data rate of 240 kbps and Gaussian minimum shift keying with 74 kbps. Additionally, EC-GSM-IoT is realized with new logical channels, called EC-Channels, over the existing GSM spectrum in order to increase the connectivity capacity up to 50,000 connected devices per cell and improve the coverage range. As a consequence, EC-GSM-IoT networks can be deployed in parallel with the existing 2G, 3G and 4G networks while taking advantage from the established GSM properties for the mobile security and privacy such as user identity confidentiality, entity authentication, mobile equipment identification, confidentiality and data integrity.

In Release 14, enhancements for radio interfaces were proposed, which resulted in an extension of the network coverage of at least 3 dBm compared to Release 13. However, the improvements in Release 15 mainly focused on the optimization of the power consumption in the idle state by proposing a delayed system information acquisition feature and a new paging indication channel, which reduces the power consumed to receive the paging block.

In order to achieve the extreme coverage with an MCL value of 164 dB, EC-GSM-IoT devices should transmit with 10 dBm higher power compared to eMTC and NB-IoT devices, which induces higher data rate and lower battery life. EC-GSM-IoT is similar to eMTC technology but in the GSM network. However, this technology has not caught as much attention as its competitors eMTC and NB-IoT. This is mainly due to the fact that more and more operators are considering taking their GSM networks out of service and reusing their associated spectrum.

## 1.3 Conclusion

### 1.3.1 Limitations of existing MTC technologies

Huge efforts have been made to satisfy the compelling requirements of MTC applications by proposing several 3GPP and non-3GPP technologies which provide relatively low data rate, low system complexity and efficient power consumption for low cost devices. However, the proposed solutions with their specific characteristics may not be adaptable for mMTC scenarios. On the one hand, the non-3GPP networks, i.e., LoRa and SigFox, are introduced as LPWAN proprietary solutions which are deployed over unlicensed ISM frequency bands with private standards and no performance guarantee. Therefore, for a large number of simultaneous connections, the system interference increases significantly leading to an important damage in

system performance [9]. In addition, their special network infrastructures are only available in particular parts of the world, which limits their availability and thereby their adaptability for wide range of services, such as devices tracking applications and connected cars, requiring a roaming service during their mobility.

On the other hand, the standardized 3GPP solutions are available worldwide with stable infrastructure and licensed frequency bands. Current communication systems focused on MTC scenarios using the existing 3GPP solutions, especially eMTC, NB-IoT and EC-GSM-IoT technologies. They have been enhanced in the successive 3GPP releases in order to improve their performances. However, the number of supported devices by these solutions is very modest compared to the mMTC target. For instance, NB-IoT and eMTC can serve a maximum number of  $\sim 50,000$  devices per cell, which is far from the mMTC challenge of 1 million connected devices per  $\text{km}^2$  [19]. Furthermore, these technologies are derived from LTE network. Therefore, they generate an excessive signaling overhead for data transmission including random access process, resource allocation, UE authentication, non-access stratum level security, access level security, data bearer setup, etc. The signaling overhead may be more important than the transmitted data size in mMTC scenarios, inducing resource waste, high energy consumption and high latency for the transmission of few bytes. For that reason, The 5G networks should bring new solutions and aspects in the incoming years in order to address the mMTC challenges.

### 1.3.2 mMTC enablers

Among the cutting-edge technologies, two have caught the attention of the scientific community :

- ▶ **Non-orthogonal multiple access schemes [20]** : This feature has been recently promoted as one of the main mMTC enablers. It was introduced to reach the connectivity requirement by increasing the number of supported users. The key idea is to allow multiple users to transmit simultaneously over the same resources using different dedicated signatures. However, this may generate an important signals interference and a complex decoding process. Therefore, robust multi-user detection techniques must be used for signals separation.
- ▶ **Grant free access** : In the existing cellular technologies, each user has to establish a resource allocation process for data transmission. The slotted-ALOHA protocol is applied at the MAC layer for the medium access control [58]-[60]. In fact, for each transmission attempt, the user which has data to transmit should manifest at the beginning of the next time slot to launch the random access channel request and the dynamic resource scheduling process, inducing an important signaling overhead, which can be a performance bottleneck in mMTC scenarios. Resource allocation is considered as one of the most expensive processes in terms of energy consumption, which may shorten the battery life. In this context, the grant free access strategy has been highlighted as a promising solution for the access issue [21]. It permits to each user to transmit its data autonomously without any beforehand communication with the

### 1.3. CONCLUSION

---

base station for resource assignment. For instance, each user can access the available resources using the slotted-ALOHA protocol, which requires a fundamental time synchronization, or the simple ALOHA protocol which offers more transmission autonomy, but at the cost of asynchronous signals at the base station [21], [40]. Grant free access is expected to greatly minimize the signaling overhead which ensures better power efficiency. Contrarily to the conventional OMA schemes, NOMA techniques allow the realization of grant free access strategy in order to take advantage of its benefits for mMTC scenarios.

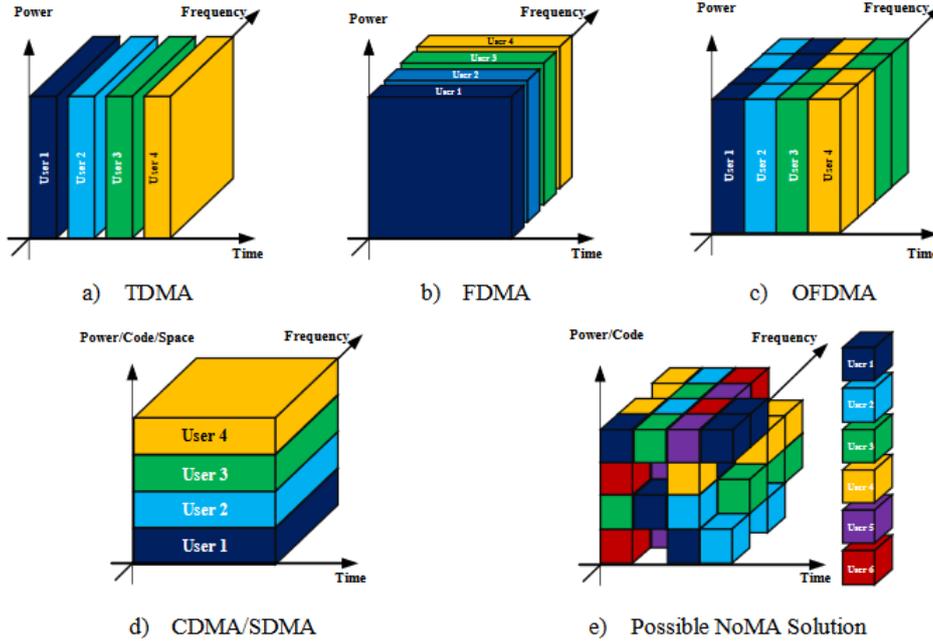


# Non orthogonal multiple access schemes

## 2.1 Introduction

The number of served users by the conventional OMA schemes is strictly limited by the number of the available orthogonal resources. More precisely, in 1G, 2G and 4G networks, each user signal is mapped to its unique dedicated orthogonal resources in time, frequency or their combination, as shown in Figure 2.1. In particular, for time division multiple access (TDMA) or FDMA schemes, user signals are uniquely transmitted over different time slots or orthogonal frequency bands, respectively. Then, they will be simply distinguished at the base station based on time or frequency domains. OFDMA scheme, used in 4G network, is considered as a combination of TDMA and FDMA schemes. The OFDMA resources are divided simultaneously and orthogonally in time and frequency. The resulting time-frequency partitions are one-to-one mapped to users. Therefore, user signals are orthogonal with no inter-user interference, leading to a simple decoding process. In code division multiple access (CDMA) scheme, used in 3G network, all users are allowed to share simultaneously all the time-frequency resources by using different spreading sequences, such as Walsh-Hadamard sequences, as seen in Figure 2.1. User signals are quasi-orthogonal with a limited multiple access interference, which offers low decoding complexity and simple multi-user detection process. However, the connection density of CDMA is limited by the maximum number of orthogonal or almost orthogonal codes.

All the previous technologies suffer from high signaling overhead for resource allocation and limited capacity of connectivity. Therefore, OMA schemes may not really be suitable to meet the mMTC requirements particularly the massive connection density issue. In order to deal with these limitations, NOMA schemes [20] have been promoted as an auspicious solution. The main idea is to overlap multiple users simultaneously and non orthogonally in the same resources, as illustrated in Figure 2.1. They permit to serve more users than the number of available orthogonal resources based on different multiplexing techniques such as power distribution or codes, but at the cost of an important inter-user interference level and high decoding



**Fig. 2.1.** Illustration of OMA and NOMA schemes [61]

complexity. As a consequence, user signals should be distinguished using advanced multi-user detection techniques. The group of NOMA schemes which apply different codes for users separation can be also seen as an improvement of the conventional CDMA scheme but with a particular code design.

Many advantages will be brought by NOMA schemes compared to the classical OMA schemes [20] :

- ▶ **Connectivity capacity** : Unlike OMA schemes and thanks to the non-orthogonality feature, the number of supported users by NOMA techniques may be higher than the number of available orthogonal subcarriers. Therefore, NOMA schemes have great potential to serve simultaneously a massive number of devices, but at the cost of significant inter-user interference and thereby high decoding complexity for signals separation .
- ▶ **Signaling overhead and transmission latency reduction** : The conventional OMA are grant-based schemes. Each user has to go through a resource allocation process for each transmission attempt. This process is very expensive in terms of transmission latency, e.g., 9.5 ms for random access in LTE [21], and signaling overhead leading to high energy consumption, especially for a huge number of connected devices. Some solutions have been recently proposed to handle this problem for OMA schemes namely 2-step RACH, which reduces the number of round trip cycles between UE and base station and thus, transmission delay and signaling overhead [62]. However, this strategy may induce asynchronous transmissions with a severe collision problem when multiple users select simultaneously the same preamble, which needs advanced techniques and new preamble formats to ensure users synchronization and detect the occurred collision [63].

The mMTC scenarios are drawn to particularly deal with a massive number of long battery life and power-limited devices. Therefore, one of the candidate enablers is to apply NOMA schemes with grant free access [51]. In fact, each UE can transmit its data as per its need without any beforehand communication with the base station for grant resource allocation, which offers reduced signaling overhead and low latency communications [20], [21]. The main feature of grant free access is to eliminate the resource scheduling process. Grant free access can be either RACH-based access, i.e., connected mode with user synchronization, or RACH-less access, inactive mode with no synchronization guarantee. In both cases, resources can be either pre-configured by the base station or randomly selected by UE. The grant free access strategy is not possible to apply with the conventional OMA schemes because each resource must be strictly allocated to a single user. However, using grant free access option with OMA schemes for mMTC scenarios may generate a severe system collision due to the limited number of resources. Hence, the collided packets should be retransmitted. The retransmission attempts are not efficient in this case and they would make the problem further complicated in terms of resources waste, transmission delay and generated signaling overhead, especially for a large number of connected devices.

- ▶ **Channel capacity and spectral efficiency** : For the uplink communications, the capacity region of NOMA schemes is larger than that of the conventional OMA schemes. As described in [64], OMA schemes generally perform sub-optimally and achieve the optimal region only at one point. However, NOMA schemes attain the maximal total achievable sum-capacity on multiple points, but at the cost of unfair user rates, particularly when the received powers are sufficiently different [20], [65]. For asymmetric channels coefficients, NOMA schemes may, in some cases, offer better rate fairness compared to OMA schemes [65]. Furthermore, applying the SIC receiver, for example, with NOMA permits to successively eliminate the inter-user interference, which improves the throughput of cell edge user and thus, allows to achieve the maximum attainable sum-capacity. In addition, users overlapping over the same time-frequency resource permits to significantly improve the spectral efficiency. It must be pointed out that NOMA schemes may be combined with MIMO technologies for further performance improvements in terms of the channel sum-capacity.
- ▶ **Relaxed channel feedback** : The channel state information (CSI) feedback is needed mainly for power allocation process. The mMTC services are not exigent in terms of reliability with PER of  $10^{-1}$ . Therefore, for NOMA power domain with a slow channel variation, the exact instantaneous CSI feedback is not necessary. In other words, an outdated CSI with tolerated imprecision may be considered in these conditions, which reduces the signaling overhead, but at the cost of an acceptable performance degradation.

Since the introduction of the IoT paradigm and the expected 5G use cases, several NOMA schemes have emerged, namely NOMA power domain (PD-NOMA),

low density spreading CDMA (LDS-CDMA), LDS-OFDM, SCMA, MUSA and Interleave division multiple access (IDMA). The main principles of these schemes will be described in details hereafter in the interest to be evaluated in the next chapters. This chapter provides also a qualitative study to a selection of the most promoted NOMA solutions to accommodate the stringent challenges of mMTC scenarios. In addition, it addresses the candidate resource access strategies for the future mMTC devices.

## 2.2 Details of NOMA schemes

### 2.2.1 Power domain NOMA

PD-NOMA scheme is one of the first proposed NOMA solutions for the future radio access network. It was originally proposed by the Japanese operator NTT DoCoMo in 2013 [36], especially for eMBB services. PD-NOMA uses a simple multiplexing aspect, i.e., power domain. The key feature is to apply a non uniform power distribution between users according to their experienced channel quality indicator. User signals are transmitted and then superposed over the same orthogonal time-frequency subcarriers, offering greater spectral efficiency than OMA schemes. The superposition coding [66], particularly for downlink transmissions, is one of the key techniques of PD-NOMA. The performance gain of this scheme compared to the classical OMA techniques is increasing with users' channel gains difference. Many algorithms have been proposed for power allocation such as, tree-search based transmission power allocation [36], [67] and fractional transmission power allocation [36], [68]. It must be pointed out that PD-NOMA may be applied in both uplink and downlink communications.

Regarding multi-user detection, PD-NOMA is typically realized with SIC receivers, which offers low decoding complexity. In fact, the SIC receiver relies on the non-uniform power distribution as well as the near-far effect for proper signals separation. Increasing the number of users leads to higher SIC decoding complexity and more severe error propagation problem, which dramatically deteriorates the performance. In addition, the power allocation process becomes more challenging for a large number of users. The performance of the PD-NOMA scheme may be also restricted by the users scheduling strategy. More precisely, for two cell center or two cell edge users overlapped on the same resource, users SINR would be very low due to the high inter-user interference, inducing a significant performance degradation. Thus, an efficient idea for PD-NOMA implementation is to attribute users with sufficiently different received powers to the same resource. Many techniques could be used for PD-NOMA waveform namely, OFDMA or SC-FDMA. This latter offers a robust performance in case of multi-path interference as well as a better adaptation to MIMO systems, which significantly increases the system throughput [36].

#### 2.2.1.1 Related work

By its apparent simplicity, PD-NOMA scheme has gained a lot of interest from the scientific community and it has been widely investigated in the literature. In [20],

[69], a survey about the main principles and aspects of PD-NOMA scheme in the uplink and the downlink communications was provided. This paper summarized the main solutions applied along with PD-NOMA namely, cooperative NOMA, MIMO systems and beamforming, user pairing, and coordinated system. Furthermore, an extensive performance comparison was presented in terms of sum-capacity, total spectral efficiency, outage probability and achievable rate compared to the conventional OMA schemes. The main challenges and perspectives of PD-NOMA scheme were also listed, opening the perspectives for future work.

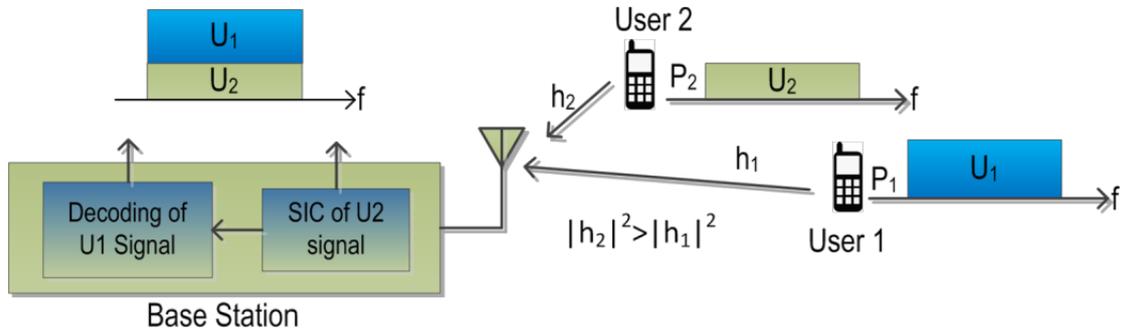
In recent years, the power optimization issue has mostly caught the attention of researchers in the interest of improving the PD-NOMA performance in terms of system sum-rate. Authors in [70] proposed an optimization problem for the weighted sum-rate maximization while considering an imperfect SIC receiver with an error propagation problem. An iterative algorithm was proposed to jointly allocate user subcarriers and power. In [71], an optimization problem was formulated for sum-rate maximization in uplink and downlink communications. For the sake of simplicity, the main problem was divided into two sub-problems. Users were firstly grouped in different clusters based on their experienced channel coefficients. Then, a power optimization process was performed for the users in each cluster under the constraints of total transmission power in the downlink, maximum transmission power in the uplink and minimal rate for each user. In order to deal with the error propagation problem, a new constraint on the difference between the received powers was also imposed. Closed-form expressions of the optimal powers were listed with the necessary conditions to obtain these values. In [72], PD-NOMA scheme was combined with beamforming technique in order to reduce the inter-cluster interference. This paper focused on the maximization of the system sum-rate while ensuring fairness between the total achieved data rates in all clusters. The optimal powers were obtained under a constraint of total system power, a constraint of minimum achievable user data rate and a constraint of minimum received power difference which permits to mitigate the inter-user interference within each cluster. In order to properly assign each user to the associated beamforming vector ensuring a maximum system sum-rate and a higher fairness index, an optimization algorithm was introduced. The presented results proved that the proposed algorithm outperforms the existing user-pairing techniques in [73], [74] thanks to its adequate power allocation strategy.

An evaluation of PD-NOMA link-level performance in terms of BER has been also addressed in certain works. For instance, the achieved BER with uplink PD-NOMA scheme using the joint maximum likelihood (ML) receiver was provided in [75]. The simulation results showed that, for a resource overload factor of 200%, the PD-NOMA could achieve a BER close to that of OFDMA scheme, but at the cost of higher decoding complexity compared to the one with SIC receiver. In addition, a joint power allocation and subcarrier assignment algorithm was proposed in this paper, which permitted to maximize the system throughput, ameliorate the spectral efficiency and increase the fairness index compared to OFDMA. In [76], an exact closed-form expression of the analytical BEP for the downlink communication and a BEP approximation for the uplink were derived while considering a resource overload of 200%, a Rayleigh fading channel and using the SIC receiver. The reported simulations proved that, under the considered conditions, the analytical expression

of the BEP corresponds well to the simulated BER, especially in the downlink. In [77] as well, an exact closed-form expression of the analytical BEP was calculated for the downlink transmissions over nakagami- $m$  fading channels with a maximum overload of 300% and a QPSK modulation scheme. The derived BEP expression was then exploited to calculate the optimal power values in two optimization problems for the minimization of the average system BEP and the establishment of system fairness with an equal BEP among users, respectively. Authors in [78] evaluated the performance of the uplink PD-NOMA with the ML receiver. This paper considered a system model with only a near and a far user from the base station using a QPSK modulation and Rayleigh fading channels. The base station was configured with multiple antenna. An exact closed-form expression of the uplink BEP was quite difficult to calculate under the considered conditions. Hence, an upper bound of the system BER was derived. The proposed expression was compared with the simulated BER and the obtained results in [76] using the SIC receiver.

However, all previous works were interested in the evaluation of PD-NOMA with time-synchronous transmissions. This hypothesis can not be always guaranteed in the practical conditions, especially in the uplink communications with grant free access. Users' signals may experience multi-path propagation, which induces time delay between users. In the asynchronous scenario, user symbols arrive at the base station with different time offset and thus, they are not aligned. As a consequence, each user symbol may be interfered by the adjacent symbols of the other users. This problem makes the signals separation more complicated with an expected performance degradation using the conventional multi-user detection techniques. Authors in [79]-[81] addressed the asynchronous transmissions issue in downlink by proposing a novel multi-user detection technique known as Triangular-SIC. The realized simulations with an overload factor of 300% and different modulation schemes proved that the proposed algorithm outperformed the conventional SIC receiver in terms of the BER for the asynchronous scenario.

### 2.2.1.2 System model



**Fig. 2.2.** Illustration of NOMA power domain for  $J=2$  [69].

We consider a multiple access channel system with  $J$  users transmitting over  $K$  subcarriers, as illustrated in Figure 2.2. Each user transmits with its attributed power  $p_{kj}$  over the  $k$ -th subcarrier, where  $\sum_{k=1}^K p_{kj} = p_j$ , and  $p_j$  is the total transmission power of the  $j$ -th user. User signals are superposed at the base station.

Therefore, the received signal at the  $k$ -th subcarrier is expressed as follow :

$$y_k = \sum_{j=1}^J h_{kj} \sqrt{p_{kj}} x_{kj} + n_k \quad (2.1)$$

where  $h_{kj} \in \mathbb{C}$  is the channel coefficient of the  $j$ -th user on the  $k$ -th subcarrier,  $x_{kj}$  is the  $j$ -th user modulated symbol, e.g., a QPSK symbol, transmitted in the  $k$ -th subcarrier with  $\mathbb{E}[|x_{kj}|^2] = 1$ ,  $n_k \sim \mathcal{CN}(0, \sigma^2)$  refers to the additive white Gaussian noise (AWGN) over the  $k$ -th subcarrier.

### 2.2.1.3 Muti user detection

PD-NOMA is typically conducted with SIC receivers [64], [69]. The latter offers a low decoding complexity, but at the cost of an increased delay especially for a massive number of devices. The key idea is to successively estimate users' symbols, reconstruct the generated interference and then subtract it from the received signal. The coexisting users in the same subcarrier are decoded in a descending order of their signal strengths. Without loss of generality, one can consider  $\text{SINR}_1 > \text{SINR}_2 > \dots > \text{SINR}_J$ . Therefore, the received signal on subcarrier  $k$  at the  $j$ -th SIC iteration is represented as follows :

$$y_k^j = \sqrt{p_{kj}} h_{kj} x_{kj} + \sum_{i=j+1}^J \sqrt{p_{ki}} h_{ki} x_{ki} + n_k^j \quad (2.2)$$

where  $n_k^j$  refers to the AWGN on subcarrier  $k$  at the  $j$ -th SIC iteration. Therefore, the corresponding SINR of the  $j$ -th user on the  $k$ -th subcarrier is expressed as  $\beta_{kj}$  [70] :

$$\beta_{kj} = \frac{p_{kj} |h_{kj}|^2}{\sum_{i=j+1}^J \sqrt{p_{ki}} |h_{ki}|^2 + \sigma^2} \quad (2.3)$$

After estimating the  $j$ -th user symbol,  $\hat{x}_{kj}$ , the generated interference is subtracted from the received signal on subcarrier  $k$  :

$$y_k^j = y_k^j - \sqrt{p_{kj}} h_{kj} x_{kj} \quad (2.4)$$

The SIC decoding process is repeated until decoding all user signals over all subcarriers.

One of the main advantages of NOMA schemes is the increase of system throughput compared to OMA schemes. In fact, for OMA systems, the bandwidth and the time are shared between users where a bandwidth partition of  $0 < \alpha_j < 1$  is assigned to user  $j$ . Hence, for a normalized bandwidth of 1 Hz, the achieved user rates, assuming capacity achieving channel coding schemes, are calculated as follows [82] :

$$R_j^{\text{OMA}} = \sum_{k=1}^K \alpha_j \log_2 \left( 1 + \frac{p_{kj} |h_{kj}|^2}{\alpha_j \sigma^2} \right) \quad (2.5)$$

However, for PD-NOMA system, the total bandwidth is simultaneously shared between users. Hence, the maximum achievable rate of each user in the  $k$ -th subcarrier is calculated as [82] :

$$\begin{aligned} R_{kj}^{\text{NOMA}} &= \log_2 (1 + \beta_{kj}) \\ &= \log_2 \left( 1 + \frac{p_{kj}|h_{kj}|^2}{\sum_{i=j+1}^J \sqrt{p_{ki}}|h_{ki}|^2 + \sigma^2} \right) \end{aligned} \quad (2.6)$$

For PD-NOMA, the maximum total achieved throughput is independent from the decoding order and, is given by [64] :

$$\begin{aligned} R^{\text{NOMA}} &= \sum_{j=1}^J \sum_{k=1}^K R_{kj}^{\text{NOMA}} \\ &= \sum_{j=1}^J \sum_{k=1}^K \log_2 \left( 1 + \frac{p_{kj}|h_{kj}|^2}{\sum_{i=j+1}^J \sqrt{p_{ki}}|h_{ki}|^2 + \sigma^2} \right) \\ &= \sum_{k=1}^K \log_2 \left( 1 + \frac{\sum_{j=1}^J p_{kj}|h_{kj}|^2}{\sigma^2} \right) \end{aligned} \quad (2.7)$$

Despite its advantages compared to the conventional OMA schemes in terms of system overload, spectral efficiency and achieved throughput, PD-NOMA may suffer from the SIC error propagation problem, which deteriorates the system performance. In addition, for an adequate power distribution and thereby a proper signals decoding process, users are recommended to be sufficiently dispersed over the cell to ensure sufficient power difference between them, ensuring a correct users detection.

#### 2.2.1.4 Conclusion

PD-NOMA scheme is one of the candidate solutions to increase the network capacity compared to the classical OMA schemes. It applies a simple multiplexing aspect, i.e., power domain, which consists in allocating a non-uniform power distribution according to users' channel conditions. PD-NOMA has a relatively low overload capacity which depends on the assigned power distribution. Moreover, for multi-user detection, PD-NOMA can be merged with SIC receiver which ensures low decoding complexity. However, user signals may experience higher error propagation problem as the number of supported users increases. PD-NOMA was originally proposed for eMBB use case with a high data rate requirement. Hence, its connectivity capacity may be limited regarding mMTC scenarios.

### 2.2.2 Low density spreading CDMA

LDS-CDMA is seen as an evolution of the conventional CDMA scheme with reduced inter-user interference [83]. CDMA is based on dense spreading sequences with non-zero elements. Therefore, all users transmit over the same chips and thereby each user's signal is affected by the interference from all other users. For non overloaded

systems, the orthogonality asset of the classic CDMA codes allows to properly separate user signals with an affordable decoding complexity. Therefore, the CDMA capacity is limited by the number of available orthogonal codes. In fact, increasing the system load, by serving more users than the number of code chips, deteriorates the codes orthogonality, which induces a high amount of inter-user interference. As a consequence, robust multi-user detection techniques with high decoding complexity are required. This type of large-length and dense sequences may not be suitable to increase the network capacity. In light of the above, LDS sequences, which are also known as sparse sequences with mostly zero components, are promoted as a promising solution. The sparsity feature permits to reduce the number of interfering users over the same chip and thus, the amount of the experienced inter-user interference. The LDS concept permits to simplify the signals separation process with lower complexity while offering a performance improvement compared to dense sequences.

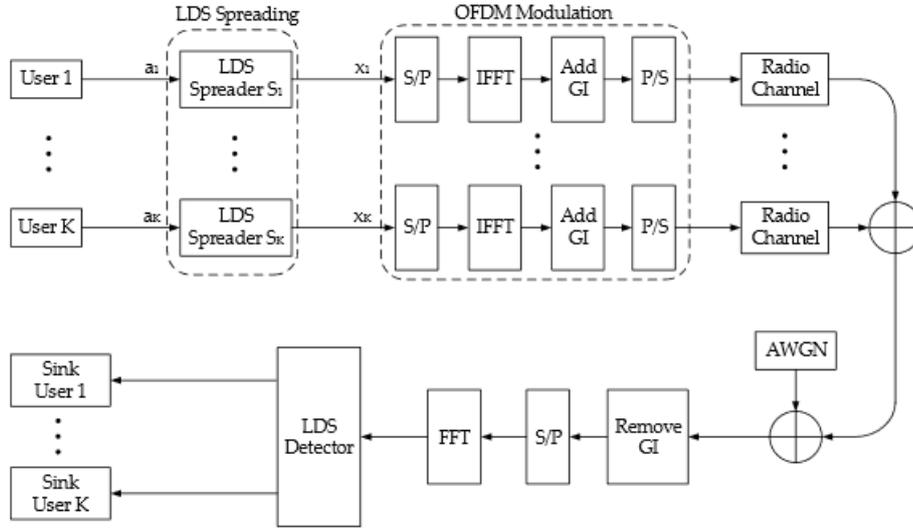
Message passing algorithm (MPA) is one of the multi-user detection techniques that can be applied with LDS-CDMA scheme. The key idea is to model the receiver system with a graph. Then, the marginal distributions of the unobserved nodes are calculated based on those of the already observed ones. This receiver offers a near-optimal decoding performance with lower complexity compared to the optimal maximum a posteriori (MAP) receiver. This is due to the sparsity asset of the used LDS sequences. The principle of the MPA receiver is detailed later with SCMA scheme.

Several works have focused on the performance evaluation of LDS-CDMA. Authors in [84] proved that LDS-CDMA combined with MPA receiver, memoryless Gaussian channel and BPSK modulation could achieve a performance close to that of a single user, without using the forward error correction techniques, for a system overload up to 200%. Furthermore, in [85] the performance of LDS-CDMA with a set of different spreading sequence matrices was investigated. The decoding performance of this scheme was evaluated using different multi-user detection techniques. Authors in [86] focused on the design of LDS sequences offering a good performance with the belief propagation receiver for an over-saturated system. The simulation results proved that a significant BER improvement was achieved compared to random sequences. In addition, the obtained BER was close to that of a single user even with a system overload of 200%. In [87], the capacity region of LDS-based schemes was investigated. Authors addressed also the impact of different density factors and the maximum number of users per chip on the achieved weighted sum-rate.

The main challenge of LDS-CDMA scheme is its application with multi-path fading channels. The generated inter-symbol interference increases the number of interfering users over the same chip and thus the density factors of the used spreading sequences. Therefore, the decoding process becomes more complex with an expected performance degradation [88]. The LDS-OFDM scheme is then proposed to deal with this issue [89].

### 2.2.3 Low density spreading OFDM

LDS-OFDM scheme is a combination between the LDS-CDMA scheme and the OFDM technique [83]. It aims to handle the performance degradation issue of



**Fig. 2.3.** Uplink LDS-OFDM system [89]

LDS-CDMA with multi-path fading channel. Using multi-carrier techniques such as OFDM may attenuate the effect of multi-path propagation [89]. The key idea of LDS-OFDM is to firstly map the transmitted symbol of each user to an associated LDS sequence. Then, the resulting sequences are spread over a set of selected subcarriers and go through an OFDM modulation, as illustrated in Figure 2.3.

Unlike the classical orthogonal schemes where the system capacity is limited by the number of subcarrier and each resource is uniquely allocated to a single user, LDS-OFDM allows to support more users than the number of resources, which significantly improves the spectral efficiency.

Similarly to LDS-CDMA, LDS-OFDM can also be merged with MPA receiver for multi-user detection. The number of interfering users in the same subcarrier is lower than the total number of active users, which minimizes the size of the system graph, and offers lower complexity. The received signal must first pass through an OFDM demodulation block before starting the iterative process using MPA. Thanks to the LDS asset and the OFDM technique, LDS-OFDM scheme can perform close in BER to single user with an overload factor of 200% [89].

LDS-OFDM can be seen as an enhancement of the multi-carrier (MC-CDMA) scheme with low density spreading sequences instead of dense ones. In the classical MC-CDMA scheme, a full exploitation of frequency diversity can be satisfied using an optimal receiver such as MAP or MPA receivers, but at the cost of high decoding complexity. Therefore, MC-CDMA is generally conducted with linear receiver in practical implementation in order to reduce the system complexity, which prevents MC-CDMA from taking advantage of the totality of the frequency diversity. Nevertheless, the satisfaction of the subcarrier spacing condition with LDS sequences enables LDS-OFDM to take full advantage of maximum frequency diversity with lower decoding complexity using an optimal signals detector compared to the conventional multi-carrier and dense spreading techniques. However, the LDS sequences may suffer from a cross-correlation issue, which induces high envelope fluctuations and thus a higher peak to average power ratio (PAPR) or cubic metric

(CM) compared to the conventional OFDMA scheme [89]. Many techniques have been proposed to address the PAPR/CM problem, such as Newman and Narahashi phases.

LDS-OFDM scheme has been widely investigated in the literature. Authors in [90] addressed the high PAPR issue of LDS-OFDM scheme. They studied the effect of the subcarriers assignment and signature phase on the system PAPR value. Two novel methods, known as Newman phases and DFT pre-coding, were proposed in order to minimize the PAPR. In [91], the design of the LDS sequences was investigated in order to optimize the resource overload. An iterative technique was proposed to address this issue using the EXIT chart. The simulation results proved that the optimized LDS sequences achieved a performance enhancement of 10 dB at BER equal to  $10^{-3}$  compared to OFDMA. Authors in [92] addressed the power allocation issue for LDS-OFDM scheme with cognitive radio networks. The main goal was to maximize the secondary users throughput by adjusting the user power coefficient at each subcarrier. A novel algorithm was proposed, which improves the secondary users throughput compared to the uniform power distribution among subcarriers. Finally, in [93], an optimization problem was formulated for the weighted sum-rate maximization by jointly allocating the subcarriers and powers. The problem was solved with an heuristic algorithm which improves the spectral efficiency and minimizes the outage probability compared to the conventional OFDM and LDS-OFDM with equal power among subcarriers.

#### 2.2.4 Sparse code multiple access

SCMA was firstly proposed by NIKOPOUR et BALIGH in [23]. It can be seen as an improved version of LDS-OFDM scheme. SCMA is well known by its theoretical optimal codebooks. The key feature of this scheme is to attribute to each user a unique dedicated codebook including  $M$  multi-dimensional codewords. The latter are mainly characterized by their sparsity asset where the number of non zero elements  $N$  should be lower than the codeword dimension  $K$ , i.e.,  $N \ll K$ . This permits to reduce the number of interfering users at each subcarrier, leading to better signals separation. Furthermore, the non zero elements should be left at the same positions among all the codewords of a given codebook, but they should vary from a codebook to another in order to reduce users collision.

The SCMA encoder performs similarly to the quadrature amplitude modulation (QAM) process but with a multi-dimensional constellation rather than simple QAM symbols. The shaping gain of the multi-dimensional constellation ensures a considerable performance enhancement compared to a simple LDS sequence with QAM symbol. User codewords are generated using a multi-dimensional complex mother constellation in  $\mathbb{C}^N$  which undergoes different operations such as mapping, dimensional permutation, conjugate, phase rotation, etc, in order to obtain the corresponding user codebooks [94]. The maximum number of possible codebooks, which can be generated while respecting the design restrictions, is calculated as :

$$C(N, K) = \frac{K!}{(K - N) N!} \quad (2.8)$$

### 2.2.4.1 Related work

The assigned codebooks are a critical feature for SCMA performance. In the last few years, several works have been interested in the codebook design issue in order to improve the performance of SCMA [23], [86], [94]-[102]. Authors in [23] and [94] underlined the basic design aspects that should be satisfied for SCMA codebooks. They also proposed a sub-optimal method for an efficient codewords construction which were compared with the low density sequences. The simulation results proved that the designed codebooks achieved a better performance in terms of block error rate (BLER) over Rayleigh fading channel. In [95], the codebook design process was carried on two steps while applying the rotation and the interleaving techniques, respectively. The proposed method outperformed the derived one in [23] and the classical low density sequences in [86].

In [96], the main idea was to maximize the achievable sum-rate through an efficient construction of the assigned users codebooks. Regarding the BER evaluation, an improvement of 1.1 dB in AWGN channel and 1.3 dB in Rayleigh fading channels was observed compared to the obtained results in [97]. In [98], star-QAM constellations were used to design the mother constellation which has to go through different operations, such as complex conjugate, phase operator or vector permutation, in order to obtain the different user codebooks. In [99], a novel algorithm, known as symbol exchange algorithm, was introduced to design near-optimal codebooks for Rayleigh fading channel with low complexity. The achieved BER with the designed codebooks using the proposed algorithm outperformed those of the derived techniques in [95], [96], [98], [100]-[102] for uplink and downlink transmissions.

### 2.2.4.2 System model

Each user picks up a  $K$ -dimensional codeword with  $N$  non-zero elements from its dedicated codebook. Then, the input data stream will be directly mapped by blocks of  $\log_2(M)$  to the selected codeword, skipping the bit-to-symbol modulation process. All the picked codewords of  $J$  users will be spread and then transmitted over the  $K$  available orthogonal subcarriers. Hence, the received signal vector at the base station on all subcarriers can be expressed as [23] :

$$\mathbf{y} = \sum_{j=1}^J \text{diag}(\mathbf{h}_j) \mathbf{x}_j + \mathbf{n} \quad (2.9)$$

where  $\mathbf{x}_j = [x_{1j}, x_{2j}, \dots, x_{Kj}]^T \in \mathbb{C}^{K \times 1}$ ,  $\mathbf{h}_j = [h_{1j}, h_{2j}, \dots, h_{Kj}] \in \mathbb{C}^{K \times 1}$  are the codeword and the channel vector of the  $j$ -th user spanning over the  $K$  subcarriers, respectively, and  $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_K)$  is the AWGN vector with  $\mathbf{n} = [n_1, n_2, \dots, n_K]$ . Due to the modulation of a limited number of subcarriers, i.e.,  $N$  over  $K$ , the decoding complexity can be reduced. Indeed, only the interference created by the overloaded users, i.e., users data transmitted on the same subcarriers, has to be dealt with [103]. Let denote  $\mathcal{N}(k) = \{j \mid x_{kj} \neq 0\}$  the set of users transmitting on subcarrier  $k$  and  $\mathcal{R}(j) = \{k \mid x_{kj} \neq 0\}$  the set of subcarriers carrying the information

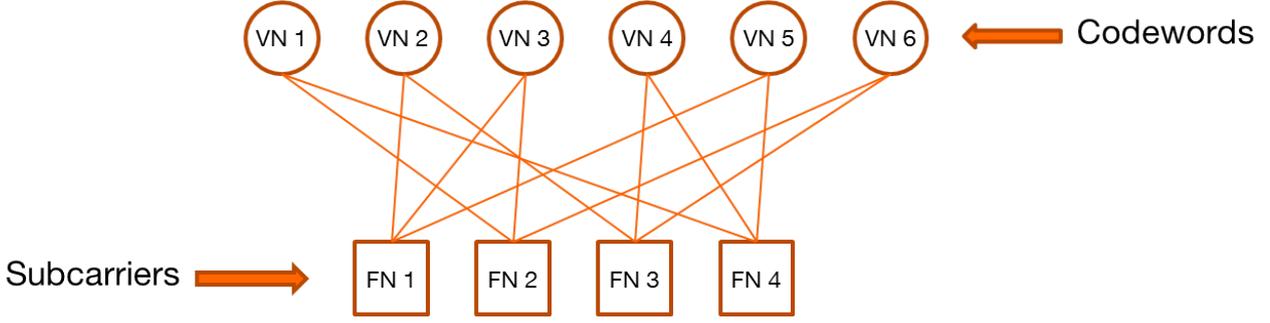


Fig. 2.4. MPA factor graph for  $J = 6$  and  $K = 4$ .

of user  $j$ . The received signal at the  $k$ -th subcarrier can be represented as [104] :

$$y_k = \sum_{j \in \mathcal{N}(k)} h_{kj} x_{kj} + n_k \quad (2.10)$$

### 2.2.4.3 Multi-user detection

The SCMA scheme has been investigated with different receivers in the literature such as SIC receivers [105], jointed SIC-MPA [105], ML [106], belief propagation (BP) [107] or the expectation propagation algorithm (EPA) [108]. However, the MPA receiver has been the most tagged one [23]. In fact, MPA is a near optimal and an iterative detection algorithm based on a bipartite factor graph modeling the receiver, as illustrated in Figure 2.4. The graph contains two different types of nodes : function nodes (FNs), representing the  $K$  subcarriers, and variable nodes (VNs), representing the  $J$  active users. The VN  $j$  is connected to the FN  $k$  when user  $j$  uses the  $k$ -th subcarrier, i.e.,  $x_{kj} \neq 0$ . The key idea of MPA is to estimate the transmitted codeword  $\hat{\mathbf{x}}_j$  given the received signal  $\mathbf{y}$  observation with an affordable computing complexity compared to the optimal MAP decoding algorithm [84]. Hence, the following optimal detection criterion [27], [84] is admitted :

$$\hat{\mathbf{x}}_j = \arg \max_{\mathbf{a} \in \mathcal{X}} \sum_{\substack{\{\mathbf{x}_i\} \in \mathcal{X}^J \\ \mathbf{x}_j = \mathbf{a}}} p(\mathbf{x}_1, \dots, \mathbf{x}_J | \mathbf{y}) \quad (2.11)$$

$$= \arg \max_{\mathbf{a} \in \mathcal{X}} \sum_{\substack{\{\mathbf{x}_i\} \in \mathcal{X}^J \\ \mathbf{x}_j = \mathbf{a}}} \prod_{i=1}^J p(\mathbf{x}_i) \prod_{k \in \mathcal{R}(j)} p(y_k | \mathbf{x}_k, l \in \mathcal{N}(k)) \quad (2.12)$$

Where  $\{\mathbf{x}_i\}$  refers to the set of corresponding users codewords,  $\mathcal{X}$  represents the  $j^{\text{th}}$  user codebook with  $|\mathcal{X}| = M$  and  $\mathcal{X}^J$  is the set of codebooks of all users. Each MPA iteration consists of two main steps : i) passing information from FNs to VNs and ii) passing information from VNs to FNs. Let  $m_{k \rightarrow j}^t(\mathbf{x}_j)$  be the extrinsic information sent from FN  $k$  to VN  $j \in \mathcal{N}(k)$  for the corresponding codeword  $\mathbf{x}_j$  at the  $t$ -th iteration. Similarly,  $g_{j \rightarrow k}^t(\mathbf{x}_j)$  is the extrinsic information passed from VN  $j$  to FN  $k \in \mathcal{R}(j)$  for the corresponding codeword  $\mathbf{x}_j$  at the  $t$ -th iteration. The MPA process is conducted on multiple steps which are defined as follows [84], [109] :

► **Step 0 : Initialization**

Before tackling the iterative process, an initialization of the system parameters is necessary. Therefore, the prior probability of having sent the codeword  $\mathbf{x}_j$  by the  $j$ -th user over its corresponding set of subcarriers  $\mathcal{R}(j)$  is initially considered as uniform, hence :

$$g_{j \rightarrow k}^0(\mathbf{x}_j) = p(\mathbf{x}_j) = \frac{1}{M} \quad \forall j = 1, \dots, J, \forall k \in \mathcal{R}(j) \quad (2.13)$$

► **Step 1 : Iterative message exchange**

The iterative information exchanges between FNs and VNs in both directions are carried out sequentially over two sub-steps as described hereafter :

– **Sub step a) : Passing information from FN to VN**

The passed message from FN  $k$  to VN  $j$  for a given codeword  $\mathbf{x}_j$  of the  $j$ -th user is calculated as a function of exponential operations with the previously received information from the related VNs, which is represented as follows :

$$m_{k \rightarrow j}^t(\mathbf{x}_j) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{\{\mathbf{x}_i | i \in \mathcal{N}(k) \setminus j\}} \exp\left\{\frac{-1}{\sigma^2} \left\| y_k - \sum_{\substack{j \\ x_{kj} \neq 0}} h_{kj} x_{kj} \right\|^2\right\} \prod_{\{i \in \mathcal{N}(k) \setminus j\}} g_{i \rightarrow k}^{(t-1)}(\mathbf{x}_i) \quad (2.14)$$

– **Sub step b) : Passing information from VN to FN**

The marginal distribution function of each VN is deduced from its previously received messages from its related FNs. Hence, the passed message from VN  $j$  to FN  $k$  for a given codeword  $\mathbf{x}_j$  of the  $j$ -th user is updated as follows :

$$g_{j \rightarrow k}^t(\mathbf{x}_j) = \frac{1}{M} \prod_{\{i \in \mathcal{R}(j) \setminus k\}} m_{i \rightarrow j}^{t-1}(\mathbf{x}_j) \quad (2.15)$$

These messages are considered as probabilities. In fact, their values should be less than 1 as well as their sum for all the corresponding codewords of each user  $j$  should be equal to one, i.e.,  $\sum_{\mathbf{x}_j \in \mathcal{X}} g_{j \rightarrow k}^t(\mathbf{x}_j) = 1 \quad \forall j = 1, \dots, J$ . Therefore, the above messages should be normalized as follows :

$$g_{j \rightarrow k}^t(\mathbf{x}_j) = \frac{\prod_{\{i \in \mathcal{R}(j) \setminus k\}} m_{i \rightarrow j}^{t-1}(\mathbf{x}_j)}{\sum_{\mathbf{x}_j \in \mathcal{X}} \prod_{\{i \in \mathcal{R}(j) \setminus k\}} m_{i \rightarrow j}^{t-1}(\mathbf{x}_j)} \quad (2.16)$$

► **Step 2 : LLR calculation and bits estimation**

The MPA algorithm needs to iterate sufficiently, i.e.,  $T$  iterations let's say, in order to converge. Then, the a posteriori probability for the corresponding codeword  $\mathbf{x}_j$  of the  $j$ -th user, defined as  $p(\mathbf{x}_j)$ , is represented as :

$$p(\mathbf{x}_j) = \frac{1}{M} \prod_{k \in \mathcal{R}(j)} m_{k \rightarrow j}^T(\mathbf{x}_j) \quad (2.17)$$

Afterwards, the bit-wise LLRs corresponding to the  $j$ -th user bits,  $b_{ij}$  with  $i \in \{1, \dots, \log_2(M)\}$ , are calculated as the ratio between the sum of the probabilities of the  $j$ -th user codewords corresponding to  $b_{ij} = 0$  and the sum of the probabilities of codewords corresponding to  $b_{ij} = 1$  :

$$\text{LLR}(b_{ij}) = \log \frac{\sum_{\{\mathbf{x}_j \in \mathcal{X} | b_{ij}=0\}} p(\mathbf{x}_j)}{\sum_{\{\mathbf{x}_j \in \mathcal{X} | b_{ij}=1\}} p(\mathbf{x}_j)} \quad \forall i \in \{1, \dots, \log_2(M)\}, \quad \forall j \in \{1, \dots, J\} \quad (2.18)$$

Finally, the  $j$ -th user bits,  $\hat{b}_{ij}$  with  $i \in \{1, \dots, \log_2(M)\}$ , are estimated by comparing the corresponding bit-wise LLR to 0 such that :

$$\begin{cases} \hat{b}_{ij} = 1 & \text{If } \text{LLR}(b_{ij}) \leq 0 \\ \hat{b}_{ij} = 0 & \text{Otherwise} \end{cases} \quad (2.19)$$

MPA receiver has a low-decoding complexity compared to the classical MAP algorithm. However, its complexity is still high compared to other simple receivers such as SIC receivers, especially for a massive number of users, due to the large number of involved exponential operations. Furthermore, the SCMA overload capacity is limited by the number of generated codebooks (2.8) which is restricted by the sparsity conditions [103]. Added to this, the SCMA codebooks attribution may induce high signal overhead and thereby high energy consumption. Therefore, all of these obstacles may reduce the chance of SCMA facing the mMTC service.

MPA is a robust multi-user detection technique with near-optimal performance, which permits to properly separate user signals. In addition, the sparsity feature of the designed codewords allows to reduce the number of interfering users in each subcarrier, which would further improve the system performance. Therefore, SCMA may be recommended for high reliable transmission requirements with a moderate number of users.

#### 2.2.4.4 Conclusion

SCMA scheme has a special transmitter process. it allocates different codebooks, including a set of multidimensional and sparse codewords, to different users, which allows the signals separation. Generally, SCMA is conducted with MPA receiver which provides a robust multi-user detection. However, it suffers from a high decoding complexity which increases exponentially with the number of users. Finally, the codebooks design complexity, distribution and the overload capacity can be major issues for this scheme, especially for mMTC scenarios.

### 2.2.5 Multi user shared access

MUSA scheme was firstly proposed by ZTE in 2016 in [24]. It can be seen as a refinement of MC-CDMA scheme with non orthogonal, low cross-correlation and short spreading sequences to support the coexistence of a large number of users in the same radio resources. MUSA scheme can be suitable for grant free access scenarios. In fact, each user can autonomously pick up a random spreading sequence within a predefined constellation with no communication with the base station or

other users. In addition, the number of non-zero elements within each sequence is not previously defined as for SCMA, which offers more degrees of freedom for the design of sequences.

Regarding spreading sequences design, each complex element of the  $j$ -th user sequence is represented as  $s_{kj} = a + ib \forall k \in \{1, \dots, K\}$  and  $\forall j \in \{1, \dots, J\}$ , where  $a$  and  $b$  are chosen in a  $M$ -ary constellation. For instance, for 3-ary constellation  $\{-1, 0, 1\}$ , the set of possible complex elements is  $\{0, 1, -1, 1i, -1i, 1+1i, 1-1i, -1+1i, -1-1i\}$  and thereby the maximum number of  $K$ -dimensional sequences is  $9^K$ , which offers a high overload capacity. An example for MUSA spreading sequences matrix,  $\mathbf{S} \in \mathbb{C}^{K \times J}$ , with  $K = 4$  and  $J = 6$  users, is represented as :

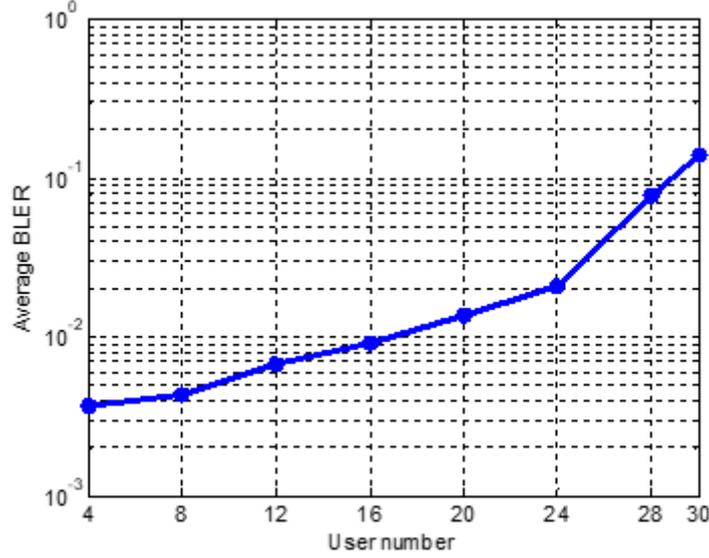
$$\mathbf{S} = \begin{bmatrix} 1+i & 1 & -1-i & 1-i & i & -1 \\ 0 & -1+i & i & 1 & -1 & 1+i \\ 1 & -i & 1+i & -1-i & 1+i & 0 \\ -1-i & 1+i & -1+i & 1 & -1-i & i \end{bmatrix}$$

Although the probability of collision is very low due to the large pool of sequences, however, if multiple users select the same sequence or different sequences with high cross-correlation, their signals can be detected using SIC receivers thanks to the inherent near-far effect of uplink transmissions.

### 2.2.5.1 Related work

Authors in [24] studied the impact of different spreading sequences design on the experienced BLER. This paper provided a comparison of the achieved performances using binary pseudo-random noise sequences, binary complex sequences with 2-ary constellations, tri-level complex sequences with 3-ary constellations and random Gaussian complex sequences in an AWGN channel and different sequence lengths. The simulations proved that pseudo-random noise sequences were the less efficient and they could not achieve a BLER of 1% using 4-dimensional sequences even for a resource occupancy of 100%. They also proved that spreading sequences length had a crucial impact on the achieved performance. In fact, the larger the spreading factor is, the lower the cross-correlation level between the sequences, but at the cost of higher decoding complexity. Therefore, a trade-off between the sequences length and the receiver complexity should be established. Moreover, they showed that for different overload factors the tri-level complex sequences had a BLER close to that of the optimal complex Gaussian sequences with the advantage of lower design complexity. This paper proved also that for fast Rayleigh fading channels, tri-level complex sequences of length 4 and two receive antennas, MUSA scheme outperformed OFDMA for high SNR values and a resources overload factor up to 400%. This is due to the offered users diversity by the experienced channel coefficients and the selected spreading sequences.

The reported results in [111] proved that MUSA scheme with only 4 available subcarriers and a single receive antenna achieved a BLER less than 10% for a resource overload up to 300% over flat fading channels and a uniformly distributed received SNR in the range of 4 ~ 20 dB presenting the near-far situation of a realistic distribution. Then, authors in [110] showed that the overload capacity can be improved while considering a multi-path fading channel and two receive antennas. They



**Fig. 2.5.** MUSA BLER for different overload factors for  $K = 4$  [110]

proved that a BLER less than 0.1 could be obtained for a system overload up to 700% for a received SNR range of 6 ~ 20 dB, as seen in Figure 2.5. Authors in [112] studied the BER performance of downlink MUSA with different power values. They showed that power allocation permitted to improve the system performance where different power distributions led to different BER values. In [113], a new MUSA version was introduced in order to increase the connectivity capacity compared to the classical MUSA. The derived scheme was applied with the generalized frequency division multiplexing technique in order to reduce the transmission latency while ensuring a high resource occupancy.

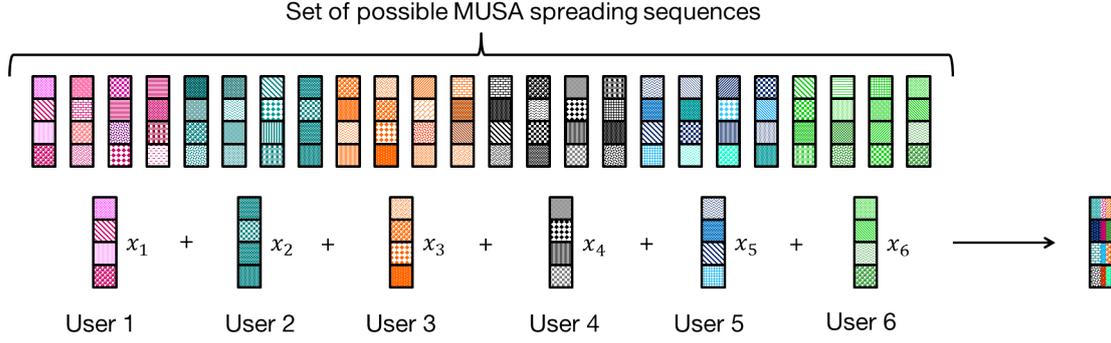
### 2.2.5.2 System model

The incoming bit stream of each user is mapped to a series of  $M$ -ary symbols where  $M$  refers to the modulation order. Then, each symbol is multiplied by all the elements of a  $K$ -dimensional random spreading sequence. The resulting code, i.e., the symbol multiplied by the spreading sequence, is spread over the  $K$  subcarriers, as illustrated in Figure 2.6. Each user may select different sequences for the transmission of different symbols, which allows to average the experienced users interference with an expected performance enhancement [83]. The received signal at subcarrier  $k$  can be given as :

$$y_k = \sum_{j=1}^J \sqrt{p_j} h_{kj} s_{kj} x_j + n_k \quad (2.20)$$

where  $s_{kj}$  refers to the  $k$ -th element of the  $j$ -th user spreading sequence  $\mathbf{s}_j = [s_{1j}, s_{2j}, \dots, s_{Kj}]^T \in \mathbb{C}^K$  and  $x_j$  is the transmitted symbol of the  $j$ -th user. Therefore, the received signal vector on all subcarriers can be expressed as :

$$\mathbf{y} = \mathbf{G}\mathbf{P}^{\frac{1}{2}}\mathbf{x} + \mathbf{n} \quad (2.21)$$



**Fig. 2.6.** MUSA scheme transmitter for  $J = 6$  and  $K = 4$ .

where  $\mathbf{x} = [x_1, x_2, \dots, x_J]^T$  is the vectorized users' symbols with  $E[\mathbf{x}\mathbf{x}^H] = \mathbf{I}_J$ ,  $\mathbf{P}^{\frac{1}{2}} = \text{diag}(\sqrt{p_1}, \sqrt{p_2}, \dots, \sqrt{p_J}) \in \mathbb{R}_+^{J \times J}$  is the users' transmission powers matrix and  $\mathbf{G} \in \mathbb{C}^{K \times J}$  is the equivalent channel matrix including the spreading sequences, which can be expressed as follows :

$$\mathbf{G} = \mathbf{H} \odot \mathbf{S} \quad (2.22)$$

where  $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_J] \in \mathbb{C}^{K \times J}$ ,  $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_J] \in \mathbb{C}^{K \times J}$  and  $\odot$  is the Hadamard product, i.e.,  $g_{kj} = h_{kj}s_{kj}$ .

### 2.2.5.3 Multi-user detection

MUSA can be combined with an ordered-SIC (OSIC) receiver for multi-user detection. The key idea is to successively estimate user's symbols while eliminating the generated interference by the already decoded users, as illustrated in Figure 2.7. This process is performed on multiple stages. In each phase, OSIC may be jointly conducted with a linear detection receiver such as minimum mean square error (MMSE) for user's symbol estimation, where the MMSE matrix can be computed as [114] :

$$\mathbf{W}^H = (\mathbf{P}^{\frac{1}{2}}\mathbf{G}^H\mathbf{G}\mathbf{P}^{\frac{1}{2}} + \sigma^2\mathbf{I})^{-1}\mathbf{P}^{\frac{1}{2}}\mathbf{G}^H \quad (2.23)$$

Based on the general case in the section (2.2.1.3), users will be sorted in a descending order according to their SINRs. For a given decoding order, e.g.,  $\text{SINR}_1 > \text{SINR}_2 > \dots > \text{SINR}_J$ , the received signal at the  $j$ -th stage, after removing the interference from the already decoded users, is represented as :

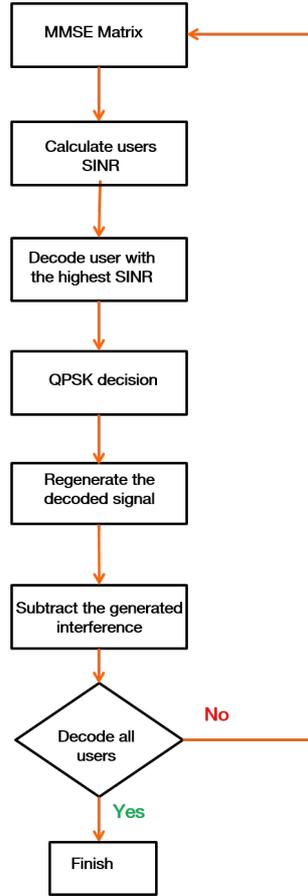
$$\mathbf{y}^j = \sqrt{p_j}\mathbf{g}_j x_j + \sum_{i=j+1}^J \sqrt{p_i}\mathbf{g}_i x_i + \mathbf{n}^j \quad (2.24)$$

where  $\mathbf{n}^j$  refers to the noise vector at the  $j$ -th iteration. Hence, the SINR of user  $j$  is [114] :

$$\beta_j(\mathbf{p}) = \frac{p_j |\mathbf{w}_j^H \mathbf{g}_j|^2}{\sum_{i=j+1}^J p_i |\mathbf{w}_j^H \mathbf{g}_i|^2 + \sigma^2 \|\mathbf{w}_j^H\|^2} \quad (2.25)$$

where  $\mathbf{p} = [p_1, \dots, p_J]^T$ ,  $\mathbf{w}_j$  and  $\mathbf{g}_j$  refer to the  $j$ -th column of the matrices  $\mathbf{W}$  and  $\mathbf{G}$ , respectively. Thus, the transmitted symbol of user  $j$  is estimated as :

$$\hat{x}_j = \mathbf{w}_j^H \mathbf{y}. \quad (2.26)$$



**Fig. 2.7.** Iterative process of MMSE-SIC receiver.

The interference of the  $j$ -th decoded symbol is regenerated and then subtracted. After the  $j$ -th iteration, the received signal is updated as follows :

$$\mathbf{y} = \mathbf{y} - \sqrt{p_j} \mathbf{g}_j \hat{x}_j. \quad (2.27)$$

Finally, the channel vector of the decoded user, i.e.,  $\mathbf{g}_j$ , is removed from the channel matrix  $\mathbf{G}$ , the MMSE matrix  $\mathbf{W}^H$  is computed again and the  $(j + 1)$ -th symbol is estimated then removed from the global signal. The process will be repeated until all users have been decoded.

MUSA scheme can be realized with other receivers such as compressive sensing MPA (CS-MPA) [115] or ML [106]. However, the SIC receiver is the most applied one as it provides an affordable decoding complexity, especially for a massive number of connected devices, but at the cost of high transmission latency. Despite its high overload capacity and grant free access asset, the performance of MUSA may be damaged by the famous error propagation issue of SIC receivers, particularly when users received powers are not sufficiently diversified, leading to close user SINRs.

#### 2.2.5.4 Conclusion

The MUSA scheme is based on complex spreading sequences in order to increase the number of possible short length and low cross-correlation sequences. They can be

generated randomly within a defined constellation. Generally, MUSA is combined with MMSE-OSIC receiver, which ensures a low decoding complexity. However, the risk of error propagation is getting more important as the resource overload increases, especially when users are received with close powers.

## 2.2.6 Pattern division multiple access

PDMA is a novel NOMA scheme which was recently proposed in order to meet the future radio access network requirements [25], [37]. More precisely, PDMA seems more adaptable for eMBB and certain mMTC scenarios. It can be realized in code, power, space domains or their combination. eMBB scenario with high data rate requirement may be satisfied by combining PDMA with certain techniques such as MIMO system, ultra-dense networking and high-frequency communication. In code domain, PDMA can be seen as a kind of MUSA scheme with binary spreading sequences, with two possible elements  $\{0, 1\}$ , where "1" indicates that the corresponding user is transmitting over this resource and "0" otherwise. Similarly to MUSA, PDMA can enable grant free access by allowing each user to randomly generate its spreading sequence, but at the cost of high collision probability. This is mainly due to the limited number of binary patterns. In fact, for  $K$  available resources, the maximum number of possible generated binary patterns is calculated as follows :

$$C(K) = \sum_{i=1}^K \binom{K}{i} = 2^K - 1 \quad (2.28)$$

where  $\binom{K}{i}$  is the number of combinations of  $i$ -elements among a set of  $K$  elements. For instance, for  $K = 4$  subcarriers, the maximum number of possible patterns is  $2^4 - 1 = 15$ . As a consequence, even under the ideal system condition, the overload capacity can not reach 400%, where the maximum set of binary spreading sequences matrix is given as :

$$\mathbf{S} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

### 2.2.6.1 Related work

Due to the linear combination of the coexisting symbols in same subcarriers using binary spreading sequences, the signals characteristics may be damaged in certain cases depending on the modulation scheme and the channel conditions, which makes the decoding process more complicated [37]. For example, considering two users on the same resource element with BPSK modulation and perfect noiseless AWGN channel, users symbols may not be distinguished correctly at the receiver in some cases. Therefore, an extended pattern matrix using the power scaling and the phase shifting was proposed in order to improve the multi-user detection process and ensure a proper symbols separation [37], [116]. Furthermore, a new aspect for pattern design was

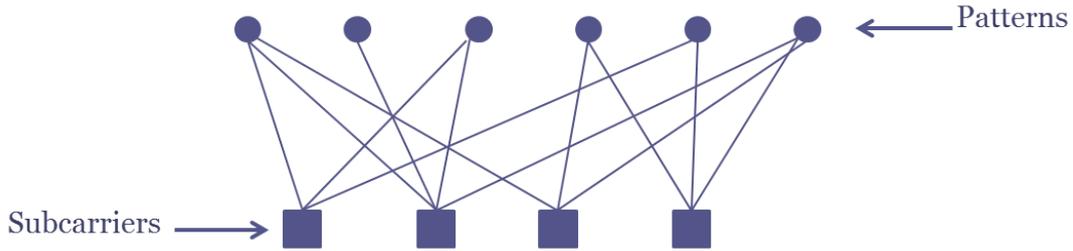
introduced in [117] with two PDMA systems. The first one is known as distributed-mapping-based PDMA and is based on non orthogonal patterns. However, The second one, called localized-mapping-based PDMA, applies quasi-orthogonal concept for pattern construction.

Authors in [116] investigated the achieved BER of PDMA scheme with the elected waveforms techniques for 5G networks such as OFDM, filter bank multi-carrier and generalized frequency division multiplexing. An analytical study of the system BER was also provided in order to justify the simulation results. In [118], an improved version of PDMA scheme using large scale antenna was proposed. The users patterns were constructed using jointly the beamforming and the power attribution aspects. Then, the multi-user detection was conducted using spatial filter and SIC receiver. Based on the reported results, the proposed PDMA scheme achieved a better performance than the conventional OMA schemes in terms of the global system sum-rate. This work was continued in [119], where two optimization problems for power allocation and beam vectors attribution were proposed in order to improve the global system sum-rate and the system connectivity, respectively. The optimal power values were obtained by the resolution of the optimization problem of the system sum-rate maximization. However, the optimal beam allocation was calculated as a solution of the minimization problem of the maximum inner product between any two beam vectors. Simulation results proved that PDMA with optimal powers and beams assignment outperformed the conventional OMA schemes and the PD-NOMA scheme in terms of the global system sum-rate.

Authors in [25] studied the performance of PDMA in terms of the achieved system sum-rate. The simulation results proved that PDMA with SIC receiver outperformed the conventional OMA schemes. In [120], the system outage probability as well as the global system throughput were analyzed for uplink transmissions. An analytical study of these metrics were also provided by calculating their closed-form expressions. The performed results in this paper showed also that the uplink PDMA achieved a better outage probability and global sum-rate compared to OMA schemes while considering different user target throughput.

### 2.2.6.2 Multi-user detection

Regarding multi-user detection techniques, PDMA scheme may be realized with SIC or MPA algorithms. However, unlike the SCMA scheme, the number of non-zero elements in PDMA patterns is variable depending on the random elements selection. Therefore, in this case, PDMA patterns can be considered as a kind of the irregular LDPC matrix [121]. Hence, considering MPA receiver, the number of connected FNs to each each VN may be different, as seen in Figure 2.8. The patterns irregularity would increase the users diversity, but at the cost of higher MPA complexity compared to the SCMA codewords, especially for a massive number of connected devices. Using SIC receiver, the decoding process is similar to that of MUSA, as already explained in the section 2.2.5.3. However, the sparsity feature of user patterns may reduce users diversity and increase the error propagation problem. The selection of the adequate receiver depends particularly on the desired trade-off between the decoding complexity and the system performance.



**Fig. 2.8.** MPA factor graph for PDMA scheme with  $J = 6$  and  $K = 4$ .

### 2.2.6.3 Conclusion

PDMA was proposed as a candidate scheme for 5G networks. User signals are differentiated based on their different patterns. Each user can select randomly and autonomously its binary pattern. Therefore, PDMA can benefit from a grant free access, which reduces the signaling overhead, the transmission delay and the energy consumption. Regarding multi-user detection, PDMA may be conducted with SIC or MPA receivers while taking advantage from the pattern sparsity and users diversity to improve the decoding process using MPA.

## 2.2.7 Interleave division multiple access

IDMA is one of the early proposed NOMA schemes, more than fifteen years ago, in order to improve the performance of asynchronous CDMA [26], [122]. Therefore, it can be seen as a kind of CDMA, but users are differentiated using different interleaver instead of different spreading sequences, as illustrated in Figure 2.9. The key idea of IDMA is to disperse the coded data sequences by chip-level interleaver, which results in approximately uncorrelated adjacent chips and thereby simpler chip-by-chip decoding process. Similarly to CDMA, IDMA shows also a robust resistance against the channel fading and it permits to alleviate the inter-cell interference.

User interleavers should be created randomly and independently. In addition, they should be long, i.e., the length of interleaver is equal to the number of coded bits, which offers a large set of possible interleavers and therefore a high overload capacity, but at the cost of high decoding complexity.

Regarding multi-user detection, IDMA signals are treated with an iterative sub-optimal receiver, known as ESE-DEC, which contains two sub-systems; an elementary signal estimator and  $J$  decoders [26]. More specifically, the elementary signal estimator handles the multiple access issue, whereas the decoders blocks treat the coding issue. These two functions provide the extrinsic LLRs about the transmitted chips as output. In fact, the detection process is performed chip by chip through an iterative operation between the elementary signal estimator block and the decoder blocks. This receiver offers a robust decoding performance, but at the cost of very high complexity due the large number of iterations. The receiver complexity increases linearly with the number of users.

According to the published results in [26], the convolutionally coded IDMA scheme with AWGN channel can support up to 64 users while achieving a BER of  $10^{-4}$  at  $\frac{E_b}{N_0} = 8$  dB, which offers a high overload capacity. However, this re-

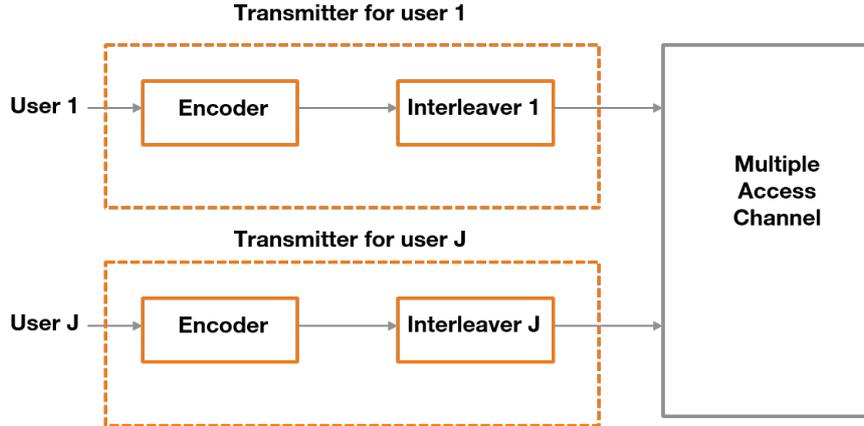


Fig. 2.9. IDMA transmitter for  $J$  users.

sults is obtained for 50 iterations for multi-user detection, inducing a high decoding complexity. The IDMA system has been re-evaluated in recent years, where several works have focused on improving its performance. Authors in [123] focused on the design of user interleavers. They proposed a new approach permitting to design the whole system interleavers instead of focusing of the design of single interleaver as in previous work [124]. This approach permits also to reduce the complexity of interleavers implementation. In [125], a performance comparison of IDMA and FDMA in the context of MTC scenarios was provided while considering two multi-carrier waveforms : OFDMA and universal filtered multi-carrier. In addition, in [126], IDMA was compared to its simplified variant which is known as resource spreading multiple access. The realized simulations under different scenarios, such as fading conditions and inexact channel estimation, proved that IDMA achieved the best performance, whereas the resource spreading multiple access scheme maintained a lower system complexity.

Despite its main advantages of high capacity of connectivity, robust decoding performance, robustness against fading, and high spectral efficiency, IDMA suffers from a very high decoding complexity. In fact, the applied ESE-DEC receiver needs a large number of iterations to converge, inducing high transmission latency. The number of iterations and therefore the system complexity and the transmission latency increase with the number of supported users.

## 2.3 Qualitative comparison

The investigated NOMA schemes are competing to meet the main mMTC challenges namely, massive connection density, low system complexity and low energy consumption. They apply different multiplexing techniques, summed up in Table 2.1. In addition, these schemes may be carried out in association with different multi-user detection techniques which leads to unequal decoding performance and complexity. As a consequence, each scheme has its own characteristics, advantages and limitations.

In this part, we aim at providing a qualitative comparison of the studied NOMA

NOMA scheme	Multiplexing techniques
PD-NOMA	Non-uniform power distribution
SCMA	Sparse, multi-dimensional and complex codebook
MUSA	Low correlation, complex and random spreading sequences
PDMA	Sparse and binary spreading sequences
IDMA	Random and independent interleavers

TABLE 2.1: Main characteristics of different NOMA schemes

schemes w.r.t the mMTC requirements regarding different performance axes namely, decoding complexity, resource design, overload capacity and adapted scenarios.

Table 2.2 provides an overview of the stressed receivers in the literature that can be used with the cited NOMA schemes. It should be mentioned that this does not mean that they are the only applicable receivers with these schemes. In fact, each scheme may be conducted with other techniques, but at the cost of different decoding performance and complexity. For instance, the MPA receiver can be applied with MUSA scheme but at the cost of very high decoding complexity due the random strategy of sequences selection leading to a large and variable number of non-zero elements. As a consequence, it is not recommended to apply such receivers with MUSA. However, the belief propagation receivers namely, MPA and EPA, can not be applied with certain schemes such as PD-NOMA because users symbols are not spread over multiple subcarriers and thus, the system can not be modeled by a factor graph.

NOMA scheme	SIC	MPA	Jointed SIC-MPA	CS-MPA	EPA	ESE-DEC	ML
SCMA	✓	✓	✓	✓	✓		✓
MUSA	✓			✓			✓
PDMA	✓	✓					
PD-NOMA	✓						✓
IDMA						✓	

TABLE 2.2: The applicable receivers for the studied NOMA schemes

### 2.3.1 Receivers complexity

This part consists in comparing the complexity of the typical receivers for the investigated NOMA schemes as summarized in Table 2.3.

The optimal ML receiver is considered as the baseline for comparison of decoding complexity and performance. It offers an optimal decoding process, but at the expense of high complexity of  $O(|\mathcal{X}|^J)$  where  $|\mathcal{X}|$  is the constellation cardinality and  $J$  is the number of active users [127], [128]. The ML complexity increases exponentially with the number of users, which may be difficult to handle especially with a massive number of users.

The MPA algorithm is a near-optimal receiver with a lower complexity than ML, i.e.,  $O(TK|\mathcal{C}|^{d_f})$ , where  $T$  is the number of iterations,  $K$  is the number of resource

### 2.3. QUALITATIVE COMPARISON

---

elements,  $d_f$  is the maximum number of users transmitting over each resource element and  $|\mathcal{C}|$  is the number of possible codewords in each codebook [94], [128]. For a massive number of users  $d_f$  is greater than 3, leading to an important system complexity.

However, the MMSE-SIC receiver generates a complexity of  $O(J^3)$ , mainly for the calculation of MMSE matrix, which grows logarithmically with the number of users  $J$  [129], [130]. This complexity is very low compared to MPA and ML.

Finally, the iterative ESE-DEC receiver has a robust decoding performance with a complexity of  $O(TJL)$ , where  $J$  is the number of users,  $L$  is the interleaver length and  $T$  is the number of iterations [50]. The ESE-DEC complexity evolves linearly with the number of users. However, since the number of iteration and the interleaver length are very large, this leads to high decoding complexity. In addition, the number of iterations increases with the number of user, making the decoding process more complicated.

	ML	MPA	MMSE-SIC	ESE-DEC
Complexity	$O( \mathcal{X} ^J)$	$O(TK \mathcal{C} ^{d_f})$	$O(J^3)$	$O(TJL)$

TABLE 2.3: Decoding complexity of typical NOMA receivers.

#### 2.3.2 Design complexity

Table 2.4 provides a comparison of the complexity of NOMA resources design such as power, code, spreading sequence and interleaver. For SCMA, user codebooks include a set of sparse and multidimensional codewords with a fixed number of non zero elements. The positions of this latter should different between codebooks in order to reduce the number of interfering users in the same subcarriers and simplify the decoding process. The design of SCMA codebooks is quite complex. In addition, they must be generated by the base station, in order to guarantee compliance with the defined conditions. After that, the codebooks will be assigned to the active users.

For MUSA and PDMA, users are separated by complex or binary spreading sequences, respectively. In fact, each user can select randomly and autonomously its short sequence with no coordination with the base station, ensuring a simple transmitter process with low complexity. Since the number of possible MUSA sequences is very high, the collision probability is lower than that of PDMA.

PD-NOMA uses a non uniform power distribution between users for their signals separation. The power allocation is the key aspect for PD-NOMA with a critical impact on the achieved performance. However, users are unable to decide their transmission powers autonomously because they have no information about the propagation environment and other users conditions, especially if no closed loop power control is used. Powers are then allocated by the base station to the connected users. The complexity of power allocation process increases with the number of users.

For IDMA, users are separated only by their different interleavers. Consequently, the achieved performance depends particularly on the allocated interleavers which should be designed efficiently with low cross-correlation. However, an excessive me-

mory cost is needed to conserve the interleaver indexes with high computational complexity.

NOMA scheme	Design complexity	Overload capacity	Decoding complexity
SCMA (MPA)	High	Medium	High
MUSA (SIC)	Low	High	Low
PDMA (SIC)	Low	Low	Low
PD-NOMA (SIC)	High	Low	Low
IDMA (ESE-DEC)	High	High	High

TABLE 2.4: Performances comparison of NOMA schemes

### 2.3.3 Overload capacity

NOMA schemes have different overloading abilities based on their respective multiplexing techniques and the system conditions, as summarized in Table 2.4. SCMA uses sparse multidimensional codewords for signals separation. However, user code-word design is restricted by several conditions namely, codeword sparsity and sub-carrier mapping, giving  $\binom{N}{K}$  possible codebooks. Therefore, SCMA has a moderate overload capacity, which may be insufficient for dense mMTC scenarios.

MUSA scheme enables a simple multiplexing techniques using randomly generated spreading sequences within a predefined constellation. MUSA can built a large pool of complex sequences using a small constellation set. As a consequence, MUSA scheme has great potential to support a huge number of devices with a high overload capacity. It can maintain up to 700% of overload over  $K = 4$  subcarriers using multi-path fading channel and two receiving antenna while achieving a BLER less than 0.1 for a uniformly distributed received SNR in the range of  $6 \sim 20$  dB [110]. However, the randomly selected sequences can generate high inter-user interference or users collision, which leads to an important performance degradation.

PDMA uses binary and sparse spreading sequences with unbalance regularity that permits to reduce the decoding complexity. However, the maximum number of possible binary sequences is limited to  $2^K - 1$ , inducing a moderate overload capacity.

PD-NOMA applies a non-uniform power distribution between users. Therefore, the overload capacity depends particularly on the allocated power values. The power allocation process becomes more and more complex as the number of users increases. Regarding the literature review and in order to ensure an acceptable performance with an affordable complexity, a low overload capacity of 200-300 % is recommended.

Finally, for IDMA, users are multiplexed using randomly generated interleavers. This scheme can support a high overload factor using robust low-rate channel codes, long interleaver and large number of iterations, which induces a very high decoding complexity [26].

### 2.3.4 Adapted scenarios

Regarding each NOMA scheme principle, system complexity, overload capacity and resource design, it may be recommended for particular scenarios under the context of mMTC. This is summarized in Table 2.5.

Given its limited number of codewords and its robust decoding performance, SCMA scheme is recommended in case of high reliability requirement or in small cells with a moderate number of simultaneously active users.

MUSA scheme, which has a high overload capacity but suffers from the SIC error propagation problem, is more suitable for use cases with a massive number of devices that have low reliability tolerance and are scattered over the cell, in order to ensure different received powers.

The PD-NOMA scheme, which is based on different power distributions, seems more suitable for eMBB scenario or certain mMTC use cases with sufficiently scattered users over a macro cell while experiencing different channel coefficients, high data traffic and moderate overload requirements, due to its medium overloading capacity.

The PDMA scheme can allow grant free access due to the simplicity of its sequences design, but at the cost of low overload capacity. As a consequence, this scheme seems more adaptable in case of low latency communications with a small number of active users. PDMA scheme combined with MPA receiver may also be used in certain scenarios with a relatively high reliability requirement.

Finally, the IDMA scheme is recommended in case of high connectivity demand, high reliable transmissions or high spectral efficiency that can tolerate high latency and system complexity.

Table 2.5 summarizes the main advantages and disadvantages of the studied NOMA schemes. It provides a global view of each scheme and its key features.

### 2.3.5 Conclusion

This section provided a qualitative comparison of the studied NOMA schemes in terms of decoding complexity, overload capacity, resource design and adaptable scenarios. However, this comparison is gathered from different publications under different system conditions. In fact, to the best of our knowledge, no work has provided a performance comparison of these schemes with respect to the main mMTC requirements while considering the same system settings. Therefore, in order to fulfill this shortage, a quantitative comparison of these NOMA schemes under the same conditions will be provided in the next chapter in the interest of selecting the most suitable one to meet the mMTC challenges.

## 2.4 Grant free access

### 2.4.1 Random access

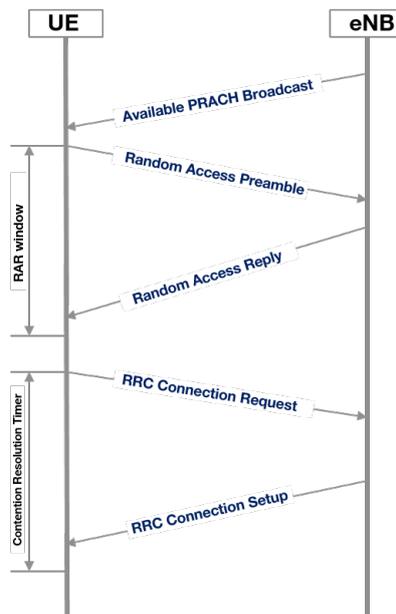
In the existing network technologies with the conventional OMA schemes, each user has to go through a resource allocation process over the orthogonal time and frequency resources. For instance, in current LTE and LTE-A network, each user uses a contention-based random access (RA) protocol for data transmission [131], [132] which may be realized on two phases : the random access channel (RACH) coordination process and the grant acquisition (GA) request. RACH consists in connecting UE with the base station. It can be triggered for different reasons such as, initial user

NOMA scheme	Advantages	Drawbacks	Compatible scenario
SCMA	<ul style="list-style-type: none"> <li>▶ High reliability</li> <li>▶ Multi-dimensional constellation shaping gain</li> <li>▶ Low inter-user interference</li> <li>▶ Sparse user codewords</li> <li>▶ Robust multi-user detection</li> </ul>	<ul style="list-style-type: none"> <li>▶ Limited overload capacity</li> <li>▶ High decoding complexity</li> <li>▶ High codebooks design complexity</li> <li>▶ High signaling overhead</li> </ul>	<ul style="list-style-type: none"> <li>▶ Small cell with limited number of active users</li> <li>▶ High reliability</li> </ul>
MUSA	<ul style="list-style-type: none"> <li>▶ Grant free access scheme</li> <li>▶ Low latency</li> <li>▶ High overload capacity</li> <li>▶ Low signaling overhead</li> <li>▶ Robust blind detection</li> <li>▶ Low energy consumption</li> <li>▶ Low system complexity</li> <li>▶ Simple spreading sequences design</li> </ul>	<ul style="list-style-type: none"> <li>▶ SIC error propagation problem</li> <li>▶ SIC latency increases with the number of users</li> <li>▶ Users received powers should be different</li> </ul>	<ul style="list-style-type: none"> <li>▶ Macro cell with large number of scattered users</li> <li>▶ Reliability tolerance scenario</li> </ul>
PDMA	<ul style="list-style-type: none"> <li>▶ Grant free access</li> <li>▶ Simple binary spreading sequences design</li> <li>▶ Low latency</li> <li>▶ Low signaling overhead</li> <li>▶ Low energy consumption</li> <li>▶ Multiple multiplexing domains : Code, power and space</li> </ul>	<ul style="list-style-type: none"> <li>▶ Low overload capacity</li> <li>▶ SIC error propagation problem</li> <li>▶ Irregular sparsity increases MPA complexity</li> </ul>	<ul style="list-style-type: none"> <li>▶ Scenarios with relatively high data rate</li> <li>▶ Low latency and grant free access scenario with limited overload requirement</li> <li>▶ Scenario with relatively high reliability</li> </ul>
PD-NOMA	<ul style="list-style-type: none"> <li>▶ Simple multiplexing principle</li> <li>▶ Low decoding complexity</li> </ul>	<ul style="list-style-type: none"> <li>▶ Low resources overload</li> <li>▶ High signaling overhead</li> <li>▶ SIC error propagation problem</li> </ul>	<ul style="list-style-type: none"> <li>▶ Scattered users over a macro cell with different channel conditions</li> <li>▶ High data traffic scenarios</li> </ul>
IDMA	<ul style="list-style-type: none"> <li>▶ High overload capacity</li> <li>▶ High spectral efficiency</li> <li>▶ Robust decoding performance</li> <li>▶ Robustness against fading</li> </ul>	<ul style="list-style-type: none"> <li>▶ High decoding complexity</li> <li>▶ High decoding latency</li> </ul>	<ul style="list-style-type: none"> <li>▶ High connectivity scenario with high transmission reliability</li> <li>▶ High spectral efficiency scenario</li> <li>▶ High latency and complexity tolerance</li> </ul>

TABLE 2.5: Advantages, drawbacks and adapted scenarios of NOMA schemes

access from idle mode, radio resource control (RRC) connection re-establishment, handover, downlink or uplink data arrival with expired timing advance. RACH process is carried out over the physical random access channel (PRACH). The information about the available PRACH resources is broadcast to all users by the base station. The RACH process can be then performed through four handshake steps [133], as seen in Figure 2.10 :

1. **Random access preamble** : For each access attempt, each user has to manifest at the next RA slot, using ALOHA or slotted-ALOHA protocols, and transmit its selected preamble over a PRACH channel. This latter is picked up uniformly random within a set of 64 orthogonal and contention-based preambles, i.e., users digital signatures. Given their orthogonality, users preambles will be successfully separated at the base station. However, if the



**Fig. 2.10.** RACH process for LTE/LTE-A networks.

same preamble is selected by multiple users, the base station will not be able to separate the signals. The preamble duration is between 1 ms and 3 ms according to the cell size. A waiting time is considered by each user after the transmission of the preamble where the response from the base station must be received within this time. Otherwise, if this time is exceeded without any response from the base station or the reception of an incorrect response that is not intended for that user, it decides to retransmit another preamble. The waiting time varies between 2 and 10 ms and is determined by the base station in a broadcast message to all users.

2. **Random access reply :** For a successfully detected preamble, the base station computes an identifier called random access radio network temporary identifier (RA-RNTI) which refers to the used RA slot for preamble transmission [134]. The random access response (RAR) message is carried out over the downlink shared channel PDSCH. It may include these items :
  - ▶ Unique identifier for the decoded preamble
  - ▶ Timing alignment for device synchronization in uplink direction
  - ▶ Dedicated uplink resources to be used by the device in the next RACH step
  - ▶ Temporary identifier called cell radio network temporary identifier (C-RNTI)
  - ▶ Optional back-off indicator which may be used in case of failure to ask the devices to wait for an extra random time value before the retransmission of a new preamble. This may serve to optimize the collision probability between the selected preambles, but at the cost of a higher transmission latency

All users transmitting preambles in the same RA slots are associated to the same RA-RNTI identifier. However, their received RAR messages include specific sub-headers to their own transmitted preambles. User RAR message, with RA-RNTI identifier corresponding to its used RA slot, may sometimes not include the correct identifier of its transmitted preamble. In this case, the user applies a random back-off time, based on the received back-off indicator, before retransmitting a new preamble in the next RA slot. In addition, due to the random selection, the same preamble may be chosen by multiple devices in the same RA slot leading to a collision between them with two possible scenarios : i) if the arrival times at the base station are different, the collision can be easily detected. As a consequence, the RAR message does not carry any information about the transmitted preamble. In this case, user picks a random back-off time and then retransmit a new preamble, as illustrated in Figure 2.11 ; ii) if the devices are equidistant from the base station, the transmitted preambles arrive at the same time, which may prevent the base station from detecting the collision. As a result, devices receive the same RAR message. In this case, the conflict will continue to the next step, i.e., RRC connection request. Therefore, the user must proceed to a collision resolution process.

3. **RRC connection request** : We consider here the random access process for the connection establishment. After a successfully received RAR message, each user launches an RRC connection request by transmitting its C-RNTI identifier using the dedicated resource received in the second RACH step. For an initial user access, this message carries also the access trigger.

If the preamble collision was not detected in the second RAR step, the user connection requests are then transmitted on the same resources, which leads to new message collisions at the base station in step 3. Therefore, users must go through a contention resolution process, as illustrated in Figure 2.12. The interested reader is referred to [134] for more details about the contention resolution process. The uncorrected users do not receive responses associated with their C-RNTI identifiers. As a consequence, after a predefined time, they consider that the procedure has failed. Then, each user schedules a new random access attempt, starting the procedure from the beginning, if it had not reached the maximum number of allowed retransmissions. More precisely, It adopts the same behavior as if it had not received a correct response to the preamble transmission, as presented in Figure 2.11.

4. **RRC connection reply** : For a successfully received connection request, the base station sends back a downlink acknowledgment with a connection setup including mainly information about the dedicated signaling radio bearers. As consequence, the user is successfully connected to the base station. However, if this message is not received, the access attempt is considered to have failed and the device must restart the entire RACH process from the beginning, as long as it has not reached the maximum number of consecutive attempts.

If the RACH process is realized successfully, the active devices are hence connected to the base station with a guaranteed uplink synchronization. Then, for each data

transmission attempt, each user has to initiate a GA process for uplink resource allocation [62] :

- ▶ Scheduling request : In order to transmit its available data, each user firstly sends a scheduling request to the base station over PUCCH channel requiring for PUSCH resources. The scheduling request is carried only on one bit to inform the base station that there is data to transmit.
- ▶ Uplink grant : After a successful detection of the scheduling request, and for the sake of optimization, the base station sends a small amount of PUSCH resources to the user in order to transmit its buffer status report (BSR) indicating the amount of its available data.
- ▶ BSR transmission : The BSR is the key information for data transmission process. At this step, using the already allocated PUSCH resources, each user transmits its BSR to the base station.
- ▶ Scheduling grant : The base station allocates the sufficient PUSCH resources to each user to transmit its available data.

In case of proceeding data transmission, the user has already allocated PUSCH resources. Therefore, in order to continue its data transmission, it transmits its BSR message directly. However, in the case of connected user to the base station without PUCCH resources, if this user has an available data to transmit, it initiates a RACH process. Then, once it receives the RAR message successfully, it receives the BSR at the third message rather than the RRC connection request.

### 2.4.2 Limitation of random access process

The random access protocol including the RACH and GA processes offers a flexible access to the available network resources. It may be suitable for LTE network and certain MTC use cases. However, for mMTC capability with its stringent requirements, the random access protocol generates a lot of critical issues namely, limited access capacity, high transmission latency, high signaling overhead and thereby reduced battery lifetime which leads to performance bottleneck. Furthermore, the contention-based RA applies Slotted-ALOHA protocol, by transmitting data at the first available opportunity, which induces a high system collision for mMTC scenarios

Many works have focused on the LTE access capacity with contention-based RA protocol [135], [136]. This latter offers a resource access session every 5 ms. Each device can select its preamble among a set of 64 ones. Therefore, in the ideal system conditions with no collision between devices, a maximum of 200 access opportunities are possible during one second with a total of 12,800 preambles [131], [135].

For mMTC scenario, the number of users wishing to simultaneously transmit their data may be more important than the offered access opportunities, leading to a large collision probability, and thus a performance degradation. Furthermore, the signaling overhead is one of the critical limitation of RA protocol. This process uses significant resources for the connection establishment for the transmission of

small data. For instance, an uplink transmission of 100 bytes, while considering RA process, security procedures, data transmission and the release of connection at the end, generates a signaling overhead of 59 bytes in uplink and 136 bytes in downlink [35]. This induces significant waste of resources, high energy consumption and shorter battery life for the transmission of small packets which are typical in many mMTC use cases.

Moreover, certain mMTC applications with less tolerance to latency may experience an unacceptable high communications latency which significantly deteriorates their QoS. In fact, a large number of simultaneous connections may imply resources congestion and thereby increase the collision probability. Devices will then initiate retransmission attempts which generate further delay, resources waste and make the problem more severe. Figure 2.13 evaluates the sources of delay in LTE system, where TTI is the data transmit time interval, SP is the signal processing time and PR is packet retransmission delay using HARQ technique. We observe that in the ideal system condition, RA is the main latency cause, during approximately 9.5 ms, compared to other operations, which would increase significantly in case of collision. In short, the radio resource access may be a source of performance bottleneck in mMTC scenario, and thus is a critical issue.

### 2.4.3 Grant free access evolution

Regarding the shortcomings of the classical grant-based access protocols with OMA schemes, the 5G stakeholders from industry and academia have made a huge effort in order to propose several enhancements and new solutions. Among them, grant-free access with NOMA has recently caught the attention of the scientific community as a promising solution for the future mMTC applications. Two possible options are entailed under the grant free access umbrella whether performing or not the RACH process. The candidate resource access strategies for 5G network are represented as follows [21], [22] :

1. **RACH-based and grant-based OMA schemes** : The conventional OMA schemes with the random access strategy, i.e., 4-step RACH and GA process, is considered as the basic uplink transmissions system LTE and 5G NR. For the sake of optimization, many improvements have been proposed to overcome the system congestion and transmission latency limitations. Access class barring mechanism [131] is one of the suggested solutions. It aims to reduce the congestion probability by controlling the access process to radio interface. Services are identified in 16 different classes with different probabilities factors. Then, for each access attempt, the device has to pick up a random variable. The latter must be less than the probability factor to launch an access session, otherwise, a random back-off time is used by the device to postpone its access. However, in case of a severe congestion, the probability factor is reduced significantly which may increase the transmission latency. Some other proposed improvements consist of dynamic or pre-configured allocation of RA slots.

Regarding the diversity of 5G services with different requirements, the single LTE RACH timeline may not be suitable for all use cases. Therefore, 3GPP proposed different RACH message timelines in order to address the different

latency exigencies [137]. Moreover, in order to minimize the congestion probability, multiple preamble formats with different lengths, i.e., short and long preambles, should be considered. For that reason, the first phase of 5G networks contains 13 different preamble formats [42].

Recently, a new minimized RACH process with only two steps is approved in the 3GPP Release 16, for the second phase of 5G networks, particularly for MTC scenarios [62], [63], [138], as illustrated in Figure 2.14. It permits to reduce the transmission latency and the signaling overhead. This solution is interesting for the cases of asynchronous devices with no need for timing advance, sporadic transmission of short packets and transmissions over unlicensed spectrum based on listen-before-talk aspect. The first message of the 2-step RACH, known as msgA, combines the preamble and payload transmissions supported in the first and the third messages of the classical 4-step RACH process, respectively. More precisely, msgA consists of transmitting a preamble over a PRACH channel and a payload over a PUSCH channel. Then, the second message, msgB, combines the second and the fourth messages of the 4-step RACH. However, the base station may have four different actions depending on the reception of msgA [63] :

- ▶ **Action 1** : When both of preamble and transmitted payload are successfully detected, the base station sends back a contention resolution message including the RAR message and the corresponding timing advance value.
- ▶ **Action 2** : A single preamble is received successfully and the payload detection fails. In this case, the msgB contains a fallback RAR requesting the UE to retransmit the payload through the third message of the 4-step RA process, i.e. the RRC connection request. Therefore, the msgB includes the corresponding timing advance and the uplink allocation of PUSCH resources for the payload transmission.
- ▶ **Action 3** : The base station detects the same preamble for multiple users. Therefore, the msgB contains a back-off indicator to be applied before retransmitting the msgA.
- ▶ **Action 4** : The base station can not detect the preamble. The user waits for a period of time before retransmitting the msgA. After  $n$  failed retransmission attempts of msgA, the user switches to the classic 4-step RACH process.

Several optimizations of the random access process with OMA schemes have been approved in releases 15 and 16 of 3GPP standardization of 5G network, such as different preamble formats and 2-step RACH, in order to reduce the signaling overhead, the access latency and the energy consumption. However, regarding mMTC scenarios with an extreme connection density of low cost devices, intermittent transmissions and very short packets, these solutions may not be the most optimal ones, especially in terms of network capacity. In this context, some other solutions have been ongoing which have gained a lot of interest from the scientific community, such as NOMA schemes and grant free access. The key idea of uplink grant free access is to transmit data without

going through the process of the dynamic resource scheduling with the base station.

2. **RACH-based and grant-based NOMA schemes :** NOMA schemes can be applied with RA protocol, 4-step RACH or 2-step RACH, while allowing multiple users to select the same preamble over the same RA slot. Then, the base station will be able to detect them based on the arrival times and/or user signatures. In addition, the number of users using the same preamble can be calculated using power control methods. The new NOMA RA access strategy improves the uplink channel capacity as well as the user throughput, but no reduction in system latency or signaling overload is guaranteed compared to OMA schemes [22].
3. **RACH-based and grant-free NOMA schemes :** The first option of grant free access consists of omitting only the GA process for devices already connected to the base station. Firstly, the device initiates a RACH process in order to ensure its alignment with the base station as well as its uplink synchronization. After that, for each data transmission attempt, it transmits its data directly to the base station without going through the resource allocation process. This option permits to cut down the signaling overhead, the latency and the energy consumption generated by the scheduling process. It must be pointed out that this option has never been possible with OMA schemes because grant free access may cause a severe collision due to the limited number of resources.
4. **RACH-less and grant-free NOMA schemes :** This option is an important evolution for resource access strategies. It reveals a total grant free access behaviour by omitting both of the RACH and GA processes. In other words, each device transmits its data as per its need without any beforehand communication with the base station, but at the cost of asynchronous uplink transmissions. However, the downlink synchronization would be somehow guaranteed. This option eliminates the signaling overhead, delay, power consumption and resources waste due to random access, i.e., RACH and GA. In addition, this strategy can be beneficial for small packet transmissions where the amount of usually exchanged signaling overhead may be more important than the transmitted data. The uplink transmission is carried over statically or semi-statically assigned resources for uplink RACH-less and grant free communications. In case of congestion, the device simply retransmits its data. However, for very bad system conditions with a serious congestion problem, the dedicated resources will be adapted.

## 2.5 NOMA standardization

The non orthogonal aspect was early introduced in the 3GPP Release 14 for LTE network as multi-user superposition transmission (MUST) scheme [139], which focuses mainly on downlink transmissions [140]. This scheme proved an important improvement in terms of the system throughput compared to the conventional OMA schemes. So far, several NOMA schemes with different multiplexing techniques have

been proposed by different companies. The 3GPP Release 15 contained the initial study of grant-based resource allocation NOMA schemes with synchronous transmission [140], [141]. Then, a study item was involved in the 3GPP Release 16 in December 2018, which focused on the investigation of the essence of uplink NOMA schemes with grant free access, synchronous transmission with 4-step RACH and asynchronous transmission with RACH-less or 2-step RACH [50]. This study entails an overview of the transmitter schemes, the candidate receivers with a complexity analysis and the possible channel structures. Furthermore, an extensive evaluation of link-level and system-level performances of the proposed NOMA schemes was provided while considering realistic conditions with near-far effect, asynchronous transmissions, frequency offset and signature collision. In fact, about 14 companies contributed in this study by providing link-level simulations in terms of BLER, for the different transmission schemes and the applied multi-user detection techniques, over a large number of different use cases, i.e., more than 35 use cases. The considered parameters by the different companies in their simulations are quite similar which allows their comparison. However, the provided system-level simulations by certain companies were conducted in different conditions which makes the performance comparison more complex.

Regarding the provided results in the studied use cases, NOMA schemes didn't prove a significant gain compared to OMA schemes with multi-user MIMO (MU-MIMO) systems, and sometimes they may be worse depending on the considered conditions. In addition, the use of uplink NOMA schemes in 5G NR is related to network implementation settings. More precisely, new features should be considered to support NOMA schemes such as, power control, users scheduling over shared resources, inter-user interference management and receiver complexity. For that reasons, NOMA schemes didn't show a worthwhile gain compared to OMA schemes with MU-MIMO techniques to be considered in 5G NR [51], [142]. Regarding the requirements of the currently supported 5G services, the conventional OMA is considered as the basic scheme for 5G networks in uplink and downlink transmissions. After the conclusion of NOMA study item, 2-step RACH strategy was approved in 3GPP work item for the second phase of 5G NR. Therefore, certain issues, included in NOMA study item, such as the channel structures should be further investigated for this access strategy [51].

The key feature of NOMA schemes is to increase the network connection density compared to OMA schemes, but at the cost of higher system complexity, latency and thus power consumption. Therefore, NOMA is left for beyond 5G scenarios, particularly the mMTC services with massive number of connection requests with great interest to apply these schemes. In fact, based on the reported results in the study item, asynchronous grant-free NOMA schemes showed a significant gain in terms of transmission latency, signaling overhead and power consumption compared to grant-based OMA, especially for low data rate. As a consequence, a work item concerning the standardization of MAC layer for short packet with asynchronous grant free was approved in the 3GPP Release 17 [51]. Among the remaining issues, robust receivers are needed to treat the asynchronous transmission with different frequency and timing offsets. In addition, the signature collision problem should be handled while providing a compromise between signaling overhead, receiver com-

plexity, power consumption and achieved performance.

## 2.6 Conclusion

NOMA schemes have been promoted as a promising solution to satisfy the massive connectivity of the mMTC services. This chapter provided an extensive study of some schemes while presenting the main principle of each technique as well as its corresponding signal model. Furthermore, it involved an exhaustive qualitative comparison of their performance regarding the stringent mMTC challenges based on the literature review. This chapter stressed also the radio resource access issue. It presented the main limitations of grant-based access strategy. Moreover, it introduced the evolution steps towards the grant free access option which is considered as an alternative strategy in order to minimize the signaling overhead, the transmission latency and the power consumption, which is highly desirable in particular for very short packets.

Several NOMA schemes have been investigated in the recent 3GPP releases. A comprehensive study of the candidate transmission schemes, typical receivers, decoding complexity and proposed channel structures is represented. In addition, a performance evaluation of the essence of NOMA schemes in terms of link-level and system level is provided with grant free access strategy. According to the reported results, NOMA schemes didn't prove a significant gain compared to OMA schemes with MU-MIMO techniques in the current 5G network. In addition, NOMA schemes require new network specifications, such as power control functionality, inter-user interference processing and resource sharing, inducing high system complexity. The new 2-step RACH strategy is approved for the second 5G phase. However, NOMA schemes and grant free access features are left for future works considering mMTC scenarios with great motivations for these concepts.

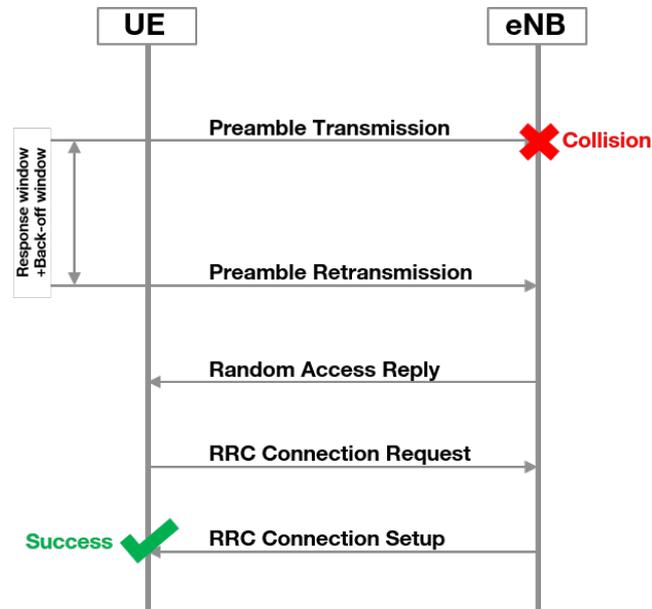


Fig. 2.11. Preamble collision resolution

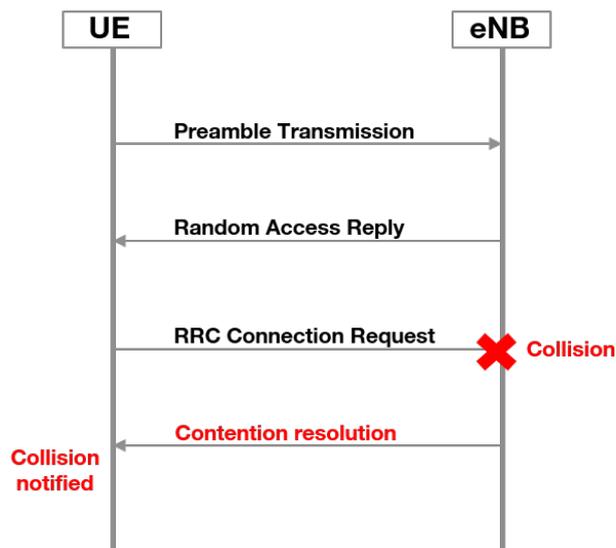


Fig. 2.12. Collision in step 3 of the random access process.

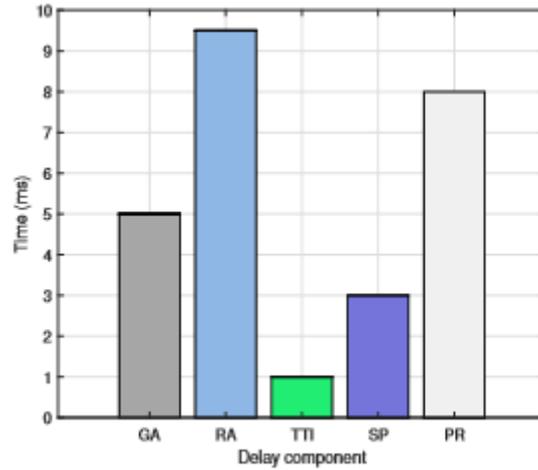


Fig. 2.13. The sources of delay in LTE system [21]

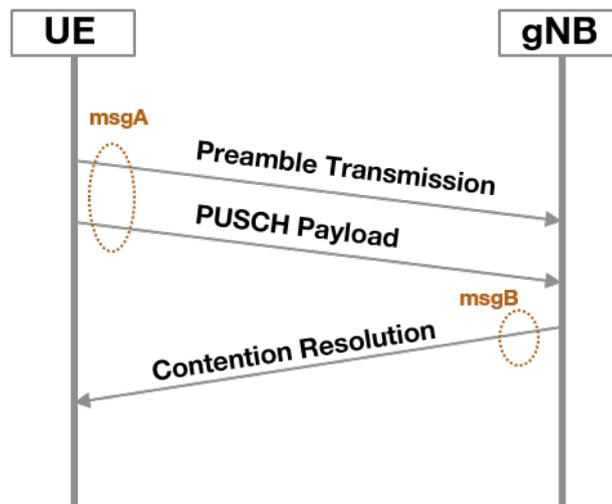


Fig. 2.14. 2-step RACH.

# NOMA schemes evaluation

## 3.1 Introduction

In chapter 2, the most popular NOMA techniques have been introduced which are, for most of them, either an improvement of the conventional CDMA with particular code design, or power-based techniques. Among the presented schemes, some of them have interesting characteristics regarding the mMTC requirements, namely SCMA [23], MUSA [24], PDMA [37] and PD-NOMA [36]. These schemes have been widely investigated and qualitatively compared in the previous chapter regarding the mMTC requirements and based on the literature review. However, this comparison is gathered from different works under different system conditions. In fact, most of published works mainly focused on the performance analysis of a single NOMA technique, only a few of them provided a simultaneous assessment of multiple schemes, which makes the relative performance evaluation of the schemes of interest more complicated.

For instance, authors in [104] provided a link-level performance evaluation of SCMA, MUSA and PDMA, but without any specific information about the used SCMA codebooks or the MUSA spreading sequences. The reported results are conducted with QPSK modulation, 150% of system overload and Rayleigh fading channel. SCMA and PDMA signals are decoded using MPA, whereas MUSA users are detected with SIC receiver. In these conditions, SCMA has the best performance, followed by PDMA and then MUSA. These results are also exploited in [143] to be compared with PD-NOMA performance. However, this paper does not provide a detailed description of system settings such as the power distribution for PD-NOMA application, which makes the performance analysis of these results unclear. In short, this paper summarizes the main principles, challenges and performance assessment of these schemes which are aggregated from different works.

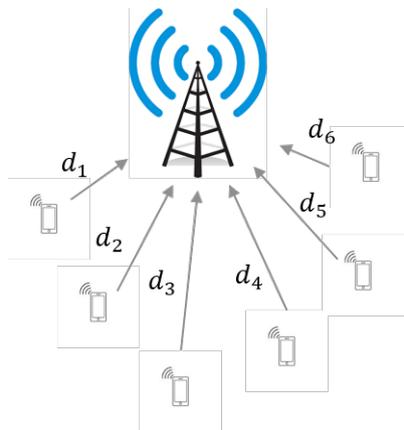
In [144], the performances of SCMA, MUSA, IDMA and OFDMA are compared in terms of BLER with different overload factors of 150%, 200% and 300% using QPSK modulation. Moreover, Rayleigh fading channels, turbo code with different rates and a transmission power of 24 dBm are used in these simulations. All the studied schemes have an equal average SNR. It can be deduced that SCMA is more resistant to rate variations and the system overload. IDMA scheme achieves

the best performance for resource overload of 150% and 200% at a low code rate of 0.2, which coincides with SCMA performance for the same conditions. However, for 200% or higher of resource overload with a moderate code rate, e.g., 0.4, IDMA experiences an error floor of at least  $8.10^{-2}$  due to error propagation. MUSA achieves a good performance under an overload factors of 150% and 200% with code rates of 0.2 or 0.4. However, for 300% of overload, a low code rate of 0.2 is necessary to alleviate the error propagation issue and achieve a low BLER value of  $10^{-2}$  at SNR=15 dB. Authors in [145], studied the BLER of SCMA, MUSA and PDMA for 150% and 200% of overload factors and different code rates. Simulations are realized with QPSK or 16-QAM modulation, two reception antennas, LTE turbo code and multipath fading channel. The reported results proved that all NOMA schemes have almost the same performances for the considered conditions.

However, we figure out that no work has studied the selected NOMA schemes, i.e., MUSA, SCMA, PDMA and PD-NOMA, together in the same system settings. In this chapter, we aim to validate the qualitative study reported in chapter 2 by providing a quantitative performance comparison for the mentioned NOMA schemes in the same system conditions. The main purpose is to select the most promising schemes for our study which achieve the best performance in terms of BER considering the mMTC requirements.

The rest of this chapter is organized as follows. A performance comparison of the investigated schemes is introduced in Section 3.2 while defining the considered system settings. Then, SCMA performance with different multi-user detection techniques is provided in Section 3.3. Finally, conclusions and on-going researches are drawn in Section 3.4.

### 3.2 Quantitative comparison



**Fig. 3.1.** Users transmission scenario for  $J = 6$  users.

A multiple access communications system of  $J$  users overlapped over  $K$  orthogonal subcarriers is considered. We investigate two different scenarios for users distribution over the cell : i) Users are equidistant from the base station with normalized path losses ; ii) Users are uniformly scattered over the cell at different distances from

### 3.2. QUANTITATIVE COMPARISON

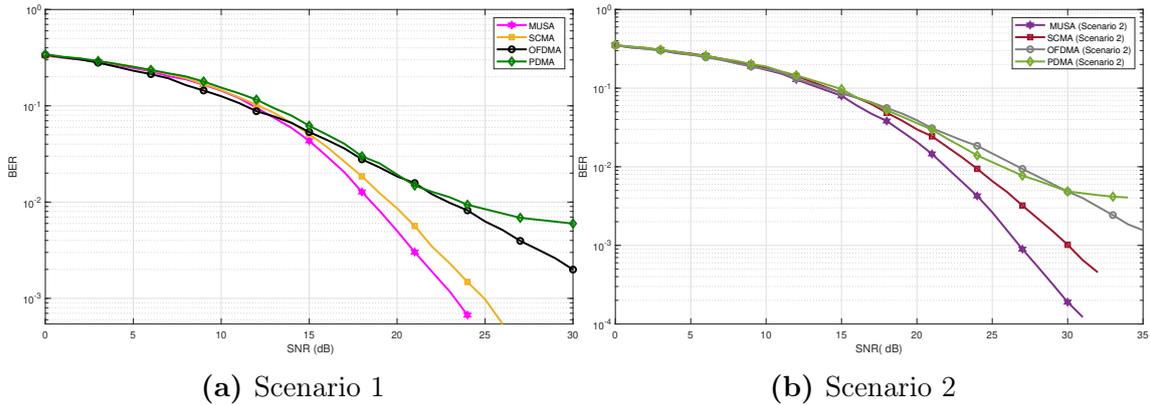
Channel model	Rayleigh fading
Channel estimation	Ideal
Antenna configuration	1Tx / 1Rx
Multiple access signatures	MUSA : Complex spreading sequences SCMA : Sparse codewords PDMA : Binary spreading sequences PD-NOMA : Channel coefficients
Number of users	$J = 4$ (100%) $J = 6$ (150%) and $J = 10$ (200%)
Number of subcarriers	$K = 4$ (100% and 150%) $K = 5$ (200%)
Overload factor	100% (SCMA, MUSA, PDMA and OFDMA) 150% (SCMA, MUSA, PDMA) 200% (SCMA, MUSA, PDMA and PD-NOMA)
Modulation	QPSK (MUSA, PDMA, OFDMA, PD-NOMA) Sparse codewords (SCMA)
Distance	Scenario 1 : Normalized path loss Scenario 2 : Users are uniformly distributed over the cell
Cell radius	3 km
Synchronization	Perfect

TABLE 3.1: Simulation settings.

the base station, leading to different path losses, as illustrated in Figure 3.1. The evaluated schemes were implemented according to the system models introduced in chapter 2 : PD-NOMA in Section 2.2.1, SCMA in Section 2.2.4, MUSA and PDMA in section 2.2.5. The simulation settings are listed in Table 3.1.

In this section, we aim to evaluate the achieved BER of SCMA, MUSA, PDMA and PD-NOMA under different resource overload factors. More precisely, we start by comparing the performances for a non overloaded system with 100% of resources occupation, using  $J = 4$  and  $K = 4$ , for SCMA, MUSA, PDMA and OFDMA, which is the equivalent of PD-NOMA scheme with only one user in each resource element. Then, we increase the overload factor to 150% using  $J = 6$  and  $K = 4$ . In this case, PD-NOMA is ignored because it is not possible to support 150% of resource capacity using this scheme. This is due to its specific multiplexing principle where one, two or more users can share the same resource element without using spreading sequences as in MUSA and PDMA or code mapping as in SCMA. At the end, the performances of SCMA, MUSA, PDMA and PD-NOMA are compared for 200% of overload factor, with  $J = 10$  and  $K = 5$ . The number of subcarriers is increased because the maximum number of possible SCMA codebooks using  $K = 4$  subcarriers is limited to 6, giving a maximum resource overload of 150%.

Unlike conventional OMA schemes, where uplink transmissions can be thought of as multiple point-to-point communications, in NOMA multiple users transmit on the same resource, resulting in multipoint-to-point communications. Therefore, the relevant statistic used to compare the different NOMA schemes is more challenging and it was the subject of several 3GPP discussions. As a consequence, several SNR definitions were proposed in [146]. According to [147], in order to compare fairly the different NOMA and OMA schemes, the total received power must be the same for all systems, and the used SNR form is calculated as the ratio between the total received power  $P_r$  and the noise power  $\sigma^2$ , i.e.,  $\text{SNR} = \frac{P_r}{\sigma^2}$ .



**Fig. 3.2.** Performance comparison of MUSA, SCMA, PDMA and OFDMA for 100% of resource occupation.

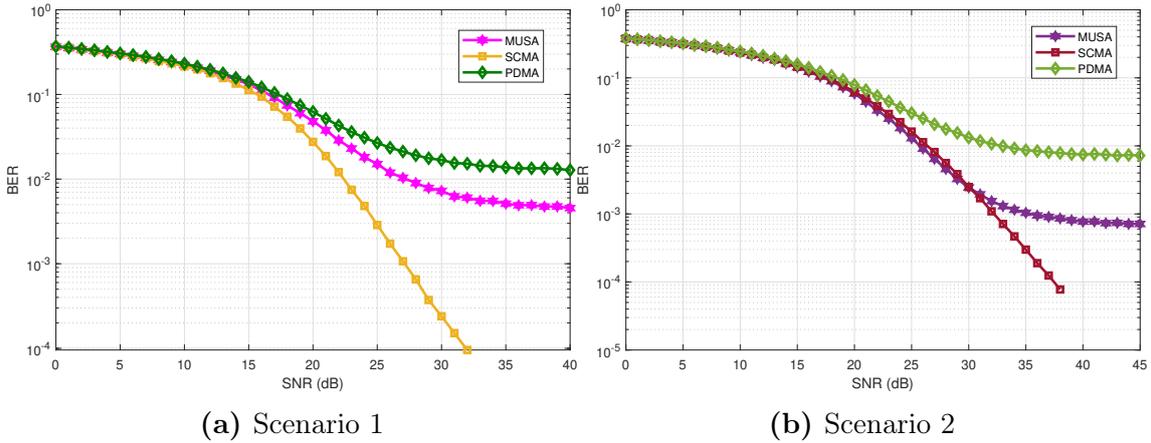
Figure 3.2 compares the performance of MUSA, SCMA, PDMA and OFDMA for a saturated network, i.e., 100% of resources occupation, in scenario 1 and scenario 2, while averaging over different network realizations. We observe that, in both scenarios, MUSA scheme has the best performance. For instance, in scenario 1, it achieves a performance gain of approximately 1.5 dB and 8 dB over SCMA and OFDMA at BER of  $2 \cdot 10^{-3}$ , respectively. In fact, the randomly generated spreading sequences and the different channel coefficients offer a high degree of diversity that permits to handle the limited inter-user interference and decode accurately user signals using the SIC receiver. However, SCMA with MPA receiver is less efficient in case of different received powers, which explains its performance degradation compared to MUSA.

For OFDMA, each user transmits over a single subcarrier. Therefore, the achievable performance is affected by the Rayleigh coefficients, which induces a performance loss compared to MUSA and SCMA where user symbols are transmitted over multiple subcarriers.

We also notice that PDMA has an error floor at BER of  $6 \cdot 10^{-3}$  in scenario 1 and  $4 \cdot 10^{-3}$  in scenario 2. This can be explained by the serious error propagation problem of the SIC receiver due to binary spreading sequences with a low degree of diversity. In addition, PDMA has a higher probability of users collision compared to MUSA and SCMA, given the limited number of possible PDMA sequences, i.e., 15 possible sequences for  $K = 4$ .

Figure 3.3a shows a performance comparison of SCMA, MUSA and PDMA in scenario 1 for 150% of network capacity. It can be observed that SCMA has the best performance thanks to its pre-allocated sparse codewords, with a reduced inter-user interference. In addition, it uses a robust multi-user detection technique, MPA, which is able to separate user signals and handle the inter-user interference through an iterative decoding process with no error propagation. However, the performance of MUSA degrades significantly w.r.t SCMA with an error floor at BER of  $4 \cdot 10^{-3}$ . This is mainly due to the error propagation problem of the SIC receiver which becomes more serious in over-saturated networks with high inter-user interference level. In this case, the selected spreading sequences with the applied channel coefficients do not provide a sufficient power difference to separate correctly user signals at high

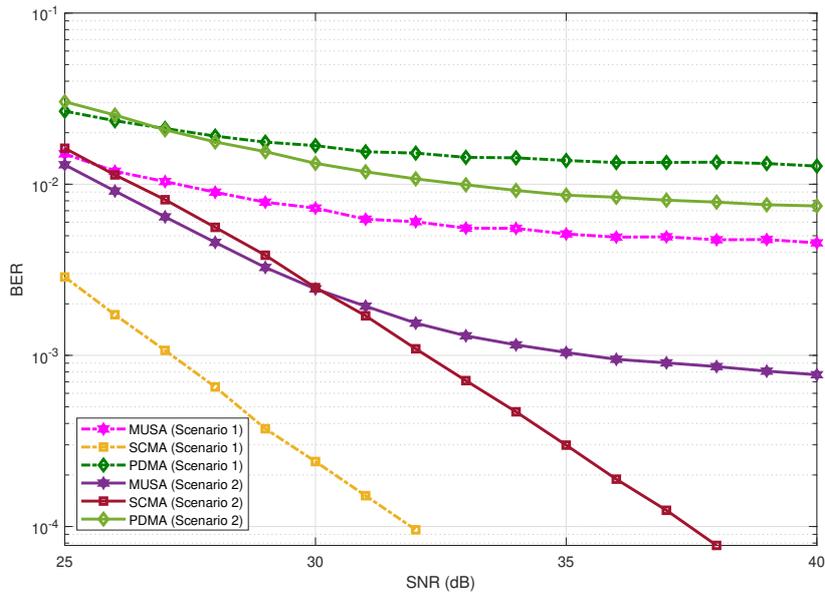
### 3.2. QUANTITATIVE COMPARISON



**Fig. 3.3.** Performance comparison of MUSA, SCMA and PDMA for 150% of resource overload.

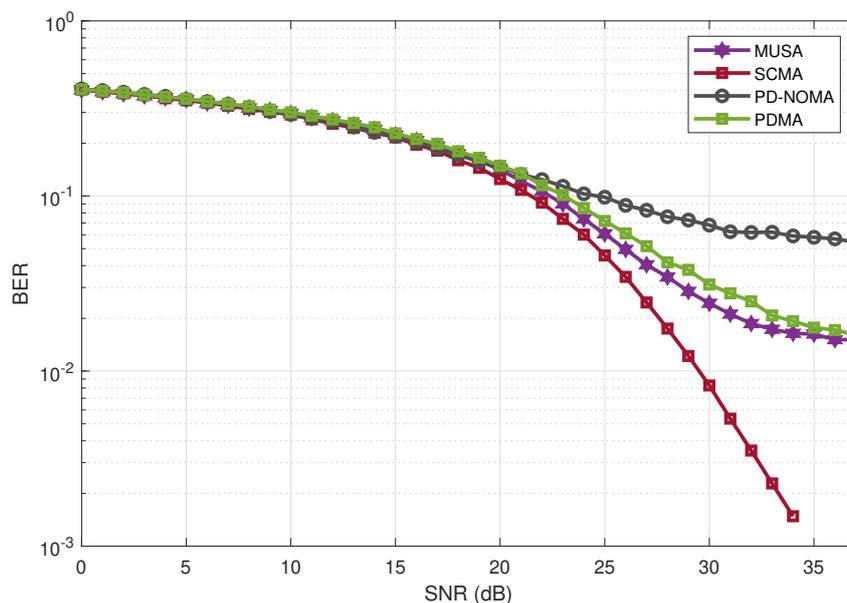
SNR values, leading to a defective decoding process. This issue is also noticed for PDMA which saturates at  $10^{-2}$ . As seen in the previous chapter, the PDMA patterns with binary elements can be considered as a subset of MUSA sequences with a lower degree of diversity and a higher probability of collision, inducing a worse signal detection using SIC receivers and thus, a performance degradation.

Figure 3.3b illustrates the achieved BER of SCMA, MUSA and PDMA in scenario 2 with 150% of resources occupation. It can be observed that PDMA and MUSA have the same behavior as in scenario 1, but with lower error floor levels, i.e.,  $7 \cdot 10^{-3}$  and  $7 \cdot 10^{-4}$ , respectively. For  $\text{SNR} \leq 30$  dB, despite its randomly generated spreading sequences, MUSA has almost the same performance as SCMA with its optimal codebooks. However, for high SNR values, i.e.,  $\text{SNR} \geq 30$  dB, SCMA has the best performance with no occurrence of error propagation phenomenon.



**Fig. 3.4.** Performance comparison of MUSA, SCMA and PDMA for 150% of resource overload in scenario 1 and scenario 2.

Figure 3.4 compares the obtained performances in both scenarios 1 and 2. It can be noticed that the SIC-based techniques, i.e., MUSA and PDMA, achieve a better performance in scenario 2 than in the case of equidistant users from the base station in scenario 1. In fact, although the experienced path losses reduce the received powers, they provide more diversity between user signals compared to scenario 1, which is the favorable conditions for SIC receivers, leading to better signals separation and thus, better system performance. However, the performance of SCMA decreases significantly in case of different path losses, e.g., a performance loss of 5 dB can be recognized at BER of  $10^{-3}$ . This can be explained by the fact that MPA is less efficient in case of different received powers and the best performance is achieved when user signals are received with equal power.



**Fig. 3.5.** Performance comparison of MUSA, SCMA, PD-NOMA and PDMA for 200% of resource overload in scenario 2.

Figure 3.5 compares the performance of MUSA, SCMA, PD-NOMA and PDMA for 200% of network capacity in scenario 2, which provides the favorable conditions for the over-saturated SIC-based schemes. It can be observed that SCMA has the best performance while achieving a BER value of  $10^{-3}$  at SNR = 32 dB. However, MUSA and PDMA both experience a severe error propagation with an error floor of  $\sim 2.10^{-2}$ . This can be explained by the increase in inter-user interference and the insufficient power difference between the received signals for a proper decoding process. Finally, PD-NOMA scheme has the worst BER level at high SNR values, i.e., an error floor of  $6.10^{-2}$ . Unlike other studied mechanisms, users in PD-NOMA are differentiated only by the experienced channel coefficients which can not guarantee a correct symbol estimation.

Regarding the two most efficient solutions SCMA and MUSA, both schemes have an acceptable performance with the excellence of SCMA regardless of the SNR value with no error propagation occurrence. Therefore, it can be noticed that SCMA appears more resistant against the variation of the propagation environment, e.g., equal or different received powers, or the increase of resource overload thanks to its

pre-allocated codebooks and robust decoding process. However, the MUSA scheme is more efficient in the case of sufficiently different received powers than in the case of close received powers while maintaining a low system complexity. But, the performance of MUSA may be affected by users collision or SIC error propagation problem at large SNR. This problem becomes more serious as the number of users increases.

SCMA and MUSA are promising solutions, but they have different trade-offs between performance and system complexity. SCMA offers the best performance, but at the cost of high system complexity. Whereas MUSA has lower system complexity with degraded performance. Since the system complexity is an important aspect for mMTC, in both uplink and downlink communications, in the following we study the performance of candidate SCMA receivers.

### 3.3 Performance study of SCMA receivers

SCMA seems an auspicious scheme with a great potential to support mMTC challenges thanks to its theoretical optimal codebooks design as well as its advanced receiver. As already said, SCMA is commonly implemented with MPA [27], [84] which is known by its robust performance in multi-user detection [105]. However, MPA experiences a high complexity due to the large number of exponential operations, which increases significantly with the number of users. mMTC use case is drawn to particularly deal with a large number of connected devices transmitting very short packets, i.e., on the order of few bytes. Therefore, high system complexity and thus, excessive power consumption are generated for the detection of a small amount of data. In this context, two simplified versions of MPA named MAX-Log-MPA [28], [148] and Log-MPA [29], inspired by the sub-optimal variations of MAP algorithm, i.e., Log-MAP and MAX-Log-MAP [30], were proposed. Log-MPA and MAX-Log-MPA algorithms operate in the logarithmic domain in order to reduce the system complexity while maintaining a satisfying performance level.

To the best of our knowledge, no work gathers in a complete study, the performances of MPA, Log-MPA and MAX-Log-MPA in an uplink SCMA scenario with additive white Gaussian noise (AWGN) and Rayleigh fading channels. Comparisons of these algorithms cannot be inferred from several articles in literature mainly because of lack of common simulation conditions or codebook design. For example, comparisons of MAX-Log-MPA with MPA or Log-MPA algorithms are provided separately in [27], [149], [150] under completely different simulations settings, e.g., the size of codebooks, the number of decoding iterations and without any precision on the codebooks themselves. Hence, a comprehensive analysis of the three algorithms is quite complicated and even impossible to deduce from the existing literature. Therefore, in this section, we aim at providing an in-depth performance comparison between the previously mentioned algorithms in AWGN and Rayleigh fading channels through extensive simulations based on the SCMA system model given in 2.2.4 in Chapter 2.

### 3.3.1 MAX-Log-MPA algorithm

MPA algorithm [27], [84], which is introduced in 2.2.4 in Chapter 2, reveals a near-optimal performance with a much lower decoding complexity compared to the optimal MAP receiver [84]. The MPA decoding process is based on a bipartite factor graph modeling the receiver. Then, the transmitted codewords are estimated through an iterative extrinsic information exchange between the FNs and the VNs representing the  $K$  system subcarriers and the  $J$  active users, respectively. However, MPA is suffering from high decoding complexity especially due to the large number of involved exponential operations through the extrinsic information exchanges. Therefore, a simplified version of MPA known as MAX-Log-MPA[28], [148], which is inspired from MAX-Log-MAP [30], was derived to deal with complexity issues.

The key idea of this simplified variation is to convert the iterative decoding process into logarithmic domain while eliminating all the exponential terms in MPA [149] using the following simplified Jacobian logarithm formula :

$$\log(e^{a_1} + \dots + e^{a_n}) \approx \max(a_1, \dots, a_n) \quad (3.1)$$

This approximation permits to replace the log function of a sum of exponential terms by a simple maximization function, inducing a lower system complexity. Let  $L_{k \rightarrow j}^t(\mathbf{x}_j)$  and  $Q_{j \rightarrow k}^t(\mathbf{x}_j)$  be the log-domain extrinsic information sent, respectively, from FN  $k$  to VN  $j \in \mathcal{N}(k)$  and from VN  $j$  to FN  $k \in \mathcal{R}(j)$  for the corresponding codeword  $\mathbf{x}_j$  at the  $t$ -th iteration. Hence, the iterative decoding process is performed as follows :

► **Step 0 : Initialization**

As explained for MPA, initially, all codewords are assumed to be uniformly distributed with equal prior probabilities to be selected at each transmission attempt. Therefore, for MAX-Log-MPA, the initialization step consists of applying (2.13) in the logarithmic domain, which becomes :

$$Q_{j \rightarrow k}^0(\mathbf{x}_j) = \log(g_{j \rightarrow k}^0) = -\log M \quad \forall j = 1, \dots, J, \forall k \in \mathcal{R}(j) \quad (3.2)$$

► **Step 1 : Iterative message exchange**

This step consists in applying the MPA equations in log-domain. Therefore, the iterative process is amended as follows :

– **sub step a) : Passing information from FN to VN**

This step consists of converting the equations (2.14) in the logarithmic domain with the sub-optimal approximation (3.1). For  $M$ -dimensional SCMA codebooks, i.e.,  $M$  codewords in each user codebook, with  $d_f$  users transmitting in the same subcarrier,  $n = M^{d_f-1}$  and  $\{a_l\} \forall l \in \{1, 2, \dots, n\}$  in (3.1) are :

$$a_l = \frac{-1}{\sigma^2} \|y_k - \sum_{\substack{j \\ x_{kj} \neq 0}} h_{kj} x_{kj}\|^2 + \sum_{i \in \mathcal{N}(k) \setminus j} Q_{i \rightarrow k}^{(t-1)}(\mathbf{x}_i) \quad \forall k = 1, \dots, K \quad (3.3)$$

Then, the log-domain transmitted message from FN  $k$  to VN  $j$  for a given codeword  $\mathbf{x}_j$  can be calculated as below :

$$L_{k \rightarrow j}^t(\mathbf{x}_j) = \max_{\{\mathbf{x}_i | i \in \mathcal{N}(k) \setminus j\}} \left\{ \frac{-1}{\sigma^2} \|y_k - \sum_{\substack{j \\ x_{kj} \neq 0}} h_{kj} x_{kj}\|^2 + \sum_{i \in \mathcal{N}(k) \setminus j} Q_{i \rightarrow k}^{(t-1)}(\mathbf{x}_i) \right\} \quad (3.4)$$

– **sub step b) : Passing information from VN to FN**

The transmitted message from VN  $j$  to FN  $k$  for the corresponding codeword  $\mathbf{x}_j$  is also calculating by transforming (2.15) in log-domain, which is calculated as follows :

$$Q_{j \rightarrow k}^t(\mathbf{x}_j) = \log\left(\frac{1}{M}\right) + \sum_{i \in \mathcal{R}(j) \setminus k} L_{i \rightarrow j}^{(t-1)}(\mathbf{x}_j) \quad (3.5)$$

For the sake of simplicity, the normalization operation, which is used for controlling the variation area of the probability value  $g_{j \rightarrow k}^t$  in (2.16), can be ignored in the logarithmic domain.

► **Step 2 : LLR calculation and bits estimation**

Finally, after  $T$  iterations, the log-domain a posteriori probability of the associated codeword  $\mathbf{x}_j$  is calculated as :

$$\log(p(\mathbf{x}_j)) = \log\left(\frac{1}{M}\right) + \sum_{k \in \mathcal{R}(j)} L_{k \rightarrow j}^T(\mathbf{x}_j) \quad (3.6)$$

The bit-wise LLRs corresponding to the  $j$ -th user bits,  $b_{ij}$  with  $i \in \{1, \dots, \log_2(M)\}$ , are calculated by combining (2.18), (3.1) and (3.6) [28] :

$$\text{LLR}(b_{ij}) = \max_{\{\mathbf{x}_j \in \mathcal{X} | b_{ij}=0\}} (\log(p(\mathbf{x}_j))) - \max_{\{\mathbf{x}_j \in \mathcal{X} | b_{ij}=1\}} (\log(p(\mathbf{x}_j))) \quad (3.7)$$

Finally, the  $j$ -th user transmitted bits,  $\hat{b}_{ij}$  with  $i \in \{1, \dots, \log_2(M)\}$ , are estimated by comparing their corresponding LLR to 0 such that :

$$\begin{cases} \hat{b}_{ij} = 1 & \text{if } \text{LLR}(b_{ij}) \leq 0 \\ \hat{b}_{ij} = 0 & \text{otherwise} \end{cases} \quad (3.8)$$

### 3.3.2 Log-MPA algorithm

After removing all the exponential terms and substituting some additions by maximizations, MAX-log-MPA has a significantly reduced complexity compared to MPA. However, it is considered as a sub-optimal algorithm due to the approximation in (3.1), which may induce a performance degradation.

The main issue in logarithmic MPA variations is to precisely compute the function  $\log(e^{a_1} + \dots + e^{a_n})$ . In that sense, an interesting improved simplification of MPA known as Log-MPA [29] was proposed. The latter is also inspired by the simplified

decoding algorithm Log-MAP [30], and aims at reaching the same performance as MPA but with a lower decoding complexity. The main idea is to add a correction term to the approximation in (3.1) in order to compensate for the performance degradation, using the following Jacobian logarithm formula :

$$\log(e^a + e^b) = \max^*(a, b) = \max(a, b) + C(a - b) \quad (3.9)$$

with  $C(x)$  is the correction term which is given as :

$$C(x) = \log(1 + e^{-|x|}) \quad (3.10)$$

This approximation can be extended to  $n$  values [151]. Assuming that the set of  $\{a_i\}_{i=1, \dots, n}$  is sorted in a descending order such that  $a_n \geq a_{n-1}, \dots, \geq a_1$ , therefore, the generalized function is calculated as :

$$\begin{aligned} \log\left(\sum_{i=1}^n e^{a_i}\right) &= \log\left(e^{a_n} \frac{\sum_{i=1}^n e^{a_i}}{e^{a_n}}\right) \\ &= \log(e^{a_n}) + \log\left(e^{-a_n} \sum_{i=1}^n e^{a_i}\right) \\ &= a_n + \log\left(\sum_{i=1}^n e^{-(a_n - a_i)}\right) \\ &= a_n + \log\left(1 + \sum_{i=1}^{n-1} e^{-(a_n - a_i)}\right) \end{aligned} \quad (3.11)$$

Thanks to (3.11), the MAX-Log-MPA approximation (3.1) is corrected and thus, the log function of a sum of exponential terms is precisely calculated as :

$$\log\left(\sum_{i=1}^n e^{a_i}\right) = \max^*(a_1, \dots, a_n) = a_j + \log\left(1 + \sum_{i \in \{1 \dots n\} \setminus j} e^{-|a_j - a_i|}\right) \quad (3.12)$$

where the generalized correction term is represented as :

$$C = \log\left(1 + \sum_{i \in \{1 \dots n\} \setminus j} e^{-|a_j - a_i|}\right) \quad (3.13)$$

with  $a_i$  are as in (3.3) and letting that  $a_j = \max(a_1, \dots, a_n)$ . Thanks to the correction term in (3.13), the loss in performance due to the use of the maximization approximation (3.1) is expected to be reduced. However, it would come at the cost of additional complexity, w.r.t. MAX-Log-MPA, due to the sum of exponential terms in (3.12) [149]. Assuming the same initial conditions as MAX-Log-MPA in (3.2), the exchanged information in Log-MPA between FNs and VNs is updated as follows :

$$L_{k \rightarrow j}^t(\mathbf{x}_j) = \max_{\{\mathbf{x}_i | i \in \mathcal{N}(k) \setminus j\}}^* \left\{ \frac{-1}{\sigma^2} \left\| y_k - \sum_{\substack{j \\ x_{kj} \neq 0}} h_{kj} x_{kj} \right\|^2 + \sum_{i \in \mathcal{N}(k) \setminus j} Q_{i \rightarrow k}^{t-1}(\mathbf{x}_i) \right\} \quad (3.14)$$

$$Q_{j \rightarrow k}^t(\mathbf{x}_j) = \log\left(\frac{1}{M}\right) + \sum_{i \in \mathcal{R}(j) \setminus k} L_{i \rightarrow j}^{(t-1)}(\mathbf{x}_j) \quad (3.15)$$

The log-domain a posteriori probability for the corresponding codeword  $\mathbf{x}_j$  is calculated as explained for MAX-Log-MPA in (3.6). Then, the bit-wise LLRs of the  $j$ -th user are computed as follows :

$$\text{LLR}(b_{ij}) = \max_{\{\mathbf{x}_j \in \mathcal{X} | b_{ij}=0\}}^* (\log(p(\mathbf{x}_j))) - \max_{\{\mathbf{x}_j \in \mathcal{X} | b_{ij}=1\}}^* (\log(p(\mathbf{x}_j))) \quad \forall i = 1, \dots, \log_2(M) \quad (3.16)$$

At the end, the users transmitted bits can be easily estimated as previously described for MAX-Log-MPA in (3.8).

### 3.3.3 Algorithmic complexity

Algorithms	MPA	MAX-Log-MPA	Log-MPA
Addition	$M^{d_f} d_f K T (1 + d_f) + T J M (N - 1)$	$M^{d_f} d_f K T (2d_f - 1) + J M N T$	$M^{d_f} d_f K T (2d_f + 1) + M T (J N - K d_f)$
Multiplication	$2M^{d_f} K T d_f (d_f + 1) + J M N T$	$M^{d_f} d_f K T (d_f + 3)$	$M^{d_f} d_f K T (d_f + 3)$
Exponential	$M^{d_f} K d_f T$	0	$M (M^{d_f - 1} - 1) d_f K T$
Maximization	0	$M d_f K T$	$M d_f K T$

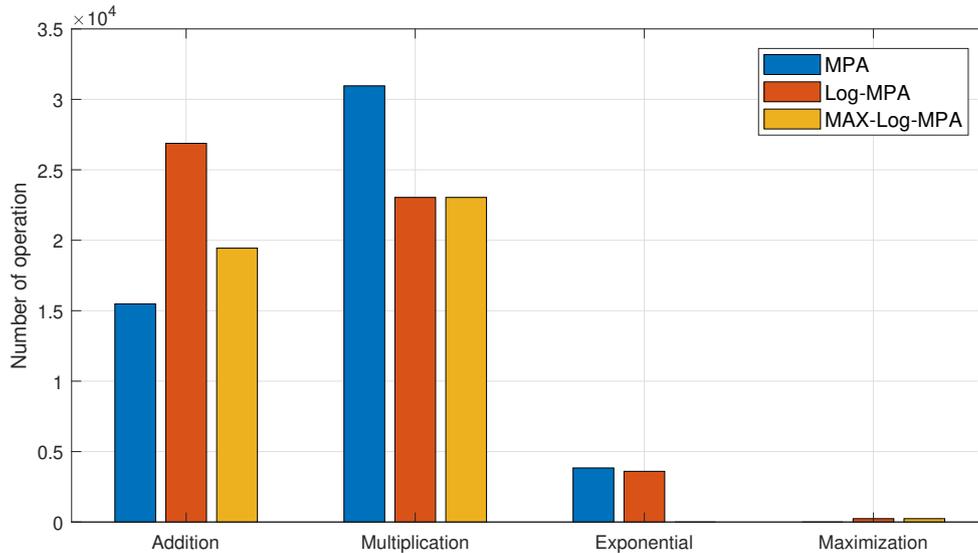
TABLE 3.2: Computational complexity of the studied algorithms.

The main purpose of the logarithmic derivations of MPA is to reduce its decoding complexity while maintaining an acceptable trade-off with the achieved performance. Table 3.2 provides a detailed comparison of the computational complexity of the different algorithms, i.e., MPA, Log-MPA and MAX-Log-MPA, for the exchanged extrinsic information between the FN and VN during T decoding iterations. In fact, the iterative process of information exchange between the function and variable nodes generates different operations namely, additions, multiplications, exponentials and maximizations. For the sake of simplicity, the division operations are considered as multiplications and the subtractions are considered as additions.

For MPA, the transmitted information from the FN to the VN in (2.14) is the most expensive step in terms of complexity. The iterative process in (2.14) and (2.15) involves  $M^{d_f} d_f K T (1 + d_f) + T J M (N - 1)$  additions,  $2M^{d_f} K T d_f (d_f + 1) + J M N T$  multiplications and  $M^{d_f} K d_f T$  exponential operations which leads to an important decoding complexity. Therefore, for MAX-Log-MPA, this equation is converted in log-domain as in (3.4) and (3.5), while using the approximation in (3.1). This simplification permits to significantly reduce the decoding complexity by removing all the exponential terms and replacing them by  $M d_f K T$  maximization functions. It includes also  $M^{d_f} d_f K T (2d_f - 1) + J M N T$  additions and  $M^{d_f} d_f K T (d_f + 3)$  multiplications. It can be observed that the total number of the generated multiplications

in MAX-Log-MPA is lower than MPA, but at the cost of a larger number of additions. However, these latter are very simple, which would not impact the decoding complexity.

For Log-MPA receiver, the addition of the correction term in (3.14), in order to improve the system performance, increases the decoding complexity compared to MAX-Log-MPA with the insertion of  $M(M^{d_f-1} - 1)d_fKT$  exponential operations and  $(2M^{d_f-1} - 1)Md_fKT$  additions. Log-MPA involves more additions than MPA, but with the advantage of a lower number of exponential terms and multiplication operations, which ensures a lower complexity.



**Fig. 3.6.** A comparison of the decoding complexity.

Figure 3.6 illustrates a comparison of the generated decoding complexity of the investigated algorithms while calculating the total number of operations as calculated in Table 3.2. We consider  $J = 6$  users and  $K = 4$  subcarriers with  $N = 2$  non zero elements in each codeword and  $d_f = 3$  users transmitting on each subcarrier. The decoding process iterated for  $T = 5$  iterations. For the sake of evaluation, the complexity of MPA is considered as the baseline.

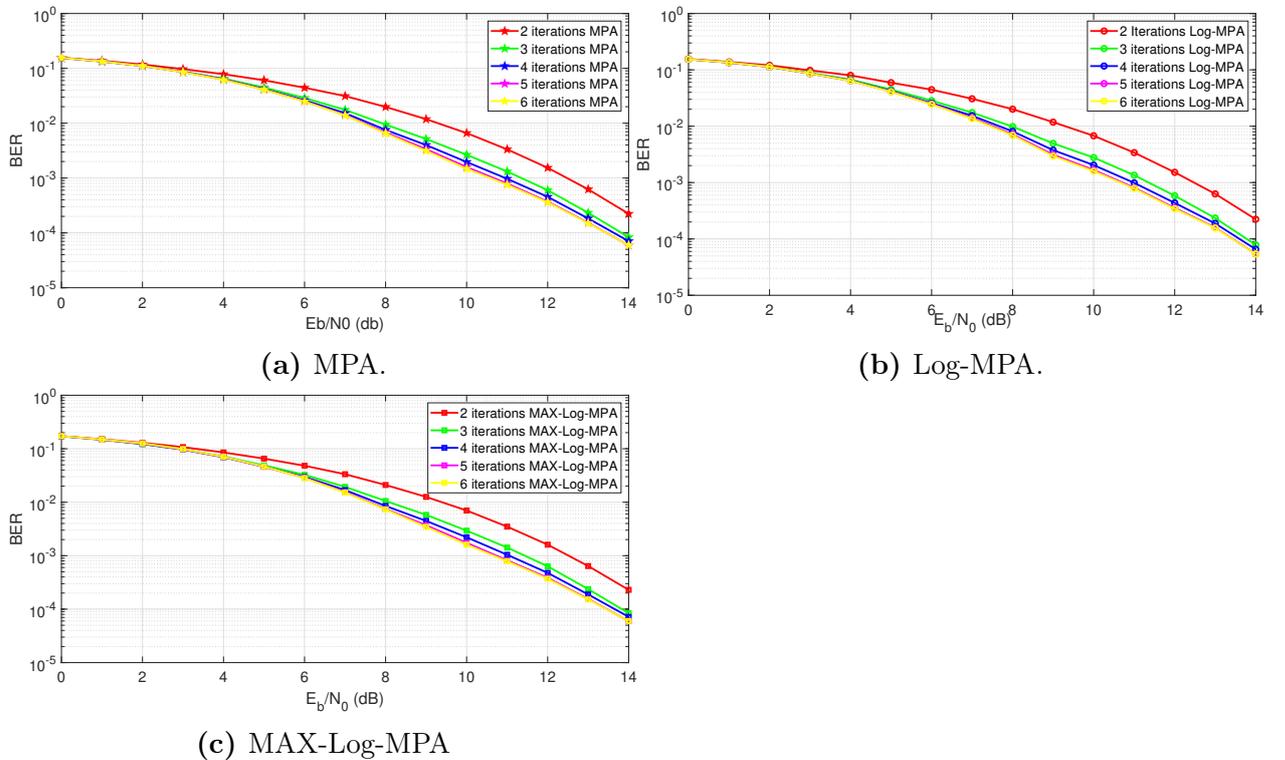
It can be observed that Log-MPA and MAX-Log-MPA generate, respectively, 73.6% and 25.5% more additions than MPA. Moreover, we can notice that both of them optimizes 25.5% of multiplications. Moreover, Log-MPA reduces 6.25% of the exponential operations compared to MPA while generating 240 new maximization functions. However, MAX-Log-MPA eliminates all the 3840 exponential terms of MPA and replaces them only by 240 maximization operations, which significantly reduces the decoding complexity.

In short, MAX-Log-MPA has the lowest decoding complexity. This optimization would come at the cost of a performance degradation compared to MPA. However, involving the correction term in Log-MPA leads to an increase in the system complexity compared to MAX-Log-MPA, but it is still lower than that of MPA. Log-MPA would improve the performance over MAX-Log-MPA, which will be verified in the next section.

### 3.3.4 Numerical results and analysis

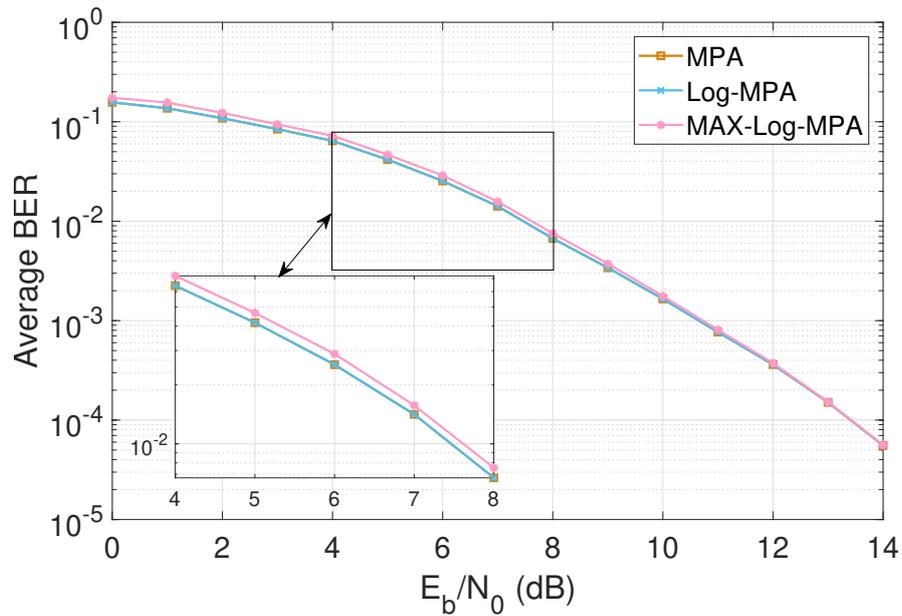
We consider an uplink communication system of  $J = 6$  users sharing  $K = 4$  subcarriers. Moreover, the codebook used are those presented in [128], where  $M = 4$  and  $N = 2$  in AWGN and Rayleigh fading channels.

Figure 3.7 illustrates the BER of MPA, Log-MPA and MAX-Log-MPA w.r.t. the number of iterations on AWGN channel. For instance, in Figure 3.7a, we observe that the performance gain of MPA is about 1.8 dB at BER of  $10^{-3}$  by passing from 2 to 5 iterations. The performance is getting better as the number of iterations  $T$  increases until  $T = 5$  where no further improvement can be noted after this limit. The same convergence behavior at  $T = 5$  is also observed for Log-MPA and MAX-Log-MPA in Figure 3.7b and Figure 3.7c, respectively. Therefore, the number of iterations of each algorithm is set to 5 in the remaining of this section.



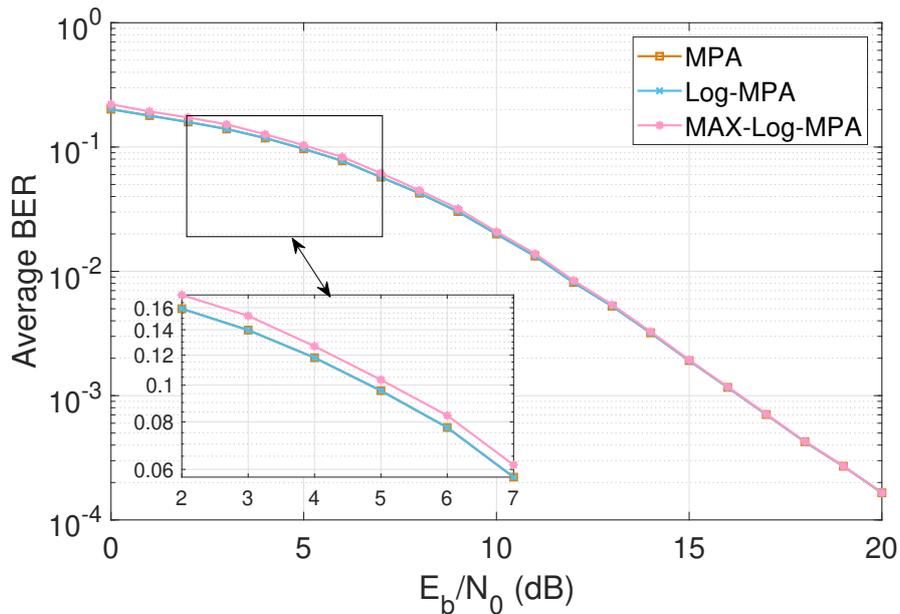
**Fig. 3.7.** Performance comparison for different number of iterations, AWGN channel,  $J = 6$  and  $K = 4$ .

Figure 3.8 shows the BER obtained by the three studied algorithms in AWGN channel. We observe that Log-MPA cannot be distinguished from MPA due to the used correction term in (3.12) allowing to compensate for the performance loss of MAX-Log-MPA. Indeed, a performance loss of 0.26 dB at BER value of  $3 \times 10^{-2}$  can be recorded between MAX-Log-MPA and MPA (and hence Log-MPA). This mismatch comes from the approximation in (3.1). Moreover, the BER obtained with MAX-Log-MPA converges to those achieved by MPA/Log-MPA as  $E_b/N_0$  increases. This behavior is due to the fact that the approximation in (3.1) is tighter for high  $E_b/N_0$ .

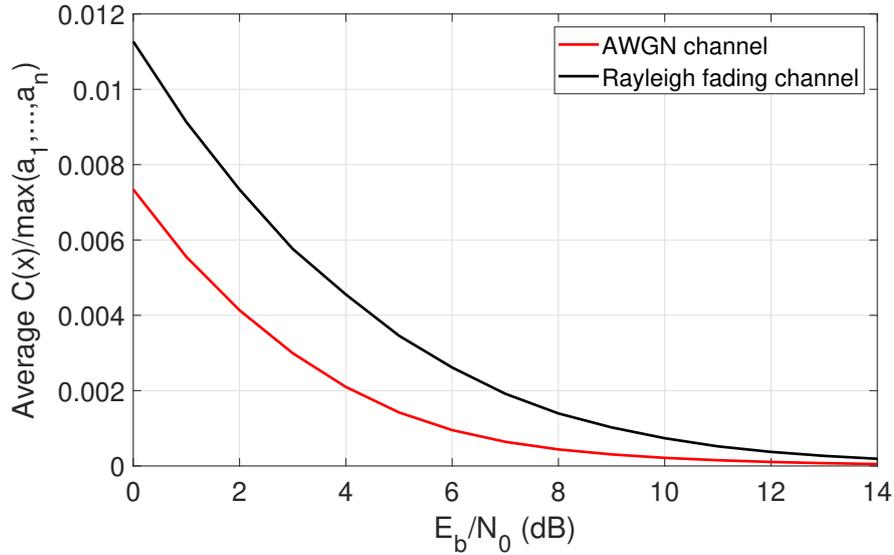


**Fig. 3.8.** MPA, Log-MPA and MAX-Log-MPA performance comparison for AWGN channel,  $J = 6$ ,  $K = 4$  and  $T = 5$ .

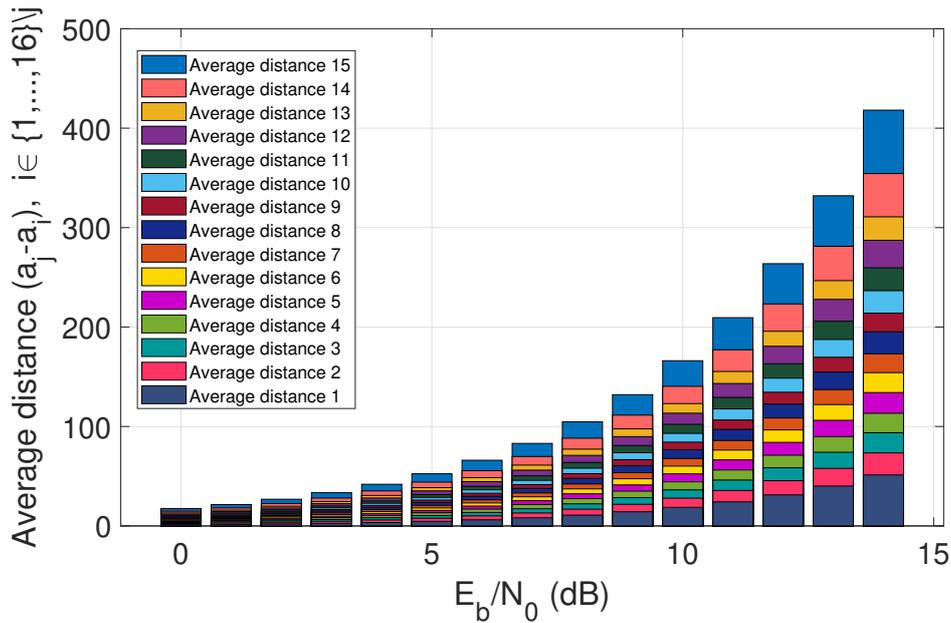
Figure 3.9 reproduces the BER achieved in Rayleigh fading channel. The performance ordering between algorithms is the same as in AWGN channel, where the error rate obtained with Log-MPA overlaps with that achieved by MPA, whereas MAX-Log-MPA experiences a low performance attenuation about 0.3 dB at BER of  $8 \times 10^{-2}$ . Similarly than in AWGN, the performance degradation attenuates when  $E_b/N_0$  increases.



**Fig. 3.9.** MPA, Log-MPA and MAX-Log-MPA performance comparison for Rayleigh fading channel,  $J = 6$ ,  $K = 4$  and  $T = 5$ .



**Fig. 3.10.** A comparison of the average ratio  $\frac{C(x)}{\max(a_1, \dots, a_n)}$  for AWGN and Rayleigh fading channels.



**Fig. 3.11.** A comparison of the average distance  $|a_j - a_i|$  for AWGN channel.

In order to further investigate the mismatch observed between MAX-Log-MPA and MPA/Log-MPA, we study in Figure 3.10, over all the possible transmitted codewords, the relative evolution of the average of the correction term  $C$  and the absolute value of the average maximization function, i.e.,  $\max(a_1, \dots, a_n)$ , used in Log-MPA algorithm in (3.12). We notice that, for both AWGN and Rayleigh fading channels, this ratio has very low values and it decreases asymptotically toward zero when  $E_b/N_0$  increases. This means that the correction term in (3.12) becomes negligible compared to the maximization term when the SNR per bit increases, which explains the low gap between the curves of MPA/Log-MPA and MAX-Log-MPA algorithms.

In addition, we realize that the Rayleigh fading channel ratio is more important than that of the Gaussian channel, which justifies the fact that the performance degradation for this latter is slightly lower than for the Rayleigh channel.

In order to better understand the algorithm behaviors, Figure 3.11 depicts the statistical evolution of the distances,  $|a_j - a_i| \forall i \in \{1, \dots, n\} \setminus j$ , averaged over all the codewords, where  $n = M^{d_f-1} = 16$  and  $a_j = \max(a_1, a_2, \dots, a_{16})$ , that appear in the correction term in (3.12) for a given user. Without loss of generality, we consider here the average distances involved in the extrinsic information (3.14) for the first user. We observe that these distances increase with  $E_b/N_0$ , making the correction term weaker when  $E_b/N_0$  increases, and hence the alleviation of the performance degradation between MAX-Log-MPA and MPA/Log-MPA.

### 3.4 Conclusion

This chapter provided a performance comparison of selected NOMA schemes namely, SCMA, MUSA, PDMA and PD-NOMA, in the same conditions in different scenarios in order to select the most appropriate one for mMTC scenario. We demonstrated that SCMA showed the best performances, but at the cost of high decoding complexity. For that reason, we provided an exhaustive study and numerical analysis of the performance of the three candidate decoding algorithms for SCMA, i.e., MPA, MAX-Log-MPA and Log-MPA, in AWGN and Rayleigh fading channels. We demonstrated that, for both channels, Log-MPA preserves strictly the same performance as MPA with a reduced complexity thanks to the added correction term. As for MAX-Log-MPA, it has a very low complexity by eliminating all the exponential operations but at the cost of a slight performance degradation. Finally, a comprehensive study of the performance evolution of Log-MPA and MAX-Log-MPA has been provided through the analysis of the behavior of Log-MPA correction term which is involved in the exchanged extrinsic information.

However, despite the MPA simplifications, the decoding complexity of SCMA is still high and may be an important constraint in mMTC, especially with a large number of connected devices. Added to this, the codebooks design complexity and their distribution among the active users remain major issues in mMTC, and these challenges are left for further work. MUSA scheme has shown a good performance particularly in the case of scattered users over the cell, showing an acceptable trade-off between the system complexity and the achieved performance, but at the cost of an error propagation problem which should be handled to improve its performance if this scheme would be considered for the future mMTC services.

# Power allocation for BER minimization in an uplink MUSA scenario

## 4.1 Introduction

As shown in the previous chapters, the MUSA scheme seems to be one of the most auspicious techniques to fulfill the binding requirements of mMTC capability thanks to its high overload capacity and its low complexity. MUSA is typically conducted with successive interference cancellation strategy for signals separation. In contrast to some receivers, such as MPA and MAP, which perform better with equal users' received powers, the SIC technique achieves its best performance for sufficiently diverse powers. Therefore, the power assignment has been considered as a major issue for MUSA which may dramatically deteriorates its performance with inappropriate allocation leading to a severe error propagation problem.

In this chapter, we seek to handle this issue by proposing an adequate power allocation model for MUSA scheme. The rest of this chapter is organized as follows. An overview of the existing works on resource allocation for NOMA is introduced Section 4.2. Then, our proposed approach for MUSA bit error rate minimization is presented in Section 4.3. After that, the numerical results and the performance analysis are provided in Section 4.4. Finally, conclusions are drawn in Section 4.5.

## 4.2 State of the art on resource allocation

All NOMA schemes are based on the idea of sharing the same resources between multiple users, which implies a high inter-user interference. Robust multi-user detection techniques, such as SIC or MPA, have been proposed for signals separation. However, they may be unable to totally correct the generated damage, which induces a performance degradation. Therefore, an optimal resource allocation, such as power allocation, optimal subcarrier attribution, or users clustering, may lend a hand to deal with this issue. Compared to the conventional OMA schemes, the re-

source allocation with NOMA schemes is more complex due to the resource sharing among multiple users, and the high inter-user interference. Several utility functions can be used for resource allocation, namely sum-rate maximization, users' fairness maximization, pairwise error probability minimization, etc,...

In literature, a lot of works have been particularly interested in the resources allocation for PD-NOMA scheme. Authors in [152] proposed power allocation strategies for downlink communications based on the experienced CSI or a minimum predefined QoS for high priority users. Each frequency band is attributed simultaneously to a group of high priority and low priority users with a total transmission power constraint. In [153], an optimization problem for power allocation in downlink scenario was proposed to calculate the minimum total system power required to ensure an equal data rate among users and to maximize the energy efficiency.

In addition, authors in [74] aimed at maximizing the system capacity through a multi-user beamforming technique. This problem is studied in two sub-steps. Firstly, each pair of users with large channels gain difference are attributed to the same cluster with a single beamforming vector, using a novel clustering algorithm. This method permits to reduce the inter-beam and inter-user interference. Then, a power allocation strategy is applied within each cluster in order to maximize the sum-capacity.

All the previous works do not take into consideration the specific conditions for a proper SIC process. Therefore, an optimization problem was presented in [71] and [154] for the maximization of the system throughput with a particular received power difference constraint. For the sake of simplicity, this problem is split into two sub-optimal problems for user clustering and power allocation under the constraints of minimum required data rate, transmission power and sufficient received power difference among users in order to avoid error propagation in the SIC receiver. Besides, authors in [70], studied the maximization of the weighted sum-rate with a perfect or imperfect SIC process for downlink transmissions. An optimal power allocation and user scheduling strategies were proposed in both cases using special conditions for signals separation.

Recently, the researchers have also addressed the resource allocation problem for SCMA scheme for performance enhancement. Authors in [155], investigated the maximization over the powers and subcarriers of the system sum-rate and the users' fairness for uplink transmissions. Both of problems are NP-hard with non convex and mixed integer nonlinear programming behavior. Therefore, two iterative algorithms have been proposed for their resolution. In [156], a power allocation approach for the downlink sum-capacity maximization was studied for a multi-cell system and under a minimum users rate constraint. The obtained solution is compared to the conventional power allocation strategies such as the uniform power allocation and the water filling algorithm. Moreover, the maximization of SCMA energy efficiency was stressed in [157] under a joint allocation of users' power and subcarriers. The problem is divided into two sub-systems for the subcarrier assignment and the power allocation, respectively.

Most of the previous works focused on the sum-rate or users' fairness maximization. However, in contrast to the eMBB service, high data rate is not a priority for mMTC scenarios with short packets. Instead, the emphasis is set on supporting

a large number of users with a moderate BER. The resource allocation for BER minimization has been widely studied for orthogonal multiple access schemes [158]. Recently, resource allocation has also been involved in the minimization of the pairwise error probability of PD-NOMA scheme for downlink transmissions in [159], [160]. In addition, authors in [112] have investigated the impact of different power coefficients on a downlink MUSA scheme using random power values. However, no work has proposed an optimal power allocation approach for an uplink MUSA scheme with OSIC receiver and inter-user interference due to resources overload.

### 4.3 Average BER minimization

In this section, we study the power allocation problem for an uplink MUSA scheme in the intention of minimizing the average system BER by mitigating the experienced error propagation while guaranteeing a sufficient power difference between concurrent received signals. Therefore, an optimization model is formulated for BER minimization under transmitted and received power constraints.

#### 4.3.1 Problem formulation

A multiple access channel system with  $J$  users transmitting over  $K$  orthogonal subcarriers is considered. Each user selects randomly a  $K$ -dimensional spreading sequence and then transmits its data over the available subcarriers, as explained in Section 2.2.5 in chapter 2. Users signals are subsequently handled at the receiver using OSIC technique in a descending order of their SINRs. Without loss of generality, we assume  $\text{SINR}_1 > \text{SINR}_2 > \dots > \text{SINR}_J$ . Considering a perfect OSIC process with a correctly decoded users' symbols, the received signal at the  $j$ -th iteration is represented as follows :

$$\mathbf{y}^j = \sqrt{p_j} \mathbf{g}_j x_j + \sum_{i=j+1}^J \sqrt{p_i} \mathbf{g}_i x_i + \mathbf{n}^j \quad (4.1)$$

where  $p_j$ ,  $x_j$  are respectively the  $j$ -th user transmission power and transmitted symbol,  $\mathbf{g}_j$  is the equivalent channel vector of the  $j$ -th user including its spreading sequence and  $\mathbf{n}^j$  is the noise vector at the  $j$ -th iteration with  $\mathbf{n}^j \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_K)$   $\forall j \in \{1, \dots, J\}$ . Therefore, the  $j$ -th user SINR is represented as [114] :

$$\beta_j(\mathbf{p}) = \frac{p_j |\mathbf{w}_j^H \mathbf{g}_j|^2}{\sum_{i=j+1}^J p_i |\mathbf{w}_j^H \mathbf{g}_i|^2 + \sigma^2 \|\mathbf{w}_j^H\|^2} \quad (4.2)$$

where  $\mathbf{p} = [p_1, p_2, \dots, p_J]$  and  $\mathbf{w}_j$  is the  $j$ -th column of the MMSE matrix  $\mathbf{W}$  in (2.23). However, the sequential decoding of users may be affected by a severe error propagation problem when the power difference between the simultaneous received signals is not significant, which leads to close users' SINR. In fact, the SIC technique performs better when users' SINR are sufficiently different, which ensures a correct decoding order of users' signal.

Therefore, in this part, we propose to minimize the average system BER under the constraint of total system or maximum individual transmission power and a further constraint on the adequate power difference between the currently decoded user and the interfering ones. We consider that users use QPSK modulation. In addition, the inter-user interference is treated as noise. Hence, a robust approximation for the error probability of each user, while eliminating the error propagation effect, is represented as [161] :

$$\text{BER}_j \approx \frac{1}{5} \exp\left(\frac{-\beta_j(\mathbf{p})}{2}\right). \quad (4.3)$$

Let  $\mathcal{J} = \{1, \dots, J\}$  denotes the set of users and  $\mathcal{J}_j = \mathcal{J} \setminus \{j\}$ . The optimization problem for the BER minimization under the defined constraint is formulated as follows :

$$P_1 \left\{ \begin{array}{l} \min_{\mathbf{p}} \quad \frac{1}{5J} \sum_{j=1}^J \exp\left(\frac{-\beta_j(\mathbf{p})}{2}\right) \quad (4.4) \\ \text{s.t.} \quad \sum_{j=1}^J p_j \leq P_{tot} \quad (4.4a) \\ p_j \leq P_{max} \quad \forall j \in \mathcal{J} \quad (4.4b) \\ p_j \|\mathbf{g}_j\|^2 - \sum_{i=j+1}^J p_i \|\mathbf{g}_i\|^2 \geq P_{\text{SIC}} \quad \forall j \in \mathcal{J}_j \quad (4.4c) \end{array} \right.$$

The inequality (4.4a) refers to the total sum power constraint, (4.4b) represents the constraint on the maximum allowed transmission power per-user and (4.4c) ensures the favorable conditions for SIC decoding while guaranteeing a sufficient difference between the received powers and thereby alleviates the error propagation issue [71]. In the uplink scenario, the practical constraint (4.4b) is imposed by the regulation authorities and the equipment design restrictions. However, the constraint (4.4a) allows to lower bound the global BER since it offers more degrees of freedom thanks to the possibility of power transfer among users [162].  $P_{\text{SIC}}$  refers to the difference between the received power of the user being decoded and the sum-power of users not yet processed, which permits to perfectly differentiate between users' signals. This parameter must be optimized in order to achieve the best decoding performance.

By studying its behavior, we remark that the optimization problem  $P_1$  is not convex on  $\mathbf{p}$  since users' power are involved in both the numerator and the denominator of the SINR (4.2). Therefore, the Karush-Kuhn-Tucker (KKT) conditions are not sufficient to ensure a global optimal solution for this problem. For ease of processing, a simplified approximation of SINR was proposed in [31]. Indeed, the SINR of the currently decoded user involves only its corresponding power while eliminating those of the interfering users. Hence, the revised SINR expression is updated as :

$$\Psi_j = \frac{p_j |\mathbf{v}_j^H \mathbf{g}_j|^2}{\sum_{i=j+1}^J |\mathbf{v}_j^H \mathbf{g}_i|^2 + \sigma^2 \|\mathbf{v}_j^H\|^2} = p_j \Gamma_j \quad (4.5)$$

where  $\mathbf{v}_j$  is the  $j$ -th column of the MMSE matrix  $\mathbf{V}$  which is calculated as in (2.23) with  $\mathbf{P} = \mathbf{I}_J$ . Hence, the simplified optimization problem becomes :

$$P_2 \begin{cases} \min_{\mathbf{p}} & \frac{1}{5J} \sum_{j=1}^J \exp\left(\frac{-p_j \Gamma_j}{2}\right) & (4.6) \\ \text{s.t.} & \text{Scenario 1 : } \begin{cases} \text{Case 1 : (4.4a)} \\ \text{Case 2 : (4.4a) and (4.4c)} \end{cases} & (4.6a) \\ & \text{Scenario 2 : } \begin{cases} \text{Case 1 : (4.4b)} \\ \text{Case 2 : (4.4b) and (4.4c)} \end{cases} & (4.6b) \end{cases}$$

$P_2$  does not correspond to  $P_1$ . Since the powers of interfering users are removed from users' SINR. In addition, the error propagation impact is not considered in the utility function (4.6). For that reason, the constraint (4.4c) is involved to handle this issue by ensuring a sufficient diversity between the received powers to limit the error propagation and improving the decoding process. The simulation results will show that the obtained optimized powers allow to achieve good results under the real system conditions with the true SINR in (4.2).

### 4.3.2 Problem resolution

The resolution process of the optimization model  $P_2$  is carried out in two different scenarios, first using the total system power constraint (4.4a), and then using the individual power constraints (4.4b). Each scenario is further investigated in two cases while considering or not the received power difference constraint (4.4c).

#### 4.3.2.1 Scenario 1

In this scenario, the optimization problem is conducted under the total system power constraint (4.4a) with or without the power difference constraint (4.4c). Regarding the problem analysis, the objective function is convex and both constraints are linear. Therefore, the KKT conditions are necessary and sufficient to come up with an optimal global solution. The resolution process starts by defining the vectorized Lagrangian function of  $P_2$  which can be represented as :

$$\mathcal{L}(\mathbf{p}, \lambda, \boldsymbol{\mu}) = P_{\text{SIC}} \mathbf{1}^T \boldsymbol{\mu} - \lambda P_{\text{tot}} + \left( \frac{1}{J} \mathbf{1}^T \mathbf{f}(\mathbf{p}) + (\lambda \mathbf{1} + \mathbf{C} \boldsymbol{\mu})^T \mathbf{p} \right) \quad (4.7)$$

where  $\lambda$  and  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_{J-1})^T$  are the Lagrange multipliers associated to the constraints (4.4a) and (4.4c), respectively. The matrix  $\mathbf{C} \in \mathbb{R}_+^{J \times J-1}$  and the vector  $\mathbf{f}(\mathbf{p}) \in \mathbb{R}_+^{J \times 1}$  are represented as :

$$\mathbf{C} = \begin{bmatrix} -\|\mathbf{g}_1\|^2 & 0 & 0 & 0 & \dots & 0 \\ \|\mathbf{g}_2\|^2 & -\|\mathbf{g}_2\|^2 & 0 & 0 & \dots & 0 \\ \|\mathbf{g}_3\|^2 & \|\mathbf{g}_3\|^2 & -\|\mathbf{g}_3\|^2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \|\mathbf{g}_{J-1}\|^2 & \|\mathbf{g}_{J-1}\|^2 & \|\mathbf{g}_{J-1}\|^2 & \|\mathbf{g}_{J-1}\|^2 & \dots & -\|\mathbf{g}_{J-1}\|^2 \\ \|\mathbf{g}_J\|^2 & \|\mathbf{g}_J\|^2 & \|\mathbf{g}_J\|^2 & \|\mathbf{g}_J\|^2 & \dots & 0 \end{bmatrix}$$

$$\mathbf{f}(\mathbf{p}) = 1/5 [e^{-p_1 \Gamma_1/2}, \dots, e^{-p_J \Gamma_J/2}]^T$$

Based on the Lagrangian vector (4.7), the KKT conditions in this scenario with the corresponding constraints, applied to the triplet  $\mathbf{p}$ ,  $\lambda$  and  $\boldsymbol{\mu}$ , are calculated as follows :

$$\frac{\partial \mathcal{L}}{\partial p_j} = \frac{-\Gamma_j}{10J} \exp\left(\frac{-p_j \Gamma_j}{2}\right) + \lambda + \|\mathbf{g}_j\|^2 \left(\sum_{i=1}^{j-1} \mu_i - \mu_j\right) = 0 \quad \forall j \in \mathcal{J}_J \quad (4.8)$$

$$\frac{\partial \mathcal{L}}{\partial p_J} = \frac{-\Gamma_J}{10J} \exp\left(\frac{-p_J \Gamma_J}{2}\right) + \lambda + \|\mathbf{g}_J\|^2 \sum_{i=1}^{J-1} \mu_i = 0 \quad (4.9)$$

$$\mu_j \left(-p_j \|\mathbf{g}_j\|^2 + \sum_{i=j+1}^J p_i \|\mathbf{g}_i\|^2 + P_{\text{SIC}}\right) = 0 \quad \forall j \in \mathcal{J}_J \quad (4.10)$$

$$\lambda \left(\sum_{j=1}^J p_j - P_{\text{tot}}\right) = 0 \quad (4.11)$$

$$P_{\text{SIC}} - p_j \|\mathbf{g}_j\|^2 + \sum_{i=j+1}^J p_i \|\mathbf{g}_i\|^2 \leq 0 \quad \forall j \in \mathcal{J}_J \quad (4.12)$$

$$\sum_{j=1}^J p_j - P_{\text{tot}} \leq 0 \quad (4.13)$$

$$\lambda \geq 0 \quad (4.14)$$

$$\mu_j \geq 0 \quad \forall j \in \mathcal{J}_J \quad (4.15)$$

where (4.8), (4.9) are the stationarity conditions, (4.10), (4.11) represent the complementary slackness conditions, (4.12), (4.13) refer to the primal feasibility and (4.14) and (4.15) introduce the dual feasibility.

**a) Case 1 :**

In this case, only the total power constraint (4.4a) is used, i.e.,  $\boldsymbol{\mu} = \mathbf{0}$ . It represents the ideal system condition where allowing a power transfer between users according to their encountered SINRs. Therefore, this case may be considered as a lower bound on the achievable BER. Based on (4.8), (4.9), (4.11), (4.13) and (4.14), the optimal solution can be readily calculated as :

$$p_j^* = \left[-\frac{2}{\Gamma_j} \ln\left(\frac{10J\lambda}{\Gamma_j}\right)\right]^+ \quad \forall j \in \mathcal{J} \quad (4.16)$$

where  $[\cdot]^+$  refers to  $\max(0, \cdot)$  and  $\lambda^*$  is computed as :

$$\lambda^* = \exp\left(-\frac{\sum_{j=1}^J \frac{2}{\Gamma_j} \ln\left(\frac{10J}{\Gamma_j}\right) + P_{\text{tot}}}{2 \sum_{j=1}^J \frac{1}{\Gamma_j}}\right) \quad (4.17)$$

**b) Case 2 :**

This case encompasses both of the total system power (4.4a) and the SIC performance guarantee (4.4c) constraints. However, no closed-form expression for users' power can be obtained using the KKT conditions. Nevertheless, an optimal solution can be reached with a gradient-based algorithm inspired by the idea of differential multiplier method, which was proposed in [163], and described in Algorithm 1. This type of problems with inequality constraints is quite intricate to solve algorithmically. Therefore, for the sake of simplicity, inequality constraints should be relaxed and converted into equalities by inserting non-negative slack variables. Adding slack variables does not change the problem. It only modifies its representation. Therefore, the optimal solution remains the same. In order to exploit all the available system power, the inequality constraint (4.4a) is directly transformed into equality, which means that this constraint will be always saturated. However, the constraints in (4.4c) are converted into equalities by introducing non-negative slack variables  $z_j^2 \forall j \in \mathcal{J}$ . The use of the squared variable  $z_j^2$ , which is always positive, is to guarantee compliance with the original inequality constraints in (4.4c), where the received power difference must be equal to or greater than  $P_{SIC}$ . Hence, (4.4a) and (4.4c) become :

$$\sum_{j=1}^J p_j = P_{\text{tot}} \quad (4.18)$$

$$p_j \|\mathbf{g}_j\|^2 - \sum_{i=j+1}^J p_i \|\mathbf{g}_i\|^2 - z_j^2 = P_{SIC} \quad \forall j \in \mathcal{J} \quad (4.19)$$

An iterative dual gradient descent and ascent algorithm is carried out on the Lagrangian function in (4.7). Users' power as well as the extra slack variables vector  $\mathbf{z} = (z_1, z_2, \dots, z_{J-1})^T$  are updated using gradient descent equations, as represented in the 6-th and 7-th rows of Algorithm 1, respectively. Whereas Lagrange multipliers are treated using an iterative gradient ascent method, in the 10-th and 11-th rows of Algorithm 1. If the slack variable  $z_j = 0$ , then the corresponding constraint (4.4c) is saturated at equality and is considered to be tight at  $\mathbf{p}$ . However, if  $z_j \neq 0$ , the corresponding constraint is a strict inequality and hence, the associated Lagrangian multiplier is equal to zero.

**4.3.2.2 Scenario 2**

This scenario reveals the practical and real conditions for an uplink transmission. It applies the maximum per-user power constraint (4.4b) while considering or not the power difference constraints (4.4c). Therefore, equivalently to scenario 1, the optimization problem in this scenario is also convex. Thus, the KKT conditions are sufficient and necessary to guarantee an optimal solution. As a consequence, the vectorized Lagrangian function is expressed as :

$$\mathcal{L}^1(\mathbf{p}, \boldsymbol{\theta}, \boldsymbol{\mu}) = \mathbf{1}^T \boldsymbol{\mu} P_{SIC} - \mathbf{1}^T \boldsymbol{\theta} P_{\text{max}} + \left( \frac{1}{J} \mathbf{1}^T \mathbf{f}(\mathbf{p}) + (\boldsymbol{\theta} + \mathbf{C}\boldsymbol{\mu})^T \mathbf{p} \right) \quad (4.20)$$

---

**Algorithm 1:** Power allocation for BER minimization

---

**Require:**  $\mathbf{p}^0, \boldsymbol{\mu}^0, \mathbf{z}^0, \lambda^0, \alpha$ ;  
**Ensure:**  $\mathbf{p}, \boldsymbol{\mu}, \lambda$   
1:  $\Delta_{\mathbf{p}}^0 = \nabla_{\mathbf{p}} \mathcal{L}(\mathbf{p}^0, \boldsymbol{\mu}^0, \lambda^0)$   
2:  $\Delta_{\mathbf{z}}^0 = \nabla_{\mathbf{z}} \mathcal{L}(\boldsymbol{\mu}^0, \mathbf{z}^0)$   
3:  $\Delta_{\boldsymbol{\mu}}^0 = \nabla_{\boldsymbol{\mu}} \mathcal{L}(\mathbf{p}^0, \mathbf{z}^0)$   
4:  $\Delta_{\lambda}^0 = \frac{\partial \mathcal{L}}{\partial \lambda}(\mathbf{p}^0)$   
5: **while**  $|\Delta_{\mathbf{p}_{i \in \mathcal{J}}}^n| > \epsilon, |\Delta_{\boldsymbol{\mu}_{i \in \mathcal{J}_J}}^n| > \epsilon$  and  $|\Delta_{\lambda}^n| > \epsilon$  **do**  
6:    $\mathbf{p}^{n+1} = \mathbf{p}^n - \alpha \Delta_{\mathbf{p}}^n$   
7:    $\mathbf{z}^{n+1} = \mathbf{z}^n - \alpha \Delta_{\mathbf{z}}^n$   
8:    $\Delta_{\boldsymbol{\mu}}^{n+1} = \nabla_{\boldsymbol{\mu}} \mathcal{L}(\mathbf{p}^{n+1}, \mathbf{z}^{n+1})$   
9:    $\Delta_{\lambda}^{n+1} = \frac{\partial \mathcal{L}}{\partial \lambda}(\mathbf{p}^{n+1})$   
10:    $\boldsymbol{\mu}^{n+1} = \boldsymbol{\mu}^n + \alpha \Delta_{\boldsymbol{\mu}}^{n+1}$   
11:    $\lambda^{n+1} = \lambda^n + \alpha \Delta_{\lambda}^{n+1}$   
12:    $\Delta_{\mathbf{p}}^{n+1} = \nabla_{\mathbf{p}} \mathcal{L}(\mathbf{p}^{n+1}, \boldsymbol{\mu}^{n+1}, \lambda^{n+1})$   
13:    $\Delta_{\mathbf{z}}^{n+1} = \nabla_{\mathbf{z}} \mathcal{L}(\boldsymbol{\mu}^{n+1}, \mathbf{z}^{n+1})$   
14: **end while**

---

where  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_J)^T$  is the Lagrange multipliers vector associated to the constraints in (4.4b). Hence, based on (4.20), the corresponding KKT conditions for this problem are defined as :

$$\frac{\partial \mathcal{L}^1}{\partial p_j} = \frac{-\Gamma_j}{10J} \exp\left(\frac{-p_j \Gamma_j}{2}\right) + \theta_j + \|\mathbf{g}_j\|^2 \left(\sum_{i=1}^{j-1} \mu_i - \mu_j\right) = 0 \quad \forall j \in \mathcal{J} \quad (4.21)$$

$$\frac{\partial \mathcal{L}^1}{\partial p_J} = \frac{-\Gamma_J}{10J} \exp\left(\frac{-p_J \Gamma_J}{2}\right) + \theta_J + \|\mathbf{g}_J\|^2 \sum_{i=1}^{J-1} \mu_i = 0 \quad (4.22)$$

$$\mu_j \left(-p_j \|\mathbf{g}_j\|^2 + \sum_{i=j+1}^J p_i \|\mathbf{g}_i\|^2 + P_{\text{SIC}}\right) = 0 \quad \forall j \in \mathcal{J} \quad (4.23)$$

$$\theta_j (p_j - P_{\text{max}}) = 0 \quad \forall j \in \mathcal{J} \quad (4.24)$$

$$P_{\text{SIC}} - p_j \|\mathbf{g}_j\|^2 + \sum_{i=j+1}^J p_i \|\mathbf{g}_i\|^2 \leq 0 \quad \forall j \in \mathcal{J} \quad (4.25)$$

$$p_j - P_{\text{max}} \leq 0 \quad \forall j \in \mathcal{J} \quad (4.26)$$

$$\theta_j \geq 0 \quad \forall j \in \mathcal{J} \quad (4.27)$$

$$\mu_j \geq 0 \quad \forall j \in \mathcal{J} \quad (4.28)$$

In this scenario, the complementary slackness conditions are represented by (4.23), (4.24), the primal feasibility is represented by (4.25), (4.26) and the dual feasibility is defined by (4.27) and (4.28).

**a) Case 1**

In this case, the minimization of the average system BER is restricted only by the maximum transmission power of each user (4.4b). Therefore, based on the corresponding KKT conditions (4.21), (4.22), (4.24), (4.26) and (4.27), each user will simply transmit with its maximum authorized power :

$$p_j = P_{\max} \quad \forall j \in \mathcal{J} \quad (4.29)$$

**b) Case 2**

This case entails both of the maximum transmission power (4.4b) and the difference between received power constraints in (4.4c). Similarly to the second case of scenario 1, the KKT conditions can not provide a closed-form expression for the optimal solution which can be then calculated by an iterative gradient-based process as in Algorithm. 1. In fact, the inequalities constraints (4.4b) and (4.4c) are converted to equalities by introducing non-negative slack variables as follows :

$$p_j + t_j^2 = P_{\max} \quad \forall j \in \mathcal{J} \quad (4.30)$$

$$p_j \|\mathbf{g}_j\|^2 - \sum_{i=j+1}^J p_i \|\mathbf{g}_i\|^2 - z_j^2 = P_{\text{SIC}} \quad \forall j \in \mathcal{J}_J$$

## 4.4 Numerical results and analysis

We consider an uplink transmission system of  $K = 4$  subcarriers and  $J = 6$  users, leading to a system overload of 150%. The propagation environment is an AWGN channel with two variants. First, users are positioned at equal distance from the base station with a normalized path loss, i.e  $h_{kj} = 1$ . Then, they are uniformly scattered over the cell while experiencing different path loss coefficients. The first case may be considered as the worst conditions for conducting SIC process specially when users transmit with equal powers. Indeed, all users signals will arrive at the base station with the same power. Hence, it will be difficult to distinguish them. The total system power and the maximum individual power are set respectively to  $P_{\text{tot}} = 10$  dB and  $P_{\max} = \frac{P_{\text{tot}}}{J}$ . The optimization problem  $P_2$  is solved at the base station knowing the corresponding users CSI and then, the obtained powers are transmitted back to users by a control channel. For devices at fixed positions, the propagation environment is considered as static, hence the power allocation process is performed only once, which induces a small signaling overhead. The derived powers from  $P_2$  are used in the real transmission chain with a MMSE-OSIC receiver.

### 4.4.1 Perfect AWGN channel

In Figure 4.1, the average BER is investigated w.r.t.  $P_{\text{SIC}} \geq 0$  when (4.4c) is involved, in both scenarios 1 and 2, for the first propagation environment with a normalized path loss, i.e.,  $h_{kj} = 1$ ,  $\text{SNR}_1 = 20$  dB and  $\text{SNR}_2 = 23$  dB. It is seen that, for

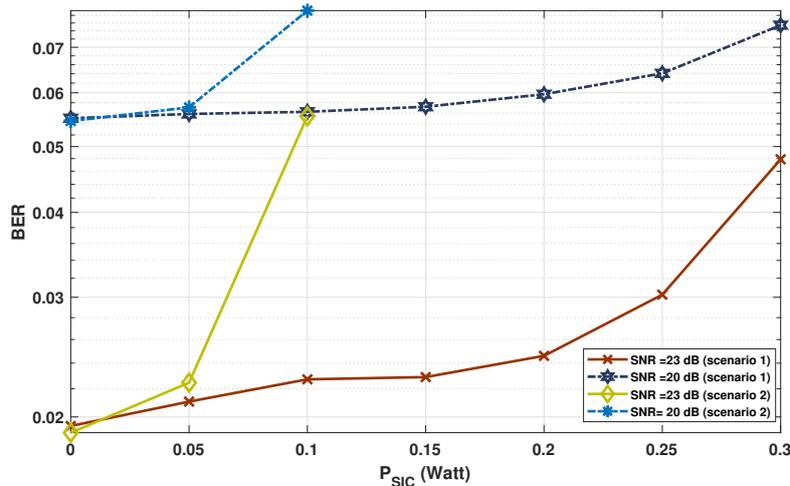


Fig. 4.1. BER w.r.t.  $P_{SIC}$  for  $SNR_1 = 20\text{dB}$ ,  $SNR_1 = 23\text{ dB}$  and AWGN channel.

both SNR values and scenarios, BER is increasing with  $P_{SIC}$ . Low  $P_{SIC}$  allows to distribute the power between users while allocating more power to the weakest users compared to higher  $P_{SIC}$  values. This latter penalizes the weakest users which has low power that significantly degrades the average system BER. Therefore, for the defined system conditions, we opt for  $P_{SIC} = 0$ .

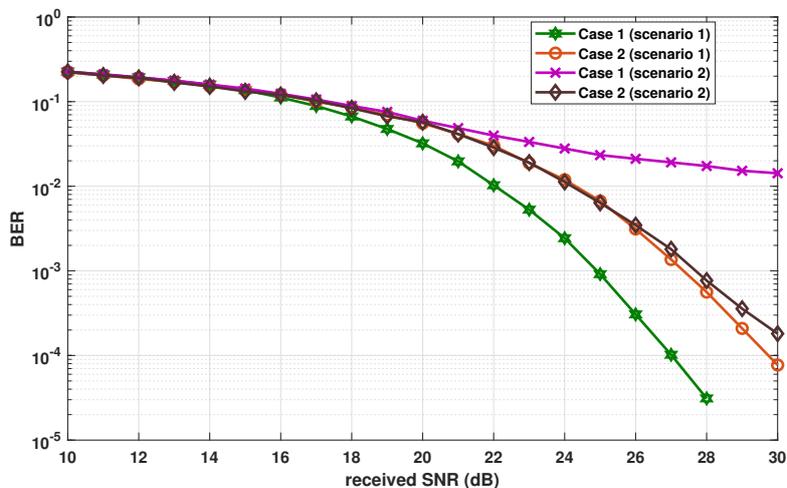


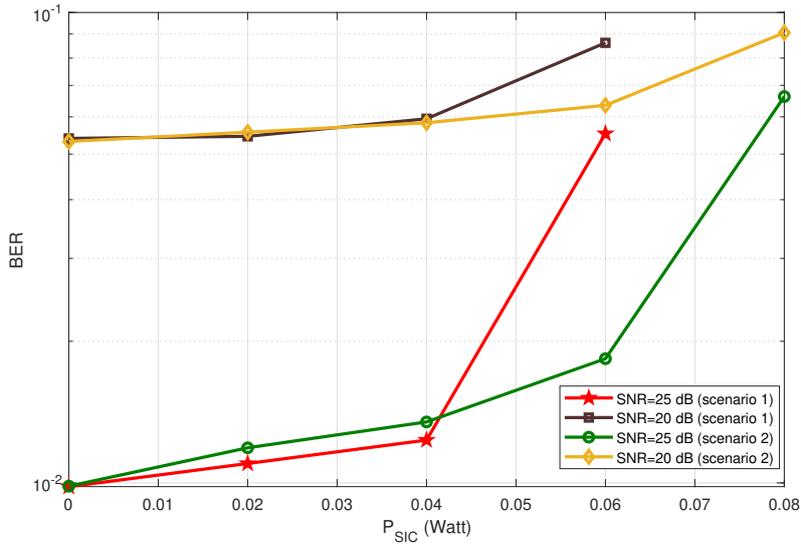
Fig. 4.2. Performance comparison for the different cases and scenarios for AWGN channel,  $J = 6$  and  $K = 4$ .

Figure 4.2 shows the comparison of the BER for the different studied constraints combinations in AWGN channel and normalized path loss. We observe that the case 1 of scenario 1, i.e., with only the total power constraint (4.4a), has the best performance. The total power constraint offers the maximum degree of freedom by allowing power transfer among users, which leads to an important improvement of the average BER. However, we remark that case 1 of scenario 2, i.e., with only the maximum individual power constraint (4.4b), has the worst BER performance. In this case, all users transmit with their maximum authorized power, and hence all signals are received with the same power. In addition, in this case, users' SINR

#### 4.4. NUMERICAL RESULTS AND ANALYSIS

are distinguished only by their selected spreading sequences which are not sufficient to ensure a correct decoding process. Therefore, the signals separation is affected by a severe error propagation, which dramatically degrades the system BER with an error floor of about  $10^{-2}$  at 30 dB. Then, adding the SIC power distribution constraint (4.4c) to both scenarios leads to similar results, not as good as case 1 of scenario 1, but achieving a significant performance improvement compared to case 1 of scenario 2. This constraint prevents all users from transmitting with their maximum power. It ensures an adequate received power distribution, leading to a reduced error propagation phenomenon and hence, a better signal detection. Moreover, it has to be noted that, in case 2 of scenario 2, the constraints combination does not allow the available power budget to be entirely consumed, then the results would be further improved considering equal total received powers to that of scenario 1. Furthermore, we notice that, for  $\text{SNR} \leq 20$  dB, there is no significant difference between all the studied cases. In fact, due to the high noise level, the iterative decoding process is damaged, which induces to an important BER.

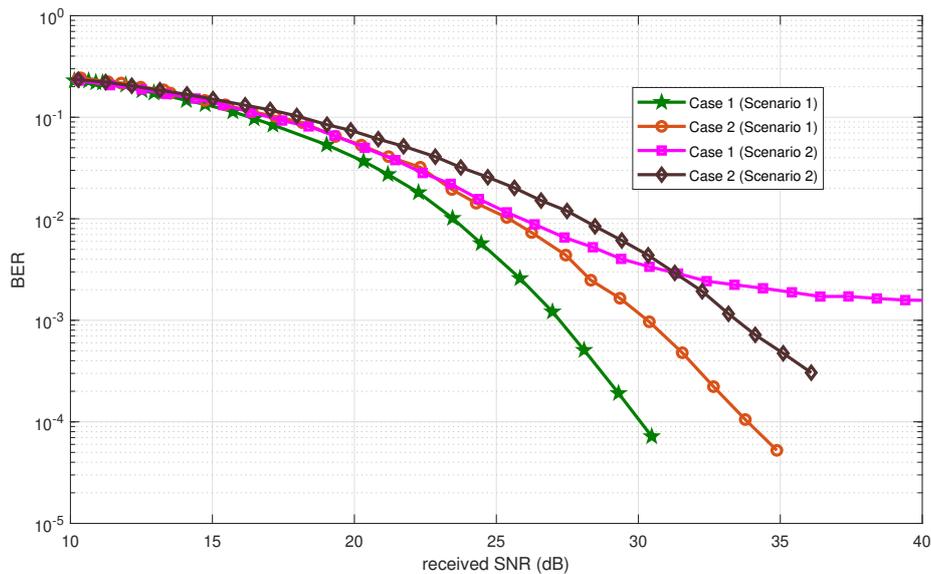
#### 4.4.2 AWGN channel with different path losses



**Fig. 4.3.** BER w.r.t.  $P_{SIC}$  for  $\text{SNR}_1 = 20\text{dB}$ ,  $\text{SNR}_2 = 25$  dB and AWGN channel with different path loss.

Figure 4.3 compares the achieved BER, averaged over AWGN channels with different path losses, in both scenarios w.r.t  $P_{SIC} \geq 0$  for  $\text{SNR} = 20$  dB and  $\text{SNR} = 25$  dB. Similarly to the cases with a perfect AWGN channel, the system BER increases with  $P_{SIC}$  value. In fact,  $P_{SIC} = 0$  leads to the best BER performance in both scenarios because it offers the fairest power distributions among users and hence a better decoding process. For the rest of simulations,  $P_{SIC}$  is set to 0.

Figure 4.4 illustrates a comparison of the achieved BER for all the reported cases in the different scenarios while averaging over different user distributions in the cell w.r.t the average received SNR. Similarly to the case of normalized path loss, the case 1 of the first scenario with only a total system power constraint (4.4a) shows



**Fig. 4.4.** Performance comparison for the different cases and scenarios for AWGN channel with different path losses,  $J = 6$  and  $K = 4$ .

the best performance due to the offered degree of freedom while allowing the power transfer among users, which provides an adequate power distribution. However, in the case 1 of scenario 2, an error floor is revealed at the BER value of  $10^{-3}$ . This can be explained by the experienced error propagation problem. Nevertheless, the diversity generated by the different users path losses permits to improve the system performance compared to the same case with normalized path loss in Figure 4.2. Then, adding the power difference constraint (4.4c) in the second case of scenario 2 leads to an acceptable power distribution with a sufficient users diversity that alleviates the error propagation issue and permits to achieve a better performance compared to the first case particularly for high SNR values, i.e.,  $\text{SNR} \geq 31$  dB. Moreover, we recognize also that the case 2 of scenario 1 with the total system power and power difference constraints has a higher BER compared to the first case of the same scenario. This is due to the fact that, in these conditions, the received power difference constraint may restrict the allocated power values and penalizes the weakest users with low power values, which induces a performance degradation.

We can observe that the performances in this case, with different path losses, have the same behavior as those of the first case with a normalized path loss in Figure 4.2. However, except in case 1 of scenario 2 with maximum per user power, the other three cases experience a performance degradation when users are at different distances from the base station. This is mainly due to the reduction in the received powers which results in lower users' SINR and consequently a degraded performance. The BER improvement in case 1 of scenario 2 with different path losses is provided by the channel diversity which helps to mitigate the SIC error propagation compared to the case of equal received power.

## 4.5 Conclusion

MUSA scheme with a typical SIC receiver is a power sensitive technique. In fact, the performance of the decoding process may be dramatically deteriorated in case of unsuitable power distribution. However, to the best of our knowledge, the problem of BER minimization for MUSA schemes with an optimal power allocation approach has not been investigated. For that reason, in this chapter, we handled this issue by proposing a novel algorithm to minimize the average system BER in two scenarios with a global system power constraint or a maximum individual power constraint. Each scenario was treated in two different cases whether the constraint on the difference between the received powers is applied or not.

The resulting problem was not convex. Therefore, in order to facilitate the resolution process, a simplified expression of user BER was applied to ensure a global optimal solution for our model. Then, we demonstrated that for the practical case with a maximum per-user power constraint, the received power difference constraint permits to alleviate the error propagation phenomenon by ensuring an adequate power distribution with a sufficient diversity between users' SINR. Therefore, each user's signal is properly decoded by successively canceling the signals of interfering users, and leads to performance improvement compared to the case with equal power allocation.

In short, thanks to the received power distribution constraint, the case with maximum per-user power achieves a performance bounded by the best performance corresponding to the ideal case of total power constraint with the advantage of power transfer among users, and the worst case corresponding to equal users power.



# Autonomous power decision for grant free access

## 5.1 Introduction

As explained in Chapter 2, in the existing technologies, users used to go through a contention-based random access protocol for data transmission. For the LTE/LTE-A network, the eNB initially broadcasts information about the available 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 for resources reservation. Therefore, this process may be very expensive in terms of signaling overhead, especially for mMTC devices.

According to [164], for the transmission of small payload of 30 bytes, 45% of the total transmitted bytes are dedicated for signaling overhead. This induces an excessive waste of resources and high energy consumption 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. As a consequence, the random access strategy may be a performance bottleneck in mMTC scenarios.

In this context, NOMA with the grant free access strategy has gained much interest, and it has been promoted by the scientific community as a promising solution to support mMTC scenarios with minimum signaling overhead, which ensures low energy consumption and long battery life [50]. Authors in [22] presented the evolution steps towards uplink NOMA schemes with grant free access. They suggested two possible options : each user can either go with RACH-based or RACH-less grant free transmission.

Authors in [165] 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 total throughput while guaranteeing a rate fairness between users. In [166], 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 by the SIC technique as a kind of user diversity. 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 grant free access is the estimation of the number of active users. This issue was stressed in [167] 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 [168], a sinusoidal code was proposed for the signals' separation. The proposed spreading sequences allow using non-iterative algorithms for multi-user detection without prior knowledge of the channel state information and the number of active users. Authors in [169] dealt with the problem of packet collisions without retransmission opportunities. A novel grant free access framework was proposed where the collision between the detected users is seen as interference by the remaining users. Moreover, the system performance was evaluated analytically by providing simplified expressions of the outage probability and the system throughput.

SCMA has also been studied with grant free access protocols. For instance, authors in [170] studied the application of SCMA with a faster than Nyquist signaling, which improved the spectral efficiency, but at the expense of 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. In [171] an iterative message passing algorithm, based on the belief propagation, was investigated for grant free access SCMA. The proposed algorithm permits jointly estimating the channel coefficients, identifying the number of active users, and detecting the transmitted data while improving the bit error rate compared to the other techniques.

Unlike the SCMA scheme, which requires prior assignment of codebooks, MUSA has great potential to enable grant free access by randomly generating the spreading sequence at UE using a predefined constellation. More precisely, 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. However, regardless RACH-based or RACH-less grant free access, some information about MUSA system should be known or detectable by the base station for a proper decoding process namely, the transmission parameters, the user identification, e.g., spreading sequence, the channel estimation and the synchronization information [21].

Generally, the UE is in sleep mode. Once it has data to transmit, it wakes up, synchronizes in downlink and receives some broadcast system information. User transmits its available data with randomly selected spreading sequence, which should be also known to the base station for signals detection. The RACH-based grant free access strategy ensures an uplink synchronization and the required information such as spreading sequences could be exchanged between UE and base station during RACH process. However, in RACH-less strategy, no prior communication is establi-

shed with the base station. In this case, two possible solutions have been proposed. The first one concerns the uplink transmission format where a preamble can be transmitted with user data [51]. The preamble may include the selected spreading sequences and some other information such as radio resources positions and retransmissions timing. It can be also used for channel estimation and timing acquisition. The second solution, which is out of the scope of this work, relies on blind multi-user detection using SIC receiver, with no prior information on user sequences, channels or number of active users, while taking advantage of the inherent near-far phenomenon and the short sequences length [111]. However, the pool of sequences must be known beforehand by the base station and the UE for a correct estimation. In order to reduce the decoding complexity of blind multi-user detection, the size of this pool should be reduced while maintaining a massive capacity of connectivity [172].

As already explained in chapter 4, the SIC technique suffers from a serious error propagation problem when the received powers are similar. Therefore, in order to ensure a proper decoding process with an attenuated error propagation, the signals must arrive at the base station with a sufficient power difference. In Chapter 4, we proposed a centralized power allocation approach where the base station knows the channel state information and the spreading sequences of all users. However, in this chapter, we are interested in the case of grant free access where each user performs a blind transmission without any information on the propagation environment and the interfering users, which makes the determination of its transmission power more complex.

The 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 [173], [174]. MAB techniques can be applied to the problem of dynamic resource allocation by balancing between the exploration and exploitation phases. At each time slot, each agent selects an arm, i.e., representing the physical resource to be shared, among an available 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 [175] 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 achieved data rates. However, these may be carried on many bits, which increases the signaling overhead ; hence, it may not really be adapted for mMTC scenarios. MAB was also merged with NOMA schemes in [176] where a distributed NOMA-based MAB approach was proposed to handle the channel access problem in cognitive radio networks. In [177], the MAB algorithms were applied in the LTE cellular network for an autonomous subcarrier 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 schemes. The characteristics of spreading sequences and the principle of the SIC receiver make the power decision further complex. Therefore, in this chapter, we aim at dealing 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. The 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 system BEP while considering the inter-user interference and the effect of error propagation. The optimal powers are obtained as the solution of the minimization problem of the global average BEP. Then, 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 to known index-based MAB algorithms adapted to the power selection problem. In this chapter, 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 chapter is organized as follows. A closed-form expression for users' bit error probability is derived in Section 5.2. Then, the proposed algorithm for autonomous power decision is described in Section 5.3. The multi-armed bandit algorithms and the studied scenarios are introduced in Section 5.4. A comparison of all power decision approaches is provided in Section 5.5. Numerical results and performance analysis are conducted in Section 5.6, and conclusions are drawn in Section 5.7.

## 5.2 BEP Analysis

The active users share the available resources with a grant free access strategy. Their signals are then detected at the base station using an MMSE-OSIC receiver. However, its critical error propagation issue may significantly deteriorate the system performance and make 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 [178]. Therefore, the average system BEP is well approximated as :

$$P_{b,\text{MMSE-SIC}} \approx \frac{1}{J \log_2(M)} \sum_{j=1}^J P_{e_j} \quad (5.1)$$

where  $P_{e_j}$  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 with no error propagation (NEP); (ii) imperfect SIC with error propagation (EP).

### 5.2.1 Perfect SIC without Error Propagation

The received signal at the  $j$ -th iteration is represented as :

$$\mathbf{y}^j = \sqrt{p_j} \mathbf{g}_j x_j + \sum_{i=j+1}^J \sqrt{p_i} \mathbf{g}_i x_i + \mathbf{n}^j, \quad (5.2)$$

where  $\mathbf{g}_j = \mathbf{h}_j \odot \mathbf{s}_j$  with  $\mathbf{h}_j$  and  $\mathbf{s}_j$  are, respectively, the channel vector and the spreading sequence of the  $j$ -th user.  $\mathbf{n}^j$  is the additive white Gaussian noise vector at the  $j$ -th iteration.

The system BEP is then calculated similarly as for the MMSE receiver while updating the MMSE matrix (2.23) at each iteration. The  $j$ -th user SINR is calculated as :

$$\beta_j^{\text{NEP}}(\mathbf{p}) = \frac{p_j |\mathbf{w}_j^H \mathbf{g}_j|^2}{\sum_{i=j+1}^J p_i |\mathbf{w}_j^H \mathbf{g}_i|^2 + \sigma^2 \|\mathbf{w}_j^H\|^2}, \quad (5.3)$$

For a QPSK modulation and assuming the inter-user interference as noise [179], the  $j$ -th user SEP is approximated as [178] :

$$P_{e_j} \approx 2Q\left(\sqrt{\beta_j^{\text{NEP}}(\mathbf{p})}\right) \left(1 - 0.5Q\left(\sqrt{\beta_j^{\text{NEP}}(\mathbf{p})}\right)\right). \quad (5.4)$$

### 5.2.2 Imperfect SIC with Error Propagation

In that case, the BEP of each user depends on the previously decoded users. In this section, we rely on the proposed approach in [179] where the SEP of the  $j$ -th user is calculated as :

$$P_{\varepsilon_j} = \sum_{i=0}^{N_j-1} P\left\{\varepsilon_j | \mathbf{b}_i^j\right\} P\left\{\mathbf{b}_i^j\right\}, \quad (5.5)$$

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 one otherwise. Each sequence refers to the state, correctly decoded or not, of all the previously  $(j-1)$  decoded users. The event  $\varepsilon_j$  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  $\mathbf{b}_i^j$ . For example, for a system with only  $J=2$  users, there are two sequences presenting the two possible decoding states of the first user :  $\mathbf{b}_0^2 = (0)$ , if the first user was correctly detected, and  $\mathbf{b}_1^2 = (1)$ , if it was erroneously detected. Therefore, the second user has two possible error probabilities  $P = \{\varepsilon_j | \mathbf{b}_0^2\}$  and  $P\{\varepsilon_j | \mathbf{b}_1^2\}$  depending on the detection of the first user.

Considering an eventual error propagation occurrence, the received signal at the  $j$ -th SIC iteration is represented as :

$$\mathbf{y}^j = \sqrt{p_j} \mathbf{g}_j x_j + \sum_{i=j+1}^J \sqrt{p_i} \mathbf{g}_i x_i + \sum_{k=1}^{j-1} \sqrt{p_k} \mathbf{g}_k (x_k - \hat{x}_k) + \mathbf{n}^j, \quad (5.6)$$

where  $\hat{x}_k$  is the estimation of  $x_k$ . The additional term compared to (5.2) 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  $\mathbb{E}\{\mathbf{n}_{eq} \mathbf{n}_{eq}^H\} = (\sum_{k=1}^{j-1} p_k \|g_k\|^2 \mathbb{E}\{\|x_k - \hat{x}_k\|^2\} + \sigma^2) \mathbf{I} = (\sum_{k=1}^{j-1} p_k \|g_k\|^2 \delta_k d + \sigma^2) \mathbf{I}$ . We define  $d$  as the square of the euclidean distance between the neighboring symbols, where  $d = 4 \sin^2(\frac{\pi}{M})$  for M-PSK modulation, and  $\delta_k = 1$  if  $x_k \neq \hat{x}_k$  and zero otherwise. As a consequence, the SINR of the  $j$ -th user, corresponding to the detection combination  $\mathbf{b}_i^j$ , is calculated as follows :

$$\beta_{j,i}^{\text{EP}}(\mathbf{p}) = \frac{p_j |\mathbf{w}_j^H \mathbf{g}_j|^2}{\sum_{i=j+1}^J p_i |\mathbf{w}_j^H \mathbf{g}_i|^2 + (\sum_{k=1}^{j-1} p_k \|g_k\|^2 \delta_k d + \sigma^2) \|\mathbf{w}_j^H\|^2} \quad (5.7)$$

Two main terms should be calculated to obtain the user SEP. Starting by the conditional probability, which is calculated according to (5.4) and (5.7), we have :

$$P\{\varepsilon_j | b_i^j\} = 2Q\left(\sqrt{\beta_{j,i}^{\text{EP}}(\mathbf{p})}\right) \left(1 - 0.5Q\left(\sqrt{\beta_{j,i}^{\text{EP}}(\mathbf{p})}\right)\right). \quad (5.8)$$

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

$$P\{\mathbf{b}_i^j\} = P\{\cap_{n=1}^{j-1} b_{i,n}^j\} = \prod_{n=1}^{j-1} P\{b_{i,n}^j | \cap_{m=1}^{n-1} b_{i,m}^j\}, \quad (5.9)$$

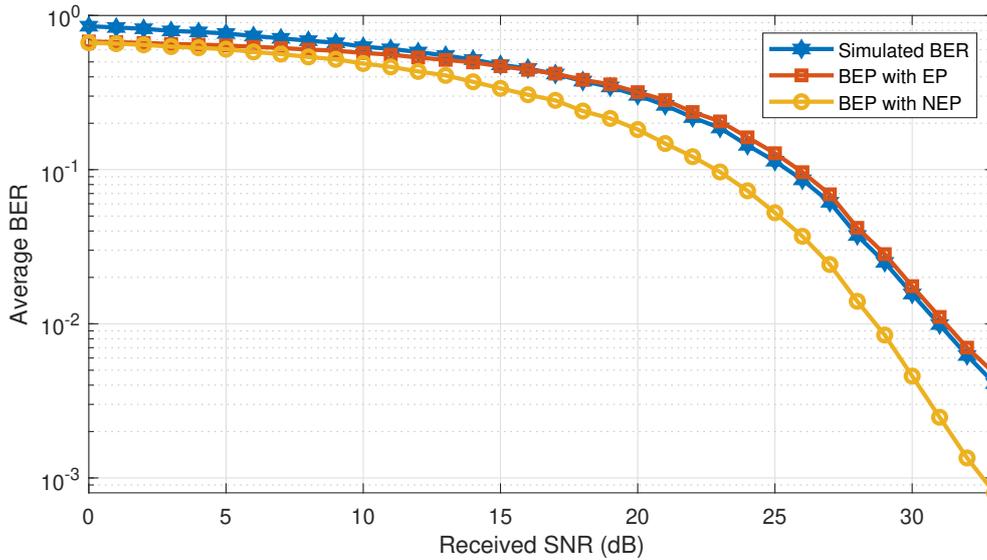
where  $P\{b_{i,n}^j | \cap_{m=1}^{n-1} b_{i,m}^j\}$  is the probability that the  $n$ -th symbol is correctly decoded or not, i.e.,  $b_{i,n}^j = 0$  or  $b_{i,n}^j = 1$ , respectively, conditioned on the estimation of the previously decoded  $(n-1)$  symbols. It is calculated as :

$$P\{b_{i,n}^j | \cap_{m=1}^{n-1} b_{i,m}^j\} = \quad (5.10)$$

$$\begin{cases} 1 - 2Q\left(\sqrt{\beta_{n,i}^{\text{EP}}(\mathbf{p})}\right) \left(1 - 0.5Q\left(\sqrt{\beta_{n,i}^{\text{EP}}(\mathbf{p})}\right)\right) & \text{if } b_{i,n}^j = 0 \\ 2Q\left(\sqrt{\beta_{n,i}^{\text{EP}}(\mathbf{p})}\right) \left(1 - 0.5Q\left(\sqrt{\beta_{n,i}^{\text{EP}}(\mathbf{p})}\right)\right) & \text{otherwise.} \end{cases} \quad (5.11)$$

Figure 5.1 compares the simulated average BER, i.e., averaged over the spreading sequences and user positions, and the analytical average BEP obtained by the proposed expression in (5.1) for 150% of overload, i.e.,  $J = 12$  and  $K = 4$ , and AWGN channel with uniformly distributed users over the cell with a radius of 5 km w.r.t. the global received SNR. We remark that the BEP expression that takes into account the error propagation phenomenon almost matches 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 (5.4) is more robust for high SNR.

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



**Fig. 5.1.** Performance comparison of the simulated BER and the analytical BEP for an AWGN channel with different users' path losses and equal transmission powers.

$$OP_1 \begin{cases} \min_{\mathbf{p}} & \frac{1}{J \log_2(M)} \sum_{j=1}^J \sum_{i=0}^{N_j-1} P \{ \varepsilon_j | \mathbf{b}_i^j \} P \{ \mathbf{b}_i^j \} \\ \text{s.t.} & p_j \leq P_{max} \quad \forall j \in \mathcal{J} \end{cases} \quad (5.12)$$

$$(5.12a)$$

where  $\mathcal{J} = \{1, 2, \dots, J\}$  is the set of active users. The derived expression of user's BEP is quite complicated to analyze theoretically with the KKT conditions. Therefore, we use an advanced optimization algorithm, e.g., particle swarm optimization [180], to solve the power allocation problem above. This algorithm is known to be efficient for complex problems [181].

### 5.3 Autonomous Power Decision

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

The key idea is to perform an iterative algorithm that takes advantage of the natural base station acknowledgment (ACK). Each user gradually updates its transmitted power from the received ACK in order to converge toward the nearest power level from  $\mathbf{p}^{opt}$ . 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 UE. Then, the base station detects the user signal and compares its transmission power with  $p^{j,opt}$ , which the 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 the user should simply transmit with its maximum authorized power  $P_{max}$  ; this case may be gainful for the cell edge users that experience bad propagation conditions ; (2) ACK = 2 if  $p^j > p^{j,opt}$  ; (3) ACK = 1 if  $p^j < p^{j,opt}$  ; and (4) ACK = 0 if  $p^j \approx p^{j,opt}$ . Each user updates its interval by shifting the  $p_{min}^j$  and  $p_{max}^j$  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 the user of this occurrence. In this case, the UE will try to initialize its power interval while taking advantage of the previously sent packets. This process is described in details in Algorithm 2.

---

**Algorithm 2:** Autonomous power decision

---

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

**Ensure:** **p**

- 1: Each user picks up its spreading sequence
  - 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  $p^{j,opt}$
  - 5: The BS compares each user power  $p^j$  with the nearest power level from  $p^{j,opt}$
  - 6: BS send an acknowledgment to each user :
    - a) If  $p^{j,opt} = P_{max} \Rightarrow \text{ACK} = 3$
    - b) If  $p^j > p^{j,opt} \Rightarrow \text{ACK} = 2$
    - c) If  $p^j < p^{j,opt} \Rightarrow \text{ACK} = 1$
    - d) If  $p^j \approx p^{j,opt} \Rightarrow \text{ACK} = 0$
  - 7: If the propagation environment is changed, the BS sends one extra bit ACK :  
Stat = 1
  - 8: Each user updates its  $p_{min}^j$  or  $p_{max}^j$  :
    - a) If ACK = 3  $\Rightarrow p^j = P_{max} > 0$
    - b) If ACK = 2  $\Rightarrow p_{max}^j = p^j \leq P_{max}$ 
      - If Stat = 1  $\Rightarrow p_{min}^j = 0$
    - c) If ACK = 1  $\Rightarrow p_{min}^j = p^j > 0$ 
      - If Stat = 1  $\Rightarrow p_{max}^j = P_{max}$
    - d) If ACK = 0  $\Rightarrow$  no update
  - 9: Return to step 2
- 

In order to ensure the convergence of the algorithm, the propagation environment and the selected spreading sequences must not change too fast. However, as will be

seen in the simulation results, the proposed algorithm converges to the near-optimal power value quite quickly.

## 5.4 Power Allocation with Multi-Armed Bandits

In this section, we revisit three known MAB algorithms, i.e.,  $\epsilon$ -greedy, upper confidence bound (UCB1), and Thompson sampling (THS), that we apply to our autonomous power selection problem. MAB is a model with  $P$  resources, called arms, each of them being associated with 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  $r_j^t(a_j)$ . Based on this reward and the number of times each arm has been played so far,  $n_j^t(a_j)$ , 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  $R_j^{MAB}$  defined as the difference between the rewards obtained using the chosen policy and the expected reward we would obtain if the best arm was always played, i.e.,  $r_j^*$ . The  $j$ -th user regret during a maximum period of  $T$  slots is calculated as follows :

$$R_j^{MAB} = Tr_j^* - \sum_{t=1}^T \mathbb{E}\{r_j^t(a_j)\} \quad (5.13)$$

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 acknowledgment 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 using the selected arm  $a_j$  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 of the other users. However, the successful decoding event of one packet using the SIC receiver depends on the successful decoding of others. Hence, every user has interest in good power selection for the other users and not only for itself. Scenario 2, we propose hereafter, takes into account this fact.

b) **Scenario 2 :**

The base station acknowledgment 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 has erroneously been 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 acknowledgment  $\{b_{2,j}^t, b_{1,j}^t\} \in \{11, 10, 00\} = \{3, 2, 0\}$ . The case where  $\{b_{2,j}^t, b_{1,j}^t\} = 01$  is not possible because  $b_{1,j}^t = 1$  means that all packets have been correctly decoded, including the  $j$ -th user packet, and hence,  $b_{2,j}^t$  is automatically equal to one. In order to meet the conditions of the convergence theorems derived in [182], the rewards should be supported in  $[0, 1]$ . Therefore, user rewards are defined as a normalization of the associated acknowledgments, 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.

### 5.4.1 Upper confidence bound

---

**Algorithm 3:** UCB1 algorithm.

---

**Require:**  $\theta$  and  $P$

Each user plays all the arms once during  $P$  plays :

**for**  $t = P + 1 : T$  **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;

      ▶  $n_j^t(a_j) = n_j^{t-1}(a_j) + 1$

      ▶  $m_j^t(a_j) = \frac{1}{n_j^t(a_j)} \sum_{i=1}^t r_j^i(a_j)$

**end**

**end**

---

UCB1 was inspired by the Agrawal's index-based policy [182]. This algorithm has a uniform logarithmic regret over time. Generally, the UCB family relies on a confidence interval on the average reward of each arm [183]. The 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  $a_j$ . The UCB1 index is defined as follows, at

time slot  $t$  :

$$\frac{1}{n_j^t(a_j)} \sum_{i=1}^t r_j^i(a_j) + \sqrt{\frac{\theta \log(t)}{n_j^t(a_j)}} \quad (5.14)$$

where  $\theta > 0$  is the exploration parameter. Originally, UCB1 was proposed with  $\theta = 2$ ; however, authors in [183] mentioned that  $\theta = 0.5$  performs better empirically.

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 3. The calculated index (5.14) 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 in the face of uncertainty”, which means that the less visited arm seems more uncertain, and thereby, it may optimistically be the best arm to play.

### 5.4.2 $\epsilon$ -Greedy

---

**Algorithm 4:**  $\epsilon$ -decreasing greedy algorithm.

---

**Require:**  $L$  and  $P$

**for**  $t = 1 : T$  **do**

**for**  $j = 1 : J$  **do**

    Select a random arm with probability  $\epsilon(t) = \min \left\{ 1, \frac{LP}{t} \right\}$

    Select with probability  $1 - \epsilon(t)$  the best arm :  $\operatorname{argmax}_{a_j} m_j^{t-1}(a_j)$

    Update the following variables :

      ▶  $n_j^t(a_j) = n_j^{t-1}(a_j) + 1$

      ▶  $m_j^t(a_j) = \frac{1}{n_j^t(a_j)} \sum_{i=1}^t r_j^i(a_j)$

**end**

**end**

---

This algorithm deals with the exploration and the exploitation dilemma randomly. At each iteration, each user either explores arbitrarily a new arm with probability  $\epsilon$  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  $\epsilon$ , the system regret evolves linearly overtime instead of being logarithmic. On the one hand, for a high  $\epsilon$  value, i.e.,  $\epsilon \approx 1$ , the 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  $\epsilon$  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  $\epsilon$  value is a critical parameter. A revised version called  $\epsilon$ -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

[175], [182] :

$$\epsilon(t) = \min \left\{ 1, \frac{CP}{d^2t} \right\} = \min \left\{ 1, \frac{LP}{t} \right\} \quad (5.15)$$

where  $L = \frac{C}{d^2} > 0$  is the exploration parameter. However, the main challenge of this policy is how to properly set the value of  $L$ . The  $\epsilon$ -decreasing greedy algorithm is described in details in Algorithm 4.

### 5.4.3 Thompson Sampling Algorithm

This approach shows a robust performance for stochastic problems and sometimes outperforms other MAB algorithms [184], [185]. The 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  $\alpha_{j,k} = \gamma_{j,k} = 2 \forall j \in \{1, \dots, J\}$  and  $\forall k \in \{1, \dots, P\}$ , where  $k$  refers to the arm index among  $P$  power levels. Then, inspired by the case where rewards follow a binomial distribution [186] 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.,  $i_{j,k} \sim \beta(\alpha_{j,k}, \gamma_{j,k}) \forall k \in \{1, \dots, P\}$  and  $\forall j \in \{1, \dots, J\}$ . The arm with the highest index, i.e.,  $\hat{i}_{j,q} = \max_{k \in \{P\}}(i_{j,k}) \forall j \in \{1, \dots, J\}$ , where  $\{P\}$  is the set of power levels, 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. This algorithm is described in details in Algorithm 5.

---

**Algorithm 5:** Thompson sampling algorithm.

---

**Require:**  $P$  and  $\alpha_{j,k} = \gamma_{j,k} = 2 \forall k = 1 \dots P$  and  $\forall j = 1 \dots J$

**for**  $t = 1 : T$  **do**

**for**  $j = 1 : J$  **do**

Select a sampled index from the beta distribution of each arm

$i_{j,k} \sim \beta(\alpha_{j,k}, \gamma_{j,k}) \forall k = 1, \dots, P$

Play the arm  $a_j$  with the highest index  $\hat{i}_{j,q} = \max_{k \in \{P\}}(i_{j,k})$

Update the following variables :

▶  $n_j^t(a_j) = n_j^{t-1}(a_j) + 1$

▶  $m_j^t(a_j) = \frac{1}{n_j^t(a_j)} \sum_{i=1}^t r_j^i(a_j)$

▶  $\alpha_{j,k} = \alpha_{j,k} + 3r_j^t(a_j)$

▶  $\gamma_{j,k} = \gamma_{j,k} + 3(1 - r_j^t(a_j))$

**end**

**end**

---

## 5.5 Complexity and Overhead Analysis

A quantitative comparison of all the examined techniques is summarized in Table 5.1. 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.

The centralized power allocation algorithm computes, at the base station, the powers to be assigned to 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 one or two 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 the 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.

	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
$\epsilon$ -decreasing greedy	Scenario1 : 1 bit	$O(P)$	Random with $\epsilon$ probability
	Scenario2 : 2 bits		
UCB1	Scenario1 : 1 bit	$O(P)$	Index-based
	Scenario2 : 2 bits		
Thompson sampling	Scenario1 : 1 bit	$O(P)$	Bayesian distribution
	Scenario2 : 2 bits		

TABLE 5.1: Quantitative comparison of the signaling overhead and the complexity at user equipment in each iteration for all algorithms.

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 the UE during this process.

All the MAB techniques have the same signaling overhead and algorithmic complexity for each transmission attempt. UCB1,  $\epsilon$ -decreasing greedy, and Thompson sampling can be seen as index-based policies. Hence, the algorithmic complexity consists of sorting  $P$  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  $O(P)$ . Furthermore, the generated signaling ove-

thead depends particularly on the applied learning scenario. In Scenario 1, the index 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 two bits. It is worth noting that the computational complexity is not considered here. The complexity of calculating a sampled index from the beta function for each arm with the Thompson sampling algorithm is higher than that of the UCB1 and  $\epsilon$ -decreasing greedy indexes.

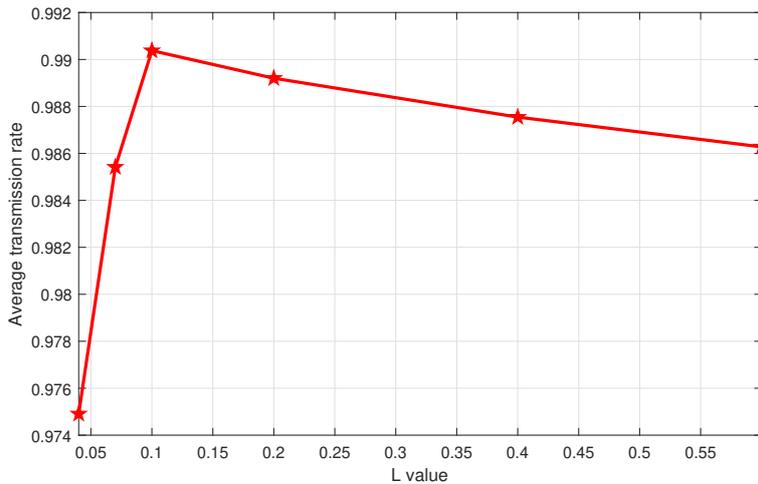
## 5.6 Numerical Results and Analysis

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  $P = 10$  possible power levels in the interest of selecting the appropriate value ensuring the best performance in both Scenarios 1 and 2. The user spreading sequences are normalized in 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 the UCB1 algorithm, the exploration of new power values is conducted by the parameter  $\theta$ . As mentioned above,  $\theta = 0.5$  provides the best performance. Therefore, this value is chosen for our simulations. The other simulation parameters are given in Table 5.2.

Channel	AWGN with path losses
Users	$J = 12$
Subcarriers	$K = 8$
Maximum individual power	20 dBm
$P$	10 levels
Power levels	[10 : 10 : 100] mW
Noise power	$\sigma^2 = -14$ dBm
$T$	1000 slots
$\theta$	0.5

TABLE 5.2: Simulation settings

The performance of the  $\epsilon$ -decreasing greedy algorithm depends on the  $\epsilon$  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  $\epsilon$ -decreasing greedy approach is to handle the exploration and the exploitation dilemma by properly setting the value of  $L$  in equation (5.15). Figure 5.2 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 terms of the average transmission rate and hence it is kept for the rest of the simulations.

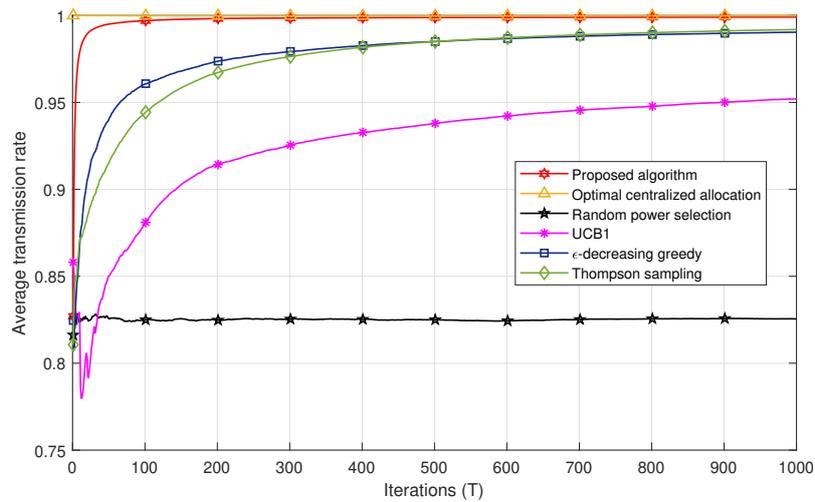


**Fig. 5.2.** Average transmission rate of  $\epsilon$ -decreasing greedy for different  $L$  values after  $T = 1000$  iterations in Scenario 1.

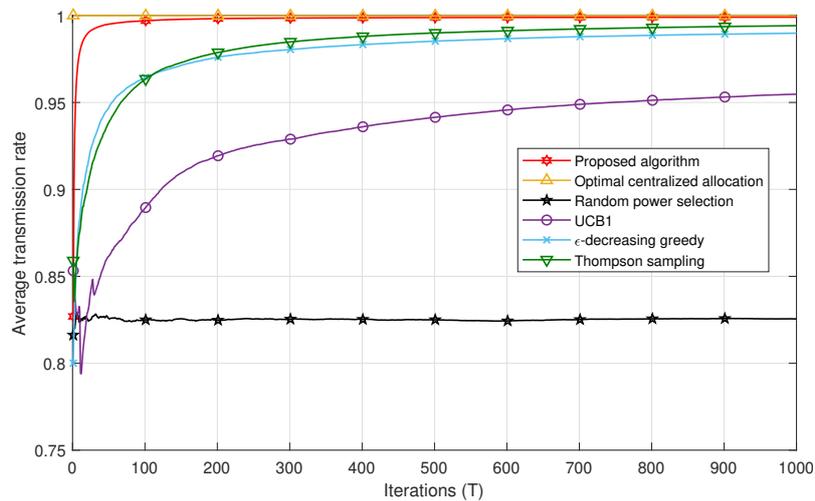
Figures 5.3 and 5.4 compare the successful transmission rate of the algorithms under study, i.e., the centralized power allocation, the proposed algorithm, the MAB algorithms ( $\epsilon$ -decreasing greedy, UCB1, and THS), and the random power selection in Scenarios 1 and 2, respectively. The proposed algorithm outperforms all the MAB techniques with a faster convergence to the optimal power in both scenarios. We also remark in Figure 5.3 that the  $\epsilon$ -decreasing greedy algorithm converges faster than the THS and UCB1 algorithms. This can be explained by the optimal selection of the value of  $L$ , which ensures a trade-off between the exploration and the exploitation phases in order to achieve the best performance. In Scenario 1, the  $\epsilon$ -decreasing greedy and THS algorithms converge to the same successful transmission rate after 400 iterations. However, the gap between  $\epsilon$ -decreasing greedy and THS is more important in Scenario 2 in Figure 5.4. In fact, after  $T = 100$  iterations, THS is slightly better than  $\epsilon$ -decreasing greedy. THS seems to take more advantage of the additional information carried by the feedback whether there is a decoding error among the users or not. However, the  $\epsilon$ -decreasing greedy algorithm has the same exploration time as in Scenario 1, the same value of  $L = 0.1$ , which gives the same speed of convergence. However, this algorithm selects better power values in Scenario 2 than in Scenario 1 thanks to the reward structure giving more information about all user packets. We notice that, both algorithms, i.e.,  $\epsilon$ -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 more errors 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,  $T$  being the number of packets sent, also known as 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 5.3 and 5.4 by shortening them to the desired value of  $T$ . In

other 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. 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 chapter 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.

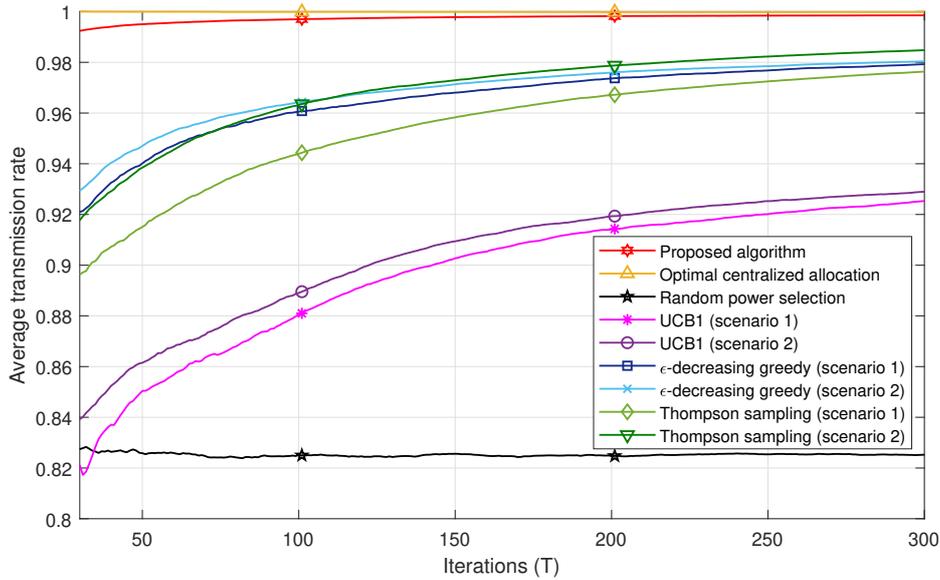


**Fig. 5.3.** Successful transmission rate comparison for all algorithms in Scenario 1.



**Fig. 5.4.** Successful transmission rate comparison for all algorithms in Scenario 2.

Figure 5.5 shows the 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  $\approx 0.94$  in Scenario 2, whereas, in Scenario 1, it attains the value of 0.91. This may



**Fig. 5.5.** Convergence comparison of all algorithms in Scenarios 1 and 2.

be explained by the fact that Scenario 2 conveys more information compared to Scenario 1 to select the best set of powers. Indeed, 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. We can see that the successful transmission rate achieved with the proposed algorithm converges to the one obtained with the optimal centralized solution only after a few 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  $\epsilon$ -decreasing greedy has a rate of 0.93. It should be noted that, after a large number of iterations, the performances of the MAB algorithms in Scenario 1 converge to those in Scenario 2.

## 5.7 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 done on this problem for the MUSA scheme in order to enhance user performance with a minimal signaling overhead. In this chapter, we address this issue by proposing a novel algorithm for autonomous power decision based on the proposed BEP approximation and the base station acknowledgments. Moreover, we study 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 on 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 for performance comparison. However, 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. The latter shows the best performance with a faster convergence rate, but also with a slightly higher signaling overhead compared to the investigated MAB algorithms, particularly for a varying propagation environment.

# Conclusion and perspectives

## Conclusion

This thesis focused on the study and the optimization of NOMA schemes to address the challenging requirements of the mMTC use case. The main contributions are summarized in Figure 6.1 :

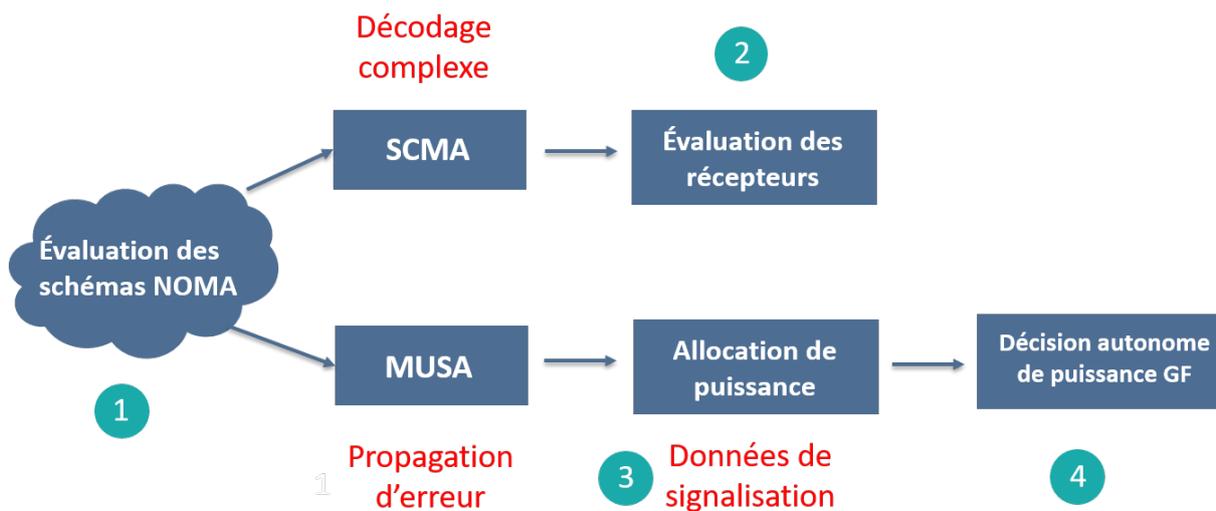


Fig. 6.1. The main contributions

1. Quantitative performance assessment : in chapter 2, we have provided a performance comparison of the most attractive NOMA schemes in order to overcome the limitation of the state of the art. This comparison permitted to choose the most appropriate schemes for mMTC. We have showed that SCMA and MUSA have the best performance compared to the other investigated schemes, but they also have some limitations that should be handled to support mMTC scenarios, namely the decoding complexity for SCMA and the power allocation for MUSA. These issues were studied subsequently.
2. SCMA decoding complexity : in chapter 3, we have tackled the problem of the high decoding complexity of SCMA while providing a comprehensive study

of the candidate receivers. More precisely, an exhaustive analysis and performance comparison of MPA with its two log-domain variants have been presented based on an in-depth investigation of the exchanged extrinsic information behavior. The performance of these algorithms have been compared in Gaussian and Rayleigh fading channels. We proved that Log-MPA achieves precisely the same performance as MPA thanks to the involved correction coefficient with a low decoding complexity compared to MPA. However, MAX-Log-MPA revealed a slight performance degradation, but with the benefit of significantly reduced decoding complexity compared to Log-MPA by removing all the exponential terms. We have proved that despite these simplifications, the decoding complexity of SCMA remains high. Therefore, we decided to study the MUSA scheme which offers a good compromise between the system complexity and the achievable performance.

3. Power allocation for BER minimization : in chapter 4, we addressed the error propagation issue of MUSA scheme with SIC receiver by proposing a novel centralized power allocation algorithm to minimize the BER in two different scenarios while considering either a maximum individual power constraint or a total system power constraint. These scenarios were further investigated with an additional constraint on the received power difference which ensures that the currently decoded user has a received power larger than the sum of power of users that have not been decoded yet. Moreover, in order to guarantee an optimal global solution for our optimization problem, a BER simplification was used. The obtained results proved that the power difference constraint was very beneficial in the practical case, with the maximum individual power constraint. We have showed that the received power difference constraint ensures an adequate power distribution among users, leading to a mitigated error propagation problem and thus, a proper decoding process.
4. Autonomous power decision for grant free access : for the sake of signaling overhead minimization, we investigated the grant free access strategy with MUSA scheme in chapter 5. In fact, we studied the ability of each user to autonomously select the adequate power ensuring the successful transmission of its packet. Firstly, a closed-form expression of the average system BEP was formulated. Then, this expression was exploited to propose a novel algorithm for autonomous power decision with minimum signaling overhead while taking advantage of the natural base station acknowledgments. Moreover, certain MAB algorithms namely,  $\epsilon$ -decreasing greedy, UCB1 and Thompson sampling, were also investigated while considering two acknowledgment scenarios. In the first scenario, each user is only informed on the reception of its own packet. In the second one, the status of all user transmissions is sent to each user. The simulations results proved that the proposed algorithm outperformed the MAB algorithms in terms of the successful transmission rate with a faster convergence to the optimal performance, but at the cost of a slightly higher signaling overhead particularly in case of fast channel variation. The studied MAB algorithms showed an acceptable performance, but with the drawbacks of longer convergence time and higher computational complexity at user equipment side.

This work contributes to a better evaluation of various promising NOMA schemes under different conditions. This allows to determine the strength and the weakness of each scheme. It also proposes some optimizations for these schemes to best meet the mMTC requirements.

## Perspectives

Despite the addressed issues of NOMA schemes in this thesis, some other issues remain unresolved and should be investigated in more details :

- ▶ Uplink synchronization : most of NOMA schemes are proposed with the assumption of users' synchronization. However, for a RACH-less grant free access strategy or outdated synchronization information, this condition is quite difficult to be guaranteed, especially for uplink transmissions with different user positions from the base station and variable propagation environments. Indeed, users' signals arrive at the base station asynchronously with different time offsets. The conventional synchronization techniques in the existing technologies, such as cyclic prefix, timing advance and channel feedback, require an information exchange between the devices and the base station in order to adjust the transmission times at which users are allowed to transmit their packets to avoid user's delays. This process involves high signaling overhead and high power consumption which may not be acceptable for the mMTC scenarios. Asynchronous NOMA schemes are more challenging than the asynchronous OMA schemes. Indeed, each user's symbol interferes with the symbols of other users sharing the same subcarrier, which makes the decoding process rather complicated. Recently, some works have addressed this issue for some NOMA schemes such as PD-NOMA [79]-[81] and SCMA [187], but this problem is not yet resolved and needs to be further investigated.
- ▶ Code design and distribution : one of the main issues of NOMA code domain schemes is the design of user codes which is the key parameter for their performances. For instance, for SCMA scheme, the design of user codebooks is quite complex. Furthermore, the sparsity asset, with a fixed number of zero elements in each codeword, may restrict the maximum number of possible generated codebooks, leading to a limited number of supported users. Moreover, since users codebooks should be generated by the base station, their distribution among the active users may induce a high amount of signaling overhead with high energy consumption. As for MUSA, despite the simplicity of the spreading sequences design with low cross-correlation condition, these sequences could also benefit from an optimization step in order to improve the system performance.

It should be mentioned that, after the conclusion of NOMA study item in 3GPP Release 16 in 2018, it was decided not to consider NOMA schemes for 5G networks. Considering the connection density of current systems, these schemes have not proven a significant gain over MIMO techniques while generating high system complexity. In addition, several implementation issues are introduced due to the

configurations of the transmitters and receivers such as power control, users scheduling over shared resources, inter-user interference management and receiver complexity. NOMA standardization activities are still ongoing. Therefore, these schemes could be adopted in future networks if a performance gain can be registered over the conventional OMA schemes while maintaining an affordable system complexity.

## Bibliography

- [1] W. Xiang, K. Zheng, and X. S. Shen, *5G mobile communications*. Springer, 2016.
- [2] H. Kim, “Enhanced mobile broadband communication systems”, in *Design and Optimization for 5G Wireless Communications*. 2020, pp. 239–302. DOI: 10.1002/9781119494492.ch7.
- [3] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, “5G Wireless Network Slicing for eMBB, uRLLC, and mMTC: A Communication-Theoretic View”, *IEEE Access*, vol. 6, pp. 55 765–55 779, 2018, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2872781.
- [4] Ericsson, “*Ericsson Mobility Report: 5G uptake even faster than expected*”, <https://www.ericsson.com/en/press-releases/2019/6/ericsson-mobility-report-5g-uptake-even-faster-than-expected>, 2019.
- [5] M. A. Siddiqi, H. Yu, and J. Joung, “5G ultra-reliable low-latency communication implementation challenges and operational issues with IoT devices”, *Electronics*, vol. 8, no. 9, p. 981, 2019.
- [6] IMT, ITUR, “*Minimum requirements related to technical performance for IMT-2020 radio interface(s)*”, <https://www.itu.int/pub/R-REP-M.2410-2017>, 2017.
- [7] C. Bockelmann et al., “Towards Massive Connectivity Support for Scalable mMTC Communications in 5G Networks”, *IEEE Access*, vol. 6, pp. 28 969–28 992, 2018. DOI: 10.1109/ACCESS.2018.2837382.
- [8] M. Shirvanimoghaddam, M. Dohler, and S. J. Johnson, “Massive Non-Orthogonal Multiple Access for Cellular IoT: Potentials and Limitations”, *IEEE Communications Magazine*, vol. 55, no. 9, pp. 55–61, Sep. 2017.
- [9] G. A. Akpakwu, B. J. Silva, G. P. Hancke, and A. M. Abu-Mahfouz, “A Survey on 5G Networks for the Internet of Things: Communication Technologies and Challenges”, *IEEE Access*, vol. 6, pp. 3619–3647, 2018, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2017.2779844.
- [10] S. K. Sharma and X. Wang, “Toward Massive Machine Type Communications in Ultra-Dense Cellular IoT Networks: Current Issues and Machine Learning-Assisted Solutions”, *IEEE Communications Surveys Tutorials*, vol. 22, no. 1, pp. 426–471, 2020. DOI: 10.1109/COMST.2019.2916177.

- 
- [11] A. Kunz, H. Kim, L. Kim, and S. S. Husain, "Machine type communications in 3gpp: From release 10 to release 12", in *2012 IEEE Globecom Workshops*, 2012, pp. 1747–1752. DOI: 10.1109/GLOCOMW.2012.6477852.
- [12] P. Reininger, "3GPP Standards for the Internet-of-Things", *Huawei, Shenzhen, China, Tech. Rep. report*, no. 3GPP, 2016.
- [13] B. S. Chaudhari and M. Zennaro, "Introduction to low-power wide-area networks", in *LPWAN Technologies for IoT and M2M Applications*, Elsevier, 2020, pp. 1–13.
- [14] R. S. Sinha, Y. Wei, and S.-H. Hwang, "A survey on LPWA technology: LoRa and NB-IoT", *Ict Express*, vol. 3, no. 1, pp. 14–21, 2017.
- [15] K. Mekki, E. Bajic, F. Chaxel, and F. Meyer, "A comparative study of LPWAN technologies for large-scale IoT deployment", *ICT express*, vol. 5, no. 1, pp. 1–7, 2019.
- [16] A. Kunz, H. Kim, L. Kim, and S. S. Husain, "Machine type communications in 3gpp: From release 10 to release 12", in *2012 IEEE Globecom Workshops*, 2012, pp. 1747–1752. DOI: 10.1109/GLOCOMW.2012.6477852.
- [17] A. Hoglund, X. Lin, O. Liberg, A. Behravan, E. A. Yavuz, M. Van Der Zee, Y. Sui, T. Tirronen, A. Ratilainen, and D. Eriksson, "Overview of 3GPP Release 14 Enhanced NB-IoT", *IEEE Network*, vol. 31, no. 6, pp. 16–22, 2017. DOI: 10.1109/MNET.2017.1700082.
- [18] S. Lippuner, B. Weber, M. Salomon, M. Korb, and Q. Huang, "Ec-gsm-iot network synchronization with support for large frequency offsets", in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, 2018, pp. 1–6. DOI: 10.1109/WCNC.2018.8377168.
- [19] 3GPP, *5g iot driven growth*, [https://www.3gpp.org/ftp/Information/presentations/presentations\\_2019/Poster\\_2019\\_08\\_v7\\_optimized.pdf](https://www.3gpp.org/ftp/Information/presentations/presentations_2019/Poster_2019_08_v7_optimized.pdf), 2021.
- [20] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A Survey of Non-Orthogonal Multiple Access for 5G", *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018. DOI: 10.1109/COMST.2018.2835558.
- [21] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey", *IEEE Communications Surveys & Tutorials*, 2020.
- [22] NTT DOCOMO, INC, "Uplink Multiple Access Schemes for NR", *3GPP TSG RAN WG1 Meeting 85 R1-165174*, 2016.
- [23] H. Nikopour and H. Baligh, "Sparse code multiple access", in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sep. 2013, pp. 332–336.
- [24] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang, and J. Xu, "Multi-user shared access for internet of things", in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, May 2016, pp. 1–5.

- [25] J. Zeng, B. Li, X. Su, L. Rong, and R. Xing, "Pattern division multiple access (pdma) for cellular future radio access", in *2015 International Conference on Wireless Communications Signal Processing (WCSP)*, 2015, pp. 1–5. DOI: 10.1109/WCSP.2015.7341229.
- [26] Li Ping, Lihai Liu, Keying Wu, and W. K. Leung, "Interleave division multiple-access", *IEEE Transactions on Wireless Communications*, vol. 5, no. 4, pp. 938–947, 2006. DOI: 10.1109/TWC.2006.1618943.
- [27] K. Xiao, B. Xiao, S. Zhang, Z. Chen, and B. Xia, "Simplified multiuser detection for scma with sum-product algorithm", in *2015 International Conference on Wireless Communications Signal Processing (WCSP)*, Oct. 2015, pp. 1–5. DOI: 10.1109/WCSP.2015.7341328.
- [28] Y. Qi, G. Wu, S. Hu, and Y. Gao, "Parallel-implemented message passing algorithm for scma decoder based on gpgpu", in *2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct. 2017, pp. 1–6. DOI: 10.1109/WCSP.2017.8170883.
- [29] F. Wei and W. Chen, "Low Complexity Iterative Receiver Design for Sparse Code Multiple Access", *IEEE Transactions on Communications*, vol. 65, no. 2, pp. 621–634, Feb. 2017, ISSN: 0090-6778. DOI: 10.1109/TCOMM.2016.2631468.
- [30] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal map decoding algorithms operating in the log domain", in *Proceedings IEEE International Conference on Communications ICC '95*, vol. 2, Jun. 1995, 1009–1013 vol.2. DOI: 10.1109/ICC.1995.524253.
- [31] Seung Hoon Nam and Kwang Bok Lee, "Transmit power allocation for an extended v-blast system", in *The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 2, Sep. 2002, 843–848 vol.2. DOI: 10.1109/PIMRC.2002.1047341.
- [32] O. Teyeb, G. Wikstrom, M. Stattin, T. Cheng, S. Faxer, and H. Do, *Evolving lte to fit the 5g future, ericsson technology review*, 2017.
- [33] E. Mohyeldin, "Minimum technical performance requirements for imt-2020 radio interface (s)", in *ITU-R Workshop on IMT-2020 Terrestrial Radio Interfaces*, 2016, pp. 1–12.
- [34] M. Wang, W. Yang, J. Zou, B. Ren, M. Hua, J. Zhang, and X. You, "Cellular machine-type communications: physical challenges and solutions", *IEEE Wireless Communications*, vol. 23, no. 2, pp. 126–135, 2016. DOI: 10.1109/MWC.2016.7462494.
- [35] H. Shariatmadari, R. Ratasuk, S. Iraj, A. Laya, T. Taleb, R. Jäntti, and A. Ghosh, "Machine-type communications: current status and future perspectives toward 5G systems", *IEEE Communications Magazine*, vol. 53, no. 9, pp. 10–17, Sep. 2015.
- [36] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (noma) for cellular future radio access", in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, Jun. 2013, pp. 1–5. DOI: 10.1109/VTCSpring.2013.6692652.

- 
- [37] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, "Pattern Division Multiple Access - A Novel Nonorthogonal Multiple Access for Fifth-Generation Radio Networks", *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3185–3196, Apr. 2017.
- [38] N. Abramson, "The aloha system: Another alternative for computer communications", in *Proceedings of the November 17-19, 1970, fall joint computer conference*, 1970, pp. 281–285.
- [39] H. A. Cozzetti and R. Scopigno, "RR-Aloha+: A slotted and distributed MAC protocol for vehicular communications", in *2009 IEEE Vehicular Networking Conference (VNC)*, IEEE, 2009, pp. 1–8.
- [40] E. Balevi, F. T. Al Rabee, and R. D. Gitlin, "Aloha-noma for massive machine-to-machine iot communication", in *2018 IEEE International Conference on Communications (ICC)*, IEEE, 2018, pp. 1–5.
- [41] 3GPP, "Technical Specification Group Services and System Aspects; Feasibility Study on New Services and Markets Technology Enablers; Stage 1 (Release 14)", *3GPP TR 22.891 V14.2.0*, 2016.
- [42] —, "Technical Specification Group Services and System Aspects; Release 15 Description; Summary of Rel-15 Work Items, Release 15", *3GPP TR 21.915 V15.0.0*, 2019.
- [43] ETSI, *5g*, <https://www.etsi.org/technologies/5G>.
- [44] M. Lauridsen, L. C. Gimenez, I. Rodriguez, T. B. Sorensen, and P. Mogensen, "From LTE to 5G for connected mobility", *IEEE Communications Magazine*, vol. 55, no. 3, pp. 156–162, 2017.
- [45] IEEE. (). What does every engineer need to know about 5g?, [Online]. Available: <https://spectrum.ieee.org/semiconductors/design/what-does-every-engineer-need-to-know-about-5g>.
- [46] 3GPP, "Technical Specification Group Services and System Aspects; Release 16 Description; Summary of Rel-16 Work Items, Release 16", *3GPP TR 21.916 V1.0.0*, 2020.
- [47] —, "Technical Specification Group Services and System Aspects; Service requirements for the 5G system; Stage 1, Release 18", *3GPP TS 22.261 V17.4.0*, 2020.
- [48] Statista, *Internet of things (IoT) connected devices installed base worldwide from 2015 to 2025 (in billions)*, <https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/>, 2015.
- [49] 5GWorldPro.com, *5g nr light in release 17*, <https://www.5gworldpro.com/5g-knowledge/5g-nr-light-in-release-17.html>.
- [50] 3GPP, "Study on non orthogonal multiple access (NOMA) for NR", *3GPP TR 38.812 v16.0.0*, Dec. 2018.
- [51] Y. Yuan, Z. Yuan, and L. Tian, "5g non-orthogonal multiple access study in 3gpp", *IEEE Communications Magazine*, vol. 58, no. 7, pp. 90–96, 2020. DOI: 10.1109/MCOM.001.1900450.

## BIBLIOGRAPHY

---

- [52] S. Tabbane, “Internet of things: A technical overview of the ecosystem”, in *Regional Workshop for Africa on “Developing the ICT ecosystem to harness Internet-of-Things (IoT)*, 2017.
- [53] Sigfox, *Sigfox*, <https://www.sigfox.com/en/node/656>, 2021.
- [54] R. L. Pickholtz, L. B. Milstein, and D. L. Schilling, “Spread spectrum for mobile communications”, *IEEE Transactions on Vehicular technology*, vol. 40, no. 2, pp. 313–322, 1991.
- [55] B. Reynders and S. Pollin, “Chirp spread spectrum as a modulation technique for long range communication”, in *2016 Symposium on Communications and Vehicular Technologies (SCVT)*, IEEE, 2016, pp. 1–5.
- [56] A. Hoglund, J. Bergman, X. Lin, O. Liberg, A. Ratilainen, H. S. Razaghi, T. Tirronen, and E. A. Yavuz, “Overview of 3GPP Release 14 Further Enhanced MTC”, *IEEE Communications Standards Magazine*, vol. 2, no. 2, pp. 84–89, 2018. DOI: 10.1109/MCOMSTD.2018.1700050.
- [57] GSMA, *3gpp low power wide area technologies-gsma white paper*, <https://www.gsma.com/iot/wp-content/uploads/2016/10/3GPP-Low-Power-Wide-Area-Technologies-GSMA-White-Paper.pdf>, 2016.
- [58] M. S. Ali, E. Hossain, and D. I. Kim, “LTE/LTE-A random access for massive machine-type communications in smart cities”, *IEEE Communications Magazine*, vol. 55, no. 1, pp. 76–83, 2017.
- [59] S. Cherkaoui, I. Keskes, H. Rivano, and R. Stanica, “Lte-a random access channel capacity evaluation for m2m communications”, in *2016 Wireless Days (WD)*, IEEE, 2016, pp. 1–6.
- [60] O. Arouk, A. Ksentini, and T. Taleb, “How accurate is the rach procedure model in lte and lte-a?”, in *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*, IEEE, 2016, pp. 61–66.
- [61] Y. Chen, A. Bayesteh, Y. Wu, B. Ren, S. Kang, S. Sun, Q. Xiong, C. Qian, B. Yu, Z. Ding, *et al.*, “Toward the standardization of non-orthogonal multiple access for next generation wireless networks”, *IEEE Communications Magazine*, vol. 56, no. 3, pp. 19–27, 2018.
- [62] S. Kim, S. Kim, J. Kim, K. Lee, S. Choi, and B. Shim, “Low latency random access for small cell toward future cellular networks”, *IEEE Access*, vol. 7, pp. 178 563–178 576, 2019. DOI: 10.1109/ACCESS.2019.2958991.
- [63] J. Kim, G. Lee, S. Kim, T. Taleb, S. Choi, and S. Bahk, “Two-step random access for 5g system: Latest trends and challenges”, *IEEE Network*, vol. 35, no. 1, pp. 273–279, 2021. DOI: 10.1109/MNET.011.2000317.
- [64] K. Higuchi and A. Benjebbour, “Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access”, *IEICE Transactions on Communications*, vol. 98, no. 3, pp. 403–414, 2015.
- [65] Z. Wei, J. Guo, D. W. K. Ng, and J. Yuan, “Fairness comparison of uplink noma and oma”, in *2017 IEEE 85th vehicular technology conference (VTC Spring)*, IEEE, 2017, pp. 1–6.

- 
- [66] T. Cover, "Broadcast channels", *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 2–14, 1972.
- [67] A. Li, A. Harada, and H. Kayama, "Investigation on low complexity power assignment method and performance gain of non-orthogonal multiple access systems", *IEICE trans. Fundamentals*, vol. 97, no. 1, 2014.
- [68] N. Otao, Y. Kishiyama, and K. Higuchi, "Performance of non-orthogonal access with sic in cellular downlink using proportional fair-based resource allocation", in *2012 International Symposium on Wireless Communication Systems (ISWCS)*, 2012, pp. 476–480. DOI: 10.1109/ISWCS.2012.6328413.
- [69] S. M. R. Islam, N. Avazov, O. A. Dobre, and K. Kwak, "Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges", *IEEE Communications Surveys Tutorials*, vol. 19, no. 2, pp. 721–742, 2017.
- [70] X. Wang, R. Chen, Y. Xu, and Q. Meng, "Low-Complexity Power Allocation in NOMA Systems With Imperfect SIC for Maximizing Weighted Sum-Rate", *IEEE Access*, vol. 7, pp. 94 238–94 253, 2019. DOI: 10.1109/ACCESS.2019.2926757.
- [71] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic User Clustering and Power Allocation for Uplink and Downlink Non-Orthogonal Multiple Access (NOMA) Systems", *IEEE Access*, vol. 4, pp. 6325–6343, 2016. DOI: 10.1109/ACCESS.2016.2604821.
- [72] A. Sayed-Ahmed and M. Elsabrouty, "User selection and power allocation for guaranteed sic detection in downlink beamforming non-orthogonal multiple access", in *2017 Wireless Days*, Mar. 2017, pp. 188–193. DOI: 10.1109/WD.2017.7918141.
- [73] S. Liu, C. Zhang, and G. Lyu, "User selection and power schedule for downlink non-orthogonal multiple access (noma) system", in *2015 IEEE International Conference on Communication Workshop (ICCW)*, 2015, pp. 2561–2565. DOI: 10.1109/ICCW.2015.7247563.
- [74] B. Kimy, S. Lim, H. Kim, S. Suh, J. Kwun, S. Choi, C. Lee, S. Lee, and D. Hong, "Non-orthogonal multiple access in a downlink multiuser beamforming system", in *MILCOM 2013 - 2013 IEEE Military Communications Conference*, 2013, pp. 1278–1283. DOI: 10.1109/MILCOM.2013.218.
- [75] M. Al-Imari, P. Xiao, M. A. Imran, and R. Tafazolli, "Uplink non-orthogonal multiple access for 5g wireless networks", in *2014 11th International Symposium on Wireless Communications Systems (ISWCS)*, 2014, pp. 781–785. DOI: 10.1109/ISWCS.2014.6933459.
- [76] F. Kara and H. Kaya, "BER performances of downlink and uplink NOMA in the presence of SIC errors over fading channels", *IET Communications*, vol. 12, no. 15, pp. 1834–1844, 2018.
- [77] T. Assaf, A. Al-Dweik, M. El Moursi, and H. Zeineldin, "Exact ber performance analysis for downlink noma systems over nakagami- $m$  fading channels", *IEEE Access*, vol. 7, pp. 134 539–134 555, 2019.

- [78] J. S. Yeom, H. S. Jang, K. S. Ko, and B. C. Jung, “BER performance of uplink NOMA with joint maximum-likelihood detector”, *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 10 295–10 300, 2019.
- [79] H. Haci, H. Zhu, and J. Wang, “A novel interference cancellation technique for non-orthogonal multiple access (noma)”, in *2015 IEEE global communications conference (GLOBECOM)*, IEEE, 2015, pp. 1–6.
- [80] H. Haci, “Non-orthogonal multiple access (noma) with asynchronous interference cancellation”, PhD thesis, University of Kent, 2015.
- [81] H. Haci, H. Zhu, and J. Wang, “Performance of Non-orthogonal Multiple Access With a Novel Asynchronous Interference Cancellation Technique”, *IEEE Transactions on Communications*, vol. 65, no. 3, pp. 1319–1335, 2017. DOI: 10.1109/TCOMM.2016.2640307.
- [82] A. J. Goldsmith and P. P. Varaiya, “Capacity of fading channels with channel side information”, *IEEE transactions on information theory*, vol. 43, no. 6, pp. 1986–1992, 1997.
- [83] L. Dai, B. Wang, Y. Yuan, S. Han, C. I, and Z. Wang, “Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends”, *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74–81, Sep. 2015, ISSN: 0163-6804. DOI: 10.1109/MCOM.2015.7263349.
- [84] R. Hoshyar, F. P. Wathan, and R. Tafazolli, “Novel Low-Density Signature for Synchronous CDMA Systems Over AWGN Channel”, *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1616–1626, Apr. 2008, ISSN: 1053-587X. DOI: 10.1109/TSP.2007.909320.
- [85] D. Guo and C.-C. Wang, “Multiuser detection of sparsely spread CDMA”, *IEEE journal on selected areas in communications*, vol. 26, no. 3, pp. 421–431, 2008.
- [86] J. Van De Beek and B. M. Popovic, “Multiple access with low-density signatures”, in *GLOBECOM 2009-2009 IEEE Global Telecommunications Conference*, IEEE, 2009, pp. 1–6.
- [87] R. Razavi, R. Hoshyar, M. A. Imran, and Y. Wang, “Information Theoretic Analysis of LDS Scheme”, *IEEE Communications Letters*, vol. 15, no. 8, pp. 798–800, Aug. 2011.
- [88] W. Xiang, K. Zheng, and X. S. Shen, *5G mobile communications*. Springer, 2016.
- [89] M. Al-Imari and M. A. Imran, “Low density spreading multiple access”, in *Multiple Access Techniques for 5G Wireless Networks and Beyond*, M. Vaezi, Z. Ding, and H. V. Poor, Eds. Cham: Springer International Publishing, 2019, pp. 493–514, ISBN: 978-3-319-92090-0. DOI: 10.1007/978-3-319-92090-0\_15. [Online]. Available: [https://doi.org/10.1007/978-3-319-92090-0\\_15](https://doi.org/10.1007/978-3-319-92090-0_15).
- [90] M. Al-Imari and R. Hoshyar, “Reducing the peak to average power ratio of lds-ofdm signals”, in *2010 7th International Symposium on Wireless Communication Systems*, IEEE, 2010, pp. 922–926.

- 
- [91] T. Huang, J. Yuan, X. Cheng, and W. Lei, "Design of degrees of distribution of lds-ofdm", in *2015 9th International Conference on Signal Processing and Communication Systems (ICSPCS)*, IEEE, 2015, pp. 1–6.
- [92] L. Meylani, I. Hidayat, A. Kurniawan, and M. S. Arifianto, "Power allocation for group lds-ofdm in underlay cognitive radio", in *2019 11th International Conference on Information Technology and Electrical Engineering (ICITEE)*, IEEE, 2019, pp. 1–5.
- [93] M. Al-Imari, M. A. Imran, R. Tafazolli, and D. Chen, "Subcarrier and power allocation for lds-ofdm system", in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, IEEE, 2011, pp. 1–5.
- [94] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "Scma codebook design", in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, Sep. 2014, pp. 1–5. DOI: 10.1109/VTCFall.2014.6966170.
- [95] D. Cai, P. Fan, X. Lei, Y. Liu, and D. Chen, "Multi-dimensional scma codebook design based on constellation rotation and interleaving", in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, 2016, pp. 1–5. DOI: 10.1109/VTCSpring.2016.7504356.
- [96] S. Zhang, K. Xiao, B. Xiao, Z. Chen, B. Xia, D. Chen, and S. Ma, "A capacity-based codebook design method for sparse code multiple access systems", in *2016 8th International Conference on Wireless Communications Signal Processing (WCSP)*, 2016, pp. 1–5. DOI: 10.1109/WCSP.2016.7752620.
- [97] G. Song, X. Wang, and J. Cheng, "Signature Design of Sparsely Spread Code Division Multiple Access Based on Superposed Constellation Distance Analysis", *IEEE Access*, vol. 5, pp. 23 809–23 821, 2017. DOI: 10.1109/ACCESS.2017.2765346.
- [98] L. Yu, X. Lei, P. Fan, and D. Chen, "An optimized design of scma codebook based on star-qam signaling constellations", in *2015 International Conference on Wireless Communications Signal Processing (WCSP)*, 2015, pp. 1–5. DOI: 10.1109/WCSP.2015.7341311.
- [99] Y.-M. Chen and J.-W. Chen, "On the Design of Near-Optimal Sparse Code Multiple Access Codebooks", *IEEE Transactions on Communications*, vol. 68, no. 5, pp. 2950–2962, 2020.
- [100] Y. Zhou, Q. Yu, W. Meng, and C. Li, "Scma codebook design based on constellation rotation", in *2017 IEEE International Conference on Communications (ICC)*, IEEE, 2017, pp. 1–6.
- [101] S. Liu, J. Wang, J. Bao, and C. Liu, "Optimized SCMA codebook design by QAM constellation segmentation with maximized MED", *IEEE Access*, vol. 6, pp. 63 232–63 242, 2018.
- [102] Z. Mheich, L. Wen, P. Xiao, and A. Maaref, "Design of SCMA codebooks based on golden angle modulation", *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1501–1509, 2018.

## BIBLIOGRAPHY

---

- [103] K. Au, L. Zhang, H. Nikopour, E. Yi, A. Bayesteh, U. Vilaipornsawai, J. Ma, and P. Zhu, “Uplink contention based scma for 5g radio access”, in *2014 IEEE Globecom Workshops (GC Wkshps)*, Dec. 2014, pp. 900–905.
- [104] B. Wang, K. Wang, Z. Lu, T. Xie, and J. Quan, “Comparison study of non-orthogonal multiple access schemes for 5g”, in *2015 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, Jun. 2015, pp. 1–5. DOI: 10.1109/BMSB.2015.7177186.
- [105] J. Zou, H. Zhao, and W. Zhao, “Low-complexity interference cancellation receiver for sparse code multiple access”, in *2015 IEEE 6th International Symposium on Microwave, Antenna, Propagation, and EMC Technologies (MAPE)*, Oct. 2015, pp. 277–282. DOI: 10.1109/MAPE.2015.7510315.
- [106] C. Wang, I. H. Liao, J. Wu, S. Fang, and J. Hsu, “Minimum Distance Optimization on Signature Code for Uplink Non-orthogonal Multiple Access”, in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, 2019, pp. 1–6. DOI: 10.1109/WCNC.2019.8886090.
- [107] S. Lim and H. Park, “Gaussian Approximation-Based Belief Propagation Receiver for Uplink SCMA Systems With Imperfect CSIR”, *IEEE Communications Letters*, vol. 22, no. 12, pp. 2611–2614, 2018. DOI: 10.1109/LCOMM.2018.2874441.
- [108] X. Meng, Y. Wu, Y. Chen, and M. Cheng, “Low complexity receiver for uplink scma system via expectation propagation”, in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, IEEE, 2017, pp. 1–5.
- [109] S. Zhang, X. Xu, L. Lu, Y. Wu, G. He, and Y. Chen, “Sparse code multiple access: An energy efficient uplink approach for 5g wireless systems”, in *2014 IEEE Global Communications Conference*, Dec. 2014, pp. 4782–4787. DOI: 10.1109/GLOCOM.2014.7037563.
- [110] ZTE, ZTE Microelectronics, “Receiver Details and Link Performance for MUSA”, *3GPP TSG RAN WG1 Meeting 86 R1-166404*, Gothenburg, Sweden, 22nd - 26th August 2016.
- [111] ZTE, “Receiver implementation for musa”, *3GPP TSG RAN WG1 Meeting 85 R1-164270*, Nanjing, China, 23rd – 27th May 2016.
- [112] Y. Xu, G. Wang, L. Zheng, R. Liu, and D. Zhao, “BER Performance Evaluation of Downlink MUSA over Rayleigh Fading Channel”, in *International Conference on Machine Learning and Intelligent Communications*, Springer, 2017, pp. 85–94.
- [113] E. Çatak, F. Tekçe, O. Dizdar, and L. Durak-Ata, “Multi-user shared access in massive machine-type communication systems via superimposed waveforms”, *Physical Communication*, vol. 37, p. 100 896, 2019.
- [114] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM wireless communications with MATLAB*. John Wiley & Sons, 2010.

- 
- [115] B. Wang, L. Dai, Y. Yuan, and Z. Wang, “Compressive sensing based multi-user detection for uplink grant-free non-orthogonal multiple access”, in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, 2015, pp. 1–5. DOI: 10.1109/VTCFall.2015.7390876.
- [116] M. Mhedhbi and F. E. Boukour, “Analysis and evaluation of pattern division multiple access scheme jointed with 5G waveforms”, *IEEE Access*, vol. 7, pp. 21 826–21 833, 2019.
- [117] X. Dai, Z. Zhang, B. Bai, S. Chen, and S. Sun, “Pattern Division Multiple Access: A New Multiple Access Technology for 5G”, *IEEE Wireless Communications*, vol. 25, no. 2, pp. 54–60, 2018. DOI: 10.1109/MWC.2018.1700084.
- [118] P. Li, Y. Jiang, S. Kang, F. Zheng, and X. You, “Pattern division multiple access with large-scale antenna array”, in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, IEEE, 2017, pp. 1–6.
- [119] Y. Jiang, P. Li, Z. Ding, F. Zheng, M. Ma, and X. You, “Joint Transmitter and Receiver Design for Pattern Division Multiple Access”, *IEEE Transactions on Mobile Computing*, vol. 18, no. 4, pp. 885–895, 2019. DOI: 10.1109/TMC.2018.2845364.
- [120] J. Zeng, D. Kong, X. Su, L. Rong, and X. Xu, “On the performance of pattern division multiple access in 5g systems”, in *2016 8th International Conference on Wireless Communications Signal Processing (WCSP)*, 2016, pp. 1–5. DOI: 10.1109/WCSP.2016.7752716.
- [121] P. Trifonov, “Design of structured irregular ldpc codes”, in *2008 IEEE Region 8 International Conference on Computational Technologies in Electrical and Electronics Engineering*, 2008, pp. 20–24. DOI: 10.1109/SIBIRCON.2008.4602580.
- [122] L. Ping, L. Liu, K. Wu, W. Leung, *et al.*, “Interleave-division multiple-access (idma) communications”, in *Proc. 3rd International Symposium on Turbo Codes and Related Topics*, Citeseer, 2003, p. 173 180.
- [123] B. Y. Kong and I. Park, “Efficient implementation of multiple interleavers in idma for 5g”, in *2018 International SoC Design Conference (ISOC)*, 2018, pp. 119–120. DOI: 10.1109/ISOC.2018.8649984.
- [124] S. Wu, X. Chen, and S. Zhou, “A parallel interleaver design for idma systems”, in *2009 International Conference on Wireless Communications Signal Processing*, 2009, pp. 1–5. DOI: 10.1109/WCSP.2009.5371716.
- [125] Y. Chen, F. Schaich, and T. Wild, “Multiple access and waveforms for 5g: Idma and universal filtered multi-carrier”, in *2014 IEEE 79th Vehicular Technology Conference (VTC Spring)*, 2014, pp. 1–5. DOI: 10.1109/VTCSpring.2014.7022995.
- [126] A. Haghghat, S. N. Nazar, S. Herath, and R. Olesen, “On the performance of idma-based non-orthogonal multiple access schemes”, in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, 2017, pp. 1–5. DOI: 10.1109/VTCFall.2017.8288410.

- [127] L. Bing and B. Bai, “Design of simplified maximum-likelihood receivers for multiuser CPM systems”, *The Scientific World Journal*, vol. 2014, 2014.
- [128] V. P. Klimentyev and A. B. Sergienko, “A low-complexity scma detector for awgn channel based on solving overdetermined systems of linear equations”, in *Problems of Redundancy in Information and Control Systems (REDUNDANCY), 2016 XV International Symposium*, IEEE, 2016, pp. 61–65.
- [129] N. I. Miridakis and D. D. Vergados, “A Survey on the Successive Interference Cancellation Performance for Single-Antenna and Multiple-Antenna OFDM Systems”, *IEEE Communications Surveys Tutorials*, vol. 15, no. 1, pp. 312–335, 2013. DOI: 10.1109/SURV.2012.030512.00103.
- [130] Y. Wu and H. Zhu, “Complexity Comparison between Two Optimal-Ordered SIC MIMO Detectors Based on Matlab Simulations”, *arXiv preprint arXiv:2003.03732*, 2020.
- [131] A. Laya, L. Alonso, and J. Alonso-Zarate, “Is the random access channel of LTE and LTE-A suitable for M2M communications? A survey of alternatives”, *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 4–16, 2013.
- [132] M. Vilgelm, S. Schiessl, H. Al-Zubaidy, W. Kellerer, and J. Gross, “On the reliability of lte random access: Performance bounds for machine-to-machine burst resolution time”, in *2018 IEEE International Conference on Communications (ICC)*, IEEE, 2018, pp. 1–7.
- [133] E. Dahlman, S. Parkvall, J. Skold, and P. Beming, *3G evolution: HSPA and LTE for mobile broadband*. Academic press, 2010.
- [134] ETSI, TS, “LTE;Evolved Universal Terrestrial Radio Access (E-UTRA);Medium Access Control (MAC) protocol specification”, *3GPP TS 36.321 V8. 10.0 Release 8*, 2011.
- [135] 3GPP, “Study on RAN improvements for machine-type communications”, *3GPP TR 37.868 V11.0.0*, Sept. 2011.
- [136] ZTE, “RACH overload solutions”, *3GPP TSG RAN WG2 70bis R2-103742*, Stockholm, Sweden, 28th Jun. 2010.
- [137] Qualcomm Incorporated, “RACH timeline consideration”, *3GPP TSG-RAN WG1 87, R1-1612035*, Nov. 2016.
- [138] Z. Li, L. Tian, Y. Yin, and W. Cao, “On contention-based 2-step random access procedure”, in *2020 International Conference on Wireless Communications and Signal Processing (WCSP)*, 2020, pp. 771–776. DOI: 10.1109/WCSP49889.2020.9299681.
- [139] H. Lee, S. Kim, and J. Lim, “Multiuser superposition transmission (must) for lte-a systems”, in *2016 IEEE International Conference on Communications (ICC)*, 2016, pp. 1–6. DOI: 10.1109/ICC.2016.7510909.
- [140] Y. Chen, A. Bayesteh, Y. Wu, B. Ren, S. Kang, S. Sun, Q. Xiong, C. Qian, B. Yu, Z. Ding, S. Wang, S. Han, X. Hou, H. Lin, R. Visoz, and R. Razavi, “Toward the standardization of non-orthogonal multiple access for next generation wireless networks”, *IEEE Communications Magazine*, vol. 56, no. 3, pp. 19–27, 2018. DOI: 10.1109/MCOM.2018.1700845.

- 
- [141] Y. Yuan and C. Yan, "Noma study in 3gpp for 5g", in *2018 IEEE 10th International Symposium on Turbo Codes Iterative Information Processing (ISTC)*, 2018, pp. 1–5. DOI: 10.1109/ISTC.2018.8625325.
- [142] B. Makki, K. Chitti, A. Behravan, and M. .-.S. Alouini, "A survey of noma: Current status and open research challenges", *IEEE Open Journal of the Communications Society*, vol. 1, pp. 179–189, 2020. DOI: 10.1109/OJCOMS.2020.2969899.
- [143] M. Hussain and H. Rasheed, "Nonorthogonal Multiple Access for Next-Generation Mobile Networks: A Technical Aspect for Research Direction", *Wireless Communications and Mobile Computing*, vol. 2020, 2020.
- [144] H. Kim, Y.-G. Lim, C.-B. Chae, and D. Hong, "Multiple access for 5G new radio: Categorization, evaluation, and challenges", *arXiv preprint arXiv:1703.09042*, 2017.
- [145] Z. Wu, K. Lu, C. Jiang, and X. Shao, "Comprehensive study and comparison on 5G NOMA schemes", *IEEE Access*, vol. 6, pp. 18 511–18 519, 2018.
- [146] CMCC, "Results and remaining issues of LLS evaluation on multiple access", *3GPP TSG RAN WG1 Meeting 86 R1-167105*, Gothenberg, Sweden, 22-26 Aug. 2016.
- [147] Huawei, HiSilicon, "LLS results for uplink multiple access", *3GPP TSG RAN WG1 Meeting 85 R1-164037*, 2016.
- [148] C. Zhang, C. Yang, W. Xu, S. Zhang, Z. Zhang, and X. You, "Efficient Sparse Code Multiple Access Decoder Based on Deterministic Message Passing Algorithm", *arXiv preprint arXiv:1804.00180*, 2018.
- [149] L. Tian, M. Zhao, J. Zhong, P. Xiao, and L. Wen, "A low complexity detector for downlink SCMA systems, year=2017", *IET Communications*, vol. 11, no. 16, pp. 2433–2439, ISSN: 1751-8628. DOI: 10.1049/iet-com.2017.0420.
- [150] H. Zhang, S. Han, and W. Meng, "Multi-stage message passing algorithm for scma downlink receiver", in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, Sep. 2016, pp. 1–5. DOI: 10.1109/VTCFall.2016.7881102.
- [151] A. P. Sokolov, S. B. Gashkov, E. E. Gasanov, P. A. Panteleev, and I. V. Neznanov, *Computation of jacobian logarithm operation*, US Patent App. 13/197,098, Jun. 2012.
- [152] M. M. El-Sayed, A. S. Ibrahim, and M. M. Khairy, "Power allocation strategies for non-orthogonal multiple access", in *2016 International Conference on Selected Topics in Mobile Wireless Networking (MoWNeT)*, 2016, pp. 1–6. DOI: 10.1109/MoWNet.2016.7496633.
- [153] J. L. Jacob and T. Abrao, "NOMA Systems Optimization to Ensure Maximum Fairness to Users", *arXiv preprint arXiv:2001.03827*, 2020.
- [154] A. Sayed-Ahmed and M. ElSabrouty, "User selection and power allocation for guaranteed sic detection in downlink beamforming non-orthogonal multiple access", in *2017 Wireless Days*, IEEE, 2017, pp. 188–193.

## BIBLIOGRAPHY

---

- [155] J. V. Evangelista, Z. Sattar, G. Kaddoum, and A. Chaaban, “Fairness and sum-rate maximization via joint subcarrier and power allocation in uplink SCMA transmission”, *IEEE Transactions on Wireless Communications*, vol. 18, no. 12, pp. 5855–5867, 2019.
- [156] S. Han, Y. Huang, W. Meng, C. Li, N. Xu, and D. Chen, “Optimal power allocation for SCMA downlink systems based on maximum capacity”, *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1480–1489, 2018.
- [157] S. Jaber, W. Chen, K. Wang, and J. Li, “Subcarrier Assignment and Power Allocation for SCMA Energy Efficiency”, *arXiv preprint arXiv:2004.09960*, 2020.
- [158] Q. Ma, P. Yang, P. Wang, L. Peng, X. He, B. Fu, and Y. Xiao, “Power allocation for ofdm with index modulation”, in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, Jun. 2017, pp. 1–5. DOI: 10.1109/VTCspring.2017.8108623.
- [159] L. Bariah, A. Al-Dweik, and S. Muhaidat, “On the performance of non-orthogonal multiple access systems with imperfect successive interference cancellation”, in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, IEEE, 2018, pp. 1–6.
- [160] L. Bariah, S. Muhaidat, and A. Al-Dweik, “Error Probability Analysis of Non-Orthogonal Multiple Access Over Nakagami-  $m$  Fading Channels”, *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1586–1599, 2019. DOI: 10.1109/TCOMM.2018.2876867.
- [161] Shengli Zhou and G. B. Giannakis, “Adaptive modulation for multi-antenna transmissions with channel mean feedback”, in *IEEE International Conference on Communications, 2003. ICC '03.*, vol. 4, May 2003, 2281–2285 vol.4. DOI: 10.1109/ICC.2003.1204281.
- [162] N. Jindal, S. Vishwanath, and A. Goldsmith, “On the duality of Gaussian multiple-access and broadcast channels”, *IEEE Transactions on Information Theory*, vol. 50, no. 5, pp. 768–783, May 2004.
- [163] J. C. Platt and A. H. Barr, “Constrained differential optimization”, in *Neural Information Processing Systems*, 1988, pp. 612–621.
- [164] S. Dawaliby, A. Bradai, and Y. Pousset, “In depth performance evaluation of LTE-M for M2M communications”, in *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2016, pp. 1–8. DOI: 10.1109/WiMOB.2016.7763264.
- [165] S. Lee, J. Kim, J. Park, and S. Cho, “Grant-free resource allocation for noma v2x uplink systems using a genetic algorithm approach”, *Electronics*, vol. 9, no. 7, p. 1111, 2020.
- [166] A. Azari, P. Popovski, G. Miao, and C. Stefanovic, “Grant-free radio access for short-packet communications over 5g networks”, in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1–7. DOI: 10.1109/GLOCOM.2017.8255054.

- 
- [167] X. Miao, D. Guo, and X. Li, “Grant-free noma with device activity learning using long short-term memory”, *IEEE Wireless Communications Letters*, 2020.
- [168] S. M. Hasan, K. Mahata, and M. M. Hyder, “Fast uplink grant-free noma with sinusoidal spreading sequences”, *arXiv preprint arXiv:2010.00199*, 2020.
- [169] R. Abbas, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, “A novel analytical framework for massive grant-free noma”, *IEEE Transactions on Communications*, vol. 67, no. 3, pp. 2436–2449, 2018.
- [170] W. Yuan, N. Wu, A. Zhang, X. Huang, Y. Li, and L. Hanzo, “Iterative receiver design for ftn signaling aided sparse code multiple access”, *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 915–928, 2019.
- [171] F. Wei, W. Chen, Y. Wu, J. Ma, and T. A. Tsiftsis, “Message-passing receiver design for joint channel estimation and data decoding in uplink grant-free scma systems”, *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 167–181, 2018.
- [172] ZTE, ZTE Microelectronics, “Discussion on Grant-free Concept for UL mMTC”, *3GPP TSG RAN WG1 Meeting 86 R1-166405*, Gothenburg, Sweden, 22nd - 26th August 2016.
- [173] E. Kaufmann, “Analyse de stratégies bayésiennes et fréquentistes pour l’allocation séquentielle de ressources”, PhD thesis, Paris, ENST, 2014.
- [174] A. Slivkins, “Introduction to multi-armed bandits”, *arXiv preprint arXiv:1904.07272*, 2019.
- [175] M. A. Adjif, O. Habachi, and J. Cances, “Joint channel selection and power control for noma: A multi-armed bandit approach”, in *2019 IEEE Wireless Communications and Networking Conference Workshop (WCNCW)*, 2019, pp. 1–6.
- [176] Z. Tian, J. Wang, J. Wang, and J. Song, “Distributed NOMA-Based Multi-Armed Bandit Approach for Channel Access in Cognitive Radio Networks”, *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1112–1115, 2019.
- [177] A. Feki and V. Capdevielle, “Autonomous resource allocation for dense lte networks: A multi armed bandit formulation”, in *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*, 2011, pp. 66–70.
- [178] J. G. Proakis and M. Salehi, *Digital communications*. McGraw-hill New York, 2001, vol. 4.
- [179] A. Zanella, M. Chiani, and M. Z. Win, “MMSE reception and successive interference cancellation for MIMO systems with high spectral efficiency”, *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 1244–1253, May 2005, ISSN: 1558-2248. DOI: 10.1109/TWC.2005.847103.
- [180] J. Kennedy and R. Eberhart, “Particle swarm optimization”, in *Proceedings of ICNN’95-International Conference on Neural Networks*, IEEE, vol. 4, 1995, pp. 1942–1948.

- [181] M. G. Sahab, V. V. Toropov, and A. H. Gandomi, “A review on traditional and modern structural optimization: Problems and techniques”, *Metaheuristic applications in structures and infrastructures*, pp. 25–47, 2013.
- [182] P. Auer, N. Cesa-Bianchi, and P. Fischer, “Distributed NOMA-Based Multi-Armed Bandit Approach for Channel Access in Cognitive Radio Networks”, *Machine learning*, vol. 47, no. 2-3, pp. 235–256, 2002.
- [183] R. Bonnefoi, L. Besson, C. Moy, E. Kaufmann, and J. Palicot, “Multi-armed bandit learning in iot networks: Learning helps even in non-stationary settings”, in *International Conference on Cognitive Radio Oriented Wireless Networks*, Springer, 2017, pp. 173–185.
- [184] Y. Zhou, J. Zhu, and J. Zhuo, “Racing thompson: An efficient algorithm for thompson sampling with non-conjugate priors”, in *International Conference on Machine Learning*, PMLR, 2018, pp. 6000–6008.
- [185] S. Agrawal and N. Goyal, “Analysis of thompson sampling for the multi-armed bandit problem”, in *Conference on learning theory*, JMLR Workshop and Conference Proceedings, 2012, pp. 39–1.
- [186] N. Gupta, O. Granmo, and A. Agrawala, “Thompson sampling for dynamic multi-armed bandits”, in *2011 10th International Conference on Machine Learning and Applications and Workshops*, vol. 1, 2011, pp. 484–489. DOI: 10.1109/ICMLA.2011.144.
- [187] Q. Yu, H. Li, W. Meng, and W. Xiang, “Sparse code multiple access asynchronous uplink multiuser detection algorithm”, *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 5557–5569, 2019. DOI: 10.1109/TVT.2019.2891771.

## AVIS DU JURY SUR LA REPRODUCTION DE LA THESE SOUTENUE

**Titre de la thèse:**

Etude des mécanismes d'accès multiples non orthogonaux pour le déploiement massif de l'IoT dans les futurs réseaux cellulaires

**Nom Prénom de l'auteur : BEN AMEUR WISSAL**

**Membres du jury :**

- Madame GOURSAUD Claire
- Monsieur ROS Laurent
- Monsieur HELARD Jean-François
- Monsieur SCHWOERER Jean
- Monsieur CANCES Jean-Pierre
- Monsieur CLAVIER Laurent
- Monsieur MARY Philippe
- Madame DUMAY Marion

Président du jury : **Monsieur CANCES Jean Pierre**  
Professeur ENSIL/ENSCI

Date de la soutenance : 25 Mai 2021



Reproduction de la these soutenue

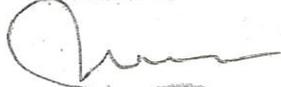
- Thèse pouvant être reproduite en l'état  
 Thèse pouvant être reproduite après corrections suggérées

Fait à Rennes, le 25 Mai 2021

Signature du président de jury



Le Directeur,



Abdellatif MIRAOU



**Titre :** Etude des mécanismes d'accès multiples non orthogonaux pour le déploiement massif de l'IoT dans les futurs réseaux cellulaires

**Mots clés :** 5G, mMTC, NOMA, allocation de puissance, accès aux ressources

**Résumé :** Les communications massives de type machine (mMTC) sont l'un des trois principaux domaines d'applications des futurs réseaux cellulaires. Les communications mMTC se caractérisent notamment par un nombre massif d'objets à connecter à faible coût, typiquement 1 million d'objets par km<sup>2</sup>, une transmission sporadique de paquets courts et une faible consommation énergétique des terminaux afin de prolonger leur durée de vie. Les schémas d'accès multiples orthogonaux conventionnels ne répondent pas aux exigences spécifiques des communications mMTC. Le nombre d'utilisateurs communiquant simultanément dépend particulièrement de la quantité de ressources disponibles et de la granularité de l'ordonnement. Cette limite peut être dépassée en renonçant à l'orthogonalité inter-utilisateur. Les mécanismes d'accès multiple non orthogonaux (NOMA) sont considérés comme une solution prometteuse pour relever ces défis des communications mMTC. Ils permettent à plusieurs utilisateurs de partager simultanément et non-orthogonalement la même ressource, ce qui augmente le nombre d'utilisateurs pris en charge. Mais, ils introduisent également de fortes interférences inter-utilisateurs, nécessitant des techniques robustes de détection de signaux avec une complexité de décodage potentiellement élevée.

Cette thèse porte sur l'optimisation des schémas NOMA en termes d'implémentation des émetteurs et des récepteurs, de complexité et de performances. Elle aborde également le problème d'accès aux canaux, c'est-à-dire avec ou sans allocation de ressources. Notre objectif est d'assurer un compromis entre les performances souhaitées et la satisfaction des exigences strictes des communications mMTC. Tout d'abord, nous menons une étude qualitative et quantitative des mécanismes NOMA les plus populaires afin de sélectionner les techniques les plus adaptées à notre cas d'étude. Ensuite, nous nous intéressons à l'optimisation des performances des mécanismes sélectionnés en proposant un nouvel algorithme d'allocation de puissance centralisé dans différents scénarios avec différentes combinaisons de contraintes. Enfin, la dernière partie consiste à proposer une nouvelle approche permettant à chaque utilisateur de choisir de manière autonome sa meilleure puissance d'émission assurant un décodage correct de ses paquets en utilisant uniquement les acquittements naturels de la station de base pour une stratégie d'accès sans allocation de ressources. Ce procédé permet d'améliorer les performances avec une signalisation minimale, une faible complexité du système et donc une consommation d'énergie réduite.

**Title :** Study of non-orthogonal multiple access schemes for the massive deployment of IoT in future cellular networks

**Keywords :** 5G, mMTC, NOMA, power allocation, resource access

**Abstract:** Massive machine type communications (mMTC) is one of the expected use cases of the future radio access network. mMTC intends to involve a massive number of low-cost devices, i.e., 1 million devices per km<sup>2</sup>, with short packet communications, low energy consumption and thereby long battery life. The conventional orthogonal multiple access (OMA) generates a limited inter-user interference, requiring simple multi-user detection techniques with low decoding complexity. However, the maximum number of supported users is limited by the number of available orthogonal resources. Therefore, OMA schemes seem incapable to address the stringent mMTC requirements. In this context, the non-orthogonal multiple access (NOMA) schemes are targeted as a promising solution to meet the mMTC challenges. They allow multiple users to share simultaneously and non-orthogonally the same resource, which increases the capacity of connectivity, but at the cost of higher inter-user interference compared to OMA schemes. As a consequence, robust multi-user detection techniques with potentially high decoding complexity are required for signals separation. This thesis focuses on the study and optimization of the most promising NOMA schemes in terms of implementation of transmitters and receivers, decoding complexity and achieved performance.

It also addresses the problem of resource access strategy, i.e., grant-based or grant free access. The main challenge is to ensure a trade-off between the desired performance and the satisfaction of the stringent requirements of mMTC. Firstly, we provide a quantitative and qualitative comparison of the most popular NOMA schemes in order to select the most suitable ones for mMTC. Then, we focus on improving the performance of the selected NOMA schemes while considering a grant-based access strategy. We propose a novel centralized power allocation algorithm in order to deal with the inter-user interference and alleviate the error propagation problem. This algorithm is investigated in different scenarios with different system constraints combinations. After that, we handle the signaling overhead issue and the power decision problem for a grant free access strategy by proposing a novel algorithm allowing each user to select autonomously its adequate power in order to properly decode all users' packets, with no information about the interfering users and based only on the natural acknowledgment from the base station. This process aims to ensure the best system performance with minimum signaling overhead, low decoding complexity and thus low energy consumption.