COLLEGE MATHS, TELECOMS
DOCTORAL INFORMATIQUE, SIGNAL
BRETAGNE SYSTEMES, ELECTRONIQUE

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES RENNES

THESE DE DOCTORAT DE

L'INSTITUT NATIONAL DES SCIENCES APPLIQUEES RENNES

ECOLE DOCTORALE N° 601
Mathématiques, Télécommunications, Informatique , Signal, Systèmes, Electronique
Spécialité : , *Télécommunications*
Par

Aya SHEHATA

Data rate compression of baseband OFDM signals for virtualized processing in cloud radio access networks

Thèse présentée et soutenue à Rennes, le 04/07/2023
Unité de recherche : Institut d'Électronique et des Technologies du numéRique (IETR) – UMR CNRS 6164
Thèse N° : 23 ISAR 18 / D23 - 18

Rapporteurs avant soutenance :

Arsenia CHORTI	Professeure, ENSEA Cergy-Pontoise
Laurent CLAVIER	Professeur, IMT Lille-Douai, Lille

Composition du Jury :

Président :	Didier LE RUYET	Professeur, CNAM, Paris
Examinateurs :	Arsenia CHORTI	Professeure, ENSEA Cergy-Pontoise
	Laurent CLAVIER	Professeur, IMT Lille-Douai, Lille
	Didier LE RUYET	Professeur, CNAM, Paris
	Claire GOURSAUD	Maîtresse de Conférences, INSA de Lyon, Lyon
Dir. de thèse :	Matthieu CRUSSIERE	Professeur, INSA de Rennes, Rennes
	Philippe MARY	Professeur, INSA de Rennes, Rennes

Acknowledgement

First, I would like to express my sincere gratitutde to my thesis advisors, Profs. Matthieu Crussiere and Philippe Mary, for their great help during my doctoral work. Their scientific ideas, their discussions, continious guidnance and the trust they put in me were the reason why I completed this journey that I will never forget. I would also like to acknowledge their human qualities that helped me through every difficult time during these three years.

I would also like to thank all my colleagues at INSA who helped me feel comfortable and welcomed me to the IETR during these three years, as well as Pascal Richard, Aurore Gouin and Justine Grommaire, who are in charge of the administration of the IETR, for facilitating the administrative and logistical processes and for always giving me advice and support in my daily life.

I would like to thank the jury members Prof. Didier LE Ruyet, Prof. Arsenia Chorti, Prof. Laurent Clavier, and Prof. Claire Goursaud for reviewing and discussing my thesis.

Without the support of my family, this thesis would definitely not have been completed. My special and sincere thanks go to my parents, Hany and Somia. My father's love, words, and belief in me, in addition to my mother's support, sacrifice, care, encouragement, and precious prayers every second of the day, were the fuel that allowed me to keep going and trying every day. This work is definitely dedicated to them.

I cannot close without thanking my brother Ahmed. He is my friend, backbone, safe net, advisor, supporter, the one who listens to me as much as possible so I can continue. I wish him the best in life because he really deserves it. Finally, the biggest thanks goes to my beautiful and joyful little daughter Sarah. Her constant smile and affection help me to overcome all the difficulties we had. She is my source of strength, inspiration and energy.

Résumé étendu

Introduction

Les opérateurs de réseaux mobiles sont confrontés à une augmentation considérable du trafic de données mobiles sur le réseau en raison de la popularité croissante des smartphones, des tablettes et des nouveaux appareils intelligents connectés qui prennent en charge un large éventail d'applications et de services gourmands en ressources. Ces appareils sont très gourmands en bande passante, en particulier les services vidéo, qui représentent les deux tiers du trafic de données mobiles. En outre, les utilisateurs ont des attentes élevées en matière de qualité, quels que soient le lieu, l'heure et la situation. En effet, le débit de données offert est passé de quelques kbps avec la norme Global System for Mobile Communications (GSM) dans les années 90 à des centaines de Mbps aujourd'hui avec les réseaux cellulaires de la dernière génération 5^{th}. Dans le même temps, les systèmes de radiodiffusion télévisuelle sont passés de l'analogique au numérique pour répondre à la demande croissante d'une expérience télévisuelle de meilleure qualité.

Pour soutenir cette croissance, il faut déployer davantage de cellules ou une plus grande capacité par cellule, ce qui entraîne une augmentation des coûts. En même temps, les utilisateurs aimeraient payer moins cher et utiliser plus de données. Par conséquent, la croissance des coûts ne peut pas suivre la même ligne de tendance que la croissance du trafic. Pour répondre aux exigences de capacité, les petites cellules peuvent être déployées dans des endroits où l'activité des utilisateurs est élevée. Dans ce cas, les interférences deviennent un défi, et des techniques telles que la coordination améliorée des interférences entre les cellules et les liaisons multipoints coordonnées sont essentielles. Enfin, pour augmenter la capacité de chaque cellule, il est possible de déployer davantage d'antennes à l'aide d'une technique massive d'entrées multiples et de sorties multiples.

Les réseaux d'accès radio virtualisé (C-RAN) constituent une nouvelle architecture de réseau qui permet de relever un certain nombre de défis auxquels sont confrontées les technologies récentes lorsqu'elles tentent de répondre aux demandes sans cesse croissantes des utilisateurs finaux. L'idée principale du C-RAN est de diviser la couche physique en deux parties : la bande de base regroupant le traitement du signal dans des serveurs banalisés, centralisés (Baseband Unit, BBU, en anglais) et de laisser seulement les têtes d'émissions sur les stations de bases (remote radio head, RRH, en anglais). Cela pourrait permettre une meilleure efficacité énergétique et des économies potentielles sur les ressources en bande de base, puisque celle-ci est mutualisée pour plusieurs cellules. En outre, la capacité du réseau est améliorée grâce à l'équilibrage de la charge et au traitement coopératif des signaux provenant de plusieurs stations

de base. Par conséquent, les stations de base à forte charge et à forte densité de déploiement peuvent être déployées dans l'architecture C-RAN.

Cependant, certains défis doivent être relevés avant que l'architecture C-RAN puisse être déployée commercialement. L'architecture C-RAN introduit une énorme surcharge sur les liaisons optiques entre le RRH et le BBU. Par rapport aux exigences de la liaison arrière, les exigences de la liaison avant devraient être 50 fois plus élevées [1]. Les énormes besoins en bande passante pour la transmission d'échantillons IQ à haute résolution générés par le traitement en bande de base dans les serveurs restent l'un des principaux défis de cette architecture. L'approche classique consiste à utiliser des fibres optiques à plus grande largeur de bande pour transmettre les échantillons IQ, mais cela a un coût très élevé. Par conséquent, la compression des données IQ avant la transmission sur la liaison fronthaul à capacité limitée est une approche intéressante pour réduire le débit de données pour l'architecture C-RAN.

Cette thèse est consacrée à ce défi. Elle vise à étudier différentes approches pour la compression du débit de données qui représentent un compromis entre la performance de compression réalisable, la distorsion du signal, la complexité de la conception et le temps de calcul. Dans la littérature, la quantification et le codage entropique sont les principales techniques de réduction du débit de données. En effet, les signaux OFDM (orthogonal frequency division multiplexing) est le schéma de modulation le plus couramment utilisé dans les normes récentes, comme par exemple le LTE (Long Term Evolution) [2], DVB-T2 (Digital Video Broadcasting - Terrestrial second generation) [3], et ATSC3.0 (Advanced Television Systems Committee version 3) [4]. Par conséquent, les études dans la littérature s'appuient principalement sur l'exploitation des caractéristiques statistiques du signal OFDM pour réduire le débit de données sur la liaison fronthaul.

Grâce au théorème centrale limite, les échantillons IQ du symbole OFDM sont gaussiens à la sortie du modulateur IFFT. Par conséquent, les études proposées dans la littérature pour la compression des échantillons supposent cette distribution. Toutefois, l'ajout de nombreuses sous-porteuses via la IFFT produit un signal caractérisé par de fortes fluctuations de puissance mesurées par le PAPR (Peak to Average Power Ratio), ce qui reste l'un des principaux inconvénients de l'OFDM. La valeur élevée du PAPR empêche les concepteurs de radiofréquences d'injecter le signal au point optimal de l'amplificateur de puissance pour réduire la consommation d'énergie. De nombreuses techniques ont été présentées dans la littérature pour réduire le PAPR de l'OFDM.

Toutefois, le pré-traitement du signal dans le domaine temporel à l'aide d'une technique de réduction du PAPR modifie la distribution des échantillons IQ, ce qui rend les quantificateurs gaussiens sous-optimaux. À notre connaissance, aucune étude n'a été réalisée sur la robustesse du quantificateur et codeur entropique gaussien à une variation de la statistique de source engendrée par les techniques de réduction du PAPR avant la compression. De même, aucun quantificateur optimisé pour les nouvelles distributions, i.e. en tenant compte de la réduction de PAPR, n'a été proposé.

Nous avons donc proposé des techniques de compression, i.e. quantificateur et codeur entropique, pour des signaux OFDM traités par des techniques de réduction du PAPR. Nous considérons les techniques d'écrêtage et de réduction du PAPR par réservation de sous-porteuse TR (Tone Reservation). L'écrêtage est la technique la

plus simple et le TR est une méthode sans distortion adoptée par plusieurs normes, par exemple DVB-T2 [3] et ATSC 3.0 [4]. Pour la méthode TR, l'algorithme QCQP (Quadratically Constrained Quadratic Problem) est étudié ; cet algorithme est considéré comme étant le meilleur pour la réduction du PAPR au sens défini par un problème quadratique. Nous évaluons le gain de performance que l'on peut obtenir lorsque le schéma de compression est optimisé en prenant en compte la statistique du signal modifiée soit par écrêtage soit par l'algorithme TR-QCQP. La métrique de performance est le MER (Modulation Error Ratio) en fonction de la résolution, i.e. le nombre de bits par échantillons.

L'obtention de la distribution du signal après modification de son PAPR peut être difficile (c'est le cas pour la technique TR-QCQP) et la mise au point du quantificateur et du codeur entropique sur cette base l'est davantage. Ainsi, l'idée de concevoir un schéma de compression prenant en compte n'importe quelle technique de réduction de PAPR à base de réseaux neuronnales fait sens. L'idée est de concevoir une structure d'auto-encodeur (AE) que l'on peut entrainer de bout en bout pour à la fois obtenir des échantillons IQ optimisés su sens du PAPR et de la compression. Nous explorons cela dans le dernier chapitre contributif de la thèse, ou nous comparons notamment les performances de l'architecture obtenue avec les schémas classiques de compression vus dans les chapitres précédents.

Chapitre 1
État de l'art et contexte

Le chapitre 1 présente le contexte de l'étude en donnant un aperçu de l'architecture C-RAN. Les avantages et les défis d'une telle architecture sont mis en évidence. Ensuite, nous présentons le problème du débit élevé des échantillons de données IQ qui sont générés dans les serveurs distants et doivent être transportés sur la liaison fronthaul jusqu'à l'émetteur. L'objectif du travail est de réduire le débit de données avant de les transmettre sur la liaison fronthaul sans affecter la qualité du signal. Nous faisons ensuite un rappel sur les schémas de modulation OFDM les plus utilisés dans les réseaux. Nous rappelons le problème du PAPR élevé caractérisant le signal OFDM, et nous donnons une description détaillée des différents schémas de compression du signal existant dans la littérature ; en particulier le processus d'élimination de la redondance dans le domaine spectral, les techniques de quantification (uniforme, non uniforme, scalaire et vectorielle) et les techniques de codage entropique.

Chapitre 2
Analyse des techniques de compression des échantillons IQ

Le chapitre 2 vise à présenter le premier problème de la thèse. Il décrit en détail l'état de l'art en matière de compression d'échantillons en bande de base à distribution gaussienne avant la transmission sur la liaison fronthaul dans l'architecture C-RAN. Il examine également les différents paramètres utilisés pour évaluer le schéma de compression en termes de compromis entre la qualité du signal et la complexité tout en obtenant un gain de compression acceptable.

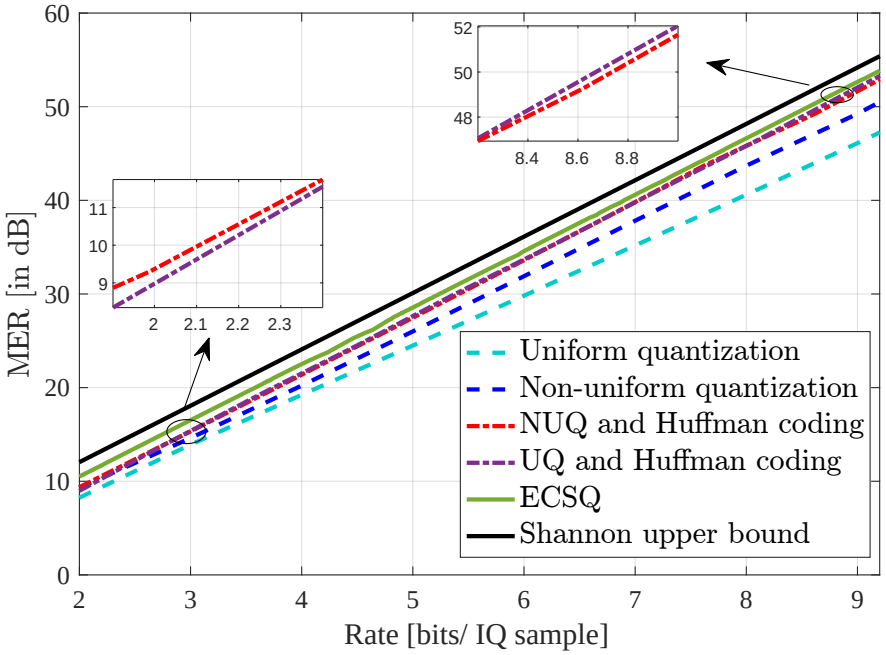


Figure 1: MER en fonction de la résolution R pour une distribution gaussienne complexe i.i.d. pour différents schémas de quantification scalaire.

La Figure 1 montre le MER en fonction de la résolution R pour différents quantificateurs scalaires : le quantificateur uniforme avec et sans codage de Huffman, le quantificateur non uniforme avec et sans codage de Huffman et le quantificateur optimisé sous contrainte d'entropie pour un signal OFDM gaussien. La borne supérieure de Shannon est également tracée pour comparaison. Le quantificateur non uniforme (NUQ) présente un net avantage par rapport au quantificateur uniforme (UQ) notamment lorsque la résolution augmente. Toutefois, l'utilisation d'un codeur entropique de Huffman à la sortie des quantificateurs uniformes et non uniformes rend les performances des deux schémas très proches. Ainsi, le quantificateur uniforme proposé en combinaison avec le codage entropique exploite la distribution statistique du signal comme dans le NUQ, mais avec une complexité de calcul moindre. En outre, le quantificateur uniforme suivi du codage de Huffman se trouve à 2.4 dB en MER de la limite supérieure de Shannon, pour $R = 9$ bits/échantillons. Le quantificateur scalaire

à contrainte entropique (ECSQ) est inférieur de 1.6 dB à la limite supérieure, ce qui est mieux que les autres schémas mais ECSQ est aussi plus complexe qu'un quantificateur uniforme avec un code de Huffman.

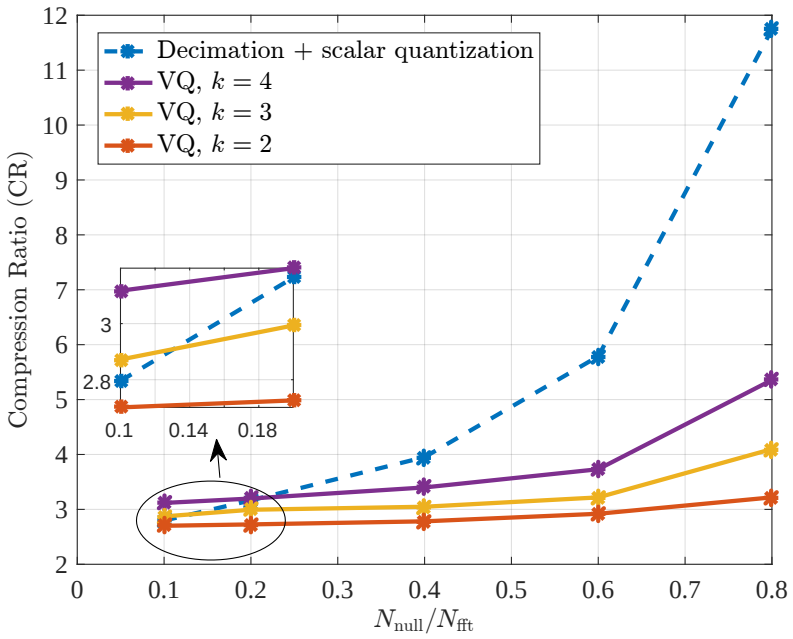


Figure 2: Taux de compression en fonction du facteur de suréchantillonnage L de deux techniques : décimation plus quantification scalaire uniforme et quantification vectorielle à longueur fixe.

La figure 2 montre le taux de compression en fonction du facteur de suréchantillonnage pour deux techniques de compression : une quantification vectorielle (VQ), où les échantillons à la sortie de la IFFT sont quantifiés par bloc de 2, 3 ou 4 échantillons, et un schéma ou l'on décime d'abord le signal pour se ramener au plus près du taux de Nyquist suivi d'une quantification scalaire. Nous avons supposé une valeur fixe de MER à 30 dB et N_{null} et N_{fft} sont le nombre de sous-porteuses nulles sur le nombre total de sous-porteuse et sont liés au facteur de sur-échantillonnage. On constate que le gain obtenu par la décimation à un faible rapport $N_{\text{null}}/N_{\text{fft}}$ est très faible en raison des faibles degrés de liberté dans ce cas et que le VQ avec une dimension vectorielle plus élevée $k = 3$ et 4 donne de meilleurs résultats car les échantillons sont plus corrélés. D'autre part, le gain de décimation augmente de manière significative en augmentant le rapport $N_{\text{null}}/N_{\text{fft}}$ et atteint un gain de compression environ 6 fois plus élevé qu'avec la quantification vectorielle avec $k = 3$ lorsque $N_{\text{null}}/N_{\text{fft}} = 0.8$. Outre les performances de compression, la complexité des deux algorithmes est également une caractéristique importante à prendre en compte. Le VQ a une complexité exponentielle en termes de dimension vectorielle et de résolution, tandis que les techniques de décimation et de quantification scalaire ont une complexité supra-linéaire en termes de longueur de filtre, mais exponentielle en termes de résolution de quantification. Ainsi, ces analyses numériques peuvent être utilisées pour sélectionner la technique de compression la plus appropriée pour une application donnée en fonction de ses spécifications.

Chapitre 3
Compression d'échantillons OFDM IQ avec réduction du PAPR

Dans le chapitre précédent, la compression des échantillons IQ dans le domaine temporel a été réalisée à la sortie du modulateur IFFT. Les échantillons IQ suivent quasiment une loi gaussienne si le nombre de sous-porteuses est suffisemment grand et donc le quantificateur gaussien peut être utilisé comme solution optimale au problème de compression. Cependant, le signal OFDM transmis est caractérisé par un PAPR élevé, et les techniques présentées pour la réduction du PAPR modifient la distribution gaussienne du signal OFDM, ce qui rend le quantificateur gaussien sous-optimale.

Par conséquent, des techniques de compression adaptées aux signaux à PAPR réduit doivent être proposées. La distribution des échantillons IQ, après avoir été traités soit par écrêtage soit par la technique QCQP, est dérivée. Ces distributions permettent ensuite de calculer les mots de codes et les seuils du quantificateur optimale ainsi que la longueur des mots de code pour le codage entropique. On peut ainsi obtenir une expression du MER asymptotique reflétant l'effet conjoint de la réduction du PAPR et de la quantification.

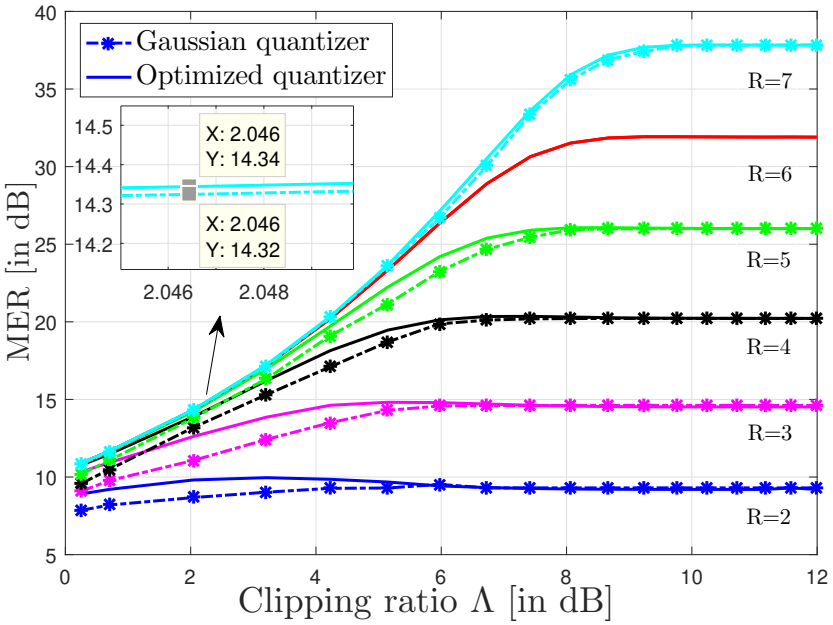


Figure 3: MER des techniques de compression gaussienne et optimisée par rapport au ratio d'écrêtage Λ pour différentes résolutions.

La figure 3 montre le MER obtenu en fonction du ratio d'écrêtage Λ pour un quantificateur gaussien et un quantificateur conçu pour les signaux écrêtés. Les performances en MER du signal écrêté sont obtenues à partir de la distribution du signal écrêté et de la formule du MER asymptotique (Fig. 3.13, Chapitre 3). On constate que l'application de techniques de compression optimisées à un signal écrêté améliore le MER à des résolutions faibles et à des rapports d'écrêtage faibles, c'est-à-dire pour un signal fortement écrêté. Toutefois, par rapport aux schémas de compression pure-

ment gaussiens, le degré d'amélioration diminue avec l'augmentation de la résolution, atteignant environ 0,02 dB lorsqu'une moyenne de 7 bits par échantillon IQ est utilisée.

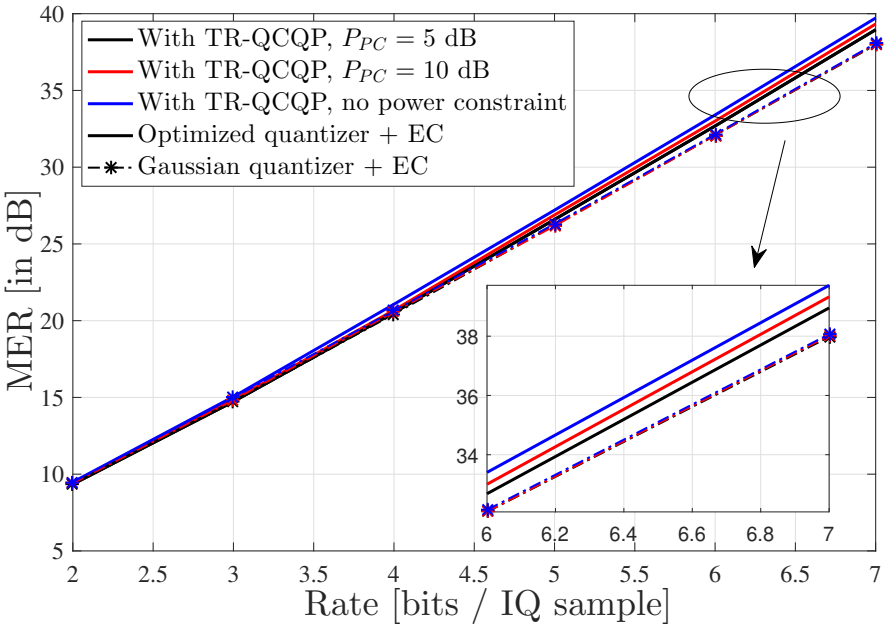


Figure 4: MER des techniques de compression gaussienne et optimisée en fonction de la résolution pour différentes contraintes de puissance P_{PC} et $N_{\mathrm{PRT}} = 1\%$.

Les figures 4 et 5 montrent l'évolution du MER en fonction de la résolution pour les techniques de compression basées sur une distribution gaussienne du signal et la compression obtenue pour un signal à PAPR réduit par réservation de sous-porteuses. De même, le MER est celui dérivé asymptotiquement (Fig. 3.16, Chapitre 3). Pour la technique QCQP, l'influence de divers paramètres est étudié à savoir, la contrainte de puissance P_{PC} allouée aux sous-porteuses, Fig. 4, et le pourcentage de sous-porteuses réservées N_{PRT}, Fig. 5. Il est démontré que l'optimisation des techniques de compression pour la distribution du signal à tonalité réservée améliore les performances du MER à des résolutions élevées. En outre, le gain de performance en MER augmente de manière significative lorsque la contrainte de puissance P_{PC} ou le pourcentage de sous-porteuses réservées N_{PRT} augmente. On constate que le quantificateur gaussien a de bonnes performances sur des signaux à PAPR réduit par réservation de sous-porteuses car sa distribution est presque gaussienne (Figs. 3.6 et 3.7, chapitre 3). Cependant, nous montrons que le quantificateur optimisé permet d'obtenir un gain significatif dans la région de quantification à haut débit ; le gain est de 1.6 dB par rapport au quantificateur gaussien à $R = 7$ bits et $P_{\mathrm{PC}} = 10$ dB et $N_{\mathrm{PRT}} = 1\%$, qui sont les paramètres d'un système DVB-T2.

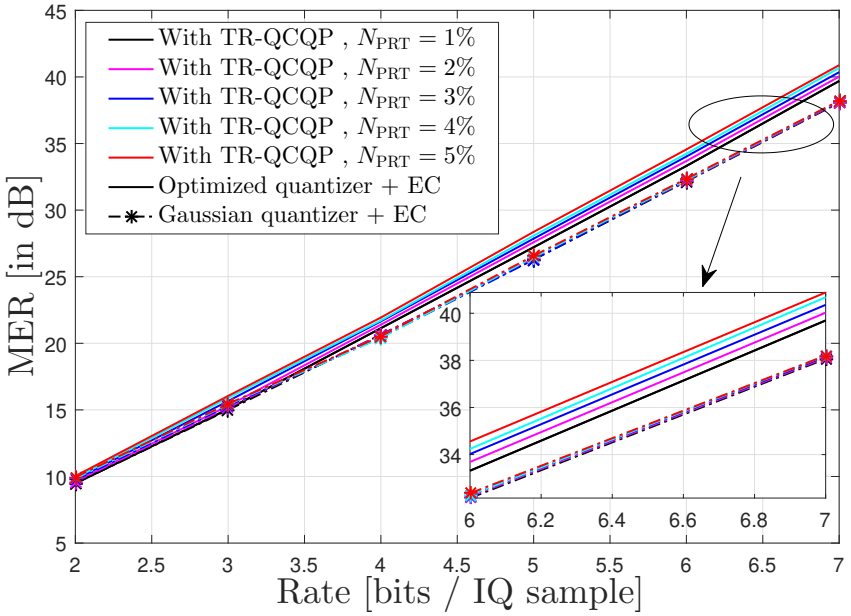


Figure 5: MER des techniques de compression gaussienne et optimisée en fonction de la résolution pour différents pourcentages d'outillage N_{PRT} et $P_{\mathrm{PC}} = 10$ dB.

Chapitre 4
Réduction conjointe du PAPR et compression du signal basée sur un réseau neuronal profond

Dans les chapitres précédents nous avons mis en évidence l'impact des techniques de réduction du PAPR dans l'optimisation des techniques de compression dans une architecture C-RAN. Dans ce chapitre, nous explorons la possibilité d'apprendre à réduire le PAPR d'un signal ainsi que le quantificateur optimale à partir d'une structure d'auto-encodeur (AE) entrainé de bout en bout à partir des échantillons IQ, sans qu'il soit nécessaire de connaître la distribution du signal d'entrée.

Dans la première partie de ce chapitre, nous explorons les architectures AE possibles pour seulement réduire le PAPR du signal OFDM. La solution proposée est comparée aux approches conventionnelles et aux techniques récentes de réduction du PAPR basées sur les réseaux neuronaux. La partie codeur de l'AE est mise en œuvre au niveau de l'émetteur sur le signal OFDM dans le domaine temporel pour réduire le PAPR du signal transmis, puis la partie décodeur est mise en œuvre au niveau du récepteur pour reconstruire le signal d'origine. Le codeur et le décodeur ont une structure de réseau identique, qui consiste en quatre couches, c'est-à-dire une couche d'entrée, deux couches cachées et une couche de sortie, et le nombre de neurones dans chaque couche cachée est de 2048. Le réseau AE est ensuite entraîné afin de réduire le PAPR du signal transmis à la sortie du codeur et pour reconstruire le signal à la sortie du décodeur sans dégrader le MER.

La figure 6 montre la CCDF du PAPR du signal à PAPR réduit $\mathbf{x}_{\mathrm{red}}$ du système proposé, qui donne la probabilité d'excéder un ratio de ψ dB en PAPR. On constate

que pour une probabilité de 10^{-2}, le système proposé améliore le PAPR de 7.7 dB par rapport au signal OFDM original. En outre, le réseau proposé est plus performant que les schémas conventionnels d'écrêtage et de TR-QCQP sans la complexité et le temps de convergence de l'algorithme TR ou la distorsion de l'algorithme d'écrêtage, ce qui sera clairement visible dans l'analyse du Taux d'erreur bits (BER). Enfin, il a surpasse légèrement l'algorithme PRNet basé sur un réseau neuronal, mais avec une complexité de calcul plus faible.

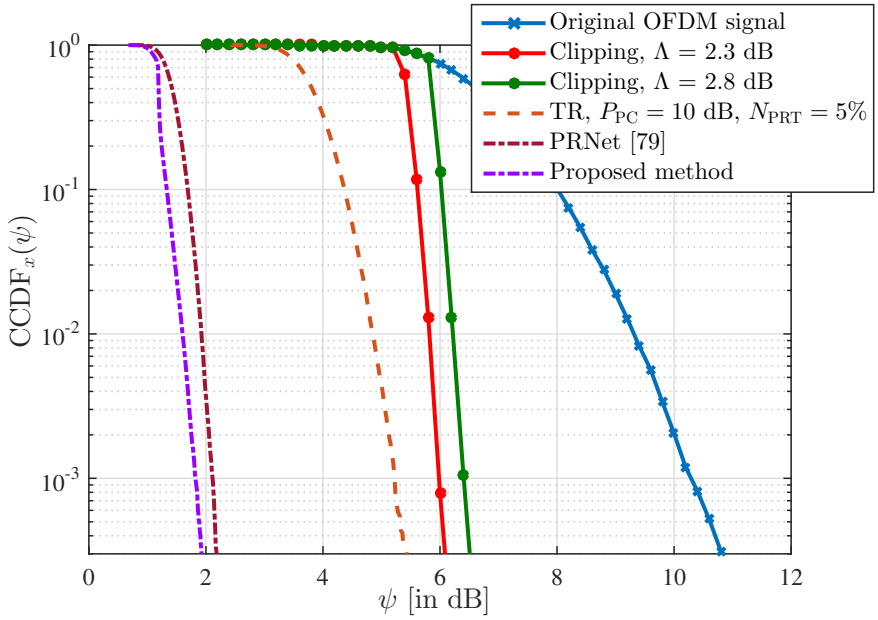


Figure 6: CCDF du PAPR pour le schéma de réduction du PAPR proposé et ceux dits conventionnels.

La figure 7 montre le BER de l'architecture proposée par rapport aux techniques classiques dans le canal à bruit blanc additif gaussien (AWGN). Compte tenu du compromis entre la réduction du PAPR et les performances en BER, le PAPR cible pour le réseau proposé est fixé à $\text{PAPR}_t = 1.2$ dB dans la fonction de perte conjointe conformément à (4.11). Bien que le schéma d'écrêtage permette de réduire le PAPR du signal OFDM, à mesure que la réduction augmente, le BER du système se dégrade en raison de la distorsion du signal par le processus d'écrétage, ce qui constitue le principale inconvénient de ce schéma. En revanche, le TR-QCQP est sans distorsion en raison de l'orthogonalité entre les sous-porteuses de données et celles réservées pour la réduction du PAPR dans le domaine fréquentiel et affiche donc les meilleures performances en termes de BER. On remarque que le BER de la solution proposée est quasiment égal à celui du schéma TR-QCQP sans distorsion, alors que PRNet et l'écrétage détériorent sensiblement les performances. En effet, dans le réseau proposé, nous avons intégré une contrainte de MER afin de garantir la reconstruction du signal contrairement à PRNet et à l'écrétage.

Dans la deuxième partie du chapitre, nous étendons l'architecture précédente pour inclure un réseau de quantification. L'auto-encodeur est entraîné de bout en bout avec une fonction de perte multiobjectif prenant en compte le MER, une contrainte

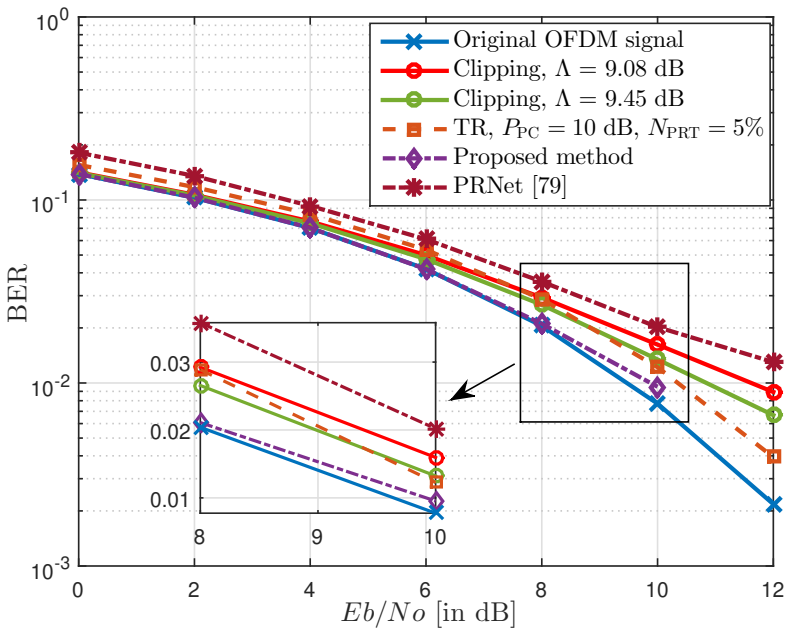


Figure 7: BER des différents algorithmes sur un canal AWGN pour une modulation 4-QAM.

de PAPR et une contrainte d'entropie pour obtenir un signal quantifié et compressé à PAPR réduit transmis sur la liaison fronthaul. Le réseau de quantification est développé en déployant l'algorithme itératif conventionnel de Lloyd en tant que couches du réseau d'apprentissage. Le codeur et les réseaux de quantification sont déployés dans la BBU pour générer un signal quantifié à PAPR réduit qui est transmis sur la liaison fronthaul. Ensuite, le décodeur est mis en œuvre dans le récepteur pour reconstruire le signal OFDM original.

La figure 8 montre le MER en fonction de la résolution R pour l'auto-encodeur avec quantification scalaire, i.e. le quantificateur travail par échantillon. En outre, les résultats sont présentés pour différentes valeurs cibles du PAPR. Nous comparons avec un système OFDM dans lequel un algorithme de réduction du PAPR, écrêtage ou TR-QCQP, est d'abord utilisé puis l'algorithme de Lloyd est utilisé pour quantifier le signal OFDM à PAPR réduit, suivi d'un codeur entropique de Huffman. Nous constatons tout d'abord que les performances du signal TR-QCQP en matière de MER sont supérieures à celles du signal écrêté, ce qui s'explique par les distorsions ajoutées par l'écrêtage. Ensuite, on constate que le MER obtenu avec l'auto-encodeur-quantificateur proposé surpasse systématiquement les performances de compression de l'algorithme de Lloyd appliqué au signal à PAPR réduit.

Le réseau proposé améliore le MER d'environ 1.1 dB par rapport à l'algorithme conventionnel de Lloyd lorsqu'une moyenne de 3 bits par échantillon IQ est utilisée. À des débits binaires plus élevés de 7 bits par échantillon IQ, l'amélioration de la performance du MER est réduite à 0.74 dB. Cela pourrait s'expliquer par le fait qu'avec l'architecture de réseau proposée, il est difficile d'apprendre efficacement les centres de quantification à grain fin pour des débits binaires plus élevés afin d'améliorer les

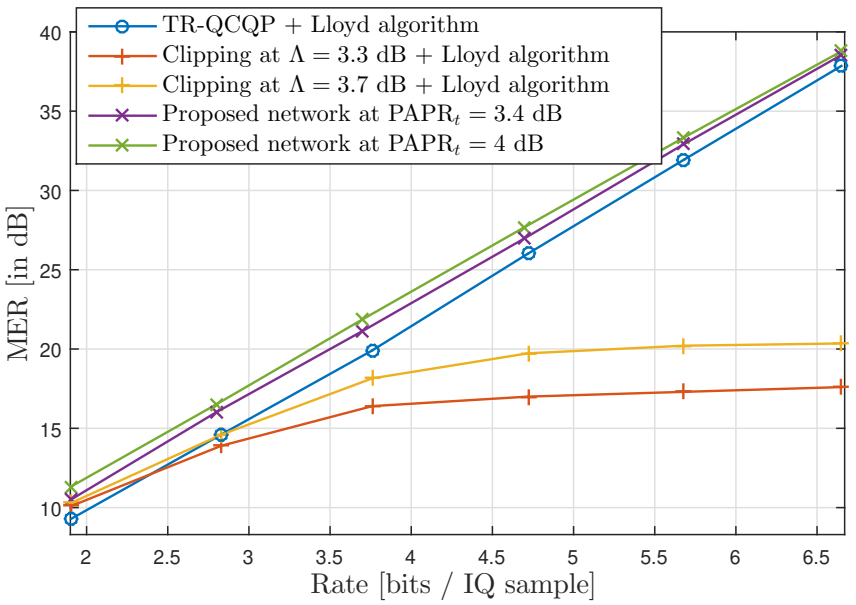


Figure 8: MER en fonction de la résolution R pour le réseau d' AE-quantification proposé appliqué au signal OFDM pour différentes valeurs cibles de PAPR.

performances de compression.

Enfin, la performance du MER s'améliore en augmentant la valeur cible du PAPR. Lorsque la valeur cible du rapport PAPR passe de 3.4 à 4 dB, le MER s'améliore de 1.90 dB à R=3 bits par échantillon et de 0.92 dB à un taux plus élevé de 7 bits par échantillon. Par conséquent, la valeur cible du rapport PAPR est fixée en fonction des exigences du système en MER. Si le système peut tolérer un MER plus faible, on peut fixer une valeur PAPR$_t$ plus petite pour améliorer l'efficacité énergétique de l'amplificateur de puissance dans une plus large mesure.

Pour une comparaison équitable et significative, la figure 9 montre la fonction de répartition complémentaire du PAPR de différentes techniques, et en particulier celle obtenue avec notre approche pour deux valeurs cibles de PAPR, i.e. PAPR$_t$ = 3.4 et 4 dB. On constate qu'à CCDF = 1×10^{-2}, le réseau peut obtenir une réduction du PAPR de 3.2 dB pour un PAPR cible de 3.4 dB par rapport au signal OFDM. Il surpasse également l'écrêtage et a approximativement les mêmes performances que l'algorithme TR-QCQP optimal, où la contrainte de puissance pour les sous-porteuses PRT P_{PC} est fixée à 10 dB et le nombre de sous-porteuses réservées à la réduction du PAPR par rapport au nombre total de sous-porteuses N_{PRT} est fixé à 5%. En revanche, pour P_{PC} = 10 dB et un rapport PRT de 1%, tel que défini dans les spécifications DVB-T2, les performances du réseau proposé en matière de PAPR sont nettement supérieures à celles de l'algorithme TR-QCQP sans distorsion. En outre, il faut tenir compte du fait que les sous-porteuses réservées pour l'algorithme TR ne transportent pas d'informations utiles, ce qui entraîne une réduction du débit du système, un inconvénient qui est évité par le réseau que nous proposons.

Si l'objectif PAPR$_t$ de notre analyse de réseau est augmenté à 4 dB pour permettre une meilleure performance de compression, comme le montre la figure 9, la performance

de réduction du PAPR du réseau reste supérieure à l'écrêtage et au TR-QCQP avec une contrainte de puissance de $P_{\text{PC}} = 10$ dB et un ratio PRT de 1%, les algorithmes.

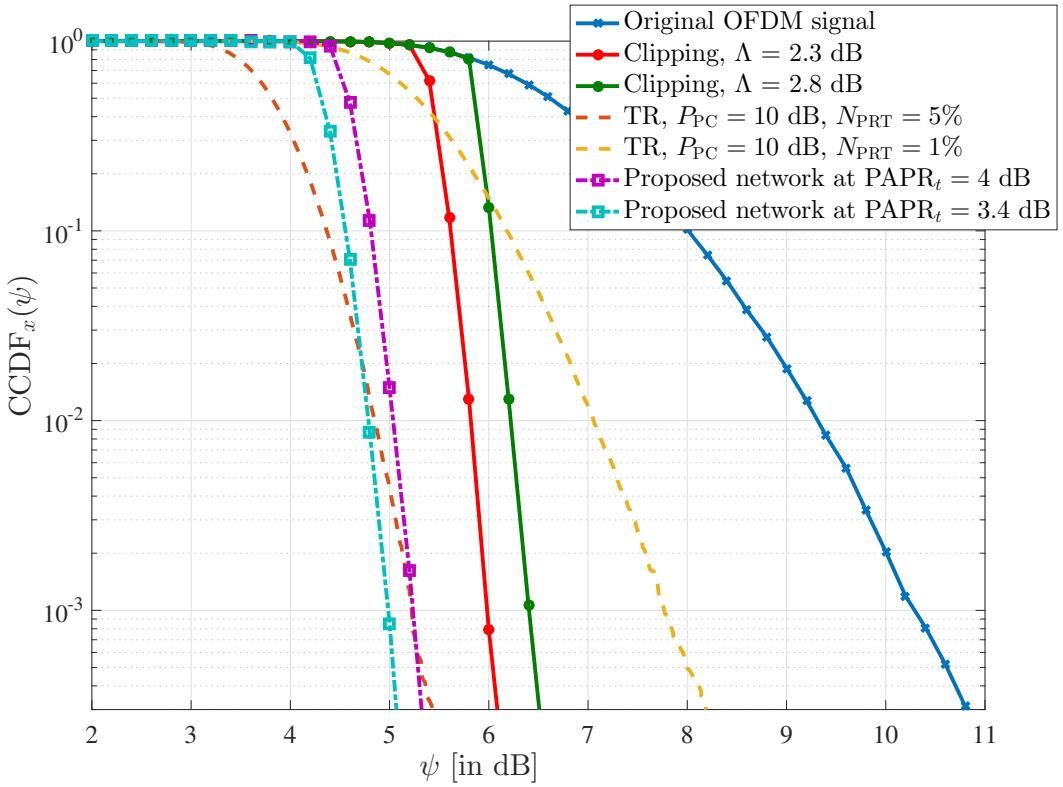


Figure 9: CCDF du PAPR pour le réseau d' AE-quantification proposé à différentes valeurs cibles de PAPR PAPR_t par rapport aux schémas conventionnels de réduction du PAPR.

Conclusions et perspectives

Dans cette thèse, nous avons étudié la compression des échantillons de données OFDM IQ avant la transmission sur la liaison fronthaul à capacité limitée dans le contexte de l'architecture C-RAN. Nous avons axé notre étude sur l'évaluation des performances des techniques de compression du débit de données sous différents angles, qui sont résumés ci-dessous :

- Tout d'abord, nous avons analysé les algorithmes de compression, y compris la quantification uniforme, non uniforme, scalaire et vectorielle, le codage entropique et les algorithmes d'élimination de la redondance dans le domaine spectral (décimation). Ensuite, nous avons proposé un algorithme simple et optimisé de quantification scalaire uniforme couplé à un codage entropique qui pourrait exploiter efficacement les caractéristiques non uniformes de la distribution gaussienne du signal OFDM. Il permet d'obtenir un gain de compression de 2.4 dB par rapport au quantificateur scalaire idéalisé contraint par l'entropie, tout en étant beaucoup moins complexe que l'approche de quantification non-uniforme. En outre, nous avons fourni une comparaison numérique entre le processus de décimation et la quantification vectorielle dans l'exploitation des corrélations entre les échantillons IQ. Nous avons montré l'effet des valeurs des paramètres pratiques appropriés dans le compromis entre la qualité requise du signal, les performances de compression et la complexité du système.

- Deuxièmement, le signal OFDM est caractérisé par un PAPR élevé, ce qui nécessite l'utilisation d'une technique de réduction du PAPR pour alimenter le signal au point optimal de fonctionnment de l'amplificateur. Cela modifie la distribution des échantillons OFDM IQ, ce qui rend les techniques de compression optimisées par la distrubution gaussienne sous-optimales. Nous avons donc étudié l'effet de la réduction du PAPR du signal OFDM, dans le domaine temporel, sur les techniques de compression des échantillons. Nous avons dérivé la distribution des échantillons IQ à PAPR réduit pour les techniques d'écrêtage et TR-QCQP. Ces distributions ont ensuite été utilisées pour optimiser le quantificateur non uniforme et le codeur entropique. Nous avons également proposé une étude asymptotique du MER à partir des distributions dérivées qui tient compte des effets de la réduction du PAPR et des opérations de quantification. Enfin, les résultats de simulations ont validé la distribution obtenue des échantillons IQ à PAPR réduit et les performances obtenues en optimisant les techniques de compression sur la base de ces distributions ont été étudiées.

- Enfin, nos analyses précédentes ont montré qu'un gain de compression significatif est obtenu lorsque les techniques de compression sont optimisées sur la base de la véritable distribution du signal d'entrée. Par conséquent, nous avons proposé une structure d'auto-encodeur permettant d'apprendre la forme du signal permettant d'atteindre un PAPR cible tout en compressant de manière efficace les échantillons IQ avant leur transmission sur le fronthaul. Nous avons d'abord étudier l'apport d'un réseau auto-encodeur pour réduire le PAPR du signal OFDM et nous avons comparé ses performances avec le meilleur réseau de neuronnes de la littérature pour cette tâche. Ensuite, nous avons fait évolué

notre auto-encodeur pour inclure un réseau de quantification, suivi d'un codage entropique pour réduire encore plus le rate. Le décodeur reconstruit le signal OFDM original au niveau du récepteur. L'auto-encodeur-quantificateur est entraîné sur la base d'une fonction de perte multi-objectifs pour produire le signal compressé à PAPR réduit. Les résultats numériques montrent que l'architecture proposée permet d'améliorer le MER du système d'environ 1 dB par rapport à l'algorithme conventionnel de Lloyd et a, en outre, approximativement les mêmes performances en termes de réduction du PAPR que l'algorithme TR-QCQP conventionnel le plus efficace.

La recherche peut être poursuivie dans de nombreuses directions. À notre avis, il serait intéressant d'examiner les aspects suivants:

- Sur le problème d'apprentissage conjoint PAPR-quantification, nous avons conçu le réseau de quantification en dépliant l'algorithme itératif conventionnel de Lloyd. L'affectation d'un mot de code à un échantillon est, par nature, une opération non-linéaire non-dérivable qui pose problème lors de l'application de l'algorithme de rétro-propagation. Nous avons résolu ce problème en fixant la dérivée de l'opérateur de mise en correspondance à 1 pendant la rétro-propagation. Cependant, il pourrait être intéressant de chercher une structure de mise en correspondance entre les échantillons et les mots de code plus souple et dérivable qui permette d'apprendre le modèle sans non-linéarité.

- D'autre part, seule une quantification scalaire a été étudié dans l'auto-encodeur. Il pourrait être intéressant d'étudier si le MER du système peut-être améliorée en examinant incluant un quantificateur vectoriel sur les échantillons IQ du signal à PAPR réduit à la sortie de l'encodeur.

- La prise en compte d'un modèle de non-linéarité de l'amplificateur de puissance lors de l'entraînement du réseau pourrait permettre d'améliorer les performances. Le décodeur serait entraîné à compenser les distorsions causées par la non-linéarité de l'amplificateur de puissance.

- Pour rendre le réseau de quantification proposé plus efficace dans des scénarios réalistes, il est essentiel d'augmenter la taille de la FFT. Dans ce cas, l'augmentation du nombre de sous-porteuses entraîne une augmentation significative de la complexité du réseau à mettre en place, en particulier pour les modes les plus importants des standards de diffusion vidéo comme le $16K$ et $32K$. Par conséquent, l'étude du passage à l'échelle est très importante si l'on souhaite pouvoir mettre en oeuvre ce type de méthode dans une architecture C-RAN.

Table of Contents

Acknowledgement i

Résumé étendu ii

Table of Contents xviii

List of Figures xxi

List of Tables xxii

Introduction 1

1 State of art and background 5
1.1 Introduction . . . 5
1.2 C-RAN architecture . . . 5
1.2.1 Base station architecture evolution . . . 6
1.2.2 C-RAN challenges - fronthaul limitations . . . 8
1.3 OFDM system . . . 11
1.3.1 Principle . . . 11
1.3.2 Peak-to-average power ratio issue in OFDM systems . . . 12
1.3.2.1 Figures of Merit . . . 13
1.4 Compression techniques . . . 15
1.4.1 Performance metrics . . . 16
1.4.2 Scalar quantization . . . 17
1.4.2.1 Preliminaries . . . 17
1.4.2.2 Uniform quantization . . . 19
1.4.2.3 Non-uniform scalar quantization with fixed-length codes 20
1.4.2.4 Entropy coding . . . 23
1.4.2.5 Non-uniform scalar quantization with variable-length codes . . . 24
1.4.3 Asymptotic distortion-rate functions for scalar quantization . . 28
1.4.4 Vector quantization . . . 31
1.4.4.1 Preliminaries . . . 32
1.4.4.2 Vector quantization with fixed-length codes . . . 32
1.4.4.3 Vector quantization with variable-length codes . . . 36
1.4.4.4 Vector quantization advantages . . . 38
1.4.5 Asymptotic distortion-rate functions for vector quantization . . 41
1.4.6 Spectral domain redundancies removal . . . 42

1.5 Conclusion 44

2 Analysis of IQ samples compression techniques 45
2.1 Introduction 45
2.2 State of the art in compression solutions for the C-RAN architecture . 45
2.2.1 Preliminary results of the comparison between state of the art compression techniques: first contribution 47
2.3 System model and simulation parameters 48
2.4 Compression based on scalar quantization analysis 49
2.5 Numerical analysis of the comparison between decimation and VQ when exploiting the redundancy between the OFDM IQ samples 53
2.6 Conclusion 55

3 Compression of PAPR-reduced OFDM IQ samples 56
3.1 Introduction 56
3.2 System model 57
3.3 PAPR reduction techniques 57
3.3.1 Clipping 58
3.3.2 Tone reservation 58
3.3.2.1 Principle of the TR technique 58
3.3.2.2 Parameters of the TR technique 59
3.3.2.3 Percentage and location of tone-reserved subcarriers . 60
3.3.2.4 Algorithms of the TR technique - optimal QCQP algorithm 60
3.4 Analytical evaluation of optimized compression techniques based on PAPR-reduced OFDM signals 61
3.4.1 Performance metrics 61
3.4.2 The impact of PAPR reduction on the amplitude and phase distributions of the OFDM signal 62
3.4.2.1 Clipping technique 63
3.4.2.2 TR-QCQP technique 63
3.5 Analytical study of the optimized compression techniques for a PAPR-reduced signal 65
3.5.1 Distribution of the PAPR-reduced IQ samples 65
3.5.1.1 Clipping technique 66
3.5.1.2 TR-QCQP technique 67
3.5.2 Optimized quantizer and entropy coder of PAPR-reduced signal 68
3.5.3 Asymptotic MER and SQNR 69
3.6 Simulation results and analysis 70
3.6.1 Simulation parameters 70
3.6.2 Validity of the distribution of the PAPR-reduced IQ samples . . 71
3.6.3 Asymptotic distortion-rate analysis 73
3.6.4 Gaussian quantizer performance evaluation 75
3.6.5 Comparison of Gaussian and optimized quantizers 77
3.7 Conclusion 84

4 Joint PAPR reduction and signal compression based on deep neural networks 85
4.1 Introduction . . . 85
4.2 PAPR reduction based on deep neural networks . . . 86
4.2.1 Structure of the feedforward neural networks . . . 86
4.2.1.1 Feedforward neural networks . . . 86
4.2.1.2 Auto-encoder feedforward neural network . . . 88
4.2.2 State of the art in PAPR reduction schemes based on deep learning 88
4.2.3 Structure of the proposed network . . . 90
4.2.4 Complexity analysis . . . 92
4.2.5 Training of the proposed network . . . 93
4.2.6 Performance analysis . . . 94
4.2.6.1 Simulation parameters and discussion of network hyperparameters . . . 94
4.2.6.2 CCDF for PAPR reduction analysis . . . 95
4.2.6.3 Bit error rate analysis . . . 95
4.3 Proposed joint PAPR reduction and compression network for C-RAN architecture . . . 98
4.3.1 Challenges . . . 98
4.3.2 The structure of the proposed joint model . . . 99
4.3.2.1 Auto-encoder structure . . . 99
4.3.2.2 ANN-based quantizer and de-quantizer models . . . 99
4.3.3 Training of the proposed joint model . . . 102
4.3.4 Performance analysis . . . 104
4.3.4.1 Simulation parameters and discussion of network hyperparameters . . . 104
4.3.4.2 Rate-distortion and PAPR analysis . . . 104
4.4 Conclusion . . . 109

Conclusions and perspectives 111

Appendix 114
A Huffman algorithm . . . 114
B The distribution of the TR-QCQP PAPR-reduced IQ samples . . . 115
C Asymptotic MER and SQNR of a compressed clipped OFDM signal . . 117

List of Acronyms 119

Mathematical Notations 121

Bibliography 124

List of Figures

1 MER en fonction de la résolution R pour une distribution gaussienne complexe i.i.d. pour différents schémas de quantification scalaire. v
2 Taux de compression en fonction du facteur de suréchantillonnage L de deux techniques : décimation plus quantification scalaire uniforme et quantification vectorielle à longueur fixe. vi
3 MER des techniques de compression gaussienne et optimisée par rapport au ratio d'écrêtage Λ pour différentes résolutions. vii
4 MER des techniques de compression gaussienne et optimisée en fonction de la résolution pour différentes contraintes de puissance P_{PC} et $N_{\mathrm{PRT}} = 1\%$. viii
5 MER des techniques de compression gaussienne et optimisée en fonction de la résolution pour différents pourcentages d'outillage N_{PRT} et $P_{\mathrm{PC}} = 10$ dB. ix
6 CCDF du PAPR pour le schéma de réduction du PAPR proposé et ceux dits conventionnels. x
7 BER des différents algorithmes sur un canal AWGN pour une modulation 4-QAM. xi
8 MER en fonction de la résolution R pour le réseau d' AE-quantification proposé appliqué au signal OFDM pour différentes valeurs cibles de PAPR. xii
9 CCDF du PAPR pour le réseau d' AE-quantification proposé à différentes valeurs cibles de PAPR PAPR_t par rapport aux schémas conventionnels de réduction du PAPR. xiii

1.1 Evolution of the base station architecture in C-RAN. 6
1.2 BBU and RRH possible functional splits [5]. 7
1.3 Base station architecture evolution to the C-RAN. 8
1.4 Digital Implementation of an OFDM modulator with N_{fft} subcarriers. . 11
1.5 Example for the power fluctuation of an OFDM signal with N_{fft} subcarriers. 13
1.6 CCDF of PAPR for a 16-QAM modulated baseband OFDM signal for different number of subcarriers N_{fft}. 14
1.7 CCDF of PAPR for a baseband OFDM signal with $N_{\mathrm{fft}} = 1024$ subcarriers and for different modulation schemes. 15
1.8 Modulation error ratio illustration. 16
1.9 Input-output function Q of a scalar quantizer. 18
1.10 Overload and granular quantization distortions. 20

1.11 The quantization levels are under-utilized by the uniform quantizer in the tail (less-distributed) areas of the PDF and over-utilized in the head area of the PDF (right). These levels are redistributed in the non-uniform quantizer based on the PDF of the signal, resulting in a lower MSE (left). 21
1.12 ECSQ for a unit variance Gaussian-distributed signal. 27
1.13 The R-D convex hull obtained by running the ECSQ algorithm for different values of the Lagrange multiplier λ. 29
1.14 Example of Voronoi regions of $\mathbb{R}^2$ dimensional space [6]. 34
1.15 Convergence of LA and 2-dimensional LBG algorithm with $R = 3$ bits per sample for a Gauss-Markov source with $\rho = 0.9$. 39
1.16 Lloyd quantization and 2-dimensional LBG quantization wth R=3 bits per sample for a unit-variance Gaussian i.i.d process. 40
1.17 Lloyd quantization and 2-dimensional LBG quantization wth R=3 bits per sample for a uniform i.i.d process. 40
1.18 Removal redundancy block. 43

2.1 The integration of IQ compression and decompression blocks in a conventional C-RAN architecture [7]. 46
2.2 Downlink system model of a C-RAN architecture. 48
2.3 MER vs loading factor γ for different resolutions. 50
2.4 MER vs resolution R for an i.i.d. complex Gaussian distribution for different scalar quantization schemes. 51
2.5 Compression system block diagram. 53
2.6 Compression ratio vs oversampling factor L of two techniques: decimation plus uniform scalar quantizer and fixed-length vector quantization. 54

3.1 Downlink system model of a C-RAN architecture considering the PAPR reduction of the OFDM signal. 57
3.2 Block diagram of the tone reservation technique. 59
3.3 PDF of the amplitude of the time-domain signal with and without clipping at $V_{\max} = 7$. 63
3.4 PDF of the amplitude of the signal with and without TR-QCQP, with different power constraints and $N_{\mathrm{PRT}} = 1\%$. 65
3.5 Simulated and analytical PDF of clipped PAPR-reduced IQ samples for different clipping ratios Λ. 71
3.6 Simulated and analytical PDF of TR-QCQP PAPR-reduced IQ samples, with and without power constraint and $N_{\mathrm{PRT}} = 1\%$. 72
3.7 Simulated and analytical PDF of TR-QCQP PAPR-reduced IQ samples, with different percentages of reserved tones and $P_{\mathrm{PC}} = 10$ dB. . . 73
3.8 $D_{\mathrm{SLB}}^{\mathrm{clip}}(R)$ and $D_{\mathrm{SLB}}^{\mathrm{QCQP}}(R)$ functions for clipped and TR-QCQP PAPR-reduced signals, at different Λ and P_{PC}, respectively, bounded above by the Gaussian distortion-rate function $D_{\mathrm{SLB}}^{\mathrm{G}}(R)$. 74
3.9 Asymptotic distortion of Lloyd quantizer and SLB for clipped and TR-QCQP PAPR-reduced signals, at different Λ and P_{PC}, respectively, vs resolution. 74

3.10 SQNR vs clipping ratio Λ of a Gaussian quantizer applied on a clipped signal vs the resolution. 76
3.11 MER vs resolution of a Gaussian quantizer applied to a TR-QCQP signal. (a) different power constraints P_{PC} and $N_{\text{PRT}} = 1\%$. (b) different percentages of PRTs N_{PRT} and $P_{\text{PC}} = 10$ dB. 78
3.12 MSE per each quantization region for different power constraints. (a) 4-level quantizer with 2 bits per sample. (b) 8-level quantizer with 3 bits per sample. 79
3.13 Simulated and analytical MER vs clipping ratio Λ of clipped optimized compression techniques for different resolutions. 80
3.14 MER of Gaussian and optimized compression techniques vs clipping ratio Λ for different resolutions. 81
3.15 Clipping PAPR-reduction and Gaussian quantizer MSEs vs clipping ratio Λ for different resolutions. 81
3.16 Simulated and asymptotic MER vs resolution of TR-QCQP optimized compression techniques for different power constraints P_{PC}. 82
3.17 MER of Gaussian and optimized compression techniques vs resolution for different power constraints P_{PC} and $N_{\text{PRT}} = 1\%$. 83
3.18 MER of Gaussian and optimized compression techniques vs resolution for different percentages of PRTs N_{PRT} and $P_{\text{PC}} = 10$ dB. 83

4.1 An example of a FFN with two hidden layers of 4 neurons each. 87
4.2 The architecture of the AE-based PAPR reduction module. 91
4.3 CCDF of PAPR for proposed and conventional PAPR reduction schemes. 96
4.4 The BER performance of the different algorithms over AWGN channel for 64-QAM modulation. 97
4.5 BER of the proposed NN trained at different E_b/N_0. 97
4.6 The architecture of the proposed joint PAPR reduction and compression network for the C-RAN architecture. 99
4.7 The architecture of the quantizer and de-quantizer. 101
4.8 MER of the proposed network at different bottleneck dimensions N_b, for $PAPR_t = 3.3$ dB, $N = 8$-levels 1-dimensional quantizer and $R_t = 2.7$ bits/sample. 105
4.9 MER vs resolution R for the proposed AE-quantization network applied on the OFDM signal for different target PAPR values PAPR_t. 106
4.10 CCDF of PAPR for proposed AE-quantization network at different target PAPR values PAPR_t compared with the conventional PAPR reduction schemes. 108
A.1 Example for the design of a Huffman algorithm. 115

List of Tables

1.1 Comparison between traditional base station, base station with RRH and C-RAN architecture [1] 9
1.2 Bandwidth specifications for LTE and DVB-T2 standards. 43

2.1 Simulation parameters for the analysis of the compression of an OFDM signal 49
2.2 Asymptotic gap between UQ and NUQ with Huffman coding 53
2.3 Hamming window FIR filter specifications 54

3.1 Number of OFDM subcarriers, active subcarriers and peak reserved subcarriers for different modes of operation in the DVB-T2 frame structure [3] 60
3.2 Simulation parameters for the analysis of the compression of a PAPR-reduced signal 71
3.3 Distortion reduction for clipped and TR-QCQP signals, relative to the Gaussian-distributed signal and the asymptotic difference in distortion between the Lloyd quantizer and SLB for each PDF. 76

4.1 Configuration of the proposed PAPR reduction model 95
4.2 Proposed deep-based neural network for the C-RAN architecture 100
4.3 Configuration of the proposed joint PAPR reduction and compression model 104
A.1 List $\mathcal{L}_{\text{symbols}}$. 114

Introduction

Context

Mobile network operators are facing tremendous growth in data traffic due to the increasing popularity of smartphones, tablets and new connected smart devices that support a wide range of resource-greedy applications and services. These devices, which provide access to a variety of services, are very bandwidth hungry, especially video services, which account for two-thirds of mobile data traffic. In addition, users have high quality expectations regardless of location, time and situation. In fact, the data rate offered has increased from a few kbps with the Global System for Mobile Communications (GSM) standard in the nineties to hundreds of Mbps today with the latest 5^{th} generation cellular networks. At the same time, the universal TV broadcasting systems have been converted from analogue to digital to meet the increasing demand for higher quality TV experience.

To support such growth, more cells or higher capacity per cell must be deployed, resulting in higher costs. At the same time, users would like to pay less and use more data. Therefore, cost growth cannot follow the same trend line as traffic growth. To meet capacity requirements, small cells can be deployed in locations with high user activity. In such a case, interference becomes a challenge, so techniques such as enhanced intercell interference coordination between cells and coordinated multipoint links are essential. Finally, to increase the capacity of each cell, more antennas can be deployed with a Massive Multiple Input Multiple Output technique.

Centralized, Cooperative, Cloud Radio Access Network (C-RAN) is a novel network architecture that can address a number of challenges faced by recent technologies as they attempt to support the ever-increasing demands of end users. The main idea behind the C-RAN is to split base stations into radio and baseband parts and pool the Baseband Unit (BBU) from multiple base stations into a centralized and virtualized BBU Pool, while the Remote Radio Head (RRH) and antennas remain at the cell sites. This enables energy-efficient network operation and potential cost savings in baseband resources. It also improves network capacity through load balancing and cooperative processing of signals originating from multiple base stations. Therefore, base stations with high load and high deployment density can be deployed in the C-RAN architecture [8].

However, there are also some challenges that need to be addressed before the commercial deployment of the C-RAN architecture. The C-RAN architecture brings a huge overhead on the links between RRH and BBU Pool. Fronthaul requirements are expected to be 50 times higher than backhaul requirements [1]. The huge bandwidth requirements for the transmission of high-resolution IQ samples generated by baseband

processing in the cloud remains one of the major challenges of this architecture. The classical approach is to use higher bandwidth optical fibers to transmit the IQ samples, but this comes at a very high cost. Therefore, IQ data compression before transmission over the capacity-constrained fronthaul link is an attractive approach to reduce the data rate over the fronthaul link for the C-RAN architecture.

This thesis is dedicated to this challenge. It aims to investigate different approaches for data rate compression that represent a tradeoff between achievable compression performance, signal distortion, design complexity and computation time. In the literature, quantization and entropy coding are the main IQ data rate reduction techniques. Indeed, the multicarrier Orthogonal Frequency Division Multiplexing (OFDM) is the most commonly used modulation scheme in recent standards, e.g., Long Term Evolution (LTE) [2], Digital Video Broadcasting - Terrestrial - second generation (DVB-T2) [3] and Advanced Television Systems Committee version 3.0 (ATSC 3.0) standard [4]. Therefore, the studies in the literature mainly rely on exploiting the statistical characteristics of the OFDM signal to reduce the data rate on the fronthaul link.

Problem and objective of the thesis

Thanks to the Central Limit Theorem (CLT), OFDM baseband IQ samples are generally assumed to be Gaussian-distributed at the output of the Inverse Fast Fourier Transform (IFFT) modulator. Therefore, the studies proposed in the recent literature for different compression blocks are optimized for a Gaussian distribution. However, the addition of many subcarriers via the IFFT results in a signal characterized by high power fluctuations measured by the Peak-to-Average Power Ratio (PAPR), which remains one of the main problems in OFDM systems. High PAPR value prevents radio-frequency designers from injecting the signal at the optimal point of the power amplifier to reduce energy consumption. Many techniques have been presented in the literature that are applied to OFDM signal in the time-domain to reduce its PAPR.

However, processing the OFDM signal in the time-domain with a PAPR reduction technique changes the distribution of the OFDM IQ samples, making Gaussian quantizers suboptimal. As far as we are aware, no studies have been performed on the robustness of Gaussian-optimized compression techniques that consider any of the PAPR reduction techniques prior to compression, nor on the proposal of an optimized quantizer for signals with modified distribution due to PAPR reduction techniques.

In this perspective, we have presented the analytical analysis of the optimized quantizer and entropy coding compression techniques for OFDM signals processed by PAPR reduction techniques. We consider clipping and Tone Reservation (TR) PAPR reduction techniques. Clipping is the simplest scheme and TR is an efficient method adopted by several standards, e.g., DVB-T2 [3] and ATSC 3.0 [4]. For the TR method, the Quadratically Constrained Quadratic Problem (QCQP) algorithm is considered, which provides the optimal performance of PAPR reduction achieved by the TR algorithm. The analytical evaluation of these techniques quantifies the performance gain in Modulation Error Ratio (MER) when using a quantizer that is optimized with respect to the true distribution of PAPR-reduced IQ samples based on various tuning factors of the clipping and TR-QCQP algorithms compared to the Gaussian-based quantizer.

The importance of optimizing the compression technique based on the true distribution of the input signal, leads us to go a step beyond conventional approaches by introducing an Auto-Encoder (AE)-based Deep Neural Network (DNN). Our idea was to implement an Auto-Encoder (AE)-based DNN that can be end-to-end jointly trained to obtain an optimized PAPR-reduced and compressed OFDM IQ samples, without requiring analytical evaluation of the distribution of the PAPR-reduced signal for compression.

Organization of manuscript

This thesis is organized in 4 chapters as follows.

Chapter 1 introduces the context of the study. First, we provide an overview of the C-RAN architecture. The advantages and challenges of such an architecture are highlighted. Then, we introduce the high data rate problem of the IQ data samples that are generated in the cloud and need to be transported over the fronthaul link to the transmitter. Then, an overview of the most commonly used OFDM modulation scheme is given. In addition, the problem of high PAPR characterizing the OFDM signal was presented. Finally, a detailed description of the various signal processing compression techniques existing in the literature to reduce the data rate on the fronthaul link was given, in particular the spectral domain redundancy removal process, also known as decimation, quantization techniques (uniform, non-uniform, scalar and vector) and entropy coding techniques.

Chapter 2 aims to introduce the first problem of the thesis. The state of the art in compressing Gaussian-distributed baseband samples prior to transmission over the fronthaul link in the C-RAN architecture is described in detail. It also examines the various parameters used to evaluate the compression scheme in terms of the trade-off between signal quality and complexity while achieving an acceptable compression gain. The state of the art shows that the studies performed at this level are insufficient. Therefore, we present a simple Gaussian-optimized uniform quantization scheme combined with entropy coding that achieves the performance of non-uniform quantization with lower computational complexity. We also present a comparison between decimation and vector quantization to exploit the spectral redundancies that characterize the OFDM signal. Finally, some practical scenarios are presented to demonstrate the importance of our proposed analysis.

Chapter 3 first describes the model of the C-RAN architecture, including the PAPR reduction and compression techniques. Then, the distribution of the PAPR-reduced OFDM IQ samples is derived when clipping and TR-QCQP schemes are applied to the IQ samples after IFFT. Then, the new IQ distribution is used to optimize the quantizer and the entropy encoder and to obtain the asymptotic MER expression reflecting the joint effect of PAPR reduction and quantization. In addition, we quantify the performance gain obtained by optimizing the compression techniques with respect to the true distribution of the PAPR-reduced IQ samples.

Chapter 4 deals with the proposed DNN-based algorithm. In the first part of this chapter, an AE-based network is proposed to reduce the PAPR of the OFDM signal compared to the conventional approaches. In the second part of the chapter, an evolution of the proposed AE network is proposed to include a quantization network

that is trained jointly with a multi-objective optimization function to obtain a PAPR-reduced and quantized signal that can be transmitted over the fronthaul link. A simulation analysis of this proposed network and comparisons with the conventional component-wise optimization algorithms are finally given.

Finally, we conclude this thesis by giving some possible directions and perspectives that emerge from our work.

List of publications

The contributions of this work have been published in the following international journals and conferences.

In preparation

(S1) Aya Shehata, Philippe Mary, and Matthieu Crussière, "Joint PAPR reduction and signal compression based on deep neural networks for OFDM systems in C-RAN architecture", to be submitted.

International journals

(J1) Aya Shehata, Philippe Mary, and Matthieu Crussière, "Analysis of Compressing PAPR-Reduced OFDM IQ Samples for Cloud Radio Access Network", in IEEE Transactions on Broadcasting, vol. 68, no. 3, pp. 765-779, Sept. 2022.

International conferences

(C1) Aya Shehata, Matthieu Crussière and Philippe Mary, "Analysis of Baseband IQ Data Compression Methods for Centralized RAN", In Proc. of the 28$^{\text{th}}$ European Signal Processing Conference (EUSIPCO 2021), Amsterdam, Netherlands (virtual conference), 18-21 January 2021, pp.1-5.

(C2) Aya Shehata, Philippe Mary, and Matthieu Crussière, "Compression of clipped OFDM IQ samples for cloud radio access network", In Proc. of the IEEE 32$^{\text{nd}}$ Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2021), Helsinki, Finland (virtual conference), 13–16 September 2021, pp.1-6.

National conferences

(N1) Aya Shehata, Philippe Mary, and Matthieu Crussière. "Compression des échantillons IQ d'un signal OFDM à PAPR réduit dans un réseau d'accès radio virtualisé", In Proc. of the Groupe d'Etudes du Traitement du Signal et des Images (Gretsi 2022), Nancy, France, 06-09 September 2022, pp. 1-4.

Chapter 1

State of art and background

1.1 Introduction

In this chapter, Section 1.2 first gives an overview of the C-RAN architecture, which is the focus of our study, and shows its advantages and limitations. The problem of the enormous data rate that must be transported on the fronthaul link between the cloud where the IQ data samples are generated and the transmitter is discussed in detail. Therefore, the data rate on the fronthaul link must be compressed using various compression techniques. However, the compression should be done in such a way that the quality of the transmitted signal is not affected too much. Next, in section 1.3, the OFDM multicarrier modulation scheme is recalled as the most commonly used modulation technique in recent communication systems. The high PAPR problem that characterizes the OFDM signal when amplified with high power nonlinear amplifiers is also presented. Finally, in section 1.4, the various signal processing compression techniques used in the literature to reduce the data rate are analyzed in detail.

1.2 C-RAN architecture

In different generations of mobile networks, GSM [9], UMTS [10], LTE [11], Long Term Evolution-Advanced (LTE-A) [12], in addition to digital television terrestrial broadcasting networks such as the DVB-T2 [3] and ATSC 3.0 [4], the base station is described with Base Transceiver Station (BTS), NodeB or Evolved NodeB (eNodeB), respectively. The base station is responsible for transmitting/receiving the signal from the user and for communicating with the core network.

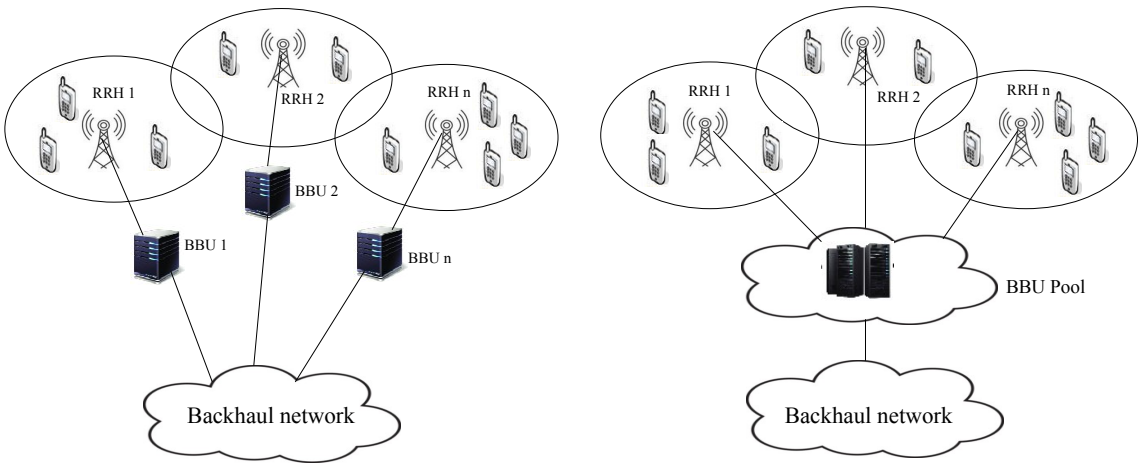
As the demand for higher capacity continues to skyrocket, some promising cell throughput improvement techniques are being used to meet the demand. For example, the complex heterogeneous network created by adding more micro cells and pico cells [13], using beam forming, integrating relays and repeaters, adopting distributed antenna system, and using the massive multiuser multiple-input multiple-output technique where a massive number of antennas simultaneously serve users in the same time-frequency resource block [14] [15]. However, this leads to increased interference between cells, complex network operation and maintenance, increased energy consumption, and much higher costs.

In practical network implementation, total costs are usually divided into CAPital EXpenditure (CAPEX) and OPerating EXpenditure (OPEX). CAPEX mainly refers to the expenses for building the network, such as site acquisition, Radio Frequency (RF) hardware, baseband hardware, software licenses, etc. OPEX, on the other hand, includes the cost of operating the network, i.e., site rental, leased lines, power, operations and maintenance, and cloud upgrades. Thus, the more bas stations, the higher CAPEX and OPEX costs.

The C-RAN is a promising network architecture that has the potential to overcome the above challenges. The concept was first introduced by IBM in [16] and described in detail in [1]. The C-RAN architecture is essentially based on "cloudification" i.e., moving baseband processing from standalone base stations to a pool of virtual base stations in a cloud data center. The architecture of C-RAN is thus based on splitting the base station functionalities into two parts, the BBU located in the cloud and the RRH located at the transmission sites. Thus, baseband processing is centralized and shared among the sites in a virtualized BBU Pool.

1.2.1 Base station architecture evolution

Thanks to the centralized network architecture, computing resources are virtualized and aggregated at a pool level in the C-RAN architecture. The BBU is connected to the core network via a backhaul Ethernet carrier, where the traffic is processed. Fig. 1.1 provides an overview of the entire C-RAN architecture.



(a) RAN with RRHs. (b) C-RAN architecture.

Figure 1.1: Evolution of the base station architecture in C-RAN.

The main functions of a base station can be divided into baseband processing and radio functionalities. The C-RAN architecture moves the physical layer functions, including baseband processing and network layer functions implemented in the BBUs, to the cloud. The RRH, on the other hand, is responsible for the radio functions. Fig. 1.2 shows a possible breakdown of the main submodules implemented in the BBU and the RRH of the C-RAN architecture.

In the early architecture, the radio and baseband processing functions are integrated in a base station. The antenna module is usually located a few meters away

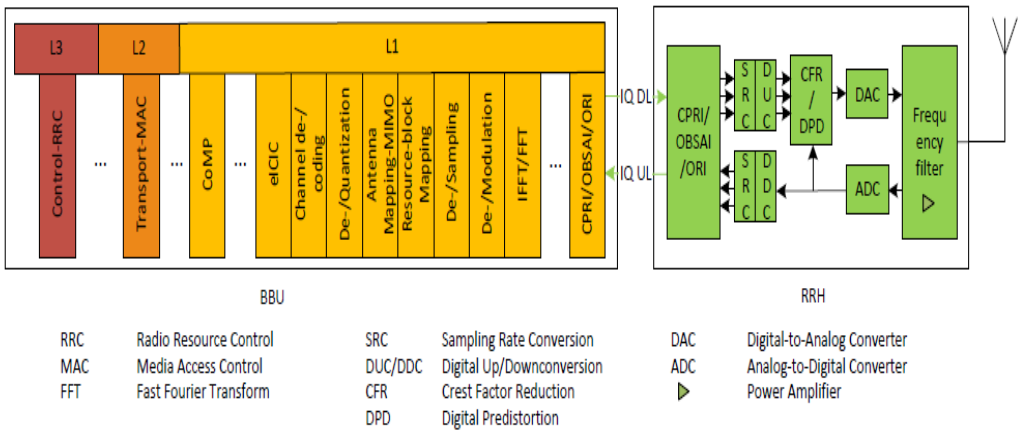


Figure 1.2: BBU and RRH possible functional splits [5].

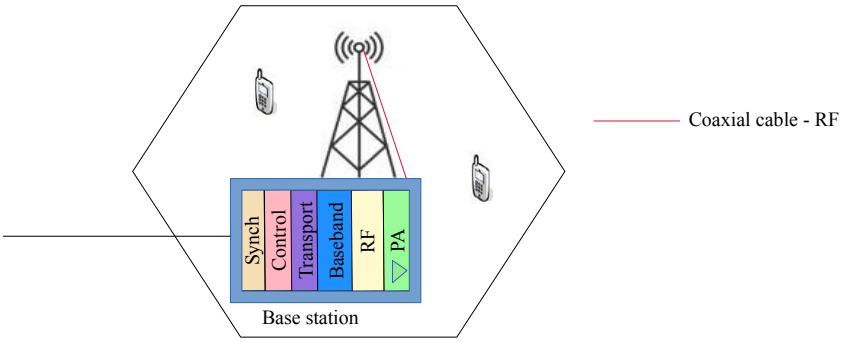
from the radio module and is connected via a coaxial cable, as shown in Fig. 1.3a. Fig. 1.3b represents a base station with RRH architecture. In this case, the base station is divided into a signal processing unit and a radio processing unit (RRH). The RRH provides the interface to the fiber and performs digital processing, digital-to-analog conversion, analog-to-digital conversion, power amplification, and filtering [17]. The division of functions between the BBU and the RRH is shown in Fig 1.2.

The connection between the BBU and the RRH can be implemented using fiber optic and microwave links, and the distance can be extended up to 40 km. Common Public Radio Interface (CPRI) [18] is the widely used radio interface protocol for IQ data transmission between RRHs and BBU. CPRI is a constant bit rate bidirectional protocol that requires accurate synchronization and strict latency control. Other protocols that can be used are open base station architecture initiative [19] and open radio equipment interface [20].

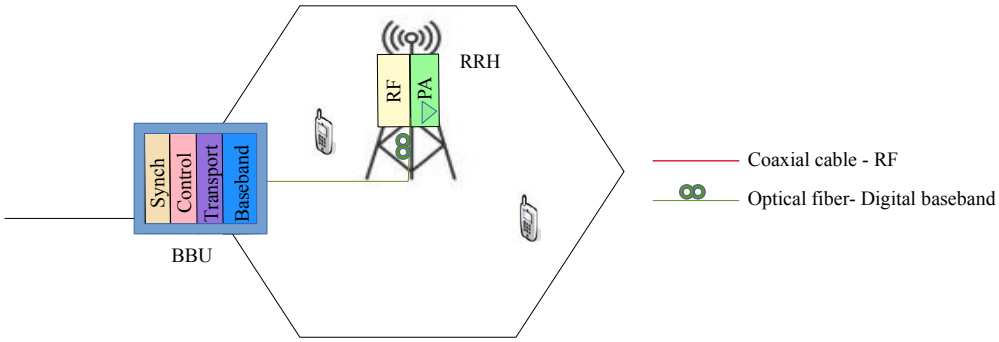
The concept of distributed wireless communication system in [21] was the introduction of C-RAN architecture by IBM in [16]. The term C-RAN is used to describe this architecture, where the letter C can be interpreted as follows: cloud, centralized processing, cooperative wireless or collaborative. In the C-RAN architecture, the BBUs are relocated from each individual cell site and centralized into one unit, the BBU Pool. This allows optimization of BBU usage between heavily and lightly loaded base stations. The BBU Pool is shared between RRHs and virtualized, as shown in Fig. 1.3c. The centralized BBU Pool results in lower cost by removing hardware from cell sites, which involves space, heating, cooling, power, and test access. In addition, the centralized topology simplifies network management, deployment, and scaling.

The virtualized cluster of the BBU Pool consists of general purpose processors for baseband processing (PHY/MAC). The X2 interface in a new form, often referred to as X2+, organizes inter-cluster communication as well as communication with other pools. The RRHs are co-located with the antennas and are connected to the high-performance processors in the BBU Pool via low-latency, high-bandwidth optical transport links called fronthaul links. Digital baseband, i.e., IQ samples, is transmitted between the BBU and the RRHs. Table 1.1 compares conventional base stations, base stations with RRH and base stations in the C-RAN architecture in terms

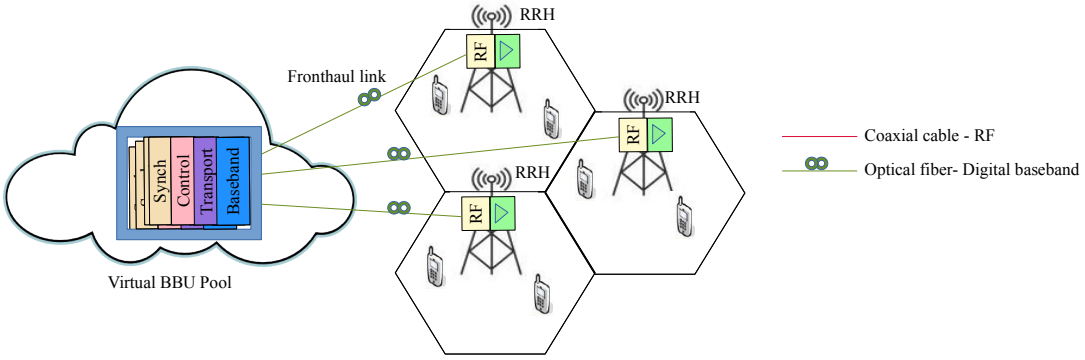
of their advantages and challenges. The C-RAN architecture offers the following advantages: improved power and cost efficiency, improved spectral efficiency, improved interference management, delay reduction, ease of maintenance and expansion, and adaptability to non-uniform traffic. A detailed description of these benefits can be found in [22].



(a) Traditional base station.



(b) Base station with RRH.



(c) C-RAN with RRHs.

Figure 1.3: Base station architecture evolution to the C-RAN.

1.2.2 C-RAN challenges - fronthaul limitations

In addition to the great advantages of the C-RAN architecture, there are also some challenges/cons that must be considered prior to implementation. These disadvantages can be summarized as follows: single-point failure is a major risk for the C-RAN

Table 1.1: Comparison between traditional base station, base station with RRH and C-RAN architecture [1]

Architecture	Baseband functionalities	Advantages	Challenges
Traditional base station	Co-located in one unit.	-	High power consumption. Resources are underutilized.
Base station with RRH	Splitted between RRH and BBU. RRH is placed with the antenna. BBU located within 20-40 km away.	Lower power consumption. Better placement of BBUs.	Resources are underutilized.
C-RAN	BBUs from many sites are co-located in the C-RAN Pool.	Much lower power consumption. Much lower number of BBUs. Cost reduction.	Considerable resources transported between RRH and BBU.

architecture, i.e., if the cloud fails, the entire network is out of service, the various virtualization techniques for the BBU Pool and the huge bandwidth overhead between the BBU and the RRH, i.e., the fronthaul constraints. In this thesis, we address the problem of fronthaul limitations, while a detailed illustration of the various challenges can be found in [22].

The digitized IQ samples are transported via the fronthaul connection between the BBU and the RRH, as shown in Fig. 1.2. This results in a huge bandwidth overhead on the optical links between the RRH and the BBU Pool. Compared to the backhaul requirements, the fronthaul link requirements are expected to be 50 times higher [23]. The main factors contributing to the size of the IQ data are: turbocoding (e.g., UMTS and LTE use 1:3 turbo coding, resulting in a threefold overhead), the IQ sample width based on the selected radio interface (e.g., CPRI) and the oversampling of the signal. For example, the sampling frequency of 30.72 MHz LTE signal is standardized for 20 MHz LTE, which is more than the 20 MHz required for sampling according to the Nyquist-Shannon theorem. The total bandwidth also depends on the number of sectors and the MIMO configuration. The factors that affect the bandwidth are summarized in (1.1) [1]. Consequently, the expected IQ throughput for a 20 MHz 4x4 MIMO, 3-sector base station exceeds 10 Gbps. The centralized BBU Pool should support 10-1000 base stations [1]. Therefore, a huge amount of data must be transmitted and the transport network must support this huge bandwidth.

$$IQBandwidth = samplingFrequency \,.\, sampleWidth. \\ 2 \,.\, MIMO \,.\, noOfSectors \tag{1.1}$$

Consequently, the classical approach is to use higher bandwidth fiber links from cell sites, i.e., RRHs at different locations, to BBU. In densified heterogeneous net-

works, this increases CAPEX costs, especially for operators who do not own the fiber networks, and limits the scope of C-RAN deployment. BBUs can be 20-40 kilometers from RRHs, and fronthaul links must serve multi-cell sites. Next-generation cell sites will have 10-20 Gbps capacity and require over 300 Gbps for the fronthaul link. For such a complex network, expanding CPRI capacity can meet the needs of the network. However, because it relies heavily on fiber fronthaul links, it will be too expensive or unavailable to deploy.

Rather than moving all baseband processing to a BBU Pool, a solution was proposed to define and extensively study a functional partitioning of the baseband processing chain to relax the most latency-complex functions in C-RAN to address the challenges of optimizing the fronthaul bit rate under the constraint of limiting CAPEX and OPEX costs [16],[24].

The structure of the C-RAN is based on BBU and RRH functionalities, which can be divided into fully centralized, partially centralized and hybrid systems. The basic classification depends on where the physical layer, Media Access Control (MAC) layer and network layer functionalities are handled. The physical layer functionalities include baseband processing, i.e., radio functions, while the MAC layer includes access control mechanisms and the network layer includes routing mechanisms. These structures can be briefly illustrated as follows:

- Fully centralized: all functions of the physical, MAC and network layers are moved to BBU. Such a structure handles all resources and therefore can benefit significantly from ease of operation and maintenance. However, the performance of such a structure is limited by the communication overhead between BBU and RRHs, as the frontahaul capabilities are not sufficient.

- Partially centralized: the physical layer functions are executed in RRHs, while the MAC and network layer functions are executed in BBUs. The overhead between BBU and RRHs is significantly reduced since the physical layer functions account for most of the computational overhead. However, this comes at the cost of complex physical and MAC communications and unsupported resource sharing between different RRHs.

- Hybrid centralized: part of the physical layer functions are performed in RRHs, while others are done in BBUs. RRH takes responsibility for user or cell specific functions mainly related to signal processing.

The fully centralized approach is the most commonly used structure in C-RAN, which requires a huge communication overhead on the fronthaul link [25]. Therefore, efficient IQ data compression before transmission over the capacity-constrained fronthaul link is an attractive approach to reduce the data rate over the fronthaul link for the C-RAN architecture. Potential solutions for compressing the digitized IQ baseband signal prior to transmission over the fronthaul link are being explored, as will be explained in Section 2.2. These compression approaches can be mixed and matched to represent a tradeoff between achievable compression performance, signal distortion, design complexity, and computational complexity.

1.3 OFDM system

The main idea of multicarrier systems is the simultaneous transmission of several data symbols in the frequency-domain, i.e., enabling parallel data transmission over so-called subcarriers. First, because the channel bandwidth is divided into narrowband flat fading subchannels, multicarrier systems are more resistant to frequency-selective channels than single-carrier systems, without the need for complex equalization filters. Second, the Inter-Carrier Interference (ICI) is avoided with the cyclic prefix or overlap-and-add operation at the receiver, whereas Inter-Symbol Interference (ISI) is avoided with any guard interval i.e., cyclic prefixes.

OFDM is one of the most popular types of multicarrier systems. OFDM is deployed in most current wired and wireless communication systems, such as the worldwide interoperability for microwave access, wireless local area networks, LTE-A and broadcasting networks such as DVB-T2 and ATSC 3.0.

1.3.1 Principle

The principle of the OFDM system is based on dividing the system bandwidth into a large number of small orthogonal subbands called subcarriers. Thus, the serial data stream is passed through a serial-to-parallel converter, which splits the serial data into a number of parallel channels in which the multicarrier modulation is implemented. The implementation complexity of the modulators/demodulators of an OFDM system with N_{fft} subcarriers is reduced by using the well-known IFFT in transmission or Fast Fourier Transform (FFT) digital processing operation during reception [26].

Therefore, the modulated data complex symbols $\tilde{\mathbf{x}} = [\tilde{x}(0), \cdots, \tilde{x}(N_{\text{fft}}-1)]^T \in \mathbb{C}^{N_{\text{fft}}\times 1}$ are the frequency-domain symbols, which modulate the N_{fft} orthogonal subcarriers and the output of the IFFT is the OFDM symbol in the time-domain, as shown in Fig. 1.4. Thus, the kth entry of the complex OFDM baseband signal vector $\mathbf{x} \in \mathbb{C}^{N_{\text{fft}}\times 1}$, can be expressed in time-domain as:

$$x(k) = \frac{1}{\sqrt{N_{\text{fft}}}} \sum_{n=0}^{N_{\text{fft}}-1} \tilde{x}(n) e^{j2\pi k \frac{n}{N_{\text{fft}}}}; \quad \forall k = [0, \ldots, N_{\text{fft}}-1]. \tag{1.2}$$

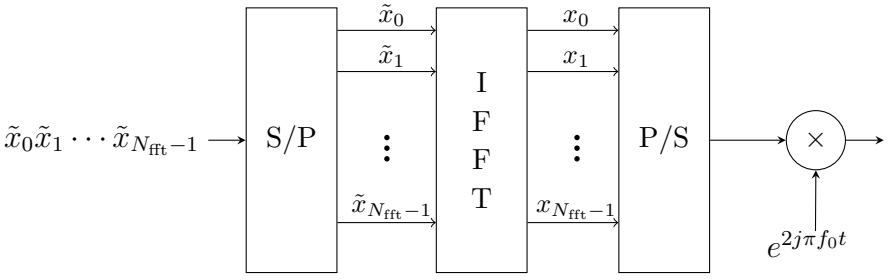


Figure 1.4: Digital Implementation of an OFDM modulator with N_{fft} subcarriers.

The IFFT/FFT blocks are chosen for their execution speed, flexibility, and precision for implementing OFDM systems. In addition, the IFFT/FFT algorithms guarantee the required orthogonality between adjacent subcarriers by their design.

The CLT states that the simultaneous addition of a large number of Independent and Identically Distributed (i.i.d) random variables leads to a Random Variable (r.v.) characterized by a zero-mean Gaussian distribution. Consequently, for N_{fft} sufficiently large, the complex OFDM sample $x(k) = \Re(x(k)) + j\Im(x(k))$, $\forall k \in \{0, \cdots, N_{\text{fft}} - 1\}$, where $\Re(x(k))$ and $\Im(x(k))$ denote the real and imaginary parts of the complex sample $x(k)$, respectively, is zero-mean complex Gaussian distributed, i.e.,

$$\Re(x(k)) \sim \mathcal{N}(0, \sigma_x^2); \quad \Im(x(k)) \sim \mathcal{N}(0, \sigma_x^2), \tag{1.3}$$

where σ_x^2 is the variance of the OFDM signal $\mathbf{x}$, i.e., $\sigma_x^2 = \mathbb{E}[|X|^2]$ and X is the r.v. representing the OFDM signal $\mathbf{x}$. Therefore, the magnitude of the OFDM signal $|x(k)|$ converges to a Rayleigh distribution, i.e.,

$$|x(k)| \sim \mathrm{f}_R^{\text{Ray}}(r) = \begin{cases} \frac{r_x}{\sigma_{\text{ray}}^2} e^{-\frac{r_x^2}{2\sigma_{\text{ray}}^2}}; & \textit{if } r \geq 0 \\ 0; & \textit{otherwise}, \end{cases} \tag{1.4}$$

where σ_{ray} is the standard deviation fitting the Rayleigh distribution and r_x depicts the amplitude of the OFDM signal $\mathbf{x}$. The phase of the OFDM signal, i.e., $\angle x(k)$ is uniformly distributed over $[-\pi, \pi]$, respectively [27].

1.3.2 Peak-to-average power ratio issue in OFDM systems

The OFDM signal is the weighted summation of independent modulated subcarriers via the IFFT operation, each multiplied by a pure phase, as shown in (1.2). Thus, the OFDM symbol is statistically seen as an addition of many i.i.d random variables. This characterizes the OFDM signal with a wide dynamic range in the time-domain. Fig. 1.5 shows the instantaneous power of an OFDM signal in the time-domain with $N_{\text{fft}} = 1024$ subcarriers. The peak and average power are indicated. From Fig. 1.5, two observations can be made. First, there are large variations in the amplitude of the OFDM signal. For most samples, the power is close to the average value, while for a few samples, the power is close to the peak value. Second, there is a large gap between the average and peak power. This gap is quantified by the PAPR metric.

Definition 1.1 (PAPR of OFDM signal)**.** *The PAPR of signal $\boldsymbol{x}$ is defined as the ratio between the maximum and average power of the signal over a given time interval, defined as follows:*

$$\text{PAPR}_x = \frac{\max\limits_{k=0,\cdots,N_{\text{fft}}-1} \left[\|x(k)\|^2\right]}{\mathbb{E}\left[\|x(k)\|^2\right]}, \tag{1.5}$$

where N_{fft} is the number of subcarriers, $\mathbb{E}[.]$ is the expectation operator and $\|.\|$ is the Euclidean norm.

Fig. 1.5 justifies the high ratio between the average and peak power of the OFDM signal, which in turn leads to the high PAPR that characterizes the OFDM system. The problem of high PAPR problem that characterizes the OFDM signal occurs during the amplification of the signal. In other words, the large dynamic range of the OFDM signal in the time-domain results in severe performance degradation when

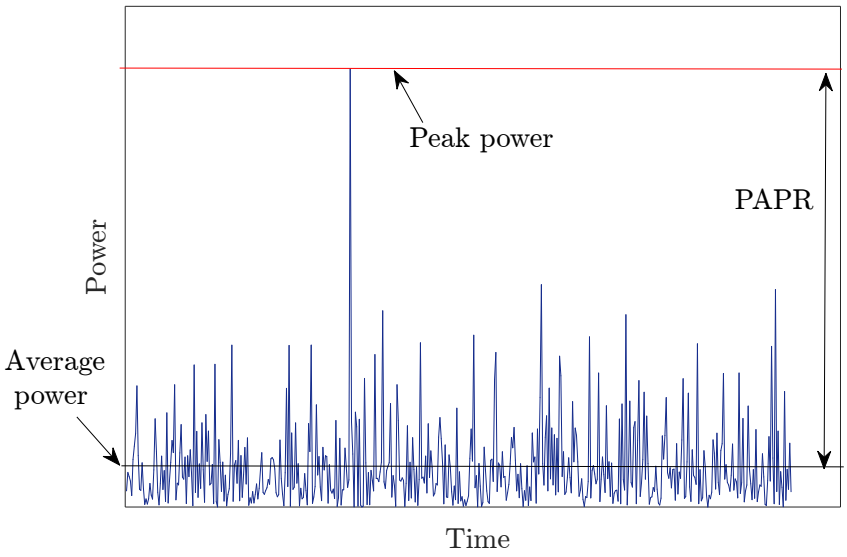


Figure 1.5: Example for the power fluctuation of an OFDM signal with N_{fft} subcarriers.

passed through a high-power amplifier, which is inherently nonlinear, leading to a reduction in the quality of the signals emitted at the output of the antenna [28]. Thus, the high PAPR of the OFDM system remains one of the main problems of the OFDM system.

1.3.2.1 Figures of Merit

CCDF of PAPR A commonly defined metric for evaluating the PAPR is the calculation of its Complementary Cumulative Distribution Function (CCDF), where PAPR is considered as a random variable defined as follows:

Definition 1.2 (CCDF of PAPR)**.** *The CCDF of PAPR is defined as the probability that the PAPR value of a randomly selected OFDM symbol exceeds a predefined threshold ψ, expressed as follows:*

$$\text{CCDF}_x(\psi) = \mathbb{P}[\text{PAPR}_x > \psi]. \tag{1.6}$$

In general, the PAPR of the signal is analyzed over an oversampled signal, i.e., with a sampling rate greater than the Nyquist rate. Since some peaks are likely to be missed at low sampling rates, the signal fluctuations are not properly visible. For an OFDM baseband signal, an analytical expression for the CCDF of the PAPR is proposed, which is valid only if the signal is transmitted in a spectrum exactly corresponding to what is required for the Nyquist rate and expressed as follows [29]:

$$\mathbb{P}[\text{PAPR}_x > \psi] = 1 - (1 - e^{-\psi})^{N_{\text{fft}}}. \tag{1.7}$$

Fig. 1.6 shows the CCDF of the PAPR for a 16-QAM modulated OFDM signal and different number of subcarriers N_{fft}. As expected, as the number of subcarriers increases, the probability that the PAPR exceeds a certain threshold increases. On

the other hand, for sufficiently large N_{fft}, the PAPR is independent of the mapping scheme of the OFDM symbols, as shown in Fig. 1.7, which shows the CCDF of the PAPR for different modulation schemes and $N_{\text{fft}} = 1024$ subcarriers.

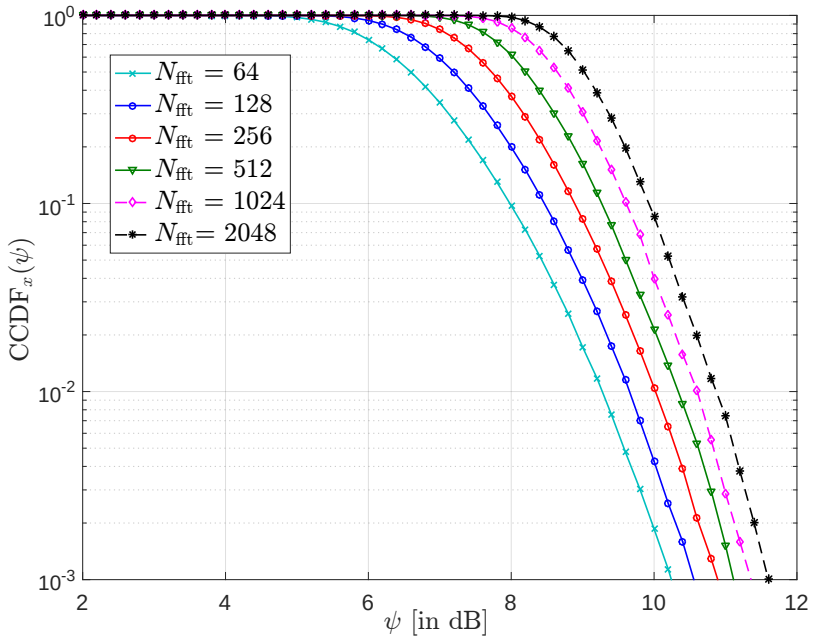


Figure 1.6: CCDF of PAPR for a 16-QAM modulated baseband OFDM signal for different number of subcarriers N_{fft}.

The CCDF of PAPR measures the clipping probability, which ensures that there is at least one peak in an observation window that has a strong amplitude and is likely to be strongly distorted. However, it does not take into account how many samples are distorted within this interval. Indeed, the quality of the transmitted signal is not affected if a single peak in an observation window is severely clipped, while even a slight distortion of a large number of samples may have unacceptable consequences. Therefore, the CCDF measurement of the PAPR is an essential metric for quantifying the performance of the OFDM system, while a complementary metric must be considered to gain full insight into the performance evaluation.

Modulation Error Ratio (MER) The MER compares the actual location of a received symbol with its ideal position and is defined as follows:

Definition 1.3 (Modulation error ratio)**.** *MER is defined as the ratio between the power of the original signal over the distortion introduced to the original signal* $\boldsymbol{x}$*. Therefore, the MER of the received signal* $\boldsymbol{x}_{\text{out}}$ *is expressed as follows:*

$$\text{MER} = \frac{\mathbb{E}\left[\|X\|^2\right]}{\mathbb{E}\left[\|X - X_{\text{out}}\|^2\right]}, \tag{1.8}$$

This can be illustrated as shown in Fig. 1.8. The numerator in MER is the average power of the complex emitted symbol and the denominator is the Mean Square Error

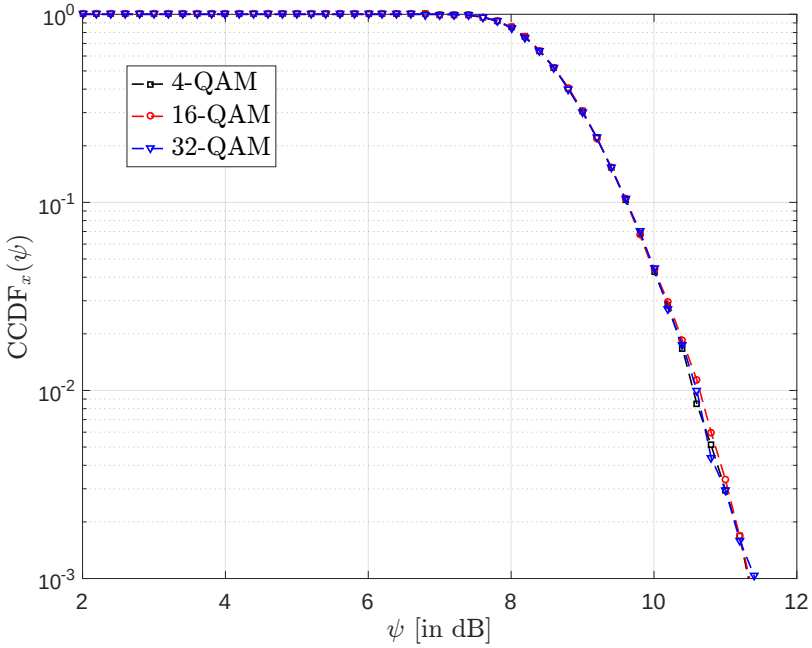


Figure 1.7: CCDF of PAPR for a baseband OFDM signal with $N_{\text{fft}} = 1024$ subcarriers and for different modulation schemes.

(MSE), which measures the distortion of the signal as the expected value of the square of the difference between the input and output signals. The MER and the PAPR CCDF are indeed complementary, where reducing one large peak can greatly reduce the PAPR while having negligible effects on MER. On the other hand, a small reduction of several peaks in the time-domain can change the PAPR slightly, while it strongly reduces the MER.

1.4 Compression techniques

In a data compression algorithm, a compressed representation is generated from an input source, requiring a smaller number of bits than the original source. The reconstruction algorithm then operates as the reversal operation on the compressed source to generate the reconstructed signal. The compression and reconstruction algorithms together are called the compression algorithm.

Based on the algorithm requirements, data compression methods can be divided into two main classes. *Lossless compression*, where no information is lost, i.e., the original data can be recovered exactly from the compressed data. This is achieved by identifying and eliminating statistical redundancies in the data. *Lossy compression*, which generally provides higher compression ratios than lossless compression, but at the cost of distortion in the signal reconstruction.

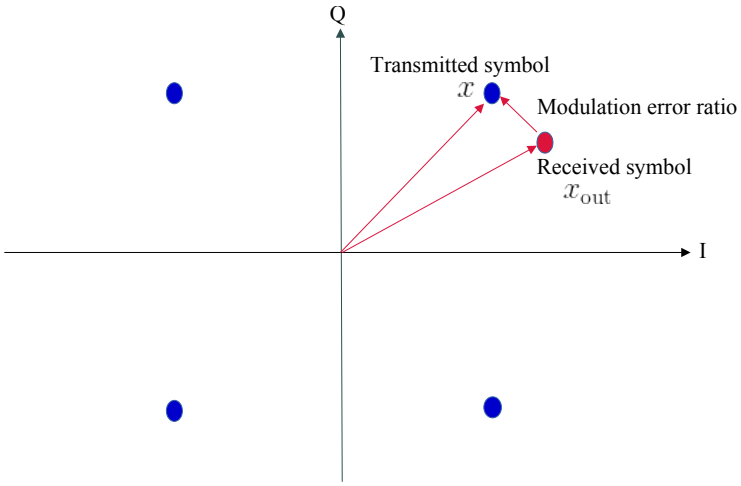


Figure 1.8: Modulation error ratio illustration.

1.4.1 Performance metrics

The performance of the compression algorithm can be evaluated in many ways, namely the amount of compression, the quality of the reconstructed data and the computational complexity. In the following, the main performance characteristics of the compression problem considered in this study are explained in more detail.

Rate In rate-distortion theory, rate R is defined as the average number of bits required to store or transmit each data sample. Entropy is a quantity defined in terms of the statistical properties of the information source. It represents a lower bound on the average number of bits required to represent the source without loss of information. According to Shannon, a large number of bits is required to represent samples with a low probability of occurrence, while samples with a high probability of occurrence are represented with a smaller number of bits [30]. In information theory, the entropy of a discrete r.v. X is defined as follows:

Definition 1.4 (Entropy [30])**.** *The entropy $H(X)$ of a discrete r.v. X is defined as follows:*

$$H(X) = -\sum_{x \in \mathcal{X}} \mathbb{P}[X = x] \log_2 \mathbb{P}[X = x], \tag{1.9}$$

where $\mathcal{X}$ is the set denoting the support of r.v. X and $\mathbb{P}[.]$ is the probability taken under the distribution of the r.v. X.

According to Shannon, entropy provides a lower bound on the reduction in the number of bits representing the source after the compression technique is applied. Shannon entropy measures information in bits/sample. If the base of the logarithm changes, the unit of entropy also changes, i.e., nat stands for $\log_e$, where e is the Euler's number and hartley stands for $\log_{10}$.

Compression ratio The most common measure used to calculate the efficiency of a compression algorithm is the Compression Ratio (CR), defined as follows:

Definition 1.5 (Compression ratio)**.** *It is the ratio between the total number of bits required to represent the uncompressed data and the compressed one, represented as follows:*

$$\mathrm{CR} = \frac{\textit{Number of bits in uncompressed data}}{\textit{Number of bits in compressed data}}. \tag{1.10}$$

CR is important to evaluate compression from a global system-level perspective, not just the reduction in the average number of bits representing each sample as defined by the rate. Therefore, rate and CR are good enough to evaluate the performance of lossless compression techniques. In lossy compression, on the other hand, the reconstructed data is different from the original data. Therefore, a performance measure is needed to evaluate the quality of the signal by measuring the degree of distortion.

As mentioned earlier, MER defined in (1.8) is the metric for evaluating signal quality. Therefore, the distortions experienced by the signal due to quantization are included in the MSE distortion measure in the denominator of the MER in (1.8). Consequently, the rate-distortion curve is a well-known analysis used to evaluate the compression performance of lossy compression techniques. It determines the minimum number of bits required to transmit a symbol, as measured by the rate R.

1.4.2 Scalar quantization

In general, quantization is a deterministic mapping of an input quantity to an output quantity. Quantization is considered as a compression algorithm because the set of achievable values at the output is finite and includes fewer elements than the set of possible values for the input. In this case, it is generally impossible to reproduce the input quantity exactly, i.e., it is an irreversible process. Quantization is thus related to *lossy compression*. If the input and output quantities are scalar, the quantization process is called *Scalar Quantization* (SQ).

1.4.2.1 Preliminaries

An N-level scalar quantizer, where N is the number of quantization levels, is defined as a nonlinear mapping such as [31]:

$$Q : \begin{array}{ccc} \mathbb{R} & \longmapsto & \mathcal{C} \\ x & \longmapsto & y \end{array} \tag{1.11}$$

- $\mathcal{C}$ is the output set, also denoted *codebook* which is a subset of $\mathbb{R}$ with size $|\mathcal{C}| = N$

  $$\mathcal{C} \triangleq \{y_1, y_2, ..., y_N\} \subset \mathbb{R}$$

  the output values y_i are said to be the codewords, also called *quantization levels*.

- $\mathbb{R}$ is the set of real values x, and is partitioned into N quantization intervals $\mathcal{R}_i$ for $i = 1, 2, \cdots, N$. The ith interval is given by

  $$\mathcal{R}_i = \{x \in \mathbb{R} : Q(x) = y_i\}; \quad \forall i = [1, 2, \cdots, N]$$

From this definition, it follows that $\bigcup_i \mathcal{R}_i = \mathbb{R}$ and $\mathcal{R}_i \bigcap \mathcal{R}_j = \emptyset \ ; \ i \neq j$. Each quantization interval $\mathcal{R}_i$ represents a single interval of the real line and the quantization levels y_i are located inside the associated quantization interval $\mathcal{R}_i$. $\mathcal{R}_i$ can also be defined by means of boundary points t_i, often called *decision levels* or *decision thresholds*. Therefore, we could write

$$\mathcal{R}_i = (t_{i-1}, t_i] = \{x : t_{i-1} < x \leq t_i\}; \quad \forall i = [1, 2, \cdots, N-1]$$

We assume that the quantization intervals include the higher but not the lower decision thresholds. The interval sizes denoted by $\Delta_i = t_i - t_{i-1}$ are called *quantization step sizes*. The first and last thresholds, i.e., t_0 and t_N, are set to $-\infty$ and $+\infty$, respectively. Thus, we denote $\mathcal{R}_0 = (-\infty, t_1]$ and $\mathcal{R}_N = (t_{N-1}, +\infty)$ as unbounded intervals or the *overload region*. If the input values are within the unbounded intervals, $\mathcal{R}_0$ or $\mathcal{R}_N$ are quantized to y_1 and y_N, respectively. The $N-2$ bounded cells, on the other hand, are known as the *granular region*. Consequently, N quantization levels and $N-1$ decision thresholds are considered in the design of the quantizer. The mapping Q of a scalar quantizer shown above can be represented by a piecewise constant input-output function, as shown in Fig. 1.9.

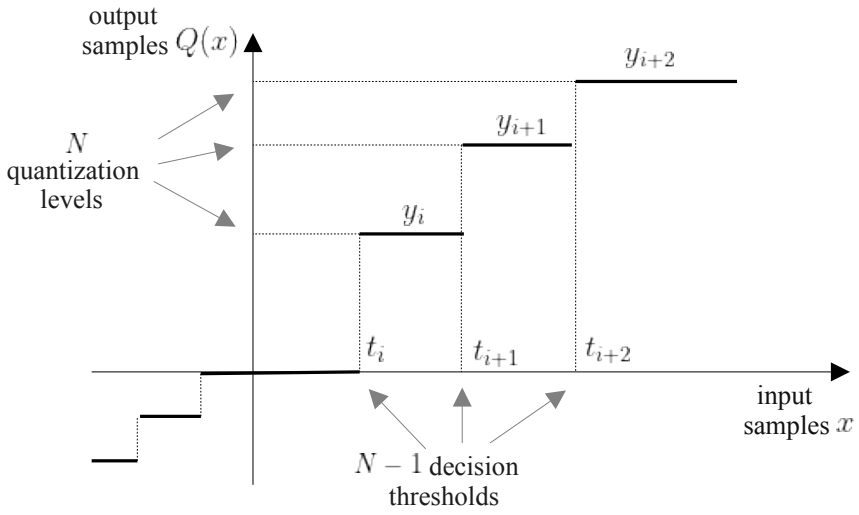


Figure 1.9: Input-output function Q of a scalar quantizer.

Quantization distortion The quantization operation Q obviously leads to a loss of information, since the reconstructed quantity obtained in (1.11) is different from the input quantity to the quantizer. The difference between the two is called the quantization error

$$q(x) = Q(x) - x. \tag{1.12}$$

The performance of the quantizer is usually measured by the resultant quantization distortion defined as follows:

Definition 1.6 (Quantization distortion)**.** *It is the average information loss for each sample caused by the quantizer. Through this context, MSE is the used distortion metric d for different algorithms due to its simplicity and practical tractability. Thus, adopting the MSE as the distortion measure, letting the quantization distortion defined as:*

$$D = \mathbb{E}[d(X, Q(X))] = \sum_{i=1}^{N} \int_{x \in \mathcal{R}_i} d(x - y_i) \mathrm{f}_X(x)\, \mathrm{d}x$$
$$= \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} (x - y_i)^2 \mathrm{f}_X(x)\, \mathrm{d}x. \tag{1.13}$$

where $\mathrm{f}_X(x)$ *is the PDF of the random input signal.*

It is obvious that the smaller the decision regions, the smaller the quantization distortion. Thus, the quantization distortion D is inversely proportional to the number of quantization levels N.

1.4.2.2 Uniform quantization

The Uniform Quantization (UQ) is a regular quantizer characterized by (i) equally spaced quantization intervals with step size Δ except for the unbounded regions, i.e., $t_i - t_{i-1} = \Delta$, $\forall i = [1, 2, \cdots, N-1]$ and (ii) the quantization levels y_i are the midpoints of the quantization intervals $\mathcal{R}_i$, for $i = 2, 3, \cdots, N-1$, and the first and last quantization levels are $y_1 = t_1 - \frac{\Delta}{2}$ and $y_N = t_{N-1} + \frac{\Delta}{2}$, respectively. Therefore, the quantization range can be defined as $B = t_{N-1} - t_1$ and could possibly be expressed as $B = \Delta \cdot (N-1)$. In this sequel, we consider the special case of the symmetric uniform quantizer defined by the condition $Q(x) = -Q(x)$ (except possibly at cell boundaries). For this special case, the range of the quantizer is then $B = 2y_N$ and the step size $\Delta = \frac{2y_N}{N}$.

The quantization distortion in (1.13) can be decomposed into two effects as follows:

$$D = \underbrace{\sum_{i=2}^{N-1} \int_{t_{i-1}}^{t_i} (x - y_i)^2 \mathrm{f}_X(x)\, \mathrm{d}x}_{D_{\text{gran.}}} + \underbrace{\int_{-\infty}^{t_1} (x - y_1)^2 \mathrm{f}_X(x)\, \mathrm{d}x + \int_{t_{N-1}}^{\infty} (x - y_N)^2 \mathrm{f}_X(x)\, \mathrm{d}x}_{D_{\text{overload}}}, \tag{1.14}$$

where $D_{\text{gran.}}$ is the distortion occurred when an input value lies in the bounded intervals, a.k.a *Granular distortion* and D_{overload} is the distortion at the first and last unbounded intervals, a.k.a *Overload distortion* as shown in Fig. 1.10.

Loading factor We can state this fact from the definitions of the granular and overload distortions. For a given Probability Density Function (PDF) $\mathrm{f}_X(x)$ and a fixed number of quantization levels N, the smaller the quantization step size Δ, the smaller the granular distortion. However, the smaller quantization step size Δ also means a smaller y_1 and y_N for a fixed N. This leads to a larger overload distortion. Therefore, the performance of the UQ should be optimized by dynamically adjusting the input signal level to achieve a tradeoff between granular and overload distortion for a fixed number of quantization levels N. On this basis, the loading factor γ is a key parameter defined as follows to describe this tradeoff:

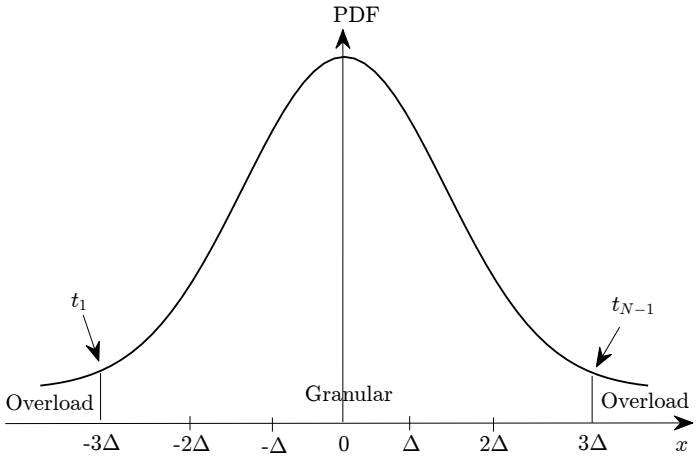


Figure 1.10: Overload and granular quantization distortions.

Definition 1.7 (Loading factor[6])**.** *The loading factor is the ratio between the maximum output amplitude of the quantizer y_N, the standard deviation of the original input signal σ, which characterizes the effective width of the signal and the adjustable factor that scales the input signal before quantization to obtain the optimal quantizer performance $s_{\text{sc.}}$*

$$\gamma = \frac{y_N}{\sigma s_{\text{sc.}}}, \tag{1.15}$$

where both y_N and σ are fixed attributes.

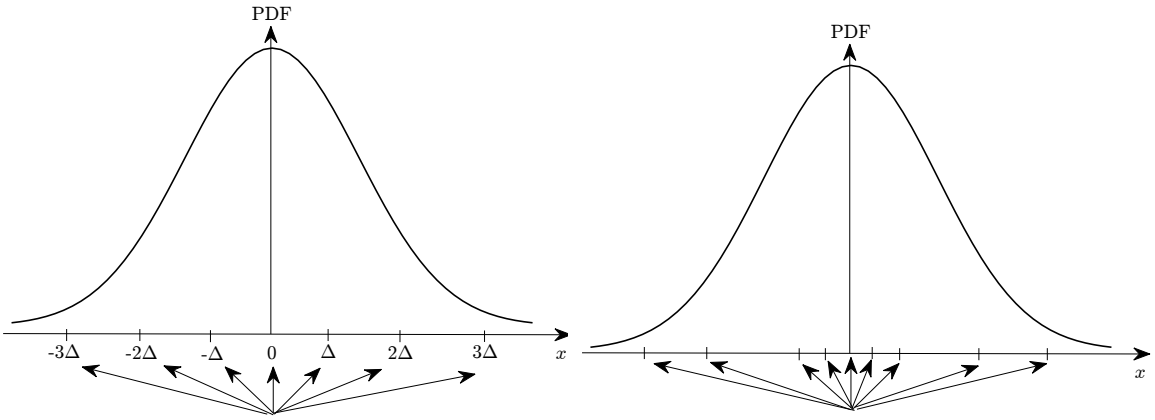
The role of the loading factor is clarified by investigating the definition in (1.15). The quantization range B is fixed for a given quantizer with N quantization levels. Therefore, a decrease in the dynamic range of the input signal, i.e., a decrease in the scaling factor $s_{\text{sc.}}$, leads to a significant decrease in the probability of overload distortion at the expense of a slight increase in granular distortion, resulting in an improvement in quantization performance. Consequently, the optimal loading factor $\gamma_{\text{opt.}}$ is the point at which the minimum quantization distortion is reached before the granular distortion becomes more significant than the overload distortion and the quantization distortion increases[1].

1.4.2.3 Non-uniform scalar quantization with fixed-length codes

Uniform quantization is optimal for minimizing distortion in the presence of uniformly distributed input signals [6]. In this case, the quantization levels are allocated uniformly, except for the first and last quantized values. However, for any other non-uniformly distributed input signals, the quantizer must be adjusted to match the distribution of the input signal. An optimal quantizer for non-uniformly distributed data samples requires more quantization levels in the region where samples have a high

[1]The loading factor is a parameter defined only for uniform quantization, since a non-uniform quantizer is optimized with respect to the PDF of the input signal, which will be explained in the next section.

probability of occurrence to reduce quantization distortion. Therefore, a Non-Uniform Quantizer (NUQ) has unequal spacing between quantization levels according to the distribution of the input signal, as shown in Fig. 1.11.



(a) Uniform quantizer. (b) Non-uniform quantizer.

Figure 1.11: The quantization levels are under-utilized by the uniform quantizer in the tail (less-distributed) areas of the PDF and over-utilized in the head area of the PDF (right). These levels are redistributed in the non-uniform quantizer based on the PDF of the signal, resulting in a lower MSE (left).

NUQ is generally optimized in order to minimize the MSE between the input and output values of the quantizer. The optimization is conducted with respect to the scalar quantizer parameters, which are the $N-1$ decision threshold ($t_i, \quad \forall i = 2, \cdots, N-1$)[2] and the N quantization levels ($y_i, \quad \forall i = 1, \cdots, N$) under the constraint of a fixed number of quantization levels N. This can be modeled mathematically as follows:

$$\{\mathcal{R}^*, \mathcal{C}^*\} = \arg\min_{\mathcal{R},\mathcal{C}} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} (x - y_i)^2 \mathrm{f}_X(x)\, \mathrm{d}x. \tag{1.16}$$

In 1957, Lloyd solved (1.16) and states the necessary conditions for designing an optimal NUQ in an unpublished paper. In 1982, Lloyd then published an effective algorithm using the derived conditions for the design of the optimal quantizer [32], known as *Lloyd algorithm*. Moreover, after few years Max independently formulated the optimal conditions and rediscovered the same algorithm proposed by Lloyd, today known as *Lloyd-Max algorithm* [33].

Optimality conditions

Centroid condition: When the decision thresholds are known $\{t_1, \cdots, t_{N-1}\}$, and by setting the partial derivative of (1.16) with respect to y_i to zero ($\frac{\partial D}{\partial y_i} = 0$), the optimal quantization levels, a.k.a. codebook levels are represented as follows:

$$y_i = \mathbb{E}(X|X \in \mathcal{R}_i) = \frac{\int_{t_{i-1}}^{t_i} x \mathrm{f}_X(x)\, \mathrm{d}x}{\int_{t_{i-1}}^{t_i} \mathrm{f}_X(x)\, \mathrm{d}x}. \tag{1.17}$$

[2] t_0 and t_N are set to $-\infty$ and ∞, respectively, for unbounded signals, On the other hand, they are set to the minimum and maximum possible values of the bounded signals, respectively.

Therefore, the optimum quantization levels are the centroids (or centers of mass) of the quantization intervals with respect to the PDF $\mathrm{f}_X(x)$ [32].

Nearest neighbor condition: When the quantization levels are known $\{y_1, \cdots, y_N\}$, and by setting the partial derivative of (1.16) with respect to t_i to zero ($\frac{\partial D}{\partial t_i} = 0$), the optimal decision thresholds are:

$$t_i = \frac{y_i + y_{i+1}}{2}. \tag{1.18}$$

Thus, the optimum decision threshold must be halfway between two adjacent levels. These results are well known and can be found in [32].

Lloyd has given a counterexample in [32] for a PDF for which the previous conditions (1.17) and (1.18) are necessary but not always sufficient to guarantee a global optimum of the quantizer (but only a local optimum). This is because the decision and quantization levels found may correspond to a saddle point, but not to the minimum we are looking for. Finally, Fleischer gave a condition for the PDF in [34] that makes (1.17) and (1.18) unique and sufficient for a globally optimized quantizer.

Lemma 1.1 (Fleischer condition [34])**.** *If the PDF of the random input signal* $\mathrm{f}_X(x)$ *satisfies the following condition*

$$\frac{d^2}{dx^2}\{\log \mathrm{f}_X(x)\} < 0; \quad , \forall x. \tag{1.19}$$

In other words, if $\mathrm{f}_X(x)$ *is log-concave, then there is only one quantizer that satisfies the optimality conditions, and this quantizer is indeed optimal. One of the examples for which the optimality conditions yield a unique solution is the Gaussian PDF.*

Lloyd algorithm The system of equations in (1.17) and (1.18) cannot be solved in closed form for $N > 2$. Therefore, the Lloyd Algorithm (LA) aims to solve (1.17) and (1.18) iteratively to achieve an optimal scalar quantization design. Its pseudocode representation is reproduced in Algorithm 1. The algorithm takes as inputs, the number of quantization levels and the PDF of the input signal, i.e., N and $\mathrm{f}_X(x)$, respectively. Then, starting from an initial codebook $\mathcal{C} = \{y_1, y_2, \cdots, y_N\}$, decision thresholds $\{t_1, \cdots, t_{N-1}\}$ are computed from (1.18), and then the codebook is updated by computing a new set of quantization levels from (1.17). The alternation between (1.17) and (1.18) is terminated when the algorithm converges or when a maximum number of iterations is reached. Convergence is achieved when the relative distortion falls below a certain predefined threshold ϵ, i.e.,

$$\frac{D^{(t-1)} - D^{(t)}}{D^{(t)}} < \epsilon, \tag{1.20}$$

where $D^{(t)}$ is the distortion at the t^{th} iteration.

Discussion on algorithm parameters The initial codebook is chosen as the basic uniform scalar codebook, which is as explained in Section 1.4.2.2. The convergence

Algorithm 1: Lloyd Algorithm

Require: N, $\mathrm{f}_X(x), \mathcal{C} = \{y_1, \cdots, y_N\}, D^{(0)}, \epsilon > 0$
Ensure: $y_i, t_i, \ \forall i \in \{1, \cdots, N\}$
while 1 **do**
  for $i = 1$ *to* N **do**
    compute y_i from (1.17)
    compute t_i from (1.18)
    compute $D^{(t)}$ from (1.13)
    if $\frac{D^{(t-1)} - D^{(t)}}{D^{(t)}} < \epsilon$ **then**
      return $y_i, t_i, \ \forall i \in \{1, \cdots, N\}$
    else
      $D^{(t)} \leftarrow D^{(t-1)}$
    end
  end
end

threshold ϵ must not be chosen too small so as not to increase the computation time, and too large values will result in a non-optimal quantizer.

Each quantization level must be represented by a binary codeword to be sent over capacity-constrained binary channels. We assume that all quantization levels have fixed-length codes in the sense that each quantization level is associated with a fixed number of bits, i.e., a codeword of the same length. The average codeword length signaled to a given quantization level y_i is denoted by l_{y_i}. Thus, if we design a scalar quantizer with a fixed number of quantization levels N, where each quantization level is signaled by a codeword of fixed-length, the average rate R is simply equal to $\log_2(N)$ bits/sample.

In this case, we naturally assume that the number of quantization levels N is a power of 2. However, fixed-length codes are not optimal in the sense of reducing the average number of bits to be sent. In other words, as we have shown previously, the quantization levels are not equally probable for a non-uniformly distributed signal. Therefore, it is not ideal to assign the same number of bits to represent quantization levels with different probabilities. Such a strategy requires variable-length codes, i.e., codes that can be adapted to the distribution of the input signal.

1.4.2.4 Entropy coding

Variable rate quantization can be achieved simply by cascading a scalar quantizer, which maps each input value to one of the finite quantization levels, and a lossless variable-length Entropy Coding (EC), which maps the output of the quantizer into a variable-length binary index such that it can be perfectly decoded by the receiver.

Entropy coding thus uses the Probability Mass Function (PMF) of quantization levels. Shannon entropy provides a way to map the set of quantization levels $\mathcal{C} = \{y_1, y_2, \cdots, y_N\}$ into codewords with variable-lengths. According to Shannon, as shown in (1.9), the average codeword length associated with the ith quantization

level is given as follows:

$$l_{y_i} = -\log_2 \mathbb{P}[Y = y_i]$$
$$\text{where,} \quad \mathbb{P}[Y = y_i] = \mathbb{P}[X \in \mathcal{R}_i] = \int_{t_{i-1}}^{t_i} \mathrm{f}_X(x)dx. \tag{1.21}$$

Thus, let L be the r.v. represents the average codeword lengths per input sample, thus, the average transmission rate of $\log_2 N$ bits per sample reduces to the average rate, i.e., entropy, which is calculated as follows:

$$R = \mathbb{E}(L) = \sum_{i=1}^{N} l_{y_i} \mathbb{P}[Y = y_i]. \quad \text{bits/sample} \tag{1.22}$$

The average distortion is the same as for scalar quantization with fixed-length codes given by (1.13). In practise, there are numerous techniques that approaches Shannon's lossless source coding theorem in computing average codeword lengths, such as Huffman coding [35], arithmetic coding [36], and Lempel-Ziv-Welch (LZW) coding [6], etc. In this study, we are interested in Huffman encoding because it is simple and slightly underestimates the true rate, i.e., in Huffman coding, the rate is always less than 1 bit per symbol away from the true entropy [35]. The Huffman algorithm is briefly explained in Appendix A.

1.4.2.5 Non-uniform scalar quantization with variable-length codes

Distortion and rate are two parameters that directly affect each other. Therefore, instead of cascading the scalar quantizer and the entropy encoder, the optimization problem in (1.16) can be modified to minimize the average distortion for a fixed entropy instead of a fixed codebook size. This leads to a scalar quantization with variable-length codes that merges quantization with entropy coding, a.k.a. Entropy Constrained Quantization (ECQ). In ECQ, the quantizer's optimization problem changes to either minimize the quantization distortion under a constraint on the maximum transmission rate or minimize the transmission rate under a constraint on the maximum tolerated distortion. This can be expressed mathematically as follows:

$$\begin{aligned} \{\mathcal{R}^*, \mathcal{C}^*, \mathcal{L}^*\} = &\underset{t_i, y_i, l_{y_i}}{\arg\min} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} (x - y_i)^2 \mathrm{f}_X(x)\, \mathrm{d}x, \\ &\text{s.t} \quad \sum_{i=1}^{N} l_{y_i} \mathbb{P}[y_i] \leq R_{\max}. \end{aligned} \tag{1.23}$$

The constrained optimization problem is formulated as a minimization of unconstrained Lagrangian function as follows:

$$\{\mathcal{R}^*, \mathcal{C}^*, \mathcal{L}^*\} = \underset{t_i, y_i, l_{y_i}}{\arg\min}\, J(t_i, y_i, l_{y_i})$$
$$\text{where} \quad J = \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} (x - y_i)^2 \mathrm{f}_X(x) + \lambda \sum_{i=1}^{N} l_{y_i} \mathbb{P}[Y = y_i], \tag{1.24}$$

where $0 \leq \lambda < \infty$ is called the Lagrangian parameter. The solution of the minimization of (1.24) is a solution of the constrained minimization problem in (1.23) in the following sense: if there exists a Lagrangian parameter λ that yields a maximally acceptable $R_{\max}$, then the corresponding distortion D is a solution of the constrained optimization problem.

In 1989, Chou, Lookabaugh, and Gray (CLG) developed a vector quantization algorithm that minimizes distortion under an entropy constraint for vector quantization [37]. The idea can be applied to the one-dimensional scalar quantizer, and the necessary conditions for designing an optimal scalar quantizer with variable-length codewords (ECQ) can be obtained by solving the unconstrained Lagrangian cost function in (1.24).

Necessary conditions

Centroid condition: From (1.23) it is evident that the rate R is independent of the values of the quantization levels y_i. Therefore, (1.24) is the same optimization problem of minimizing the distortion in the Lloyd algorithm (1.16), where the optimal quantization levels are the centroid of the quantization interval with respect to the signal PDF $\mathrm{f}_X(x)$, shown in (1.17)

Nearest neighbor condition: To derive a condition for the decision thresholds, assume that the quantization levels $\{y_1, \cdots, y_N\}$ and the codeword lengths $\{l_{y_1}, \cdots, l_{y_N}\}$ are known. Then setting the partial derivative of (1.24) with respect to t_i to zero, we get:

$$t_i = \frac{y_i + y_{i+1}}{2} + \frac{\lambda}{2} \frac{l_{y_{i+1}} - l_{y_i}}{y_{i+1} - y_i}. \tag{1.25}$$

The decision thresholds in (1.25) show that the optimal thresholds are no longer midway between the two adjacent levels, but is shifted towards the level with the longer average codeword length (less probable level). This is an important property in the development of ECQ and will be discussed in detail later.

Finally, the optimal average codeword length associated with each quantization level l_{y_i} only depends on the decision thresholds t_i. For given decision thresholds and hence probabilities $\mathbb{P}[Y = y_i]$, the average codeword lengths l_{y_i} can be directly determined as follows:

$$\begin{aligned} & l_{y_i} = -\log_2 \mathbb{P}[Y = y_i], \\ & \text{where,} \quad \mathbb{P}[Y = y_i] = \int_{t_{i-1}}^{t_i} \mathrm{f}_X(x)\, \mathrm{d}x, \end{aligned} \tag{1.26}$$

and the average transmission rate is approximately calculated in the design of the quantizer from the entropy of the random input signal given by (1.22). It is therefore called an entropy-constrained scalar quantizer.

Entropy-constrained Lloyd algorithm To design an optimal variable-length scalar quantizer using the optimality conditions in (1.17) and (1.25) and using the entropy of the source in (1.22) to measure the average rate, an iterative algorithm similar to

the Lloyd algorithm is constructed, known as the *Entropy Constrained Scalar Quantization (ECSQ)* and its pseudocode is presented in Algorithm 2.

The algorithm takes as inputs, the initial number of quantization levels of the quantizer N and the PDF of the random input signal $\mathrm{f}_X(x)$. Then, starting from an initial codebook $\mathcal{C} \equiv \{y_1, y_2, \cdots, y_N\}$, an initial codeword lengths $\{l_{y_1}, \cdots, l_{y_N}\}$, and a given Lagrange multiplier λ. The algorithm starts by computing the decision thresholds $\{t_1, \cdots, t_{N-1}\}$ from (1.25) and then updates the codebok quantization levels and their average codeword lengths from (1.17) and (1.26), respectively. The alternation between (1.17), (1.25) and (1.26) is stopped when the relative decrease of the Lagrangian cost function between two iterations falls below a certain predefined threshold ϵ

$$\frac{J^{(t-1)} - J^{(t)}}{J^{(t)}} < \epsilon. \tag{1.27}$$

Algorithm 2: Entropy-Constrained Lloyd Algorithm

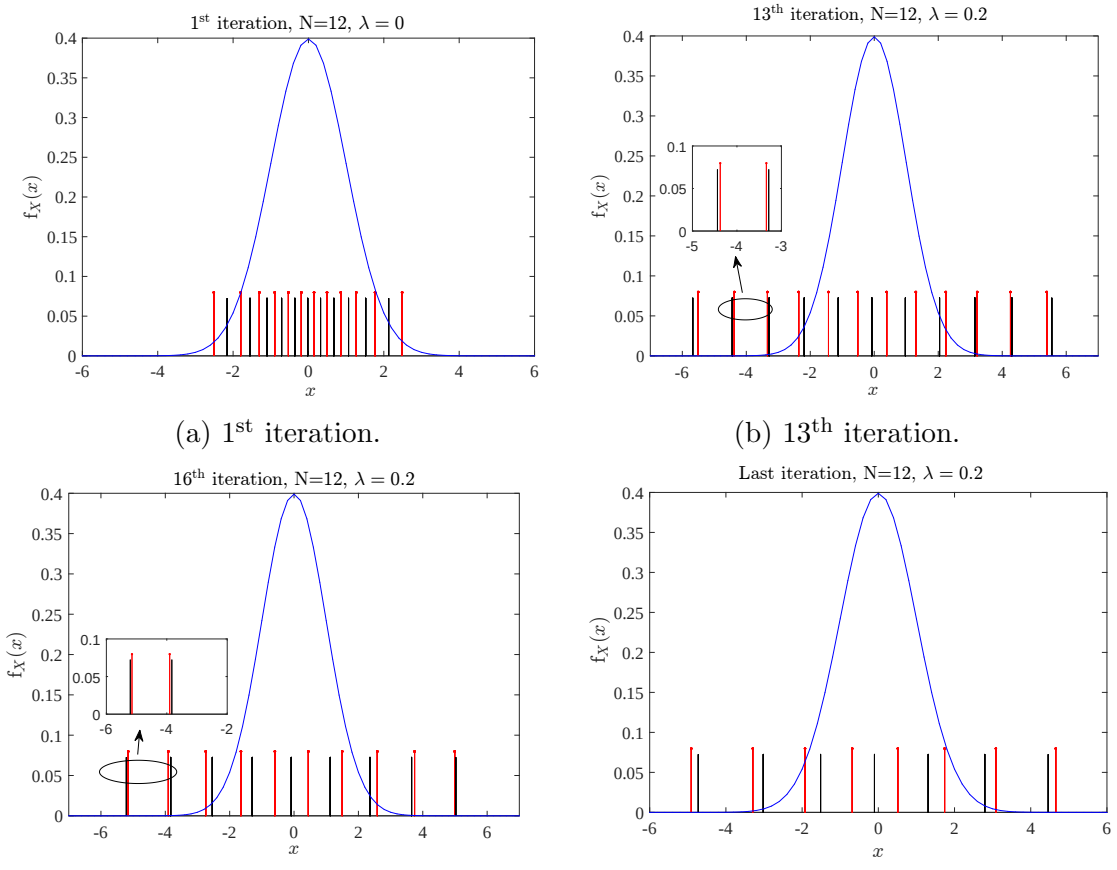
Require: $\mathrm{f}_X(x), N, \mathcal{C} = \{y_1, \cdots, y_N\}, \{l_{y_1}, \cdots, l_{y_N}\}, J^{(0)}, \epsilon > 0, \lambda \geq 0$
Ensure: y_i, l_{y_i} and $t_i, \quad \forall i \in \{1, \cdots, N\}$
$M \leftarrow N$
while 1 **do**
  for $i = 1$ *to* M **do**
    compute y_i from (1.17)
    compute l_{y_i} from (1.26)
    compute t_i from (1.25)
    for $j = 1$ *to* M **do**
      if $\mathbb{P}(Y = y_j) == 0$ **then**
        remove y_j
        $M \leftarrow M - 1$
      end
    end
    compute $j^{(t)}$ from (1.24)
    if $\frac{J^{(t-1)} - J^{(t)}}{J^{(t)}} < \epsilon$ **then**
      return $y_i, l_{y_i}, t_i, \ \forall i \in \{1, \cdots, N\}$
    else
      $J^{(t)} \leftarrow J^{(t-1)}$
    end
  end
end

Discussion on algorithm parameters The initial codebook is chosen to be the optimal codebook obtained by a fixed-length Lloyd algorithm (where $\lambda = 0$). Thus, the initial codeword lengths are set to $\log_2 N$ bits/sample. ϵ and $J^{(0)}$ are set similarly to the Lloyd algorithm, and how the Lagrange multiplier is chosen is briefly explained in the Lagrange minimization subsection.

As mentioned earlier, (1.25) shows that the decision threshold is shifted towards the lower probability level (longer codeword length). As the algorithm proceeds, the probability of the rarely chosen quantization level is further reduced until it is small

enough to no longer be chosen. To better explain this, we apply ECSQ to a unit-variance distributed Gaussian signal with initial number of levels $N = 12$ and Lagrange multiplier $\lambda = 0.2$.

Fig. 1.12a shows the resultant distribution of the quantization levels of the Lloyd algorithm. The blue curve represents the PDF of the signal, and the red and black lines represent the codebook quantization levels and decision thresholds, respectively. Figure 1.12b shows the 13$^{\text{th}}$ iteration. It can be seen that the second quantization region includes two quantization levels, since the decision threshold t_i is shifted to the less probable level during optimization. This makes it meaningless to represent a quantization interval with two representation levels, leading to the elimination of the quantization level with zero empirical probability. This can also be seen in the first quantization region in the 16^{th} iteration in Fig. 1.12c. Finally, the final codebook after convergence is shown in Fig. 1.12d. The quantizer ends up with 8 quantization levels, which is less than the initial number of levels $N = 12$. Consequently, the initial number of quantization levels N is crucial for the convergence of the algorithm (explained in more detail in [37]).



(a) 1$^{\text{st}}$ iteration.

(b) 13$^{\text{th}}$ iteration.

(c) 16$^{\text{th}}$ iteration.

(d) Last iteration.

Figure 1.12: ECSQ for a unit variance Gaussian-distributed signal.

Lagrange minimization The transformation of the constrained problem in (1.23) into an unconstrained problem in (1.24) raises the question of choosing the Lagrange

multiplier $\lambda \geq 0$ that satisfies the rate constraint $R_{\max}$ and represents the tradeoff between rate and distortion. The optimal Lagrange multiplier λ^* must satisfy the following conditions.

$$J = D + \lambda^*(R - R_{\max}) = 0. \tag{1.28}$$

This operation is an *outer* algorithm to the ECSQ algorithm and can be performed with any root finding algorithm such as the bisection method. However, we will focus on the algorithm used in [37], which aims to find a good initialization for the ECSQ algorithm to obtain an optimal quantizer. Let us first consider the solution of the algorithm for $\lambda_0 = 0$ (which gives the fixed-length codebook generated by the Lloyd algorithm with no rate constraint). This generates the optimal quantizer by optimizing $\arg\min_{t_i,y_i} \mathbb{E}[d(X, Q(X))]$ and the average rate is at most $\log_2 N$ bits per symbol. Then choose a slightly higher value of λ along with the obtained codebook at $\lambda_0 = 0$ to minimize $J_{\lambda_1} = D + \lambda_1 R$. Then continue with the obtained codebook and with a higher value λ_2 to minimize $J_{\lambda 2} = D + \lambda_2 R$. This procedure continues until λ^* satisfies the rate constraint $R_{\max}$. The formula for incrementing λ in this procedure is expressed as follows [37]:

$$\lambda_{n+1} = \frac{\lambda_n - \lambda_{n-1}}{R_{n-1} - R_n} \frac{R_0}{K}, \tag{1.29}$$

where R_n is the average rate at the n^{th} iteration, K is the total number of Lagrange multipliers to be tested. For the first iteration $\lambda_0 = 0$, as mentioned before and λ_1 is calculated as $\lambda_1 = \kappa \frac{D_0 - D_\infty}{R_0 - R_\infty}$. Where, D_0 and R_0 are the distortion and rate at $\lambda = 0$, respectively. D_∞ and R_∞ are the distortion and rate at $\lambda = \infty$[3], respectively. κ is a very small value (tends to zero). The obtained average rates and distortions tuples obtained at each $\lambda \geq 0$, i.e., $[(R_0, D_0), (R_{\lambda_1}, D_{\lambda_1}), \cdots, (R_\infty, D_\infty)]$, are constructing the entire convex hull of the rate-distortion R-D function, as shown in Fig. 1.13.

1.4.3 Asymptotic distortion-rate functions for scalar quantization

For most random processes, no analytical expression for the distortion-rate function can be given. Shannon gave a fundamental lower bound, known as *Shannon Lower Bound (SLB)* for the distortion-rate function in any lossy compression technique [30]. It gives the minimum acceptable distortion in receiving a signal without exceeding a certain rate, i.e., $D_{\text{SLB}}(R)$ or the minimum required transmission rate for a source without exceeding a certain distortion, i.e., $R_{\text{SLB}}(D)$. Let $h(X)$ be the differential entropy of an absolutely continuous r.v. defined as follows.

Definition 1.8 (Differential entropy)**.** *The differential entropy of an absolutely continuous r.v. X is defined as follows:*

$$h(X) = -\int_{\mathcal{X}} \mathrm{f}_X(x) \log_2 \mathrm{f}_X(x)\, \mathrm{d}x, \tag{1.30}$$

where $\mathcal{X}$ is the set denoting the support of r.v. X.

[3] $(\lambda = \infty)$ corresponds to a rate-zero codebook, where there is only one quantization level being the centroid of the entire input space. So we can say that $R_\infty = 0$ and D_∞ is the quantization distortion due to a one-level quantizer.

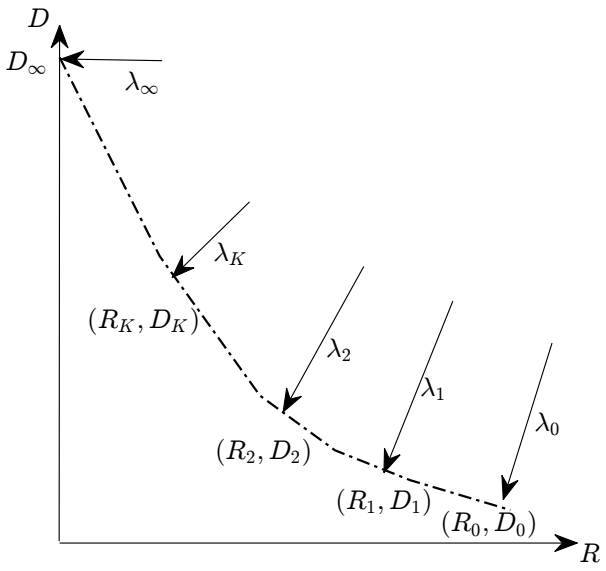


Figure 1.13: The R-D convex hull obtained by running the ECSQ algorithm for different values of the Lagrange multiplier λ.

Shannon's lower bound is thus defined as follows

Definition 1.9 (Shannon lower bound [30])**.** *For any stationary i.i.d source and MSE distortion, Shannon lower bound for the rate-distortion function is expressed as follows:*

$$R_{\mathrm{SLB}}(D) = h(X) - \frac{1}{2}\,\log_2(2\pi e D), \tag{1.31}$$

and equivalently the distortion-rate function can be expressed as follows:

$$D_{\mathrm{SLB}}(R) = \frac{1}{2\pi e} 2^{2h(X)} 2^{-2R}, \tag{1.32}$$

where e is the exponential, R and D are the average rate and distortion, respectively, of the lossy compression technique.

For a Gaussian-distributed signal with zero mean and σ^2 variance, Shannon's lower bound is as follows:

$$D^{\mathrm{G}}_{\mathrm{SLB}}(R) = \sigma^2\, 2^{-2R}. \tag{1.33}$$

The representation of SLB in (1.32) in dB produces the well-known *6-dB rule*. The 6-dB rule states that at high rates, each additional bit in the quantization of the input signal reduces the MSE distortion by 6 dB. In general, it is difficult to derive an analytical expression for the rate-distortion of an optimal quantizer and one way to deal with this is to consider the *High-Rate Approximation (HRA)* or *Asymptotic quantization*. This is the case when we assume a very high rate ($R \to \infty$) or, correspondingly, a small distortion ($D \to 0$). Therefore, for high rates, the performance of a quantization design can be evaluated by comparing the high-rate approximation of the distortion-rate function with Shannon's lower bound.

The general assumption used to derive the HRA is that the number of quantization levels N is very large. This results in two important facts: (i) the quantization step size Δ is very small compared to the granular region, which makes the $\mathrm{f}_X(x)$ within each quantization interval $\mathcal{R}_i$ is nearly constant, (ii) the overload distortion is very small and thus negligible. These assumptions simplify the derivation of the analytical performance and the accuracy of the approximated expressions derived under these assumptions increases as the rate approaches infinity.

Uniform quantizer The design of a uniform quantizer for a uniformly distributed signal is one of the few exceptions in which the rate-distortion function can be derived analytically. For a uniformly distributed signal, $\mathrm{f}_X(x) = \frac{1}{2a}$ in the range between $[-a, a]$, the quantization error is also uniformly distributed, $\mathrm{f}_Q(q) = \frac{1}{\Delta}$ for $\frac{-\Delta}{2} \leq q \leq \frac{\Delta}{2}$. Thus, the average distortion can be calculated simply as follows:

$$D = \int_{-\Delta/2}^{\Delta/2} q^2 \mathrm{f}_Q(q) dq = \frac{\Delta^2}{12}, \tag{1.34}$$

where Δ is the quantization step size. For a UQ, Δ is constant and defined as $\Delta = \frac{2a}{N}$, $N = 2^R$, and the variance of a uniformly distributed signal with zero mean is $\frac{a^2}{3}$. By substituting Δ into (1.34), the asymptotic distortion-rate function of a uniform quantizer can be expressed as follows.

Proposition 1.1. *The asymptotic distortion-rate formula for a uniform scalar quantizer is expressed as:*

$$D_{\mathrm{UQ}}(R) = \sigma^2 2^{-2R}, \tag{1.35}$$

where σ is the standard deviation of the input signal, $R = \log_2(N)$ bits/sample and N is the number of quantization levels.

Proposition 1.1 can be represented in terms of the loading factor γ as follows:

$$D_{\mathrm{UQ}}(R) = \frac{\gamma^2 \sigma^2}{3N^2}. \tag{1.36}$$

Scalar quantizer with fixed-length codes (LA) The asymptotic distortion-rate function of a scalar quantizer with fixed-length codes as a function of the PDF of the input signal and the number of quantization levels N is given as follows.

Proposition 1.2 (Panter and Dite formula [38])**.** *The asymptotic distortion-rate formula for a non-uniform scalar quantizer with fixed-length codes is expressed as:*

$$D_{\mathrm{LQ}}(R) \sim \frac{2^{-2R}}{12} \left(\int_{-\infty}^{\infty} \sqrt[3]{\mathrm{f}_X(x)}\, \mathrm{d}x \right)^3, \tag{1.37}$$

where $\mathrm{f}_X(x)$ is the PDF of the quantizer input signal, $R = \log_2(N)$ bits/sample and N is the number of quantization levels.

The asymptotic distortion-rate function expressed in Proposition 1.2 was derived by Panter and Dite in [38] and is known as the "*Panter and Dite formula*".

Scalar quantizer with variable-length codes (ECSQ) Finally, the asymptotic distortion-rate function of a scalar quantizer with variable-length codes is expressed as follows.

Proposition 1.3 (Gish and Pierce formula [39])**.** *The asymptotic distortion-rate formula for a non-uniform quantizer with variable-length codes is expressed as:*

$$D_{\text{ECSQ}}(R) = \frac{1}{12} 2^{2h(X)} 2^{-2R}, \tag{1.38}$$

where $h(X)$ is the differential entropy of the input signal $\boldsymbol{x}$.

This was first proved by Gish and Pierce in [39] and then in [40] and is known as the "*Gish and Pierce formula*". Gish and Pierce conclude the famous theory that the uniform quantizer is the optimal high-rate (high-resolution) entropy constrained scalar quantizer, independent of the source statistics for the MSE criterion. In other words, the uniform quantizer is asymptotically optimal at minimizing the mean squared quantization error when the quantization cell size approaches zero, i.e., when the rate approaches infinity. Expressing Propositions 1.2 and 1.3 in dB, the aforementioned 6 dB rule becomes apparent.

Moreover, the high-resolution approximations can be used to illustrate how far an optimal achievable variable-length scalar quantizer (ECSQ) is from SLB in the rate-distortion tradeoff by comparing them as follows:

$$R_{\text{ECSQ}} - R_{\text{SLB}} = [h(X) - \frac{1}{2}\log_2(12D)] - [h(X) - \frac{1}{2}\log_2(2\pi eD)] = 0.255 \text{ bits}. \tag{1.39}$$

$$\text{SNR}_{\text{SLB}} - \text{SNR}_{\text{ECSQ}} = 1.53 \text{ dB/sample}. \tag{1.40}$$

Finally, as shown above, the distortion of an optimal scalar quantizer with variable-length codes is only 1.53 dB larger than Shannon's lower bound at high rates. This observation is due to Gish and Pierce, who asymptotically generate the rate-distortion function of the ECSQ [39].

For a Gaussian-distributed signal with zero mean and variance σ^2, the asymptotic distortion-rate function in Propositions 1.2 and 1.3 are expressed as follows:

$$D^{\text{G}}_{\text{LQ}}(R) = \frac{\pi\sqrt{3}}{2}\sigma^2 2^{-2R}, \quad \text{and} \quad D^{G}_{\text{ECSQ}}(R) = \frac{\pi e}{6}\sigma^2 2^{-2R}. \tag{1.41}$$

1.4.4 Vector quantization

Scalar quantization is by far the most popular form of quantization. However, it is generally impossible to reach Shannon's rate-distortion bound with scalar quantization [31, section 9.9]. This bound can only be achieved if a vector of input samples is jointly quantized to a corresponding output vector. In this case, the output and input quantities are vectors and the quantization procedure is called Vector Quantization (VQ). The Shannon limit is reached asymptotically with VQ as the vector dimension increases, i.e., as the number of samples in the input and output vectors approaches infinity. However, finding the optimal vector quantizer also becomes a prohibitively complex task as the vector dimension increases [41].

Vector quantization generalizes scalar quantization from one-dimension to multiple dimensions. Therefore, it can always be considered as a redundancy elimination technique. The use of multiple dimensions introduces new concepts, techniques, and advantages over scalar quantization in terms of the tradeoff between rate and distortion. It improves system performance over scalar quantization. This performance improvement is achieved by exploiting the linear dependence (correlation) or the nonlinear dependence between the samples of the input vector that is created by oversampling the signal [41].

1.4.4.1 Preliminaries

VQ is a nonlinear mapping where a k-dimensional input vector $\mathbf{x} = \{x_0, x_1, \cdots, x_{k-1}\}$ is mapped to one of N quantized vectors. Thus, an N-level vector quantizer is defined as a mapping function such as [6]:

$$Q : \begin{array}{ccc} \mathbb{R}^k & \longmapsto & \mathcal{C} \\ \mathbf{x} & \longmapsto & \mathbf{y} \end{array} \tag{1.42}$$

- $\mathcal{C}$ is the output set of finite vectors, denoted as *codebook*. It is a subset of $\mathbb{R}^k$ with size $\|\mathcal{C}\| = N$

  $$\mathcal{C} \triangleq \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N\} \subset \mathbb{R}^k$$

  the output vectors $\mathbf{y}_i$ are the *quantization levels*. For a finite N, each output vector can be represented by $\log_2 N$ bits/vector. Here, each sample (vector component) is represented at the rate $R = \frac{\log_2 N}{k}$ bits/sample.

- $\mathbb{R}^k$ is the set of real vectors $\mathbf{x}$ divided into N ($k-$dimentional) regions (cells) S_i, $\forall i = 1, 2, \cdots, N$. The ith region is given by

  $$\mathcal{S}_i = \{\mathbf{x} \in \mathbb{R}^k : Q(\mathbf{x}) = \mathbf{y}_i\}; \quad \forall i = [1, 2, \cdots, N]$$

  From this definition it follows that $\bigcup_i \mathcal{S}_i = \mathbb{R}^k$ and $\mathcal{S}_i \bigcap \mathcal{S}_j = \emptyset \; ; \; i \neq j$.

Thus, VQ is completely defined by a set of quantization regions $\{\mathcal{S}_0, \cdots, \mathcal{S}_N\}$, and the corresponding codebook $\mathcal{C} \equiv \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N\}$. Similar to scalar quantization, the unbounded cell is called an overload cell and the volume of all overload cells is called *overload region* . All other cells bounded between $k-$dimensional intervals form the *granular region*.

1.4.4.2 Vector quantization with fixed-length codes

In 1980, Linde, Buzo, and Gray extended the optimality conditions derived by Lloyd to the design of vector quantizers [42]. Similarly, the goal is to find a codebook and partition design that minimizes distortion under the condition of a fixed number of output quantization levels N. The necessary conditions can be derived similarly to the Lloyd quantizer in Section 1.4.2.3.

The average quantization distortion per sample between the quantized input and output vectors is generalized to the multiple dimensions as follows:

$$D = \frac{1}{k} \sum_{i=1}^{N} \int_{\mathbf{x} \in \mathcal{S}_i} d_k(\mathbf{x} - \mathbf{y}_i) \mathrm{f}_{\mathbf{X}}(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \tag{1.43}$$

where $\mathrm{f}_{\mathbf{X}}(\mathbf{x})$ denotes the k-dimensional joint PDF of the input vector $\mathbf{x}$ and d_k is the type of the distortion measure. for MSE measure of distortion, we have:

$$D = \frac{1}{k}\sum_{i=1}^{N}\int_{x_1\in\mathcal{S}_i}\cdots\int_{x_k\in\mathcal{S}_i}\|\mathbf{x}-\mathbf{y}_i\|_2^2\,\mathrm{f}_{X_1,\cdots,X_k}(x_1,\cdots,x_k)\,\mathrm{d}x_k\cdots\mathrm{d}x_1. \tag{1.44}$$

Each quantization level $\mathbf{y}_i$ is represented by a codeword. The length of the codeword $l_{\mathbf{y}_i}$ is the number of bits representing the quantization vector $\mathbf{y}_i$. Thus, using the entropy of the input source, the average codeword length in bits per sample is expressed as follows:

$$R = \frac{1}{k}\sum_{i=1}^{N} l_{\mathbf{y}_i}\mathbb{P}[\mathbf{Y}=\mathbf{y}_i], \tag{1.45}$$

where $l_{\mathbf{y}_i}$ is the codeword length assigned to the quantization vector $\mathbf{y}_i$ and $\mathbb{P}[\mathbf{Y}=\mathbf{y}_i]$ is the probability of the quantization vector given by

$$\mathbb{P}[\mathbf{Y}=\mathbf{y}_i] = \int_{\mathcal{S}_i}\mathrm{f}_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x}. \tag{1.46}$$

Necessary conditions

As mentioned earlier, the necessary conditions for the design of the vector quantizer are generated by minimizing the average quantization distortion represented in (1.44). The quantization levels and quantization regions are thus the generalization of the scalar-based centroid and nearest neighbor constraints presented in (1.17) and (1.18), respectively to the k-dimensional space as follows:

$$\mathbf{y}_i = \frac{\int_{\mathcal{S}_i}\mathbf{x}\mathrm{f}_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x}}{\int_{\mathcal{S}_i}\mathrm{f}_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x}}. \tag{1.47}$$

$$\mathcal{S}_i \subset \left\{\mathbf{x} : \|\mathbf{x}-\mathbf{y}_i\|_2^2 \le \left\|\mathbf{x}-\mathbf{y}_j\right\|_2^2; j\neq i\right\}; \quad \forall i=[1,\cdots,N]. \tag{1.48}$$

Similarly to the scalar case, according to (1.48), the optimal quantizer is the one that assigns the input vector $\mathbf{x}$ to the closest quantization level $\mathbf{y}_i$ that minimizes its distortion. Thus, the quantization region $\mathcal{S}_i$ is the set of input vectors closer to $\mathbf{y}_i$ than any other codebook level.

In the scalar case, these regions were simply real line intervals. However, in $\mathbb{R}^k$ space, the N regions partitioning the $k-$dimensional space have a more complex structure called *Voronoi regions* [6]. Each of the N regions is represented by one of the N codebook levels $\{\mathbf{y}_1\cdots\mathbf{y}_N\}$. Fig. 1.14 shows the Voronoi region of a $2-$ dimensional space.

Tie-breaking rules

- If an input vector $\mathbf{x}$ has more than one quantization vector that minimizes its distortion, the codeword with the lowest index is chosen [32, 42].

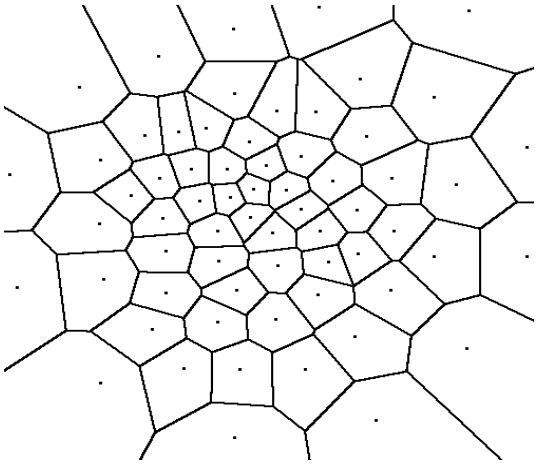


Figure 1.14: Example of Voronoi regions of $\mathbb{R}^2$ dimensional space [6].

- Empty-cell problem: Cells with zero probability during optimization[4] will have a meaningless undefined centroid. Therefore, the empty cell is removed and other cell with positive probability is split into two cells to keep the goal of minimizing the distortion under the constraint of a fixed number of levels. Section 1.4.4.2 discusses how to apply the empty cell algorithm in designing the codebook [6].

In vector quantization, differentiation does not always work because of the potentially complicated surfaces of the partition cells, i.e., ignorance of the multivariate PDF of the input vector components. Some simple shapes can be handled, as mentioned in [42]. Therefore, in vector quantization, the quantizer is usually developed offline on a large Training Sequence (TS) (a finite collection of samples generated from the source distribution to represent it statistically) instead of the probabilistic description of the k-dimensional input random vector $\mathbf{x}$. The generated codebook is applied to all future data not used in the training sequence. Finally, a stationary ergodic process operating on a sufficiently long TS can guarantee good performance on any other test sequence, as shown in [42, 43].

The Linde-Buzo-Gray Algorithm (LBG) is a sample version of the Lloyd algorithm and is defined as follows. The optimality conditions in (1.47) and (1.48) are solved iteratively to obtain an optimal vector quantizer. In case of a long sequence of training vectors representing the signal, the average quantization distortion, centroid condition, and nearest neighbor condition in (1.44), (1.47), and (1.48) are empirically calculated as follows, respectively:

$$D = \frac{1}{k}\left[\frac{1}{n}\sum_{j=1}^{n}\|\mathbf{x}_j - \mathbf{y}_i\|_2^2\,|(\mathbf{x}_j \in \mathcal{S}_i)\right]; \quad \forall i = [1, \cdots, N]. \tag{1.49}$$

$$\mathbf{y}_i = \frac{1}{|\mathcal{S}_i|}\sum_{j=1}^{|\mathcal{S}_i|}\mathbf{x}_{i,j}; \quad \forall i = [1, \cdots, N],$$
$$\text{where,} \quad \mathbf{x}_{i,j} = \{\mathbf{x}_j \in \mathcal{S}_i\}; \quad \forall j = [1, \cdots, n], \tag{1.50}$$

[4]Cells with zero probability are the cells that were not selected as codebook level by any of the input vectors.

and

$$\mathcal{S}_i \subset \left\{\mathbf{x}_j : \|\mathbf{x}_j - \mathbf{y}_i\|_2^2 \leq \|\mathbf{x}_j - \mathbf{y}_m\|_2^2 ; m \neq i\right\}; \quad \forall i = [1, \cdots, N], \tag{1.51}$$

where $\{\mathbf{x}_j,\ j = 1, \cdots, n\}$ is the training set of input vectors, n is the length of the TS and $|\mathcal{S}_i|$ is the number of input vectors assigned to cell $\mathcal{S}_i$.

Algorithm 3 is the pseudocode representation of the LBG algorithm. One starts the algorithm with a fixed number of quantization levels N, a TS and an initial codebook. First, each vector in the training sequence is mapped to the nearest quantization vector from (1.51), then the codebook is updated by computing the quantization levels from (1.50). The iterations between (1.50) and (1.51) are stopped when the relative distortion between one iteration and the previous one falls below a certain predefined threshold ϵ. This algorithm guarantees either a reduction of the average distortion or its maintenance. Therefore, it helps to start the algorithm with a good initial codebook to optimize the performance of the quantizer.

Algorithm 3: Linde-Buzo-Gray Algorithm

```
Require: {x_j, j = 1, ··· , n}, N, C = {y_1, ··· , y_N}, D^(0), ε > 0
Ensure: y_i,  ∀i ∈ {1, ··· , N}
while 1 do
    for i = 1 to N do
        Assign each training vector to the nearest neighbor from (1.51)
        compute y_i from (1.50)
        compute D^(t) from (1.49)
        if (D^(t−1) − D^(t)) / D^(t) < ε then
            return y_i, S_i,  ∀i ∈ {1, ··· , N}
        else
            D^(t) ← D^(t−1)
        end
    end
end
```

Discussion of algorithm parameters During optimization, in the case of an empty cell (cell with zero probability), the cell with the highest distortion is selected to split into two cells, which reduces the distortion under the constraint of a fixed number of levels. As already mentioned, ϵ must be chosen not too small in order not to increase the computation time, nor too large to ensure sufficient distortion.

Initial codebook designs The iterative LBG algorithm guarantees either improvement or at least maintenance of the average distortion. If one starts with a good initial codebook, the computation time consumed in improving the codebook can be reduced, i.e., the convergence rate is faster. Furthermore, if the initial codebook is good enough, it may be sufficient not to perform any further improvement algorithms. In the literature, a variety of codebook designs are used for either general clustering analysis or vector quantization. In general, a good clustering algorithm requires more time and a higher computational cost. Therefore, there is always a tradeoff between the performance of the clustering algorithm and its complexity. In our context, we will show

the approaches used in our analysis, while an overview of most clustering algorithms can be found briefly in [6].

Random coding Random coding is the simplest approach to fill the initial codebook. It is based on filling the codebook with the first N vectors of the input training sequence. This method does not always lead to a good codebook because the first Nk samples of a training sequence are statistically insufficient, except for large k or N values. As a small improvement to this method, instead of selecting the first N vectors, one can take a vector every M_{vec} training vectors. This technique attempts to maintain the simplicity of the random coding method, in addition to obtaining separate vectors from the input signal that can better characterize the properties of the signal when the data is highly correlated.

Product codes The product codes shape the codebook as follows. The idea is to generate k *scalar quantizers*, each with 2^R codewords for each dimension separately and then the Cartesian product of the scalar quantizers generates a vector quantizer. More specifically, if we want to design a vector quantizer with 2^{kR} quantization levels to quantize a k-dimensional vector with a resolution of R bits per individual sample, we first design a scalar quantizer denoted by $q(x)$, and then the quantization of the input vector $\mathbf{x}$ is as follows: $Q(\mathbf{x}) = Q(x_0, \cdots, x_{k-1}) = (q(x_0), \cdots, q(x_{k-1}))$. In other words, the codewords of the vector quantizer are the Cartesian product of the scalar quantizers. Henceforth, product codes are only possible if R is an integer, as required by scalar quantization.

It is worth mentioning that other clustering approaches have been used in the design of vector quantizers, e.g., the modified k-means algorithm [44]. The different approaches differ in their complexity, storage limit, and convergence rate. However, the standard LBG algorithm explained is the most widely used algorithm for designing a VQ codebook.

1.4.4.3 Vector quantization with variable-length codes

In 1989, Chou, Lookabaugh, and Gray developed a vector quantization algorithm that minimizes distortion under an entropy constraint, i.e., Entropy Constrained Vector quantization (ECVQ) [37]. The goal is to minimize the distortion under a rate constraint similar to the ECSQ optimization problem defined in (1.24) but generalized to k-dimensional space and based on this, the necessary conditions that generate the ECSQ in (1.17) and (1.25), respectively are generalized to k-dimensional space.

The Chou-Lookabaugh-Gray algorithm (CLG/ECVQ): The nearest neighbor and centroid conditions are solved iteratively to achieve an optimal entropy-constrained vector quantization design. In the considered case of a long training sequence representing the signal $\{\mathbf{x}_j;\ j = 1, \cdots, n\}$, the average distortion and centroid condition are mathematically formulated as in (1.49) and (1.51) respectively. The average rate and nearest neighbor conditions are formulated mathematically as

follows, respectively:

$$R = \frac{1}{k}\left[\frac{1}{n}\sum_{j=1}^{n}(l_{\mathbf{y}_i}|\mathbf{x}_j \in \mathcal{S}_i)\right]; \quad \forall i = [1, \cdots, N].$$
$$\text{where,} \quad l_{\mathbf{y}_i} = -\log_2 \mathbb{P}[\mathbf{Y} = \mathbf{y}_i] \quad \text{and} \quad \mathbb{P}[\mathbf{Y} = \mathbf{y}_i] = \frac{|\mathcal{S}_i|}{n}. \tag{1.52}$$

and

$$\mathbf{x}_j \in \mathcal{S}_i; \quad \text{if} \;\; J(\mathbf{x}_j, \mathbf{y}_i, l_{\mathbf{y}_i}) \leq J(\mathbf{x}_j, \mathbf{y}_m, l_{\mathbf{y}_m}); \quad \forall m \neq i,$$
$$\text{where,} \quad J(\mathbf{x}_j, \mathbf{y}_i, l_{\mathbf{y}_i}) = |\mathbf{x}_j - \mathbf{y}_i|_2^2 + \lambda l_{\mathbf{y}_i}, \tag{1.53}$$

where n is the length of the TS.

Algorithm 4 is the pseudocode representation of the CLG algorithm. The algorithm takes as inputs, the initial number of quantization levels N, a training sequence from the source distribution $\{\mathbf{x}_j,\ j = 1, \cdots, n\}$, an initial codebook $\mathcal{C} = \{\mathbf{y}_1, \cdots, \mathbf{y}_N\}$, an initial set of average codeword lengths $\{l_{\mathbf{y}_1}, \cdots, l_{\mathbf{y}_N}\}$ and a given Lagrange multiplier λ. First, each vector in the training sequence is assigned to the nearest quantization cell from (1.53), then the codebook quantization levels and their average codeword lengths are updated from (1.50) and (1.52), respectively. The iterations between (1.50), (1.52) and (1.53) are stopped when the relative decrease of the Lagrangian cost function between two iterations falls below a certain predefined threshold ϵ.

$$\frac{J^{(t-1)} - J^{(t)}}{J^{(t)}} < \epsilon. \tag{1.54}$$

Algorithm 4: Chou-Lookaubaugh-Gray Algorithm

Require:
$\{\mathbf{x}_j, j = 1, \cdots, n\}, N, \mathcal{C} = \{\mathbf{y}_1, \cdots, \mathbf{y}_N\}, \{l_{\mathbf{y}_1}, \cdots, l_{\mathbf{y}_N}\}, J^{(0)}, \epsilon > 0, \lambda \geq 0$
Ensure: $\mathbf{y}_i, l_{\mathbf{y}_i}$ and $\mathcal{S}_i, \quad \forall i \in \{1, \cdots, N\}$
while 1 **do**
  for $i = 1$ *to* N **do**
    assign each training vector to the nearest codeword from (1.53)
    compute $\mathbf{y}_i$ from (1.50)
    compute $l_{\mathbf{y}_i}$ from (1.52)
    compute $J^{(t)}$ from (1.24)
    if $\frac{J^{(t-1)} - J^{(t)}}{J^{(t)}} < \epsilon$ **then**
      return $\mathbf{y}_i, l_{\mathbf{y}_i}, \mathcal{S}_i, \ \forall i \in \{1, \cdots, N\}$
    else
      $j^{(t)} \leftarrow j^{(t-1)}$
    end
  end
end

Discussion of algorithm parameters The initial codebook is a crucial parameter for an optimal quantization design. Therefore, it is chosen as the optimal codebook generated from a fixed-length LBG algorithm (where $\lambda = 0$), and the average codeword length assigned to each quantization level is $\log_2 N$ bits/vector. The choice of ϵ and $J^{(0)}$ is similar to the previous algorithms, and the determination of the Lagrange multiplier is evaluated as explained in Section 1.4.2.5.

The empty cell problem encountered in the LBG algorithm is solved by splitting the cell with the highest distortion into two cells, thereby reducing the distortion and maintaining the fixed number of levels constraint. In the CLG algorithm, on the other hand, the goal is to minimize the distortion with limited entropy, and splitting the cell with the highest distortion minimizes the distortion but may increase the entropy. Therefore, in CLG/ECVQ, the empty cells are removed directly, and the algorithm continues with a smaller codebook size [37].

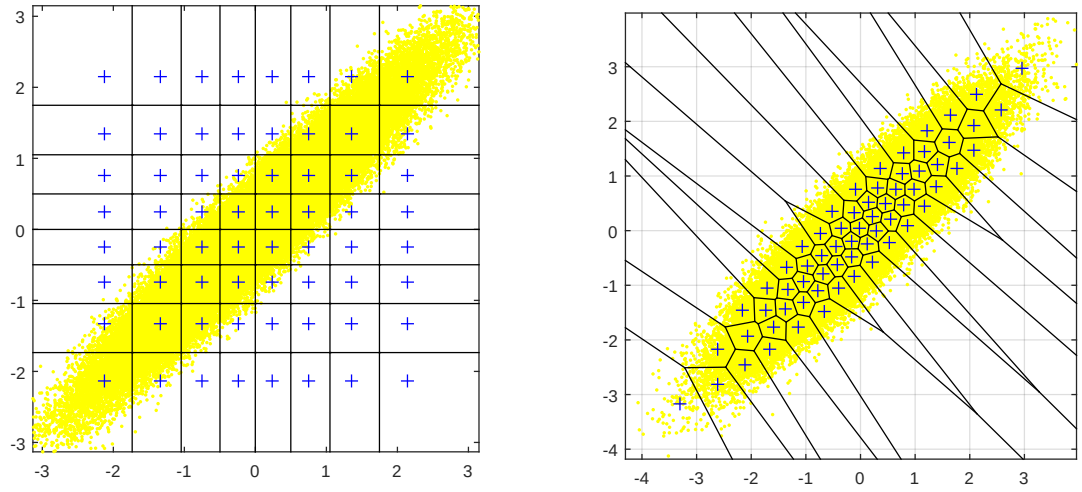
1.4.4.4 Vector quantization advantages

Vector quantization improves compression gain compared to scalar quantization. This performance improvement is shown to be achieved due to three different effects: *space-filling advantage* (corresponding to dimensionality), *shape advantage* (corresponding to the shape of the input signal PDF), and *memory advantage* (corresponding to linear and nonlinear dependencies between vector components). These advantages are briefly discussed below, while a deeper and more detailed analysis of the quantitative gain achieved by each attribute can be found in [41].

Memory advantage The memory advantage is the advantage that exploits the linear and nonlinear dependence between the components of the input vector in sources with memory. In VQ designs, the k-dimensional quantization cells are optimally partitioned and compacted to exploit these dependencies.

Fig. 1.15 shows a simple two-dimensional example illustrating the memory advantage of VQ, where the blue crosses represent the vectors of the codebook representation and the yellow dots represent the input training vectors associated with each quantization region. It shows the results of a fixed-length scalar Lloyd algorithm and a two-dimensional vector quantizer, i.e., LBG algorithm with $k = 2$ and $R = 3$ bits per sample for a Gauss-Markov process with a correlation factor $\rho = 0.9$. It can be seen that the VQ in Fig. 1.15b can efficiently exploit the dependence between the components of the TS, unlike the SQ in Fig. 1.15a, which only exploits the PDF of the one-dimensional TS without exploiting the dependence. Thus, for sources with strong statistical dependencies between samples, the memory gain is much larger than the shape and space filling gains.

Space-filling advantage The space-filling advantage refers to the gain obtained by the efficiency of convergence of the vector quantization algorithm to the most efficient polytope, i.e., the geometry of quantization domains, that can fill the $\mathbb{R}^k$ space and minimize the distortion. In the case of scalar quantization, it is clear that the k-dimensional space is partitioned into k-dimensional hyperrectangles, since they are the Cartesian products of one-dimensional intervals. However, this does not represent the most efficient packing of the $\mathbb{R}^k$ space.



(a) LA with k=1 and R=3 bits/sample.

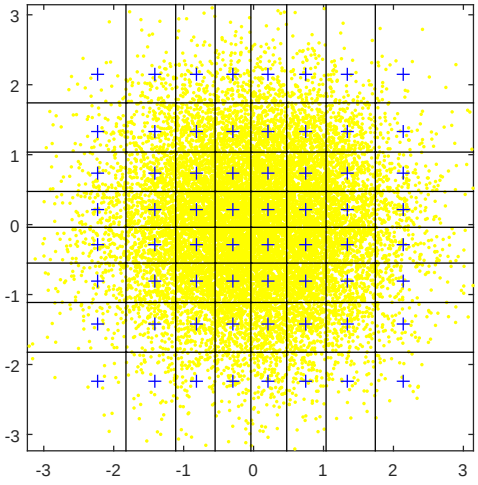
(b) GLA with k=2 and R=3 bits/sample.

Figure 1.15: Convergence of LA and 2-dimensional LBG algorithm with $R = 3$ bits per sample for a Gauss-Markov source with $\rho = 0.9$.

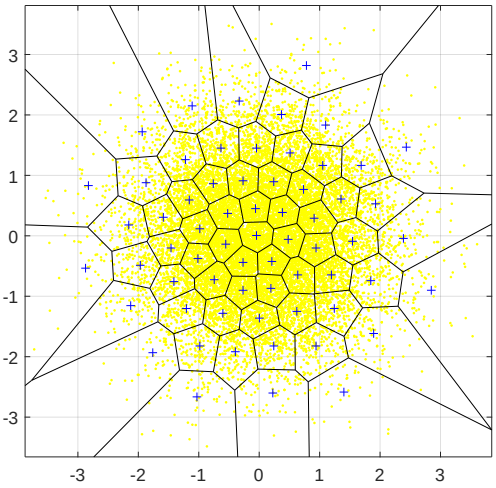
To illustrate this advantage, a scalar and a two-dimensional vector quantizer, i.e., the Lloyd and LBG algorithms with k=2, are used to generate the quantizer codebook for a Gaussian process with unit variance and a uniform i.i.d. process with $R = 3$ bits per scalar sample. Fig. 1.16 shows the Voronoi tessellation of the $\mathbb{R}^2$ space for the Gaussian process. Fig. 1.16a first shows the partitioning of the $\mathbb{R}^2$-space using the scalar LA, i.e., its Cartesian product in two-dimensional space, resulting in a rectangular shape of the quantization regions. Fig. 1.16b shows the partitioning of the $\mathbb{R}^2$ space after the convergence of the LBG algorithm. The quantization regions are approximated to hexagonal polytopes and the quantization distortion is reduced from 15.42 dB for LA to 14.67 dB with the LBG algorithm. Also for the uniform PDF in Fig. 1.17, the partitioning converges similarly to a hexagonal polytope and the quantization distortion decreases from 20.02 to 19.7 dB with vector quantization.

Consequently, the space-filling gain is independent of the PDF of the input source (either of its shape or of the statistical dependence between its components), it is always recognized as the fundamental advantage of vector quantization over scalar quantization. The space-filling gain is proved to be asymptotically limited to 1.53 dB or 0.25 bits per sample at high rates and vector dimensions approaching infinity. This bound corresponds exactly to the gab between the most complex scalar quantization design, i.e., scalar quantization with entropy constraint and the Shannon rate-distortion bound at high rates [41].

Shape advantage: The convergence of the GLA to the appropriate polytope form covering the $\mathbb{R}^k$ is the space-filling advantage and the densification of the highly dependent k-dimensional space with quantization cells based on the multi-dimensional PDF of the input signal is the memory advantage. Similarly, the shape advantage describes how well the algorithm converges the size of the quantization cells to fit the shape of the marginal PDF describing each component of the input vector, i.e., a non-uniform quantization

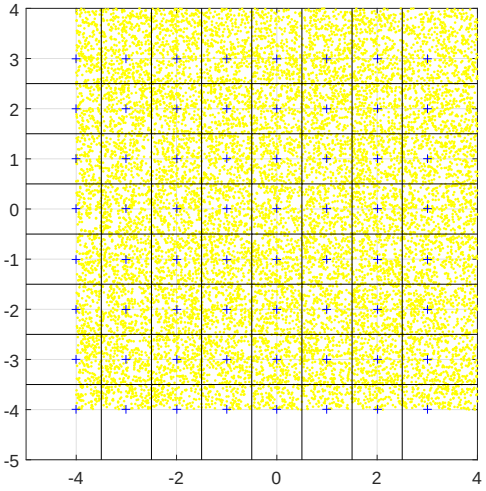


(a) LA with k=1 and R=3 bits/sample.

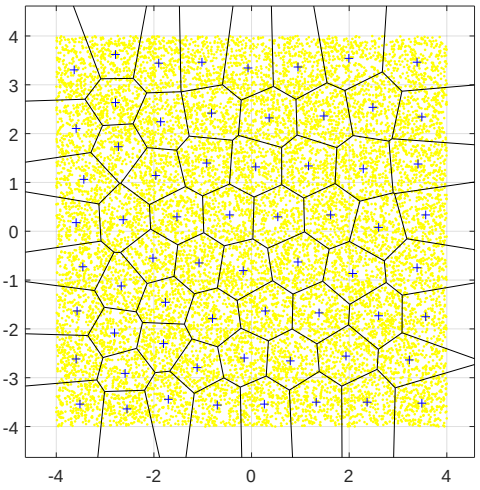


(b) GLA with k=2 and R=3 bits/sample.

Figure 1.16: Lloyd quantization and 2-dimensional LBG quantization wth R=3 bits per sample for a unit-variance Gaussian i.i.d process.



(a) LA with k=1 and R=3 bits/sample.



(b) GLA with k=2 and R=3 bits/sample.

Figure 1.17: Lloyd quantization and 2-dimensional LBG quantization wth R=3 bits per sample for a uniform i.i.d process.

1.4.5 Asymptotic distortion-rate functions for vector quantization

In this section, we discuss the high-resolution (asymptotic) distortion-rate functions for vector quantization. Again, these results were proved under the assumption of *high-rate quantization approximations/asymptotic quantization*, which requires a very high number of quantization levels and vector dimensions, i.e., $R, k \to \infty$. Thus, for a high number of quantization levels N, the quantization regions $\mathcal{R}_i$ are small enough to assume a uniform PDF for each region. Moreover, similar to scalar analysis, the distortion due to the overload region is small enough to be neglected, and only the granular region is considered.

First, a lower Shannon bound for the rate-distortion function for a k-dimensional input source and the MSE distortion measure is given as follows.

Proposition 1.4 ([45])**.** *Shannon lower bound for the rate-distortion function for a k-dimensional input source and MSE distortion is expressed as follows:*

$$R_{\mathrm{SLB-MD}}(D) = h(\boldsymbol{X}) + \frac{k}{2}\log_2\frac{1}{D} - \frac{k}{2}\log_2\left(\frac{2}{k}\left(V_k\Gamma\left(1+\frac{k}{2}\right)\right)^{\frac{2}{k}}\right), \tag{1.55}$$

or equivalently:

$$D_{\mathrm{SLB-MD}}(R) = \frac{k}{2}\left(V_k\Gamma\left(1+\frac{k}{2}\right)\right)^{\frac{2}{k}} 2^{\frac{2}{k}h(\mathbf{X})}2^{-\frac{2}{k}R}, \tag{1.56}$$

where $h(\mathbf{X})$ is the differential entropy of a multi-dimensional PDF, k is the number of components per vector, $\Gamma(.)$ is the gamma function and V_k denotes the volume of the k-dimensional unit sphere, i.e., $V_k = \frac{\pi^{\frac{k}{2}}}{\Gamma(\frac{k}{2}+1)}$.

The high-rate quantization theory provides tractable equations for the performance of vector quantizers for any vector dimension. For VQ, more complex results exist, thus, in this context, we briefly show the results and for deeply detailed analysis, please refer to [46, 47].

Zador proved in [46] that the asymptotic distortion-rate function of a vector quantizer with fixed-length codes is given as follows:

$$D_{\mathrm{GLA}}(R) = C(k)\left[\int_{-\infty}^{\infty}[\mathrm{f}_{\mathbf{X}}(\mathbf{x})]^{\frac{k}{k+2}}\,\mathrm{d}\mathbf{x}\right]^{\frac{k+2}{k}} N^{-\frac{2}{k}}, \tag{1.57}$$

where $\mathrm{f}_{\mathbf{X}}(\mathbf{x})$ is the multivariate PDF of the input signal and N is the number of quantization levels, equal to $N = 2^{kR}$ where R is the number of bits per scalar dimension. $C(k)$ is a term called the quantization coefficient, which is independent of the input signal PDF. The quantization coefficient depends on the optimal shape of the quantization regions covering the k-dimensional space. The difficulty in using the above expression lies in the quantization coefficient $C(k)$, which is known exactly only for very few cases [41, 48]. Based on this coefficient, an upper and lower bounds for the distortion-rate functions are given as follows.

Proposition 1.5 ([41, 46]). *Asymptotically, the distortion-rate formula for a vector quantizer with fixed-length codes is lower and upper bounded for the MSE distortion measure, respectively, as follows:*

$$D_{\mathrm{GLA}}(R) \geq \frac{k}{k+2}\left(\frac{2\pi^{\frac{k}{2}}}{k\Gamma(\frac{k}{2})}\right)^{-\frac{2}{k}}\left[\int_{-\infty}^{\infty}[\mathrm{f}_{\boldsymbol{X}}(\boldsymbol{x})]^{\frac{k}{k+2}}\,\mathrm{d}\boldsymbol{x}\right]^{\frac{k+2}{k}}N^{-\frac{2}{k}}, \tag{1.58}$$

and

$$D_{\mathrm{GLA}}(R) < \frac{2\Gamma(\frac{2}{k})}{k(k+2)}\left(\frac{2\pi^{\frac{k}{2}}}{k\Gamma(\frac{k}{2})}\right)^{-\frac{2}{k}}\left[\int_{-\infty}^{\infty}[\mathrm{f}_{\boldsymbol{X}}(\boldsymbol{x})]^{\frac{k}{k+2}}\,\mathrm{d}\boldsymbol{x}\right]^{\frac{k+2}{k}}N^{-\frac{2}{k}}. \tag{1.59}$$

Similar to scalar quantizers, the asymptotic distortion-rate function of a vector quantizer with variable-length codes, where the entropy and not the number of levels is bounded, is given as follows.

Proposition 1.6 ([46]). *Asymptotically, the distortion-rate formula for a vector quantizer with variable-length codes and MSE distortion measure is given as follows:*

$$D_{\mathrm{ECVQ}}(R) = C(k)2^{-\frac{2}{k}[H_{\min}-h(\boldsymbol{X})]}, \tag{1.60}$$

where $C(k)$ is the quantization coefficient that does not depend on the input signal PDF. $h(\boldsymbol{X})$ is the differential entropy of the multivariate PDF of the input signal and $H_{\min}$ is the entropy of the entropy-constrained quantization outputs.

1.4.6 Spectral domain redundancies removal

IFFT is used to generate OFDM signals. The smallest appropriate IFFT size is typically chosen according to the Nyquist criterion. In practice, however, the OFDM signal is transmitted in a broader spectrum than the Nyquist sampling rate, i.e., the OFDM signal is an oversampled signal. Oversampling is used to prevent out-of-band radiation from interfering with the useful signals in the OFDM system due to implementation complexity. Therefore, the OFDM signal is oversampled by a factor L, which is defined as follows:

$$L = \frac{f_s}{f_m}, \tag{1.61}$$

where f_s is the sampling rate of the OFDM signal and f_m is the system bandwidth. Table 1.2 shows the transmission bandwidth specifications for LTE and DVB standards. This can be exploited to compress the signal without degrading its quality by removing the redundant spectrum and leaving a signal with a lower sampling rate f_{ds} before the transmission over the fronthaul link.

Decimation is the process responsible for reducing the sampling rate of the OFDM signal. The process by which the OFDM signal is decimated by a rational factor is depicted in Fig. 1.18 and explained as follows. First, the input signal is upsampled by a factor m, where m zero-valued samples are inserted between the original samples to increase the sampling rate, mathematically represented as.

$$x_{\mathrm{up}}(i) \triangleq \begin{cases} x(i/m); & \textit{if} \quad \frac{i}{m} \in \mathbb{Z} \\ 0; & \text{otherwise,} \end{cases} \tag{1.62}$$

Table 1.2: Bandwidth specifications for LTE and DVB-T2 standards.

	LTE (downlink)		DVB-T2
Channel bandwidth (MHz)	10	20	8
Sampling frequency (MHz)	15.36	30.72	9.14
Occupied channel bandwidth (MHz)	9.015	18.015	7.61
FFT mode	$1K$	$2K$	$8K$
FFT size (N_{fft})	1024	2048	8192
Occupied subcarriers	601	1201	6817
Guard subcarriers	423	847	1375

where $x_{\text{up}}(i)$; $\forall i \in \{1, \cdots, N_{\text{fft}}\}$ is the upsampled signal. Then, the upsampled signal is filtered using a Finite Impulse Response (FIR) low-pass filter with a bandwidth limited to $[-f_{ds}/2, f_{ds}/2]$. Finally, the filtered signal is downsampled by a factor p, selecting every pth sample and discarding the remaining samples, mathematically represented as:

$$x_{\text{dec}}(i) \triangleq x_{\text{fil.}}(pi); \quad i \in \mathbb{Z}, \tag{1.63}$$

where $x_{\text{fil.}}(i)$ and $x_{\text{dec}}(i)$; $\forall i \in \{1, \cdots, N_{\text{fft}}\}$ are the filtered and decimated signals, respectively.

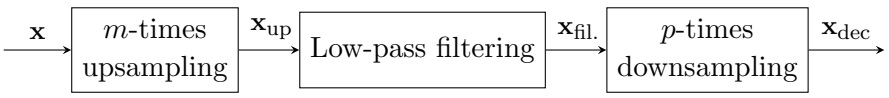


Figure 1.18: Removal redundancy block.

The role of the FIR filter is to act as an interpolation filter after the upsampling process and as an anti-aliasing filter before the downsampling process. Upsampling requires a low-pass filter after increasing the data rate to smooth the discontinuities created by adding the zeros, i.e., an interpolation filter. Downsampling, on the other hand, requires a low-pass filter before decimation to reduce the problem of high-frequency signal components caused by decimation, i.e., anti-aliasing filter. Therefore, both operations can be performed using a single FIR filter with the lower of the two cutoff frequencies. The decimation factor F is the key parameter for the performance evaluation and is defined as:

$$F = \frac{f_s}{f_{ds}} = \frac{p}{m} \geq 1. \tag{1.64}$$

The value of f_{ds} depends on the amount of redundant spectrum to be removed. It is important to note that the decimation factor F is limited to the oversampling factor L to preserve the useful data, i.e., $F \leq L$. The filter coefficients must be designed to ensure good amplitude and phase response in the passband, and the filter length N_w must be chosen to compromise between spectral shaping requirements and filter complexity. In practical implementation, specific parameters such as the upsampling factor, the downsampling factor, and the filter length vary according to different system specifications.

1.5 Conclusion

In this chapter, the basic concept of the C-RAN architecture was introduced. The advantages and limitations of such an architecture are then discussed and the fronthaul limitations faced by the C-RAN architecture were explained. Based on this, compression of the digitized IQ data before transmission over the capacity-constrained fronthaul link was presented as a potential solution to the fronthaul limitations. Then, the basic concept of OFDM system as the most commonly used multicarrier modulation scheme is presented. Also, an overview of the high PAPR problem that characterizes the OFDM signal during amplification is given. The main compression techniques proposed in the literature for data rate compression in the C-RAN architecture are described in detail in this chapter, especially quantization and entropy coding.

More specifically, we have discussed the Lloyd algorithm, which produces a quantizer with a fixed number of quantization levels N and a rate defined simply as $\log_2 N$ bits per sample, for a non-uniformly distributed signal. However, the fact that the quantization levels are not equally probable, leads to the use of an entropy encoder to reduce the rate of $\log_2 N$ to the entropy utilizing the PMF of quantization levels. Then, the idea was further improved to reformulate the optimization problem to minimize the distortion considering a constraint on the entropy rate, known as an entropy-constrained quantizer. Finally, the advantage of quantizing vectors instead of scalar samples using vector quantization is discussed.

In the next chapter, the state of the art of signal processing compression techniques of the digitized baseband IQ samples will be presented in order to reduce the data rate over the fronthaul link, which has been studied essentially in C-RAN context. Moreover, this analysis will be described to show the inadequacy of the studies performed for this level. Thus, the first objective of the thesis will be detailed and analyzed through different practical scenarios.

Chapter 2

Analysis of IQ samples compression techniques

2.1 Introduction

In this chapter, we discuss the state of the art for compressing complex baseband samples prior to transmission over the fronthaul link in the C-RAN architecture. Section 2.2 analyzes the main IQ data compression techniques exploiting temporal and spectral signal characteristics studied in this thesis. We consider the parameters for evaluating the compression scheme to achieve a tradeoff between the required signal quality and complexity while achieving an acceptable compression gain. Finally, we propose a simple, optimized uniform quantization scheme combined with entropy coding that achieves non-uniform quantization performance at much lower computational complexity by exploiting the temporal statistical properties of the signal in Section 2.4. Furthermore, we present a comparison between efficiently removing the statistical redundancy characterizing the oversampled OFDM signal through decimation and exploiting the correlation between the oversampled IQ OFDM samples through vector quantization, and analyze the tradeoff between these two techniques in terms of compression, signal quality, and computational complexity in Section 2.5.

2.2 State of the art in compression solutions for the C-RAN architecture

In recent years, studies have mainly focused on signal processing techniques for data reduction in the fronthaul link for the C-RAN architecture. However, most of these studies have been conducted considering LTE-like signals and based on their characteristics. In these studies, the compression techniques are integrated as separate functional blocks on both sides of the fronthaul link, i.e., in the BBU and the RRH. Fig. 2.1 depicts the proposed implementation of the IQ compression and decompression algorithms within the conventional architecture of the C-RAN in most of the recent studies. They rely mainly on exploiting the statistical properties of the OFDM signal to reduce the data rate on the fronthaul link.

At the recent state of the art, some authors propose to downsample the OFDM

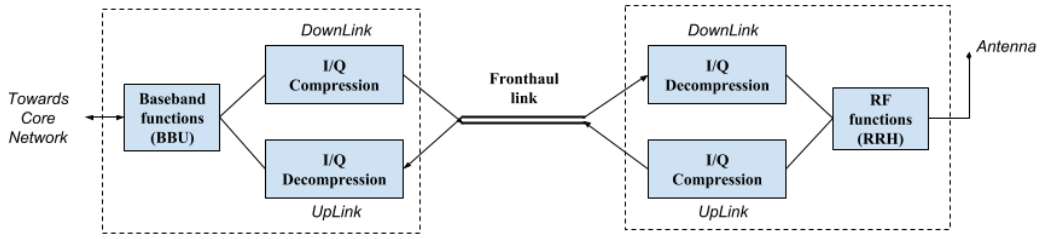


Figure 2.1: The integration of IQ compression and decompression blocks in a conventional C-RAN architecture [7].

signal to the sampling frequency required by the Shannon-Nyquist theorem, taking advantage of the oversampling of the OFDM signal. In addition, the guard interval of the OFDM symbol before transmission can be saved if it is assumed that the fronthaul channel does not generate multipath effects. Finally, it is assumed that the OFDM IQ samples at the output of the IFFT modulator are Gaussian-distributed thanks to the CLT. Therefore, the compression blocks, i.e. quantization and entropy coding, are optimized for a Gaussian distribution.

The authors in [7] propose to use the multi-rate filter previously explained in Section 1.4.6 to decimate the original OFDM signal by a rational factor to approach the Nyquist sampling rate. They illustrate their method using a 10 MHz LTE signal sampled at 15.36 MHz. Thus, the decimation method can save about one-third of the required transmission bandwidth. To deal with signals that have a large dynamic range due to small and large-scale fading, the authors propose a block scaling function. Typically, time-domain IQ samples are transmitted with Q_s bits per complex component. If all these samples are directly compressed to Q_q bits, low-amplitude signal components will be affected by high quantization noise.

Thus, block scaling is based on segmenting the IQ samples into blocks. In each block, a scaling factor is selected for the IQ samples within the block. This scaling factor is chosen as the largest absolute value within the block and each sample in the block is further scaled based on this scaling factor. However, the scaling factor must be transmitted over each block as an overhead. The scaling factor is quantized over the conventional sample resolution Q_s, i.e., 15 bits based on the CPRI protocol [18]. On the other hand, the scaled IQ samples of each block are quantized over Q_q bits to minimize the quantization error. Then, they propose a non-uniform quantization for the IQ samples based on a gradient-based algorithm. Although this technique has been shown to perform well, the authors have not been able to show that it is optimal, i.e., minimizes the mean square error of the quantization distortion.

The authors also study the case where there are multiple physical and/or logical transport connections between BBU and RRH. They use the possibility of assigning a different quality of service to each connection to improve the compression ratio. The idea was to transmit the information flow over different links with different quality of service requirements. They propose a scheme that consists in transmitting on the first link the compressed IQ samples with the scaling factor explained above. Meanwhile, the quantization error of the compressed samples is calculated with respect to the original IQ samples. Then, this error is scaled with a different scaling factor and quantized over a different number of bits and sent over the second link. This procedure

is repeated for the subsequent links.

The performance of this scheme is based on the idea that successful transmission over each link helps to improve the quality of the composite decompressed signal. Finally, the authors use the idea of introducing a pseudo-random noise with a different realization to the signal to be quantized on each considered link. This technique allows the quantization error on each link to be independent and more robust to transmission errors.

In [49, 50], the authors consider decimation, block scaling combined with vector quantization instead of scalar quantization to exploit temporal correlation between IQ samples and increase compression gain, but at the cost of higher computational complexity. The authors consider different vectorization schemes and initial codebook choices to optimally exploit the temporal correlation between samples and choose an initial codebook that does not lead to a local optimum. Finally, they use entropy coding after the vector quantizer to optimally assign different codeword lengths to each quantization level, taking advantage of the different probabilities of the codebook levels.

Moreover, in [51, 52], the authors explore linear predictive coding to investigate the temporal correlation between samples and solve the complexity problem posed by vector quantization. The idea can be simply illustrated as follows. They create a prediction of the current sample to be quantized using P_{sam} of previously quantized samples. The error between the original and predicted samples is then quantized, followed by entropy coding to reduce the average number of bits used to represent the error and transmitted through the fronthaul link. In the decoder, the error and the predicted signal are used to reconstruct the uncompressed signal in the same steps as in the encoder.

Finally, in [53], the authors propose a compression scheme based on the Discrete Cosine Transform (DCT). The DCT is a Fourier-dependent time-frequency transform characterized by its strong energy compaction property. Therefore, after DCT, the signal can be represented by coefficients in the frequency-domain, which can be divided into blocks with high and low frequency components. On this basis, the conventional compression methods, e.g., Lloyd quantization and Huffman coding, are used to quantize and encode the high-frequency components with a smaller number of bits and the low-frequency components with more bits. In general, quantization and entropy coding are the main data rate reduction techniques for the C-RAN architecture.

2.2.1 Preliminary results of the comparison between state of the art compression techniques: first contribution

An overview of the most recent techniques for reducing the data rate in the fronthaul link has been presented. In this study, we first investigate the well-known lossy and lossless compression methods that exploit the scalar and vector-based properties of the samples to find a solution that balances compression performance, complexity, and signal quality. We first propose to exploit the statistical signal characteristics by using a simple optimized uniform quantizer in combination with entropy coding, rather than developing a more complex non-uniform quantizer as it is common in literature. Our proposed approach results in significant compression gain and lower

computational complexity.

Moreover, the OFDM IQ baseband samples are correlated due to oversampling, as described in Section 1.4.6, leading to the idea of exploiting this statistical redundancy either by vector quantization or decimation. In this study, we investigate how to efficiently exploit the correlation between signal samples from the perspective of the compression objective. According to the wireless system specifications, we show how to achieve optimal performance in terms of compression, signal quality and complexity by (i) removing the OFDM signal oversampling overhead through decimation and (ii) exploiting these time-correlated IQ baseband signals through vector quantization. In contrast to the method proposed in [49], where decimation is followed by vector quantization, we rather compare both strategies, since the former eliminates the main memory advantage of the latter by reducing signal correlation. Then, a performance/complexity analysis is performed and discussed when the oversampling factor varies.

2.3 System model and simulation parameters

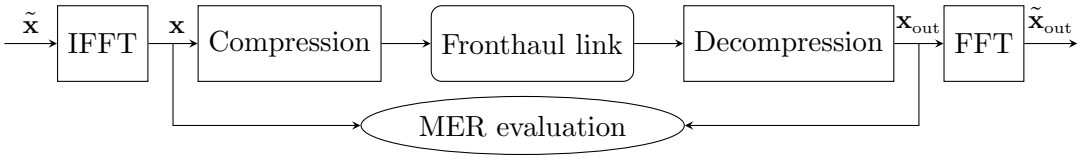


Figure 2.2: Downlink system model of a C-RAN architecture.

In Fig. 2.2, the downlink OFDM transmission chain considered in our analysis is shown. An OFDM system with N_{fft} subcarriers and M-QAM modulation is considered. The time-domain IQ samples are compressed before being transmitted over the fronthaul link. Let $\tilde{\mathbf{x}} = [\tilde{x}(0), \cdots, \tilde{x}(N_{\text{active}} - 1)] \in \mathbb{C}^{N_{\text{active}} \times 1}$ be the QAM data symbols in the frequency domain, modulating the subcarriers, where N_{active} is the number of loaded subcarriers and $\mathbf{x} \in \mathbb{C}^{N_{\text{fft}} \times 1}$ is the complex OFDM signal in the time-domain loaded with $N_{\text{null}} = N_{\text{fft}} - N_{\text{active}}$ empty subcarriers, i.e., guard subcarriers. Thus, $\mathbf{x}$ can be expressed as follows:

$$\mathbf{x} = \mathbf{F}^H \mathbf{M} \tilde{\mathbf{x}}, \tag{2.1}$$

where $\mathbf{F} \in \mathbb{C}^{N_{\text{fft}} \times N_{\text{fft}}}$ is the DFT matrix and $\mathbf{M} \in \mathbb{C}^{N_{\text{fft}} \times N_{\text{active}}}$ is a mapping matrix decomposed into $\frac{N_{\text{null}}}{2} \times N_{\text{active}}$ null matrices and identity matrix of size N_{active} as follows:

$$\mathbf{M} = \begin{bmatrix} \mathbf{0}_{\frac{N_{\text{null}}}{2} \times N_{\text{active}}} \\ \mathbf{I}_{N_{\text{active}}} \\ \mathbf{0}_{\frac{N_{\text{null}}}{2} \times N_{\text{active}}} \end{bmatrix},$$

where $\mathbf{0}_{n \times p}$ and $\mathbf{I}_n$ is the $n \times p$ null matrix and the square identity matrix of size n, respectively. Finally, $\mathbf{x}_{\text{out}} \in \mathbb{C}^{N_{\text{fft}} \times 1}$ is the decompressed output vector and $\tilde{\mathbf{x}}_{\text{out}}$ is its frequency-domain representation. We consider an 8-MHz DVB-T2 frame structure in our simulations, whose parameters are listed in Table 2.1 unless otherwise stated.

Table 2.1: Simulation parameters for the analysis of the compression of an OFDM signal

Metrics	Values
Channel bandwidth in MHz	8
Sampling frequency in MHz	9.14
FFT size	$32K$
Number of data subcarriers	26265
Number of guard subcarriers	5503
Number of OFDM symbols	1000
Constellation	64-QAM
No. of bits for representation of IQ samples (R_0)	15 bits per each real and imaginary component [18].

The number of guard subcarriers within the OFDM symbol determines the oversampling factor that characterizes the signal. The use of the oversampling factor as the main parameter in our analysis is essential in view of the different wireless system specifications. For example, in LTE cellular technology, LTE signals are roughly upsampled by a factor $L = 1.5$, as shown in Table 1.2. This factor can either be reduced to stick to the Shannon-Nyquist requirement before transmission over the fronthaul link, or the statistical redundancy between samples is exploited with either VQ or linear predictive coding. In other standards, such as DVB-T2 broadcasting technology, where almost all subcarriers are loaded, i.e., the signal is sampled slightly larger than required by the Nyquist criterion, about $L = 1.16$, as shown in Table 1.2. Therefore, the output signal should not have a large redundancy between samples to be exploited.

2.4 Compression based on scalar quantization analysis

We investigate the behavior of the scalar quantization described in Section 1.4.2. First, we analyze the optimized UQ by dynamically adjusting the level of the input signal, represented by the loading factor γ, to achieve the tradeoff between granular and overload distortions explained in Section 1.4.2.2. We derive the granular and the overload distortion, denoted by $D_{\text{gran.}}$ and D_{overload}, respectively, composed the quantization distortion according to (1.14) as a function of the loading factor γ, as shown in the following proposition.

Proposition 2.1. *The granular distortion $D_{\text{gran.}}$ can be expressed asymptotically as $R \to \infty$ as follows [6]:*

$$D_{\text{gran.}} = \frac{\gamma^2 \sigma^2}{3N^2}, \tag{2.2}$$

and a closed-form expression for the overload distortion D_{overload} is derived for a zero-mean Gaussian-distributed signal in terms of γ as follows:

$$D_{\text{overload}} = \left[\sigma^2(1+\gamma^2)Q\left(\gamma\left(1-\frac{1}{N}\right)\right) - \frac{\gamma\sigma^2}{\sqrt{2\pi}}\left(1+\frac{1}{N}\right)e^{-\frac{\gamma^2}{2}\left(1-\frac{1}{N}\right)^2}\right], \tag{2.3}$$

where $Q(u) = (1/\sqrt{2\pi}) \int_u^\infty exp(-v^2/2)\,\mathrm{d}v$ *is the Q-function.*

According to the overload distortion in (2.3) in Proposition 2.1, which consists of a constant and a monotonically decreasing function, the overload distortion decreases as γ increases. In contrast, the granular distortion in (2.2) increases with γ. Thus, the optimal γ that balances the overload and granular distortions is

$$\gamma_{\text{opt.}} = \arg\min_{\gamma} D. \tag{2.4}$$

Unfortunately, a closed form solution cannot be reached and a numerical search is performed. Therefore, the choice of the optimal loading factor $\gamma_{\text{opt.}}$ is achieved numerically for each quantization resolution, as shown in Fig. 2.3 clearly.

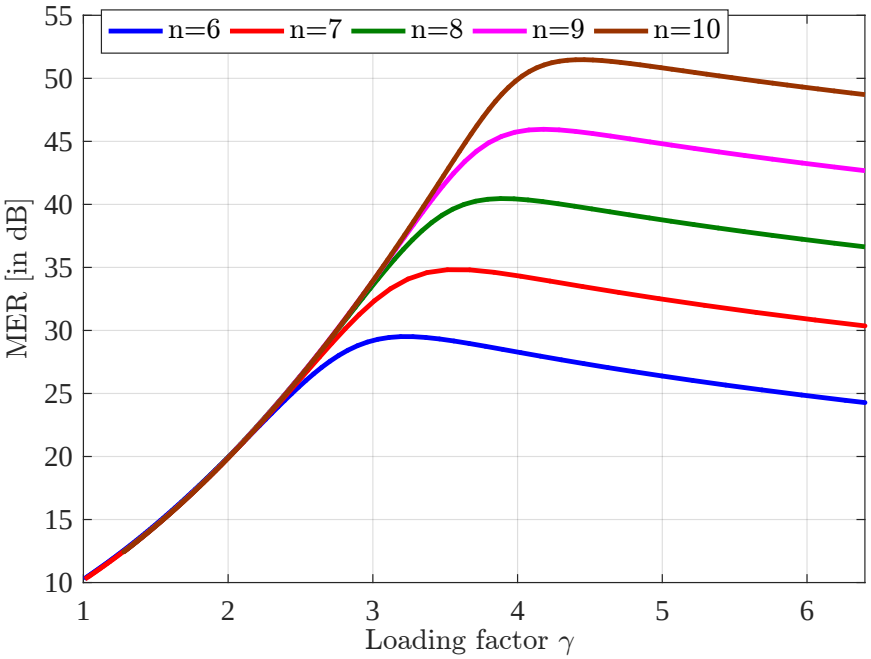


Figure 2.3: MER vs loading factor γ for different resolutions.

Fig. 2.3 shows the MER as a function of the loading factor γ fo different resolutions. We observe that the MER admit a unique value for the loading factor $\gamma_{\text{opt.}}$, for which we obtain the highest MER at which the best tradeoff between granular and overload distortions is achieved. For a given R, the graph can be divided into three main regions:

- Region 1: driven by low values of the loading factor, i.e., high values of the scaling factor), where the characteristic is slightly linear with respect to γ. This region is characterized by significant overload noise, which decreases exponentially inversely to γ, as shown by the derived overload distortion in (2.3). Meanwhile, the granular noise increases as the dynamic range of the input signal is reduced. However, the power of the granular noise is still lower than that of the overload distortion, i.e., the overload distortion is still significant, which explains the observed linear increase of MER in this range.

- Region 2 (optimum region): where the MER increases to reach its maximum value at the optimal loading factor $\gamma_{\text{opt.}}$, which trades between the granular and overload distortions. Then, the MER starts to decrease when the granular noise becomes significant. Thus, the optimized uniform quantization curve is obtained by inserting a scaling block before the uniform quantizer corresponding to the simulated $\gamma_{\text{opt.}}$ to change the dynamic range of the OFDM input signal before quantization.

- Region 3: MER gradually decreases with γ. The power of the signal is lower because the scaling factors are lower in this range. The granular distortion becomes more significant than the overload distortion, resulting in a lower MER. On the other hand, there is no more overload distortion, which leads to the following relevant consequence: the measure of MER for two consecutive numbers of bits is equal to 6.02 dB, which satisfies the famous asymptotic 6-dB rule.

Fig. 2.4 shows the MER as a function of resolution R for different scalar quantization designs, i.e., uniform quantizer in addition to fixed-length and entropy-constrained non-uniform quantizers optimized for a Gaussian-based input signal. First of all, for all quantization designs, the higher the resolution R, the higher the achievable MER. This is explained by the fact that larger values of R extend the range of the quantizer. This allows larger values of signal amplitudes to be quantized with lower quantization error, which in turn increases the MER.

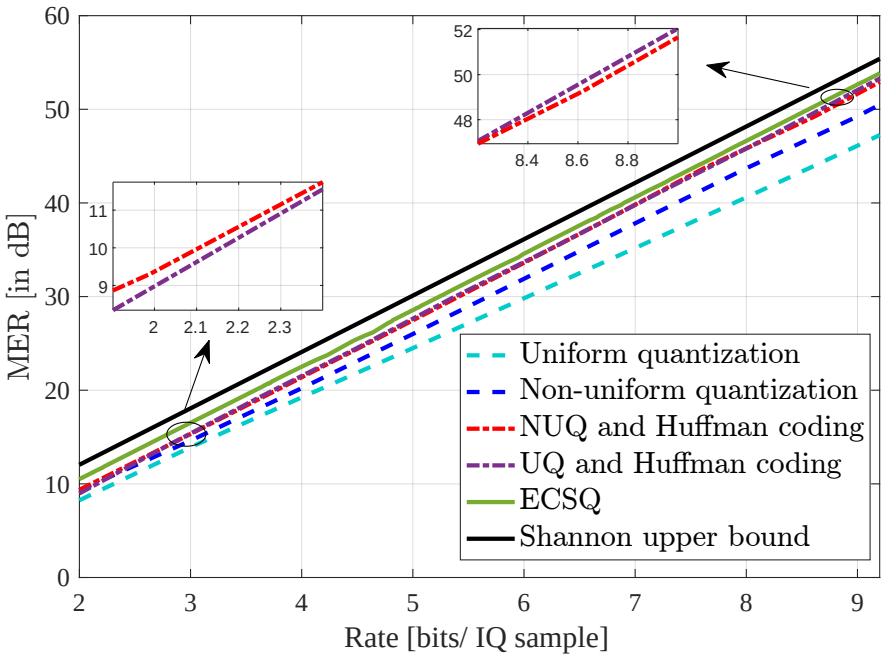


Figure 2.4: MER vs resolution R for an i.i.d. complex Gaussian distribution for different scalar quantization schemes.

The solid black line in Fig. 2.4 represents Shannon's upper bound for an i.i.d source derived for a Gaussian distribution with zero mean shown in (1.33). The dashed cyan and blue curves represent the optimized uniform quantizer and the Lloyd-based

non-uniform quantizer, respectively, represented with fixed-length codewords of $\log_2 N$ bits/sample. The Lloyd-based quantizer, represented by the dashed blue curve, is obtained using the closed-form formulas for the optimal quantization levels and the average number of bits allocated to each quantization level in (1.17) and (1.21), respectively, which are derived as follows for a Gaussian-distributed signal with zero mean:

$$y_i = \frac{\sigma}{\sqrt{2\pi}} \frac{\left(e^{-\frac{t_{i-1}^2}{2\sigma^2}} - e^{\frac{-t_i^2}{2\sigma^2}} \right)}{\mathrm{Q}\left(\frac{t_{i-1}}{\sigma}\right) - \mathrm{Q}\left(\frac{t_i}{\sigma}\right)}, \tag{2.5}$$

$$l_{y_i} = -\log_2 \left(Q\left(\frac{t_{i-1}}{\sigma}\right) - Q\left(\frac{t_i}{\sigma}\right) \right). \tag{2.6}$$

In summary, it can be seen from Fig. 2.4 that the non-uniform quantizer has a clear advantage over the optimized uniform quantizer and the difference becomes larger at higher resolutions.

As explained earlier, the use of Huffman coding at the output of both the uniform and non-uniform quantizers allows for greater compression by assigning variable-length codewords to different quantization levels according to their probability of occurrence. As can be seen from the red and violet dashed-dotted lines in Fig. 2.4, the UQ and NUQ algorithms are nearly consistent when using the entropy coder. As shown in the zoomed regions in Fig. 2.4, NUQ slightly outperforms UQ at low resolutions by about 0.4 dB at 2 bits/sample. However, UQ improves by almost 0.45 dB at 9 bits as the number of bits per sample increases.

This behavior can be explained as follows. The optimal loading factor $\gamma_{\text{opt.}}$ increases with increasing number of bits per sample, as shown in (1.15) and illustrated in Fig. 2.3. According to (1.15), a high γ leads to a low scaling factor, so the scaled signal amplitudes at the input of the quantizer decrease and are quantized to values close to zero. The Huffman algorithm encodes these levels with short codeword lengths due to their higher probability under the Gaussian distribution. Thus, the optimized uniform quantizer in combination with entropy coding exploits the statistical distribution of the signal as in non-uniform quantization, but with lower computational complexity.

This can be illustrated by an analysis of the asymptotic gap between the entropy of the uniform and the non-uniform quantization outputs given by the asymptotic distortion-rate functions stated for the Gaussian distribution in (1.36) and (1.41) , respectively. This gap is expressed as follows:

$$\lim_{R\to\infty} R_{UQ} - R_{LQ} = \frac{1}{2}\log_2\left(\frac{3\sqrt{3}\pi}{2\gamma^2}\right). \tag{2.7}$$

If we take the optimal loading factors $\gamma_{\text{opt.}}$ from Fig. 2.3 for different resolutions into (2.7), we obtain the values of the asymptotic gap between the two techniques, which are summarized in Table 2.2. Table 2.2 shows the improvement of UQ over NUQ followed by entropy coding in both cases by increasing the number of resolution bits, i.e., a larger compression gain can be obtained with UQ for the same MER.

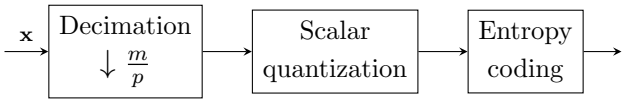
Finally, the solid green curve in Fig. 2.4 represents the ECSQ performance. It is about 1.6 dB lower than the Shannon upper limit. On the other hand, the MER obtained with UQ followed by Huffman coding is about 2.4 dB below the upper bound.

Table 2.2: Asymptotic gap between UQ and NUQ with Huffman coding

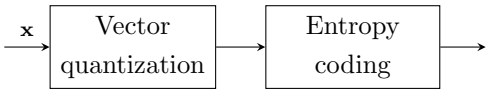
Resolution bits	2	3	4	5	6	7	8
$\gamma_{\text{opt.}}$	1.80	2.33	2.76	3.21	3.52	3.89	4.09
$R_{UQ} - R_{LQ}$	0.66	0.29	0.05	-0.17	-0.30	-0.43	-0.52

Thus, depending on the system requirements, we can decide whether the gain of about 1 dB achieved with ECSQ over UQ and Huffman code is worth the additional complexity of the ECSQ implementation.

2.5 Numerical analysis of the comparison between decimation and VQ when exploiting the redundancy between the OFDM IQ samples



(a) Decimation followed by scalar quantization.



(b) Vector quantization.

Figure 2.5: Compression system block diagram.

In this section, we compare the compression performance between vector quantization and decimation when exploiting the statistical redundancy that characterizes the OFDM signal. Figs. 2.5a and 2.5b illustrate both compression schemes. In Fig. 2.5a, an optimized uniform scalar quantizer with acceptable complexity is implemented, thanks to its reasonable performance shown in the previous section. For the vector quantization in Fig. 2.5b, a fixed-length vector quantizer is implemented using the LBG algorithm.

The compression ratio is calculated as $\text{CR} = \frac{R_0}{\frac{m}{p} R_{\text{sq}}}$ and $\text{CR} = \frac{R_0}{R_{\text{vq}}}$ for decimation plus scalar UQ and VQ, respectively. Where R_{sq} and R_{vq} are the number of bits per sample calculated from $\log_2 N$ and $\frac{\log_2 N}{k}$ for scalar and vector quantizations, respectively.

As the oversampling factor L increases, the number of guard zero-padded subcarriers $N_{\text{null}} = N_{\text{fft}} - N_{\text{active}}$ within an OFDM symbol increases. Therefore, the ratio $\frac{N_{\text{null}}}{N_{\text{fft}}}$ is considered as an oversampling parameter and the compression ratio increases with the increase of the oversampling factor. In the case shown in Fig. 2.5a, higher oversampling values lead to more redundancy in the signal. Therefore, a higher compression ratio can be achieved by increasing the decimation factor F, which is limited by the oversampling factor, as shown in Section 1.4.6. For the vector quantization in

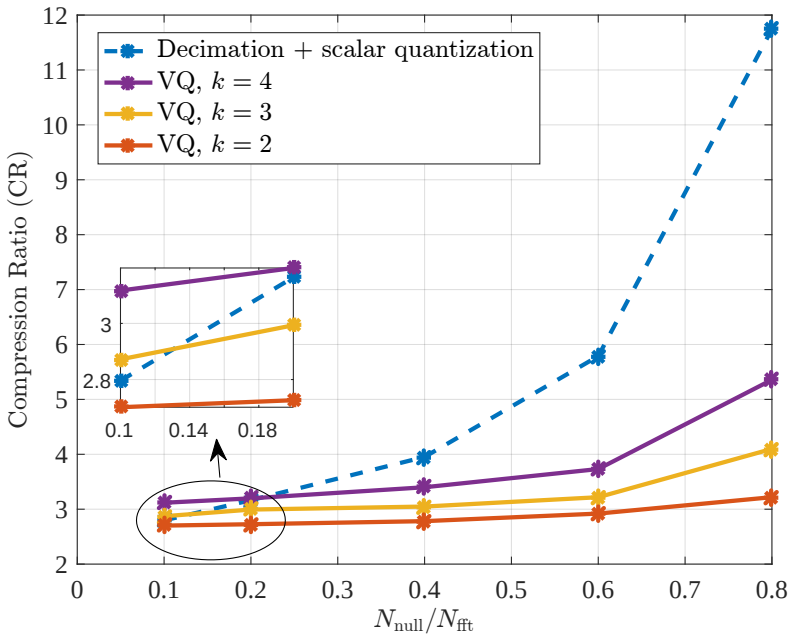


Figure 2.6: Compression ratio vs oversampling factor L of two techniques: decimation plus uniform scalar quantizer and fixed-length vector quantization.

Fig. 2.5b, CR increases thanks to the increase of the temporal correlation between samples with the oversampling factor. The compression gain is less rapid than in the first solution for a given vector length k. However, by increasing the vector dimension k, additional compression gain can be obtained due to the quantization memory advantage.

In Fig. 2.6, the compression ratio is shown as a function of the oversampling factor represented by $\frac{N_{\text{null}}}{N_{\text{fft}}}$. We assume a fixed MER = 30 dB to compare the compression achieved by the two schemes. For decimation, we implement a Hamming-windowed FIR filter. Table 2.3 shows the chosen decimation parameters (m, p) and filter length N_w, which yield an acceptable tradeoff between signal quality and complexity for different oversampling factors.

The simulation results show that the gain obtained by decimation at low oversampling factor is very small due to the low degrees of freedom in this case, and VQ with $k = 3$ or 4 performs better. On the other hand, the decimation gain increases significantly by increasing the oversampling factor and reaches about 6 times higher compression gain than with the vector quantizer with $k = 3$ at $L = 0.8$.

Table 2.3: Hamming window FIR filter specifications

N_{null}/N_f	0.1	0.2	0.4	0.6	0.8
Decimation factor (F)	16/15	6/5	3/2	11/5	9/2
Filter length (N_w)	730	470	150	310	310

In addition to compression performance, the computational complexity of the two

algorithms is also an important feature to consider. The complexity of the first algorithm is $\mathcal{O}(N_w \log N_w + 2^{R_{sq}})$, where N_w is the filter length and R_{sq} is the resolution of the quantizer. The complexity of VQ is $\mathcal{O}(2^{KR_{vq}})$, where k is the vector dimension and R_{vq} is the number of bits for each sample. VQ has an exponential complexity in terms of vector length and resolution, while the decimation and UQ techniques have a sup-linear complexity in terms of filter length but an exponential one in terms of quantization resolution. Therefore, the above numerical analysis can be used to select the most appropriate compression technique for a given application. For illustration, without loss of generality, DVB-T2 or ATSC3.0 broadcasting technologies have empty subcarriers in approximately $\sim 4/25$ of their spectrum. Therefore, to achieve more compression, a VQ with higher vector lengths is preferable, given its complexity. While an LTE-like system has about $\sim 2/5$ empty subcarriers, decimation with UQ can achieve an acceptable compression gain at 30 dB MER with acceptable complexity according to Table 2.3.

2.6 Conclusion

This chapter presents the state of the art of known lossy and lossless compression techniques that exploit the scalar-based and vector-based characteristics of IQ samples for data rate reduction in the fronthaul link for the C-RAN architecture. More specifically, the main compression algorithms are described in detail: spectral domain redundancies removal, i.e., decimation to remove the statistical redundancy of the oversampled OFDM signal, quantization and entropy coding as the main compression blocks to explore the statistical characteristics of the signal. In addition, vector quantization was investigated, which exploits time-correlated samples, which is not possible with scalar quantization. In this context, we show the inadequacy of the state of the art studies and highlight the first objective of this thesis.

Finally, we have shown through our analysis that a simple optimized scalar uniform quantization algorithm coupled with entropy coding over the resulting scalar codebook efficiently exploits the non-uniform statistical distribution of the OFDM signal. This proposed strategy achieves a significant compression gain, about 2.4 dB below the idealized scalar quantizer, with much lower complexity compared to the non-uniform quantization approach. Second, we have provided a comparison between decimation and vector quantization by exploiting the correlations between the IQ samples. We have shown that there is a smooth tradeoff between the required signal quality, compression performance and system complexity based on the choice of appropriate practical parameter values.

Chapter 3

Compression of PAPR-reduced OFDM IQ samples

3.1 Introduction

In the previous chapter, the compression of the time-domain IQ samples was analyzed before transmission over the fronthaul link in the C-RAN architecture. These analyzes were performed for the OFDM multicarrier waveform at the output of the IFFT modulator, which follows a quasi-Gaussian distribution thanks to the CLT. Thus, the conventional Gaussian-based optimized quantizer can be used as the optimal solution for the compression problem

However, the transmitted OFDM signal is characterized by a high PAPR, which is one of the main problems of the OFDM system. Many techniques have been presented in the literature to reduce the PAPR. As mentioned earlier, in a fully centralized solution of the C-RAN architecture, PAPR reduction techniques are regularly used in BBU to reduce the fluctuations of the OFDM IQ samples, which changes their distribution. This fact makes the Gaussian quantizer suboptimal.

Therefore, in this chapter we study the impact of reducing the PAPR of the OFDM signal before compression in the C-RAN architecture through rate-distortion analysis. We consider clipping and tone reservation PAPR reduction algorithms presented in Section 3.3. The former is the simplest approach to PAPR reduction, while the latter is one of the most effective algorithms used in broadcasting standards such as DVB-T2 and ATSC 3.0.

First, the distribution of PAPR-reduced OFDM IQ samples is derived and validated considering clipping and QCQP-based TR techniques in Section 3.4. The obtained IQ distribution is further exploited to optimize the compression blocks, i.e., quantization and entropy coding, and to obtain an asymptotic expression for the MER that reflects the joint effect of PAPR reduction and quantization operations in Section 3.5. Finally, in Section 3.6, we investigate the robustness of a Gaussian-optimized quantizer to a change in the signal distribution at its input due to the application of a PAPR reduction technique. Based on these analyzes, we quantify the performance gain in MER, when using a quantizer optimized with respect to the true distribution of PAPR-reduced IQ samples based on various tuning factors of the clipping and TR-QCQP algorithms compared to the Gaussian quantizer.

3.2 System model

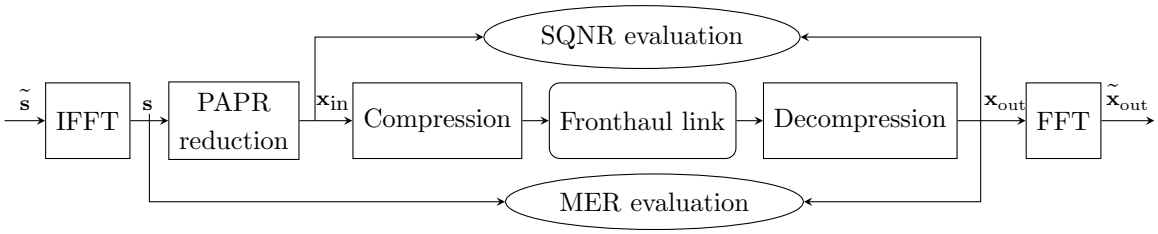


Figure 3.1: Downlink system model of a C-RAN architecture considering the PAPR reduction of the OFDM signal.

Let us consider the downlink OFDM transmission chain depicted in Fig. 3.1. A PAPR reduction technique is applied to the OFDM signal at the output of the IFFT modulator and then compressed before being transmitted over the fronthaul link. In the frequency domain, let the original OFDM signal be $\tilde{\mathbf{s}} = [\tilde{s}(0), \cdots, \tilde{s}(N_{\text{fft}} - 1)] \in \mathbb{C}^{N_{\text{fft}} \times 1}$ of size N_{fft}.

Each of the N_{fft} subcarriers is loaded with M-QAM modulated symbols. In the time domain, $\mathbf{s}$, $\mathbf{x}_{\text{in}}$, and $\mathbf{x}_{\text{out}} \in \mathbb{C}^{N_{\text{fft}} \times 1}$, represent respectively, the original complex OFDM signal, the PAPR-reduced IQ samples before compression, and after decompression. In our analysis, we consider for compression the scalar-based non-uniform quantizer with fixed-length codes, i.e., the Lloyd quantizer, followed by the entropy coding technique described in Sections 1.4.2.3 and 1.4.2.4, respectively.

3.3 PAPR reduction techniques

Many works proposed in the literature have studied the problem of PAPR characterizing the OFDM signal. A lot of them are applied to the OFDM baseband signal in the time domain, e.g., [28, and references therein]. PAPR reduction techniques are generally classified into three categories: Signal distortion techniques, probabilistic techniques, and signal addition techniques.

Signal distortion techniques are performed for the OFDM baseband signal in the time-domain. They are characterized by their simplicity and great performance in reducing the PAPR of the signal. However, these techniques introduce in-band and out-band distortions. Signal distortion techniques include clipping [54], windowing [55], and companding [56]. The simplicity of these techniques therefore comes at the cost of degrading the end-to-end performance of the signal, which limits their practical application.

In the probabilistic techniques, different copies of the original OFDM baseband signal are made by changing the amplitude, phase, or position of the subcarriers. Then, the copy with the minimum PAPR is selected and transmitted. Probabilistic methods include selective mapping and partial transmit sequence [57]. These methods have the advantage of being distortionless and providing significant PAPR reduction. On the other hand, they require that side information be transmitted to the receiver to indicate

which sequence was chosen for transmission. This reduces spectral efficiency and increases computational complexity, especially as the number of subcarriers increases.

Finally, the addition signal techniques consist in adding an additional signal $\mathbf{c}$ to the signal of interest $\mathbf{s}$ in order to reduce its PAPR. These techniques are characterized by a good PAPR reduction performance without introducing signal distortion or requiring the receiver to know secondary information. TR [58], tone injection [59], and Active Constellation Extension (ACE) [60] are examples of this kind of techniques. In our study, we are interested in PAPR reduction techniques using clipping and tone reservation. Therefore, they are explained in detail with their main properties.

3.3.1 Clipping

Clipping is the simplest PAPR reduction technique. It consists in clipping the high amplitudes of the signal to a certain threshold $V_{\max}$ without affecting the phase of the signal. It is a widely used PAPR reduction technique that provides a high reduction gain. The clipped time-domain signal $x_{\text{in}}^{\text{clip}}(k)$ is [54]

$$x_{\text{in}}^{\text{clip}}(k) = \begin{cases} s(k); & if \ \ \|s(k)\| < V_{\max} \\ V_{\max} e^{j\theta_s}; & if \ \ \|s(k)\| \geq V_{\max}. \end{cases} \tag{3.1}$$

Despite the simplicity of clipping, it introduces potentially high in-band and out-of-band distortions to the signal. Out-of-band distortion can be reduced by filtering the clipped signal. However, this comes at the cost of peak growth, i.e., the signal exceeds the clipping threshold $V_{\max}$ after clipping and filtering.

This can be solved by an iterative process between clipping and filtering operations until a certain clipping threshold is reached, a.k.a., clipping and filtering technique for PAPR reduction [61]. Finally, in-band distortion caused by clipping remains a major problem limiting the practical implementation of the clipping PAPR reduction scheme.

3.3.2 Tone reservation

Tone reservation is considered as one of the most efficient distortionless PAPR reduction techniques. It is an addition signal technique presented in [58] and is considered as the most efficient PAPR reduction technique without distortion. It is based on reserving a subset of subcarriers, named the Peak Reserved Tones (PRTs), to cancel the large peaks of the signal. These tones carry no useful information and are evaluated to reduce the PAPR of the OFDM signal in the time-domain. Due to the orthogonality between the subcarriers, TR is a distortionless PAPR reduction technique since the PRTs are ignored during reception. The TR technique has been adopted by several standards, e.g., DVB-T2 and ATSC 3.0.

3.3.2.1 Principle of the TR technique

The TR technique can be seen in the time-domain as an addition of a peak cancellation signal $\mathbf{c} \in \mathbb{C}^{N_{\text{fft}} \times 1}$ orthogonal to the original OFDM signal $\mathbf{s}$, as shown in Fig. 3.2. Thus, the tone-reserved PAPR-reduced signal $\mathbf{x}_{\text{in}}^{\text{TR}}$ is expressed as follows:

$$x_{\text{in}}^{\text{TR}}(k) = s(k) + c(k); \quad \forall k = [0, \ldots, N_{\text{fft}} - 1]. \tag{3.2}$$

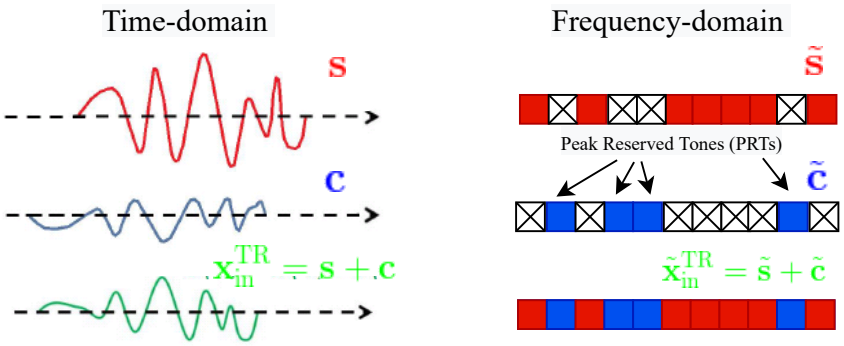


Figure 3.2: Block diagram of the tone reservation technique.

Let $\tilde{\mathbf{c}}$ be the frequency-domain vector consisting of N_{fft} subcarriers composed of υ tones reserved for peak reduction. These tones belong to the subset $\mathcal{B}$ such that $\mathcal{B} = \{k \in [0, \ldots, N_{\text{fft}} - 1] : \tilde{s}(k) = 0\}$ and $\mathcal{B}^c = \{k \in [0, \ldots, N_{\text{fft}} - 1] : \tilde{c}(k) = 0\}$. Thus, $\tilde{\mathbf{s}}$ and $\tilde{\mathbf{c}}$ lie in disjoint frequency subspaces, which ensures that the TR technique is a distortionless process.

The aim of the TR scheme is to compute the optimal values of the PRTs in the peak cancellation signal $\mathbf{c}$ for each data vector $\mathbf{s}$ such that there is a tradeoff between the efficiency of PAPR reduction and computational complexity. Therefore, the problem can be mathematically structured to find $\mathbf{c}$ that minimizes the maximum peak values of the resulting OFDM signal $\mathbf{s}$, i.e.,

$$\underset{c(n), n\in\beta}{\arg\min} \quad \|s(k) + \mathcal{F}^{-1}\{\tilde{c}(k)\}\|_{\infty}^{2}, \tag{3.3}$$

where $\mathcal{F}^{-1}\{.\}$ is the inverse Fourier transform operation and $\|\cdot\|_{\infty}$ is the infinite norm.

3.3.2.2 Parameters of the TR technique

The PAPR reduction efficiency of the TR algorithm is evaluated by two main parameters, namely the percentage of subcarriers dedicated to PAPR reduction out of the total number of subcarriers, denoted N_{PRT}, and the power allocated to the PRTs. With a constant total transmit power, the power allocated to the PRTs does not contribute to the transmission of useful data and at the same time reduces the power allocated to the carriers transmitting the data. Typically, standards such as DVB-T2 add a peak power constraint to the PRTs, denoted P_{PRT}, to limit the power dedicated to PAPR reduction compared to the power of the data subcarrier, denoted P_{data}.

Let P_{PC} be the power control in dB, which is the difference between the maximum PRT power and the average data subcarrier power. Thus, the power constraint condition can be expressed mathematically as [62]:

$$\max_{n\in\beta} \quad \|\tilde{c}(n)\|_{\infty}^{2} \leq \varsigma . P_{\text{data}}, \tag{3.4}$$

where $\varsigma = 10^{\frac{P_{\mathrm{PC}}}{10}}$ and $P_{\mathrm{PRT}} = \varsigma.P_{\mathrm{data}}$.

3.3.2.3 Percentage and location of tone-reserved subcarriers

As shown earlier, the reserved subcarriers for reducing the PAPR of the OFDM signal carry no useful information. Therefore, in order not to significantly affect the throughput of the system, the percentage of the total number of subcarriers reserved for PAPR reduction must be small. The positions of the reserved subcarriers are usually characterized by being uniform and distributed over the N_{fft} subcarriers to exploit all degrees of freedom.

For example, in the DVB-T2 broadcasting standard, the percentage of dedicated PRTs subcarriers is $N_{\mathrm{PRT}} = 1\%$. Moreover, the position of the reserved subcarriers is fixed and defined in the standard (cf. Appendix A in [63]). Without loss of generality, Table 3.1 summarizes the number of total subcarriers N_{fft}, active subcarriers N_{active}, data subcarriers N_{data} and reserved subcarriers υ for the different FFT modes used in our analysis.

Table 3.1: Number of OFDM subcarriers, active subcarriers and peak reserved subcarriers for different modes of operation in the DVB-T2 frame structure [3]

mode	N_{fft}	N_{active}	N_{data}	υ
$1K$	1024	853	843	10
$2K$	2048	1705	1687	18
$4K$	4096	3409	3373	36
$8K$	8192	6817	6745	72

3.3.2.4 Algorithms of the TR technique - optimal QCQP algorithm

The TR was presented by Tellado in [62]. As illustrated, the idea is to load a subset of the reserved tones with complex values and add them to the original signal, resulting in a transmitted time-domain signal with reduced PAPR. Thus, the difference between the various TR algorithms lies in the method of calculating the values that are loaded into the peak cancellation signal $\mathbf{c}$. The algorithms used to solve the problem should be a compromise between PAPR reduction efficiency and implementation complexity.

The TR optimization problem can be formulated and solved optimally as a convex QCQP defined as [62]:

$$\begin{aligned} &\min_{\mathbf{c}} \quad \tau \\ &\text{subject to } \|\mathbf{s}+\mathbf{c}\|_{\infty}^{2} \leq \tau, \end{aligned} \tag{3.5}$$

where τ is the maximum power of the time-domain PAPR-reduced signal $\mathbf{x}_{\mathrm{in}}^{\mathrm{TR}}$. For cases considering a power constraint on the PRTs, as in DVB-T2 and ATSC3.0 for instance, (3.4) is added as an additional constraint to (3.5).

The QCQP solution achieves the upper bound of the achievable PAPR reduction level with the TR technique but with a high computational complexity [62]. Thus, the main drawback of the optimal QCQP solution is its high complexity, which limits

its practical implementation despite its optimal performance. However, adopting the C-RAN architecture increases the acceptance of the complexity of the QCQP solution by moving the implementation to the cloud. In recent years, many researchers have investigated a suboptimal but simpler solution to the PRT assignment problem in [64, and references therein]. In this thesis, we will consider the optimal QCQP-based tone reservation algorithm for our analysis.

3.4 Analytical evaluation of optimized compression techniques based on PAPR-reduced OFDM signals

The quantizer can be optimized using either a known probabilistic distribution or an observed long training sequence of data. The main drawbacks of the latter method is

- An N-level quantizer trained on a k-length training sequence requires computing $N.k$ distances and assigning k data points at each iteration until convergence, as shown previously.
- A sufficiently long training sequence is essential so that the resulting quantizer will hopefully work well for future data.
- The algorithm must be repeated multiple times with different initial training sequences to mitigate dependence.

Such intensive computations make Lloyd's algorithm slow, especially for large data sets. Therefore, prior knowledge of the probabilistic description of the source is more efficient and requires less computational complexity. On the other hand, Huffman coding suffers from a deviation from the entropy bound when the statistical characteristics of the source are unknown.

Therefore, in the next sections, we first derive the distribution of the PAPR-reduced IQ samples, i.e., $\mathrm{f}_X(x)$ and $\mathrm{f}_Y(y)$ for the in-phase and Quadrature-phase components, respectively. The parameters of the compression blocks, i.e., quantization and entropy coding, for the new distribution of the PAPR-reduced signal are analyzed. Finally, the joint effect of PAPR reduction and compression distortion is studied through the MER analysis.

3.4.1 Performance metrics

The Rate R, MER and Signal to Quantization and Noise Ratio (SQNR) are the metrics used to evaluate the performance of the model under study, as shown previously. The evaluation of the MER defined in (1.8) takes into account the signal distortions caused by the PAPR reduction and compression, i.e., quantization. Entropy coding is indeed distortionless. Therefore, we define the SQNR metric to evaluate only the performance of the compression techniques. In this case, $\mathbf{x}_{\text{in}}$ is considered as the

reference signal and the SQNR is defined as

$$\mathrm{SQNR} = \frac{\mathbb{E}\left[\|X_{\mathrm{in}}\|^2\right]}{\mathbb{E}\left[\|X_{\mathrm{in}} - X_{\mathrm{out}}\|^2\right]}, \tag{3.6}$$

where X_{in} is the r.v. representing the PAPR-reduced samples.

In that perspective and taking into account the technique used for the PAPR reduction, i.e., either distortion or distortionless, the following facts must be considered for a proper system evaluation.

- If clipping is considered as a PAPR reduction technique that leads to signal distortion, both clipping and quantization operations contribute to system performance degradation. Therefore, MER and SQNR metrics are used to evaluate the system performance in detail.

- In the TR PAPR reduction technique, the received signal $\tilde{\mathbf{x}}_{\mathrm{out}}$ is followed by removing the TR added signal, thanks to the orthogonality between the data and the tone-reserved signals in the frequency-domain. Thus, the distortion of the OFDM signal is caused only by the quantization. Consequently, the MER between $\mathbf{s}$ and $\mathbf{x}_{\mathrm{out}}$ can be evaluated after removing the added TR signal in the frequency-domain or between the intermediate signals $\mathbf{x}_{\mathrm{in}}$ and $\mathbf{x}_{\mathrm{out}}$. We consider the latter case, where the system MER can be computed directly from (3.6), taking into account only the quantization distortion.

- The denominator of (3.6) is the quantization distortion. Through our analytical analysis, we consider the asymptotic scalar quantization distortion with fixed-length codes, i.e., Lloyd algorithm, stated in proposition 1.2 (Panter and Dite formula).

3.4.2 The impact of PAPR reduction on the amplitude and phase distributions of the OFDM signal

As mentioned before, the amplitude of the OFDM signal $\mathbf{s}$ at the input of the PAPR reduction block converges to a Rayleigh-distributed r.v., thanks to the CLT as follows:

$$\mathrm{f}_{R_s}(r) = \begin{cases} \frac{r}{\sigma^2_{\mathrm{ray}}} e^{-\frac{r^2}{2\sigma^2_{\mathrm{ray}}}}; & \textit{if } r \geq 0 \\ 0; & \textit{otherwise}, \end{cases} \tag{3.7}$$

where R_s is the r.v. representing the amplitude of the input signal and σ_{ray} is its mode. The phase, on the other hand, is uniformly distributed.

However, the distribution of the signal changes after PAPR reduction, depending on the technique used. There are few studies in the literature on the distribution of signal amplitude after PAPR reduction [65]. Therefore, we will illustrate in detail the modified amplitude distribution of the reduced PAPR signal using clipping and QCQP-based TR schemes, which are the simplest and most promising PAPR reduction techniques considered in this study, respectively.

3.4.2.1 Clipping technique

Clipping is applied to the signal amplitude as in (3.1) and changes the amplitude distribution of the OFDM signal without affecting the phase, as reminded in the following lemma.

Lemma 3.1 ([66]). *The time-domain amplitude distribution of the clipped PAPR-reduced signal, is expressed as follows:*

$$\mathrm{f}_{R_{\mathrm{in}}}^{\mathrm{clip}}(r) = \begin{cases} \mathrm{f}_{R_s}(r); & \text{if } r < V_{\max} \\ e^{-\Lambda}\delta(r - V_{\max}); & \text{if } r = V_{\max} \end{cases} \tag{3.8}$$

where R_{in} is the r.v. representing the amplitude of the PAPR-reduced signal $\boldsymbol{x}_{\mathrm{in}}$, $\mathrm{f}_{R_s}(r)$ is the Rayleigh distribution characterized the amplitude of the original OFDM signal $\boldsymbol{s}$ and $\delta(r)$ is the Dirac distribution. $e^{-\Lambda}$ is the probability that the amplitude of $\boldsymbol{s}$ exceeds $V_{\max}$, i.e., $\int_{V_{\max}}^{\infty} \mathrm{f}_{R_s}(r)dr = e^{-\Lambda}$ and Λ is the clipping ratio defined as $\Lambda = \frac{V_{\max}^2}{2\sigma_{\mathrm{ray}}^2}$.

Fig. 3.3 depicts both the amplitude distributions of the OFDM signal before and after clipping, denoted $\mathrm{f}_{R_s}(r)$ and $\mathrm{f}_{R_{\mathrm{in}}}^{\mathrm{clip}}(r)$, respectively. $\mathrm{f}_{R_{\mathrm{in}}}^{\mathrm{clip}}(r)$ follows $\mathrm{f}_{R_s}(r)$ for $r < V_{\max}$ and off-samples above this threshold are gathered in the peak at $V_{\max}$.

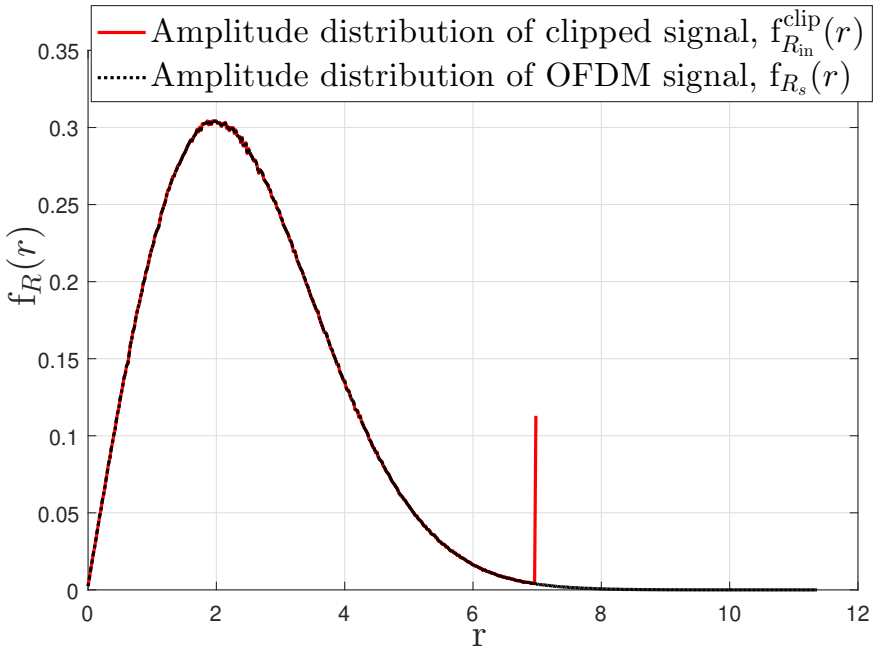


Figure 3.3: PDF of the amplitude of the time-domain signal with and without clipping at $V_{\max} = 7$.

3.4.2.2 TR-QCQP technique

The authors in [67] evaluate the amplitude distribution of the tone-reserved signal solved by the optimal QCQP algorithm. They have shown that it is possible to model the amplitude distribution as a superposition of two separate modes, which is stated in the following lemma.

Lemma 3.2 ([67]). *The time-domain amplitude distribution of the tone-reserved PAPR-reduced signal, is expressed as follows:*

$$\mathrm{f}_{R_{\mathrm{in}}}^{\mathrm{QCQP}}(r) = (1-p)\ \mathrm{f}_{R_{\mathrm{in1}}}(r) + p\ \mathrm{f}_{R_{\mathrm{in2}}}(r), \tag{3.9}$$

where

$$\mathrm{f}_{R_{\mathrm{in1}}}(r) = \begin{cases} q\ \mathrm{f}_R^{\mathrm{Ray}}(r)(1-\mathrm{F}_R^{\mathrm{GEV}}(r)); & \textit{if } r \geq 0 \\ 0; & \textit{if } r < 0, \end{cases} \tag{3.10}$$

and

$$\mathrm{f}_{R_{\mathrm{in2}}}(r) = \begin{cases} \mathrm{f}_R^{\mathrm{GEV}}(r); & \textit{if } r \in D_2 \\ 0; & \textit{otherwise} \end{cases} \tag{3.11}$$

where R_{in1} and R_{in2} are the r.v.s representing the samples drawn from the first and the second modes, respectively, p is a scaling parameter between the first and the second mode and q is a normalization parameter that ensures $\int \mathrm{f}_{R_{\mathrm{in1}}}(r)\,\mathrm{d}r = 1$. $\mathrm{f}_R^{\mathrm{Ray}}(r)$, $\mathrm{F}_R^{\mathrm{GEV}}(r)$ and $\mathrm{f}_R^{\mathrm{GEV}}(r)$ are the Rayleigh PDF, the Generalized Extreme Value (GEV) Cumulative Distribution Function (CDF) and the GEV PDF, defined respectively as follows:

$$\mathrm{f}_R^{\mathrm{Ray}}(r) = \begin{cases} \frac{r}{\sigma_{\mathrm{ray}}^2} e^{-\frac{r^2}{2\sigma_{\mathrm{ray}}^2}}; & \textit{if } r \geq 0 \\ 0; & \textit{otherwise}, \end{cases} \tag{3.12}$$

$$\mathrm{F}_R^{\mathrm{GEV}}(r) = \begin{cases} e^{-\left(1+k_1\left(\frac{r-\mu_1}{\sigma_1}\right)\right)^{-1/k_1}}; & \textit{if } r \in D_1 \\ 0; & \textit{otherwise}, \end{cases} \tag{3.13}$$

and

$$\mathrm{f}_R^{\mathrm{GEV}}(r) = \begin{cases} \frac{1}{\sigma_2} t(r)^{k_2+1} e^{-t(r)}; & \textit{if } r \in D_2 \\ 0; & \textit{otherwise} \end{cases} \tag{3.14}$$

$$\textit{with } t(r) = \left(1 + k_2\left(\frac{r-\mu_2}{\sigma_2}\right)\right)^{-1/k_2}, \tag{3.15}$$

where σ_{ray} is the standard deviation fitting the Rayleigh distribution, μ_1, σ_1, $k_1 < 0$, μ_2, σ_2 and $k_2 < 0$ are respectively the location, scale and shape parameters that fit the GEV CDF $\mathrm{F}_R^{\mathrm{GEV}}(r)$ and the PDF $\mathrm{f}_R^{\mathrm{GEV}}(r)$, respectively and $D_1 = \left]-\infty, \mu_1 - \frac{\sigma_1}{k_1}\right]$ and $D_2 = \left]-\infty, \mu_2 - \frac{\sigma_2}{k_2}\right]$ are their respective domains.

The parameters of the Rayleigh and GEV distributions are calculated in [67] using the probability weighted moments estimation method. Fig. 3.4 shows the amplitude distributions of the OFDM signal before and after TR-QCQP, i.e., $\mathrm{f}_{R_s}(r)$ and $\mathrm{f}_{R_{\mathrm{in}}}^{\mathrm{QCQP}}(r)$, respectively. As mentioned earlier, the $\mathrm{f}_{R_{\mathrm{in}}}^{\mathrm{QCQP}}(r)$ PDF can be modeled with a bimodal distribution whose parameters depend on the power control P_{PC}. The QCQP algorithm modifies the signal so that high amplitude samples have a lower probability of appearing compared to Rayleigh-distributed samples. In this way, QCQP concentrates the high amplitude samples at some levels that form a bump at the tail of the Rayleigh distribution, while the other samples continue to follow a Rayleigh amplitude distribution.

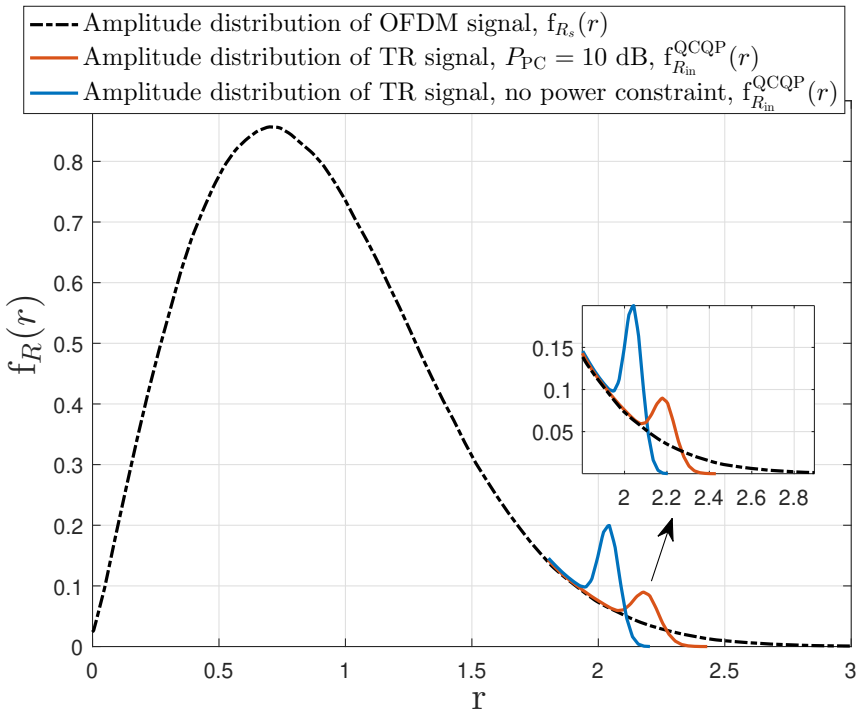


Figure 3.4: PDF of the amplitude of the signal with and without TR-QCQP, with different power constraints and $N_{\text{PRT}} = 1\%$.

3.5 Analytical study of the optimized compression techniques for a PAPR-reduced signal

In this section, we derive the distribution of the PAPR-reduced IQ samples, i.e., $f_{X_{\text{in}}}(x)$ and $f_{Y_{\text{in}}}(y)$ for the in-phase and Quadrature-phase components, respectively.

3.5.1 Distribution of the PAPR-reduced IQ samples

The conversion from polar to Cartesian coordinates of the complex samples is such as $(x, y) = \varphi(r, \theta)$ with:

$$\begin{aligned} x &= r\cos\theta & r &= \sqrt{x^2+y^2} \\ y &= r\sin\theta & \theta &= \varphi_2^{-1}(x,y)\,, \end{aligned} \tag{3.16}$$

where $\varphi : \zeta \mapsto \Upsilon$ is a one-to-one mapping from the polar coordinates domain ζ to the Cartesian coordinates domain Υ and $\varphi_2^{-1} : \Upsilon \mapsto [-\pi, \pi]$ with:

$$\varphi_2^{-1}(x,y) = \begin{cases} \tan^{-1}\left(\frac{y}{x}\right); & if\, x > 0 \\ \tan^{-1}\left(\frac{y}{x}\right) - \pi; & if\, x < 0, y < 0 \\ \pi - \tan^{-1}\left(\frac{y}{x}\right); & if\, x < 0, y \geq 0 \\ \pi/2; & if\, x = 0, y > 0 \\ -\pi/2; & if\, x = 0, y < 0. \end{cases} \tag{3.17}$$

According to the change of variables theorem, the joint PDF of the couple (X, Y) is

$$\mathrm{f}_{X_{\mathrm{in}},Y_{\mathrm{in}}}(x,y) = |J(x,y)|\mathrm{f}_{R_{\mathrm{in}},\Theta_{\mathrm{in}}}\left(r(x,y),\theta(x,y)\right), \tag{3.18}$$

where $|J(x,y)|$ is the modulus of the Jacobian of the mapping φ computed as

$$|J(x,y)| = \left|\left|\begin{matrix}\frac{\partial r}{\partial x} & \frac{\partial r}{\partial y}\\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y}\end{matrix}\right|\right| = \frac{1}{\sqrt{x^2+y^2}}; \quad \forall\,(x,y) \neq (0,0). \tag{3.19}$$

As mentioned earlier, the PAPR reduction techniques under consideration change the signal amplitude without affecting the phase. Thus, $\forall k \in \{0, \cdots N_{\mathrm{fft}} - 1\}$, $r_{\mathrm{in}}(k) = |z_{\mathrm{in}}(k)|$ and $\theta_{\mathrm{in}}(k) = \angle z_{\mathrm{in}}(k)$ are samples drawn from two independent random variables, i.e., R_{in} and Θ_{in}, respectively. Therefore, the joint PDF of $(R_{\mathrm{in}}, \Theta_{\mathrm{in}})$ is the product of the marginal PDFs, i.e., $\mathrm{f}_{R_{\mathrm{in}},\Theta_{\mathrm{in}}}(r,\theta) = \mathrm{f}_{R_{\mathrm{in}}}(r)\mathrm{f}_{\Theta_{\mathrm{in}}}(\theta)$.

3.5.1.1 Clipping technique

Based on the clipping operation defined in (3.1), the respective domains of polar and Cartesian coordinates of the clipped PAPR-reduced signal become $\zeta = [0, V_{\max}] \times [-\pi, \pi]$ and $\Upsilon =]-V_{\max}, V_{\max}[^2$, respectively. Using the amplitude distribution of the clipped signal in Lemma 3.1, in addition to the uniform distribution of the phase $\mathrm{f}_{\Theta_{\mathrm{in}}}(\theta) = \frac{1}{2\pi}$, the joint PDF of the polar coordinates can be expressed as

$$\mathrm{f}^{\mathrm{clip}}_{R_{\mathrm{in}},\Theta_{\mathrm{in}}}(r,\theta) = \begin{cases} \frac{r}{2\pi\sigma^2_{\mathrm{ray}}} e^{-\frac{r^2}{2\sigma^2_{\mathrm{ray}}}}; & if\, r < V_{\max} \\ \frac{e^{-\Lambda}}{2\pi}\delta(r - V_{\max}); & if\, r = V_{\max}. \end{cases} \tag{3.20}$$

Thus, substituting (3.20) and (3.19) into (3.18), the marginal PDF of the Cartesian coordinate of a clipped r.v. is obtained as

$$\mathrm{f}^{\mathrm{clip}}_{X_{\mathrm{in}}}(x) = \int_{-V_{\max}}^{V_{\max}} \left(\frac{1}{2\pi\sigma^2_{\mathrm{ray}}} e^{-\frac{x^2+y^2}{2\sigma^2_{\mathrm{ray}}}} + \frac{e^{-\Lambda}}{2\pi\sqrt{x^2+y^2}} \delta\left(\sqrt{x^2+y^2} - V_{\max}\right) \right) \mathrm{d}y. \tag{3.21}$$

Thus, a closed-form expression of the PDF of the clipped IQ samples can be obtained as in the following theorem.

Theorem 3.1. *The probability distribution of the clipped PAPR-reduced IQ samples after IFFT, is expressed as follows:*

$$\mathrm{f}^{\mathrm{clip}}_{X_{in}}(x) = \frac{1}{\sqrt{2\pi\sigma^2_{\mathrm{ray}}}} e^{-\frac{-x^2}{2\sigma^2_{\mathrm{ray}}}} \mathrm{erf}\left(\frac{V_{\max}}{\sqrt{2\sigma^2_{\mathrm{ray}}}}\right) + \frac{e^{-\Lambda}}{2\pi} \frac{1}{\sqrt{{V_{\max}}^2 - x^2}}; \quad x \in]-V_{\max}, V_{\max}[\tag{3.22}$$

where $\mathrm{erf}(u) = (2/\sqrt{\pi}) \int_0^u \exp(-t^2)\, dt$ *is the error function.*

Proof. Using the integral identity in [68, eq. (1), Sec. 3.321, p. 336] and applying the change of variable $z = \sqrt{x^2+y^2}$, then adequately rearranging the terms leads to the given distribution of the In-phase component $\mathrm{f}^{\mathrm{clip}}_{X_{\mathrm{in}}}(x)$. □

From (3.22), one can check that without clipping, i.e., when $V_{\max} \to \infty$ and hence $\Lambda \to \infty$, the second term vanishes and the error function tends to one. This leads to the zero-mean Gaussian distribution characterizing the OFDM signal.

3.5.1.2 TR-QCQP technique

As for the QCQP-based tone-reserved signal, the respective domains of the polar and Cartesian coordinates are $\zeta = [0,\infty) \times [-\pi,\pi]$ and $\Upsilon = \mathbb{R}^2$. The joint PDF $\mathrm{f}^{\mathrm{QCQP}}_{R_{\mathrm{in}},\Theta_{\mathrm{in}}}(r,\theta)$ is the product of the amplitude distribution of the TR-QCQP signal in Lemma 3.2 and the phase distribution $\mathrm{f}^{\mathrm{QCQP}}_{\Theta_{\mathrm{in}}}(\theta) = \frac{1}{2\pi}$. Thus, substituting the joint PDF in (3.18) together with the Jacobian determinant in (3.19), we obtain the marginal PDF of the Cartesian coordinate of the tone-reserved r.v. from

$$\mathrm{f}^{\mathrm{QCQP}}_{X_{\mathrm{in}}}(x) = \underbrace{\int_{-\infty}^{\infty} \frac{(1-p)q}{2\pi\sigma^2_{\mathrm{ray}}} e^{-\frac{x^2+y^2}{2\sigma^2_{\mathrm{ray}}}} \left(1 - \mathrm{F}^{\mathrm{GEV}}_R(r(x,y))\,\mathrm{d}y\right.}_{\mathrm{f}_{X_1}(x)} + \underbrace{\int_{-\infty}^{\infty} \frac{p}{2\pi\sigma_2\sqrt{x^2+y^2}} G(x,y)^{k_2+1} e^{-G(x,y)}\,\mathrm{d}y}_{\mathrm{f}_{X_2}(x)}, \tag{3.23}$$

$$\text{where, } \mathrm{F}^{\mathrm{GEV}}_R(r(x,y)) = e^{-\left(1+\frac{k_1}{\sigma_1}\left(\sqrt{x^2+y^2}-\mu_1\right)\right)^{-1/k_1}}, \tag{3.24}$$

$$\text{and } G(x,y) = \left(1 + \frac{k_2}{\sigma_2}\left(\sqrt{x^2+y^2} - \mu_2\right)\right)^{-1/k_2}, \tag{3.25}$$

where X_1 and X_2 are the r.v.s representing the samples drawn from the first and the second mode of the In-phase Cartesian distribution, respectively. Finally, one can show that the distribution of the In-phase component, $\mathrm{f}^{\mathrm{QCQP}}_{X_{\mathrm{in}}}(x)$[1], can be expressed as in the following theorem.

Theorem 3.2. *The probability distribution of the TR-QCQP PAPR-reduced IQ samples after IFFT, is expressed as follows:*

$$\mathrm{f}^{\mathrm{QCQP}}_{X_{in}}(x) = \mathrm{f}_{X_1}(x) + \mathrm{f}_{X_2}(x), \quad x \in]-d^*, d^*[\tag{3.26}$$

where

$$\mathrm{f}_{X_1}(x) = \frac{(1-p)q}{\sqrt{2\pi}\sigma_{\mathrm{ray}}} e^{-\frac{x^2}{2\sigma^2_{\mathrm{ray}}}} \mathrm{erf}\left(\frac{\sqrt{r_s^2 - x^2}}{\sqrt{2\sigma^2_{\mathrm{ray}}}}\right), \quad \textit{with} \quad r_s = \mu_1 + \frac{\sigma_1}{k_1}\left((\ln 2)^{-k_1} - 1\right), \tag{3.27}$$

and

$$\mathrm{f}_{X_2}(x) = \frac{-p}{\pi} \sum_{p_1=0}^{\infty} \sum_{p_2=0}^{p_1} \xi\left[\Gamma\big(k_2(p_1+p_2+1)+1, \psi_1\big) - \Gamma\big(k_2(p_1+p_2+1)+1, \psi_2\big)\right], \tag{3.28}$$

$$\textit{with,} \quad \xi = \binom{-\frac{1}{2}}{p_1}\binom{p_1}{p_2}\left[c_1^{\frac{-(1+2p_1)}{2}} c_2^{p_1-p_2} c_3^{p_2}\right], \tag{3.29}$$

[1]The distribution of the Quadrature-phase component, $\mathrm{f}^{\mathrm{clip}}_{Y_{\mathrm{in}}}(y)$ and $\mathrm{f}^{\mathrm{QCQP}}_{Y_{\mathrm{in}}}(y)$, are obtained from the same steps as in Theorems 3.1 and 3.2, respectively, given that the phase is uniformly distributed.

$$c_1 = \frac{\sigma_2^2}{k_2^2}, \qquad c_2 = \frac{2\mu_2\sigma_2}{k_2} - 2c_1, \qquad c_3 = -c_1 - c_2 + \mu_2^2 - x^2,$$
$$\psi_1 = \left[1 + \frac{k_2}{\sigma_2}(x - \mu_2)\right]^{-\frac{1}{k_2}}, \quad \psi_2 = \left[1 + \frac{k_2}{\sigma_2}\left(\sqrt{x^2 + d^{*2}} - \mu_2\right)\right]^{-\frac{1}{k_2}}. \tag{3.30}$$

and

$$d^* = \max\left(r_s, \mu_2 - \frac{\sigma_2}{k_2}\right). \tag{3.31}$$

Where $\Gamma(u, v) = \int_v^\infty e^{-t}t^{u-1}dt$ *is the upper incomplete Gamma function,* $\binom{r}{k}$ *is the generalized binomial coefficient defined as* $\binom{r}{k} = \frac{(r)_k}{k!}$ *where,* $(r)_k = \prod_{n=0}^{k-1}(r-n)$ *is the falling factorial and* p, q, $\sigma_{\text{ray}}, \mu_1, \sigma_1, k_1, \mu_2, \sigma_2$, *and* k_2 *are the fitted parameters of the amplitude distribution* $\mathrm{f}_{R_{\text{in}}}^{\text{QCQP}}(r)$ *in Lemma 3.2.*

Proof. See Appendix B. □

We can see that the distribution of the TR-QCQP IQ samples takes the form of a series expansion with gamma functions and depends on the estimated fitted parameters of the GEV and Rayleigh distributions that model the amplitude of the tone-reserved signal in Lemma 3.2.

3.5.2 Optimized quantizer and entropy coder of PAPR-reduced signal

The parameters of the optimized Lloyd-based quantizer and entropy encoder are the quantization levels $y_i, \forall i = 1, \cdots, N$ and the decision thresholds $t_i, \forall i = 1, \cdots N-1$, which are used to design the quantizer in Section 1.4.2.3, and the average number of bits associated with each quantization level given by the entropy encoder according to (1.21). Based on the PDF of the clipped PAPR-reduced IQ samples in Theorem 3.1, the parameters of the optimized Lloyd quantizer and the entropy encoder can be derived by injecting (3.22) into (1.17) and (1.21), respectively, as shown in the following proposition.

Proposition 3.1. *The codebook quantization levels for the real and imaginary parts of a clipped signal and the average codeword length assigned to each quantization level in the entropy coder stage can be expressed by*

$$y_i^{\text{clip}} = \frac{\frac{\sigma_{\text{ray}}}{\sqrt{2\pi}}\text{erf}\left(\frac{V_{\max}}{\sqrt{2\sigma_{\text{ray}}^2}}\right)\left[e^{-\frac{t_{i-1}^2}{2\sigma_{\text{ray}}^2}} - e^{-\frac{t_i^2}{2\sigma_{\text{ray}}^2}}\right] + \frac{e^{-\Lambda}}{2\pi}\left[\sqrt{V_{\max}^2 - t_{i-1}^2} - \sqrt{V_{\max}^2 - t_i^2}\right]}{\text{erf}\left(\frac{V_{\max}}{\sqrt{2\sigma_{\text{ray}}^2}}\right)\left[\text{Q}\left(\frac{t_{i-1}}{\sigma_{\text{ray}}}\right) - \text{Q}\left(\frac{t_i}{\sigma_{\text{ray}}}\right)\right] + e^{-\frac{\Lambda}{2\sigma_{\text{ray}}^2}}\left[\arcsin\left(\frac{t_{i-1}}{V_{\max}}\right) - \arcsin\left(\frac{t_i}{V_{\max}}\right)\right]}, \tag{3.32}$$

$$L_{y_i}^{\text{clip}} = -\log_2\left(\text{erf}\left(\frac{V_{\max}}{\sqrt{2\sigma_{\text{ray}}^2}}\right)\left[\text{Q}\left(\frac{t_{i-1}}{\sigma_{\text{ray}}}\right) - \text{Q}\left(\frac{t_i}{\sigma_{\text{ray}}}\right)\right] + e^{-\frac{\Lambda}{2\sigma_{\text{ray}}^2}}\left[\arcsin\left(\frac{t_{i-1}}{V_{\max}}\right) - \arcsin\left(\frac{t_i}{V_{\max}}\right)\right]\right). \tag{3.33}$$

where $Q(u) = (1/\sqrt{2\pi})\int_u^\infty exp(-v^2/2)dv$ *is the Q-function.*

Proof. Applying the adequate change of variables $u = V_{\max}^2 - x^2$ and $x = V_{\max} \sin u$ for the numerator and the denominator of (1.17), respectively, along with the integral identity in [68, eq. (1), Sec. 3.321, p. 336], y_i^{clip} and $L_{y_i}^{\text{clip}}$ are obtained. □

Without clipping, i.e., $V_{\max} \to \infty$ and thus $\Lambda \to \infty$, the error function tends to one and the exponential term in (3.32) and (3.33) tends to zero and we obtain the parameters of a quantizer and an entropy encoder optimized for a zero-mean Gaussian distribution given in (2.5) and (2.6), respectively.

The PDF of the tone-reserved IQ samples obtained in Theorem 3.2 remains complex because special functions with complex arguments are involved. Therefore, closed form expressions for the codebook quantization levels and the average codeword length associated with each quantization level are unattainable for PAPR-reduced signals with tone-reservation technique and they are computed numerically by injecting $\mathrm{f}_{X_{\text{in}}}^{\text{QCQP}}(x)$ in (1.17) and (1.21), respectively.

3.5.3 Asymptotic MER and SQNR

A closed-form expression of the MER and SQNR in (1.8) and (3.6), respectively, can be evaluated asymptotically for the clipping PAPR reduction technique based on Theorem 3.1 as stated in the following lemma. In this case, the asymptotic SQNR is only important to evaluate the distortion caused by the quantizer.

Lemma 3.3. *The asymptotic MER and SQNR expressions of a compressed clipped OFDM signal, considering a scalar NUQ and EC can be expressed by:*

$$MER^{\text{clip}} = \frac{P_s}{D_C + D_{\text{LQ}}^{\text{clip}}} = \frac{2\sigma_{\text{ray}}^2}{D_C + D_{\text{LQ}}^{\text{clip}}}, \tag{3.34}$$

and

$$SQNR^{\text{clip}} = \frac{P_{z_{\text{in}}}^{\text{clip}}}{D_{\text{LQ}}^{\text{clip}}}, \tag{3.35}$$

where D_C, $D_{\text{LQ}}^{\text{clip}}$ and $P_{z_{\text{in}}}^{\text{clip}}$ are the clipping distortion, asymptotic quantization distortion and the power of the complex clipped IQ samples, represented respectively as follows:

$$D_C = 2\sigma_{\text{ray}}^2 \Gamma\left(2, \frac{V_{\max}^2}{2\sigma_{\text{ray}}^2}\right) - 2^{\frac{3}{2}} \sigma_{\text{ray}} V_{\max} \Gamma\left(\frac{3}{2}, \frac{V_{\max}^2}{2\sigma_{\text{ray}}^2}\right) + V_{\max}^2 e^{-\frac{V_{\max}^2}{2\sigma_{\text{ray}}^2}}, \tag{3.36}$$

$$D_{\text{LQ}}^{\text{clip}} = \frac{2^{-2R}}{6}\left[\frac{2\sqrt[3]{a} V_{\max}}{\sqrt{6b\sigma_{\text{ray}}^2}} {}_1F_1\left(\frac{1}{2}; \frac{3}{2}; \frac{-V_{\max}^2}{6\sigma_{\text{ray}}^2}\right) - \frac{26cV_{\max}}{25\sqrt{3\sigma_{\text{ray}}^2}\sqrt[3]{a^2}} {}_1F_1\left(\frac{1}{2}; 1; \frac{V_{\max}^2}{3\sigma_{\text{ray}}^2}\right)\right]^3, \tag{3.37}$$

$$P_{z_{\text{in}}}^{\text{clip}} = \frac{2\sigma_{\text{ray}}^2}{\sqrt{\pi}} \text{erf}\left(\frac{V_{\max}}{\sqrt{2\sigma_{\text{ray}}^2}}\right)\left[\sqrt{\frac{\pi}{2}} \text{erf}\left(\frac{V_{\max}}{\sqrt{2\sigma_{\text{ray}}^2}}\right) - \frac{V_{\max}}{\sqrt{2\sigma_{\text{ray}}^2}} e^{-\frac{V_{\max}^2}{2\sigma_{\text{ray}}^2}}\right] + \frac{V_{\max}^2}{4} e^{-\frac{V_{\max}^2}{2\sigma_{\text{ray}}^2}}. \tag{3.38}$$

with $a = \frac{1}{\sqrt{2\pi\sigma_{\text{ray}}^2}} erf\left(\frac{V_{\max}}{\sqrt{2\sigma_{\text{ray}}^2}}\right)$, $b = \frac{1}{2\sigma_{\text{ray}}^2}$, $c = \frac{e^{-\Lambda}}{2\pi}$ and ${}_1F_1$ is the confluent hypergeometric function.

Proof. See Appendix C. □

Finally, it comes that the MER depends only on the input signal standard deviation σ_{ray}, the clipping amplitude threshold $V_{\max}$ and the number of bits per sample R, which are the key parameters of the PAPR reduction and compression techniques.

For TR-QCQP, the asymptotic MER of the compressed tone-reserved OFDM signal can be computed numerically from the MER definition in (1.8) based on the PDF of the tone-reserved IQ samples $\mathrm{f}_{X_{\text{in}}}^{QCQP}(x)$ in Theorem 3.2. The numerator of (1.8) is the average power of a complex tone-reserved IQ sample, denoted by $P_{z_{\text{in}}}^{\text{QCQP}}$ and is calculated from $P_{z_{\text{in}}}^{\text{QCQP}} = 2\int_{-d*}^{d*} x^2 \mathrm{f}_{X_{\text{in}}}^{QCQP}(x)dx$. As shown earlier, quantization is the only source of distortion in the TR PAPR reduction technique. Therefore, the denominator of (1.8) is the scalar quantization distortion computed asymptotically from Proposition 1.2.

3.6 Simulation results and analysis

In this section, the distributions of the IQ samples obtained according to the clipping and tone reservation techniques derived in the theorems 3.1 and 3.2 are validated by simulations. These expressions are then exploited to analyze the effect of reducing the PAPR of the OFDM signal before compression in a C-RAN architecture.

3.6.1 Simulation parameters

In our simulations and without loss of generality, we consider an 8-MHz DVB-T2 frame structure with 64-QAM constellations[2]. The $32K$ FFT mode is considered for the analysis of the clipping PAPR reduction technique with 32768 active subcarriers conveying the QAM symbols. On the other hand, only the $8K$ mode with 8192 subcarriers is considered for TR-QCQP due to the prohibitive computational complexity of QCQP for larger FFT sizes. It can be emphasised that an N_{fft} size of $8K$ is sufficiently large to ensure the convergence of the real and imaginary parts of the complex OFDM signal to a Gaussian distribution. This ensures a fair comparison between the clipping and TR analysis.

In TR, when analyzing the impact of the different power control P_{PC} applied to PRT subcarriers compared to data subcarriers, a percentage of dedicated subcarriers $N_{\text{PRT}} = 1\%$ is set, and the PRT positions are those used in the DVB-T2 frame structure stated in (Appendix A in [63]). In the case of a different percentage of subcarriers dedicated for the TR algorithm N_{PRT}, we set a power control $P_{\text{PC}} = 10$ dB and randomly generate the positions of the PRTs. The simulation parameters used are summarized in Table 3.2.

[2]Note that the order of the constellations has a very limited impact on the distribution of the samples in the time-domain. Hence, simple 64-QAM is used in the simulation instead of more sophisticated constellation such as rotated 256-QAM, which is used in DVB-T2 and ATSC3.0 for instance.

Table 3.2: Simulation parameters for the analysis of the compression of a PAPR-reduced signal

Metrics	Clipping analysis	TR-QCQP analysis
Channel BW in MHz	8	8
Sampling frequency in MHz	9.14	9.14
FFT size	$32K$	$8K$
Number of data subcarriers	26265	6817
Number of guard subcarriers	5503	1375
Number of OFDM symbols	1000	1000
Constellation	64-QAM	64-QAM

3.6.2 Validity of the distribution of the PAPR-reduced IQ samples

First, we validate the obtained time-domain PAPR-reduced IQ samples distribution for both clipping and TR PAPR reduction techniques.

Clipping technique

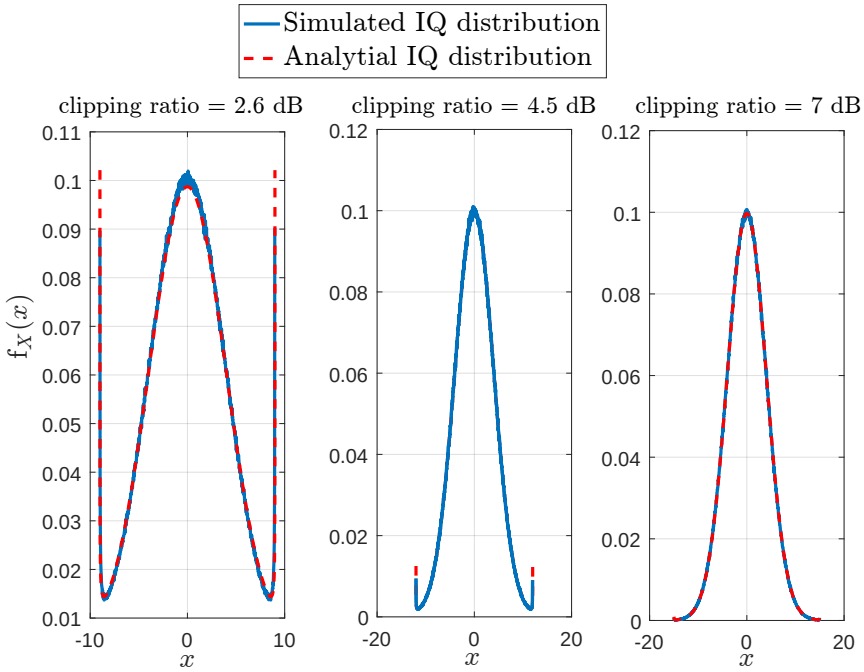


Figure 3.5: Simulated and analytical PDF of clipped PAPR-reduced IQ samples for different clipping ratios Λ.

We begin by validating the obtained clipped PAPR-reduced IQ samples PDF in Theorem 3.1. Fig. 3.5 shows the PDF for different clipping ratios of 2.6 dB, 4.5 dB, and 7 dB, respectively. The analytical curves perfectly match the simulated ones,

confirming the statement of Theorem 3.1. It is worth noting that the distribution of the clipped IQ samples converges to a zero-mean Gaussian distribution as the clipping threshold is increased. In fact, the distortion decreases while increasing the clipping threshold, as expected from (3.22).

TR-QCQP technique

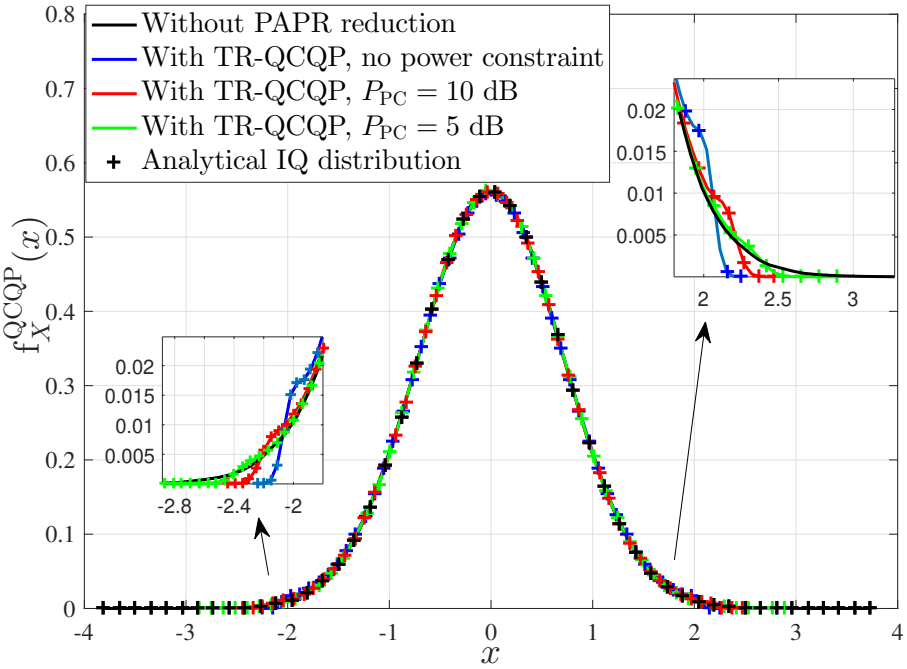


Figure 3.6: Simulated and analytical PDF of TR-QCQP PAPR-reduced IQ samples, with and without power constraint and $N_{\text{PRT}} = 1\%$.

We validate the distribution of the TR-QCQP PAPR-reduced IQ samples expressed in Theorem 3.2. Fig. 3.6 shows the analytical [3] and the simulated IQ PDF for a set of 1% PRT over the total number of subcarriers. The power constraint applied to PRT subcarriers compared to data subcarriers is set to 5, 10 dB, or no constraint. In addition, in Fig. 3.7, the simulated and analytical PDF are shown for different percentages of reserved tones over the total number of subcarriers and with a power constraint $P_{\text{PC}} = 10$ dB.

It is important to note that the TR algorithm introduces a power increase on reserved subcarriers compared to data subcarriers, according to the power constraint defined in (3.4). Therefore, in order to ensure a fair comparison, the energy of the tone-reserved PAPR-reduced signals is normalized to 1. Finally, our findings in Theorem 3.2 are perfectly validated by Figs. 3.6 and 3.7. For our upcoming analysis, it is important to investigate how the distribution of the tone-reserved IQ samples depends on the tuning factors P_{PC} and N_{PRT}.

[3] We use the fitted parameters of the amplitude distribution of the TR-QCQP signal estimated in [67].

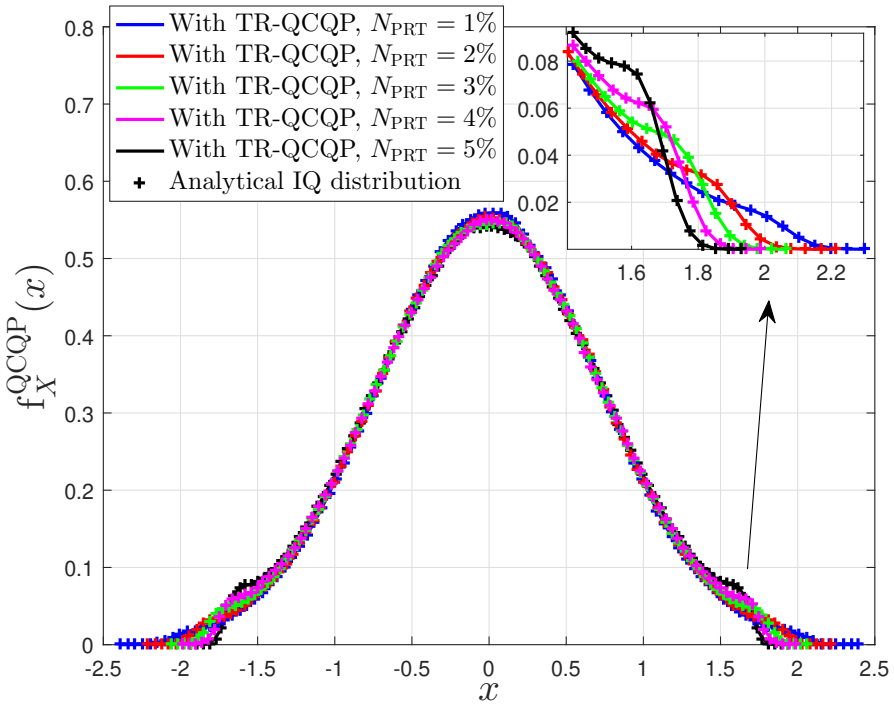


Figure 3.7: Simulated and analytical PDF of TR-QCQP PAPR-reduced IQ samples, with different percentages of reserved tones and $P_{\mathrm{PC}} = 10$ dB.

Impact of power control P_{PC} The distribution of the signal before applying the tone reservation algorithm is Gaussian, as observed in Fig. 3.6. On the other hand, after TR, the PDF is a multimodal distribution with a main mode and two minor modes in the form of "shoulders" on each side of the main mode. The probability of the minor modes is very small compared to the main mode and depends on the value of the power control P_{PC}. The main mode still follows a Gaussian distribution for small signal values, and the power assigned to the PRTs allows the tail of the distribution to change as follows. When additional power is added, the tail of the distribution decreases and the concentration of the minor mode becomes more significant. More precisely, when no power constraint is applied, the tail of the distribution is reduced as much as possible.

Impact of percentage of PRTs N_{PRT} The larger the number of reserved tones, the more the PAPR is reduced. For this reason, and as before, the tail of the distribution decreases along with the growth of the minor modes as N_{PRT} increases, as shown in Fig. 3.7.

3.6.3 Asymptotic distortion-rate analysis

In this section, we show the gain in the performance of the lossy compression technique, i.e., quantization, resulting from the change of the distribution of the OFDM signal when it is subjected to a PAPR reduction technique. These analyses are presented asymptotically based on Shannon distortion-rate theory. Shannon gave a fun-

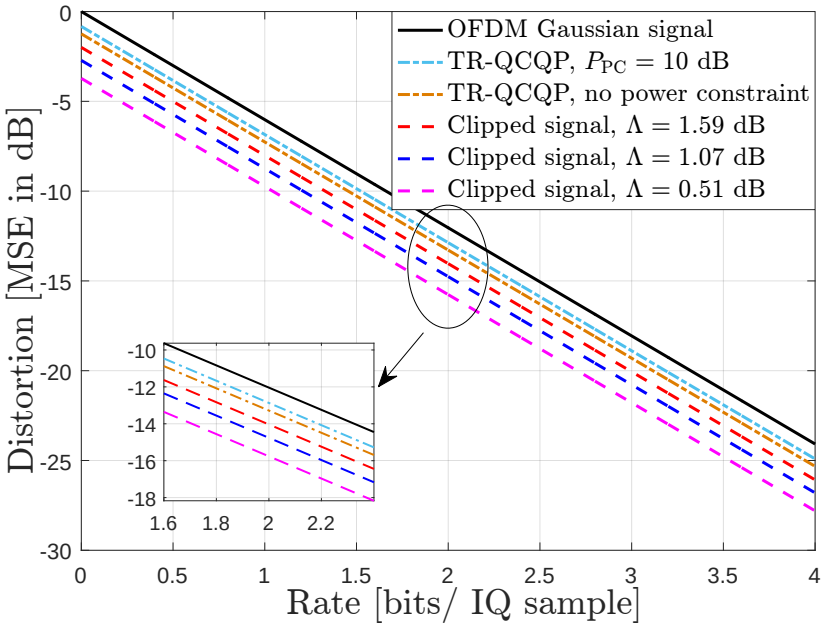


Figure 3.8: $D_{\mathrm{SLB}}^{\mathrm{clip}}(R)$ and $D_{\mathrm{SLB}}^{\mathrm{QCQP}}(R)$ functions for clipped and TR-QCQP PAPR-reduced signals, at different Λ and P_{PC}, respectively, bounded above by the Gaussian distortion-rate function $D_{\mathrm{SLB}}^{\mathrm{G}}(R)$.

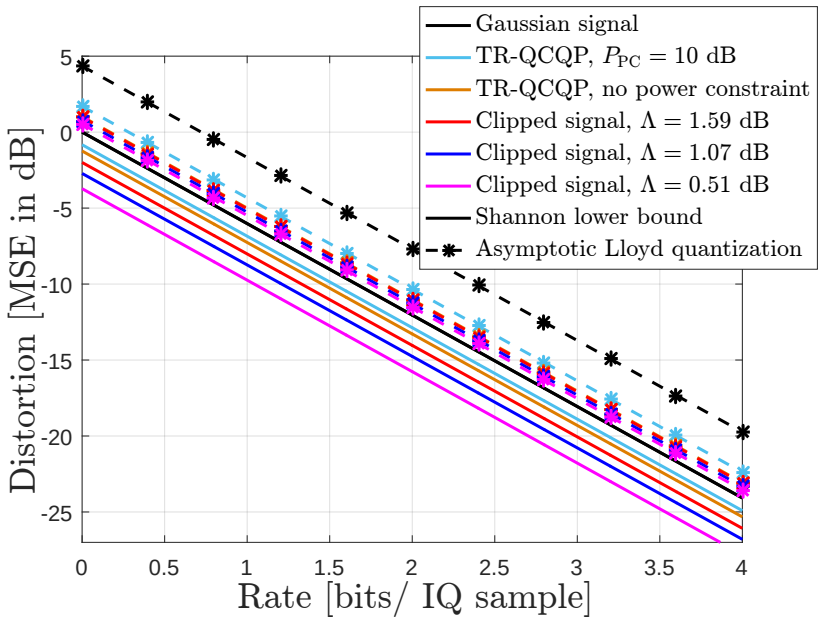


Figure 3.9: Asymptotic distortion of Lloyd quantizer and SLB for clipped and TR-QCQP PAPR-reduced signals, at different Λ and P_{PC}, respectively, vs resolution.

damental lower bound (SLB) for the distortion-rate function $D_{\mathrm{SLB}}(R)$, which is shown in (1.32).

Although the distortion-rate function has a simple expression, it cannot be evaluated analytically except in some special cases. However, it is known that the Gaussian source maximizes the differential entropy and hence the SLB for a given variance [30]. Therefore, for a r.v. X with a given variance σ^2, $h(X)$ is upper bounded by

$$h(X) \leq h_G(X) = \frac{1}{2}\log_2(2\pi e\sigma^2), \tag{3.39}$$

and the distortion-rate is in turn upper bounded as follows:

$$D_{\mathrm{SLB}}(R) \leq D^G_{\mathrm{SLB}}(R) = \sigma^2 2^{-2R}, \tag{3.40}$$

where $h_G(X)$ is the differential entropy of a Gaussian random variable with variance σ^2 and $D^G_{\mathrm{SLB}}(R)$ is SLB for a Gaussian source.

Fig. 3.8 shows the distortion-rate Shannon bounds for the distributions of clipped signals with different clipping ratios, denoted $D^{\mathrm{clip}}_{\mathrm{SLB}}(R)$, and for tone-reserved signals with various power constraints, denoted $D^{\mathrm{QCQP}}_{\mathrm{SLB}}(R)$, together with the Gaussian upper bound $D^{\mathrm{G}}_{\mathrm{SLB}}(R)$. The Gaussian upper bound $D^{\mathrm{G}}_{\mathrm{SLB}}(R)$ is plotted from (3.40), while the distortion-rate functions $D^{\mathrm{clip}}_{\mathrm{SLB}}(R)$ and $D^{\mathrm{QCQP}}_{\mathrm{SLB}}(R)$ are computed numerically using the corresponding PDFs in Theorems 3.1 and 3.2 respectively, in Shannon bound in (1.32).

It can be seen that the highest distortion usually results from the Gaussian distribution. At low clipping ratios, the deviation of $D^{\mathrm{clip}}_{\mathrm{SLB}}(R)$ from the Gaussian distribution $D^{\mathrm{G}}_{\mathrm{SLB}}(R)$ is evident. As the clipping ratio increases and the signal converges to a zero-mean Gaussian-distributed signal, this deviation decreases. Moreover, $D^{\mathrm{QCQP}}_{\mathrm{SLB}}(R)$ is slightly lower than the Gaussian upper limit. This is due to the fact that the TR-QCQP algorithm preserves the Gaussian distribution except for a slight change at the tail of the distribution. Indeed, when the power control P_{PC} increases, the tail of the TR-QCQP PDF decreases and the contribution of the minor modes increases. Therefore, the $D^{\mathrm{QCQP}}_{\mathrm{SLB}}(R)$ of a tone-reserved signal is farther from the Gaussian upper bound.

Moreover, for the sake of clarification and comparison, Fig. 3.9 shows the asymptotic Lloyd quantizer limit defined in Proposition 1.2, i.e., *Panter and Dite formula* for Gaussian, clipped and tone-reserved distributions together with their SLB. The asymptotic scalar quantization distortion is calculated from the following closed-form expression for the Gaussian signal: $D^{\mathrm{G}}_{\mathrm{LQ}} = \frac{\pi\sqrt{3}}{2}\sigma^2 2^{-2R}$ [48], derived in (3.37) for the clipped signal and computed numerically for the tone-reserved signal. The difference between the asymptotic quantization distortion and SLB for each PDF is summarized in column 3 of Table 3.3. Obviously, the Lloyd quantizer can achieve lower MSE, i.e., better performance by approaching SLB for clipped and tone-reserved signals than for the Gaussian signal.

3.6.4 Gaussian quantizer performance evaluation

In this section, the robustness of Gaussian-optimized compression techniques is evaluated when applied to a PAPR-reduced signal. The results are given thanks to the SQNR to highlight the effect of the quantization block.

Table 3.3: Distortion reduction for clipped and TR-QCQP signals, relative to the Gaussian-distributed signal and the asymptotic difference in distortion between the Lloyd quantizer and SLB for each PDF.

PDF	$D^{\mathrm{G}}_{\mathrm{SLB}}(R)$-$D_{\mathrm{SLB}}(R)$	$D_{\mathrm{LQ}}(R)$ - $D_{\mathrm{SLB}}(R)$
Gaussian PDF	0	4.34 dB
TR-QCQP, $P_{\mathrm{PC}} = 10$ dB	0.83 dB	2.52 dB
TR-QCQP, no power limit	1.24 dB	2.26 dB
Clipped, ($\Lambda = 1.59$ dB)	2.0 dB	2.97 dB
Clipped, ($\Lambda = 1.07$ dB)	2.72 dB	3.47 dB
Clipped, ($\Lambda = 0.51$ dB)	3.72 dB	3.60 dB

Clipping technique

Fig. 3.10 shows the simulated SQNR w.r.t. the clipping ratio Λ for different resolutions. The solid lines represent the optimal performance of the Gaussian quantizer applied to a Gaussian OFDM signal. The dashed lines represent the SQNR obtained when the Gaussian quantizer is applied to the clipped IQ samples. We first conclude that applying a Gaussian quantizer to a clipped signal results in a severe degradation of the SQNR. Moreover, the performance loss increases as the resolution of the quantizer increases. Finally, the higher the clipping threshold Λ, the lower the performance loss, as the distribution of the clipped IQ samples converges to a zero-mean Gaussian distribution. Therefore, the SQNR curves reach a constant ceiling at high clipping thresholds.

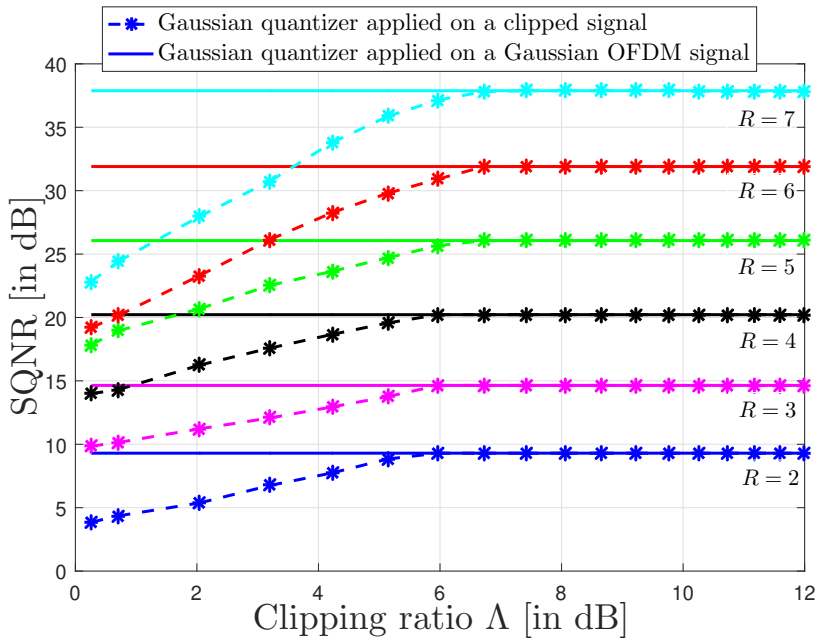


Figure 3.10: SQNR vs clipping ratio Λ of a Gaussian quantizer applied on a clipped signal vs the resolution.

TR-QCQP technique

Fig. 3.11 shows the simulated MER as a function of resolution R for different values of P_{PC} and N_{PRT}, respectively. The solid line represents the optimal performance of a Gaussian quantizer applied to a Gaussian OFDM signal. The dashed lines represent the MER obtained when the Gaussian quantizer is applied to the tone-reserved, PAPR-reduced IQ samples. We first note that applying a Gaussian quantizer to a tone-reserved signal does not degrade the performance; on the contrary, it slightly improves the performance of the MER compared to the Gaussian signal. Moreover, the MER improves when P_{PC} or N_{PRT} is increased. Based on our previous analysis, this can be explained as follows.

The main mode follows the Gaussian distribution. Therefore, the Gaussian codebook quantization levels are optimally adapted in this region. On the other hand, the tail of the distribution is reduced when P_{PC} or N_{PRT} is increased, which means that high amplitude values occur with a quasi-null probability, resulting in less quantization distortion compared to purely Gaussian-distributed samples. In Fig. 3.12a and Fig. 3.12b, the partial quantization distortion is shown for each quantization region for 4 and 8 level quantizers, respectively. The quantization distortion of tone-reserved signals is reduced in the first and last regions compared to the distortion of the OFDM signal, while it is the same in the other regions because the distributions are quite similar.

3.6.5 Comparison of Gaussian and optimized quantizers

Although the Gaussian quantizer has been shown to be very robust to the modification in the distribution of samples after PAPR reduction in some situations, this does not mean that no gain can be expected from an optimized quantizer. In this section, we therefore use the obtained distribution of PAPR-reduced IQ samples to investigate the gain in system performance obtained by optimizing the compression techniques.

Clipping technique

In Fig. 3.13, the analytical expression of the MER given in Lemma 3.3 is validated by plotting the analytical and simulated MER as a function of the clipping ratio Λ for different resolutions. Due to the precision of the asymptotic quantization distortion formula in Proposition 1.3, the analytical expression of MER agrees perfectly with the simulation results at high resolutions as long as $R \geq 4$ [48]. Moreover, the difference between the theoretical and the simulated MER does not exceed 0.25 dB at low resolution, which validates the proposed MER expression.

Fig. 3.14 shows the MER for both Gaussian and clipped optimized compression techniques obtained from the proposed analytical study as a function of the clipping ratio Λ for different resolutions. First, MER increases while increasing the clipping ratio, and at high clipping ratios only quantization distortion affects system performance. Second, the higher the clipping ratio, the closer the clipped distribution approaches the Gaussian one. Thus, the two types of compression techniques merge to achieve an upper bound, which is the MER when only the quantization distortion of a Gaussian-

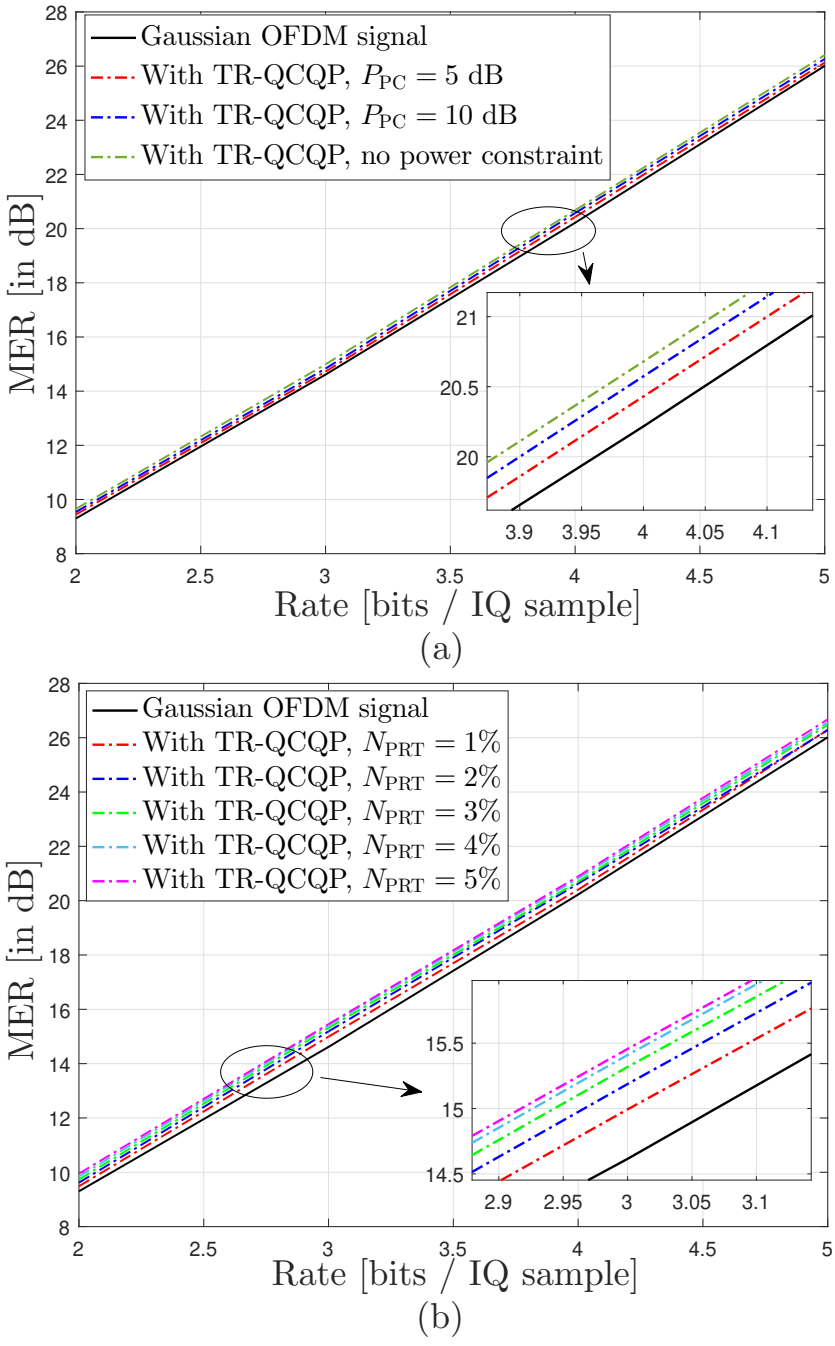


Figure 3.11: MER vs resolution of a Gaussian quantizer applied to a TR-QCQP signal. (a) different power constraints P_{PC} and $N_{\mathrm{PRT}} = 1\%$. (b) different percentages of PRTs N_{PRT} and $P_{\mathrm{PC}} = 10$ dB.

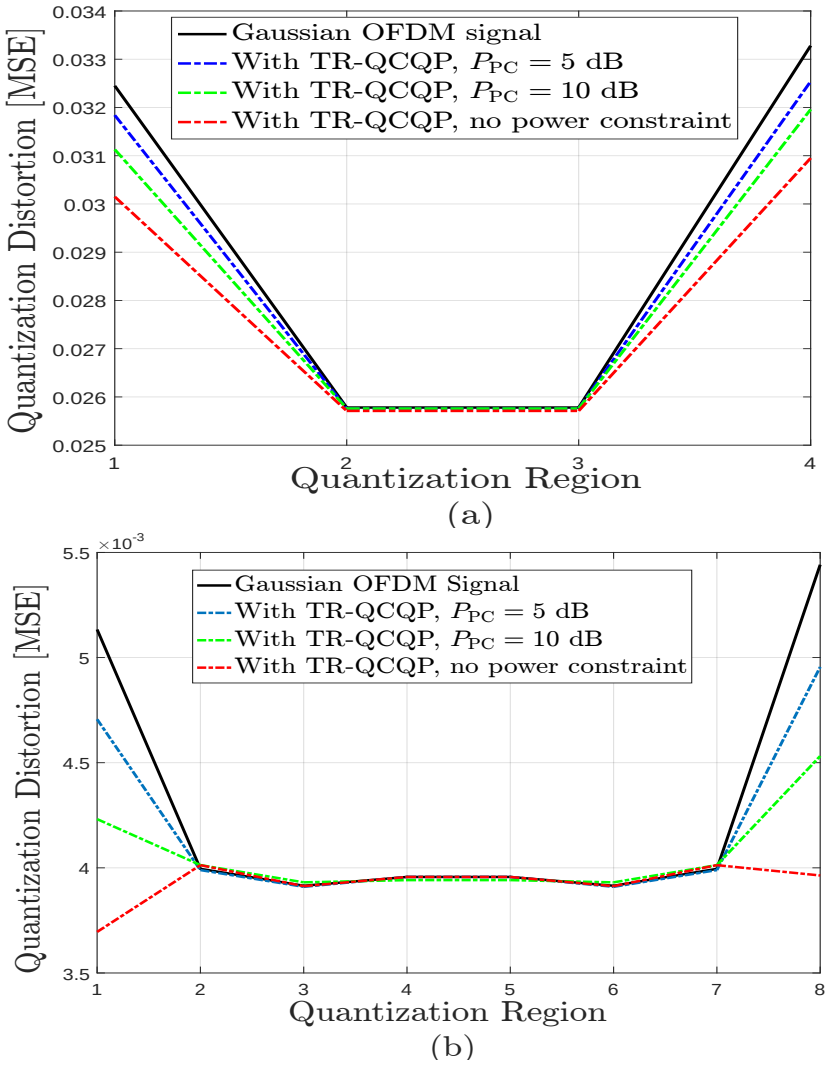


Figure 3.12: MSE per each quantization region for different power constraints. (a) 4-level quantizer with 2 bits per sample. (b) 8-level quantizer with 3 bits per sample.

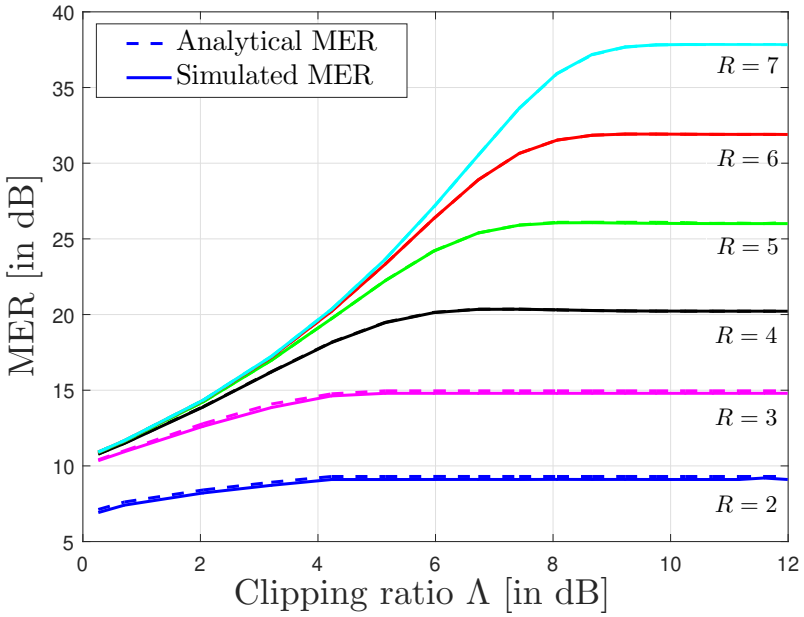


Figure 3.13: Simulated and analytical MER vs clipping ratio Λ of clipped optimized compression techniques for different resolutions.

distributed signal at a given rate is considered. This explains the 6-dB gap between the different rates at high clipping ratios and verifies the famous 6-dB quantization rule [6].

In conclusion, applying optimized compression techniques to a clipped PAPR-reduced signal improves the performance of the MER at low resolutions and low clipping ratios, i.e., a highly clipped signal. However, compared to pure Gaussian compression schemes, the degree of improvement decreases as resolution increases, reaching about 0.02 dB when an average of 7 bits per IQ sample is used.

This can be explained by exploring Fig. 3.15, where the clipping and Gaussian quantization distortions are plotted separately for different resolutions. It can be observed that although the Gaussian quantizer is not optimized for the clipped input signal, the quantization distortion is negligible compared to the clipping distortion at high rates. This is because a large number of quantization levels in a finite quantization region effectively reduces the quantization distortion to almost 0.006, even if it is not an optimized quantizer. Thus, in both cases, clipping distortion becomes the only dominant fixed distortion compared to the quantization distortion.

TR-QCQP technique

We confirm the asymptotic MER numerically using the obtained PDF of the TR-QCQP IQ samples in Theorem 3.2. Fig. 3.16 compares the asymptotic theoretical MER with the values obtained by simulations for different power constraints on the reserved tones. The theoretical values are represented by bars bordered with dashed lines while the values obtained by simulations are bars bordered by solid lines. They agree acceptably at the high rates. More precisely, the deviation between the asymptotic and simulated results is less than 0.18 dB for $R \geq 5$. This validates our theo-

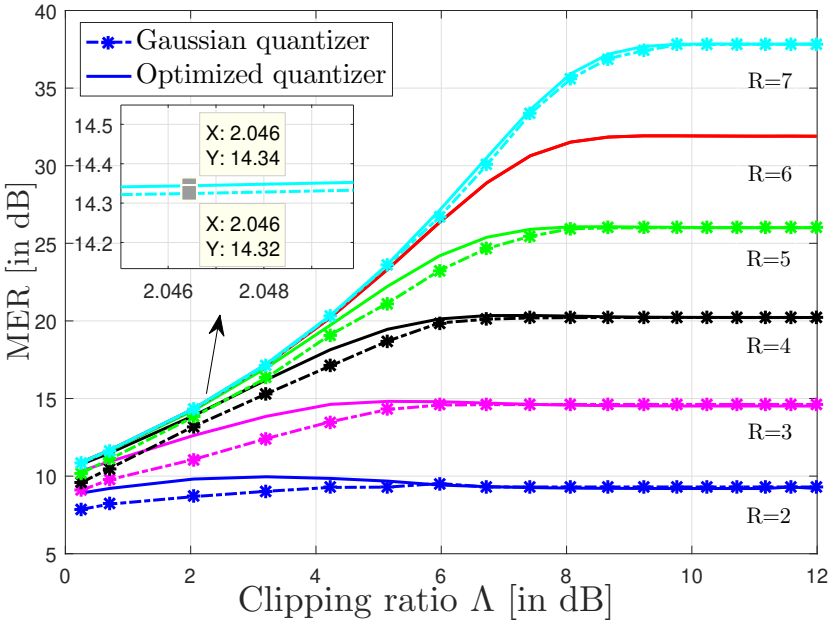


Figure 3.14: MER of Gaussian and optimized compression techniques vs clipping ratio Λ for different resolutions.

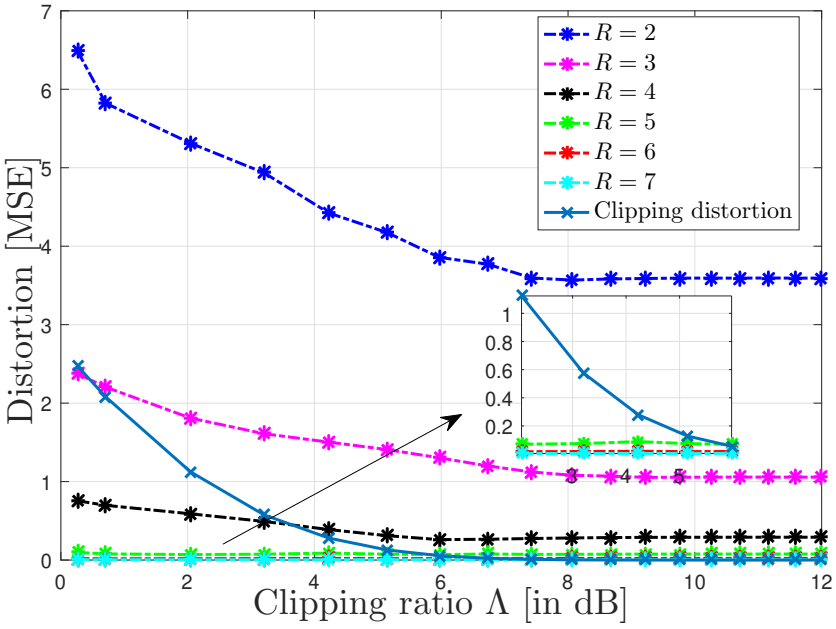


Figure 3.15: Clipping PAPR-reduction and Gaussian quantizer MSEs vs clipping ratio Λ for different resolutions.

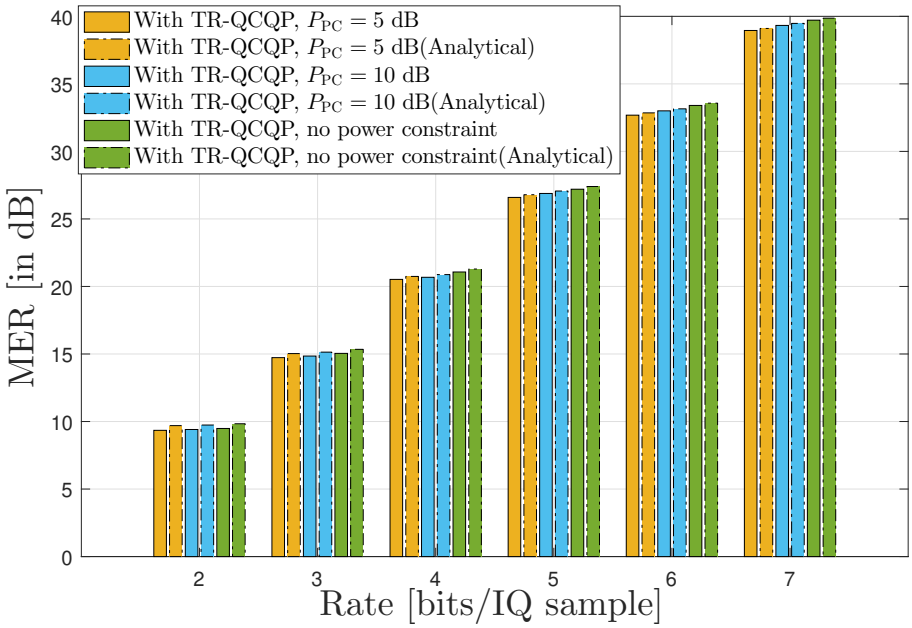


Figure 3.16: Simulated and asymptotic MER vs resolution of TR-QCQP optimized compression techniques for different power constraints P_{PC}.

retical findings in Theorem 3.2. Therefore, we can use the derived IQ PDF together with the estimated parameters of the amplitude distribution in [67] to evaluate the performance of different compression schemes without having to perform complex and time-consuming numerical simulations.

Figs. 3.17 and 3.18 show the MER of both Gaussian and tone-reserved optimized compression techniques obtained by the numerical integration of the IQ distribution obtained in Theorem 3.2 as a function of the resolution R for different P_{PC} and N_{PRT}, respectively. A zoom to the high-rate region is also included. It is shown that optimizing the compression schemes for the distribution of the tone-reserved signal improves the MER performance at high resolutions. Moreover, the performance gain of the MER increases significantly as the power constraint P_{PC} or the percentage of reserved tones for TR N_{PRT} increases.

This can be explained by the fact that higher rates, i.e., a larger number of quantization levels covering the finite quantization region, reflect the ability of the optimized codebook quantization levels to exploit the statistical changes in the tail of the distribution. Indeed, the role of the optimized compression techniques becomes more efficient when the contribution of the minor modes becomes more significant by increasing the P_{PC} or the N_{PRT}. For a $P_{\mathrm{PC}} = 10$ dB and a PRT ratio of 1%, as defined in the DVB-T2 specifications, the optimized compression techniques improve the MER by about 1.6 dB compared to the purely Gaussian compression technique, when on average 7 bits per IQ sample are used.

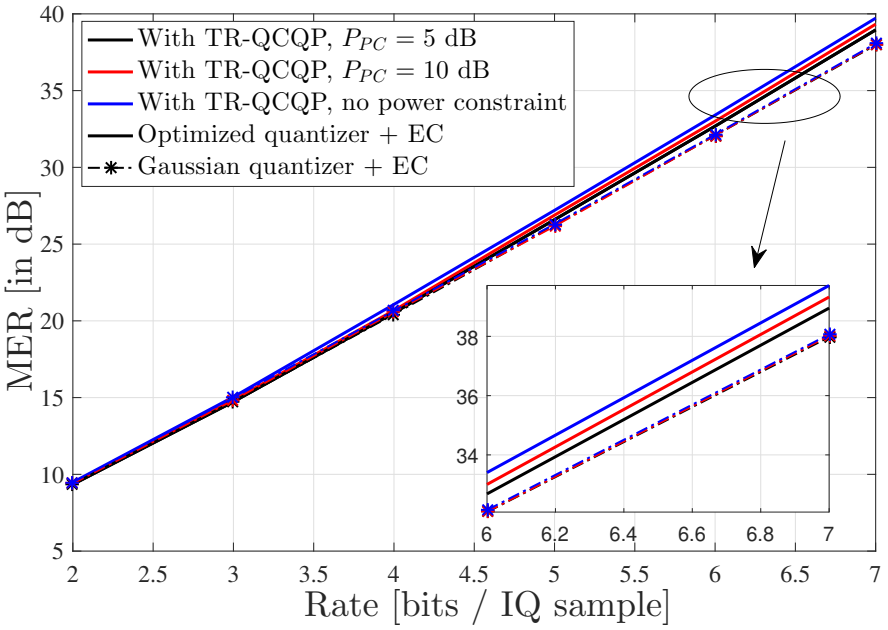


Figure 3.17: MER of Gaussian and optimized compression techniques vs resolution for different power constraints P_{PC} and $N_{\mathrm{PRT}} = 1\%$.

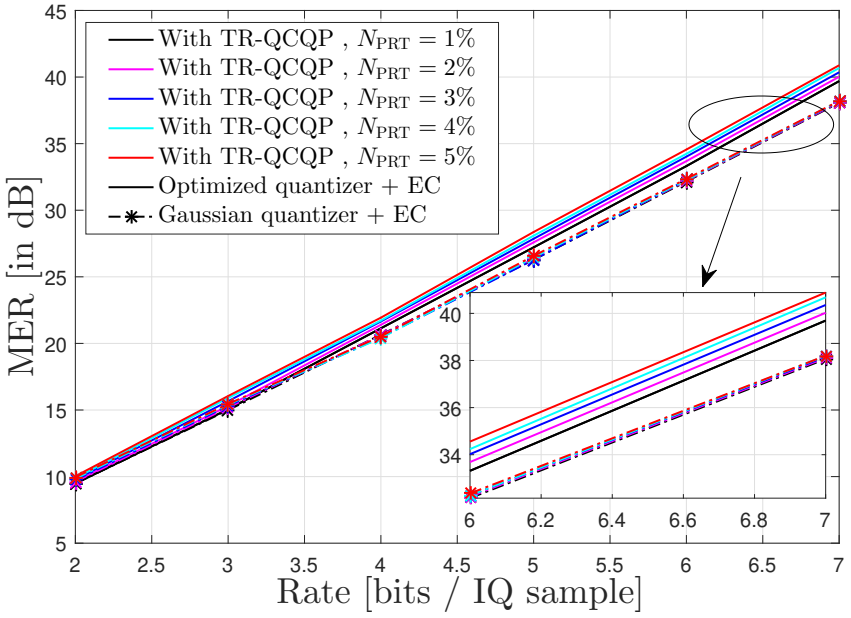


Figure 3.18: MER of Gaussian and optimized compression techniques vs resolution for different percentages of PRTs N_{PRT} and $P_{\mathrm{PC}} = 10$ dB.

3.7 Conclusion

In this chapter, we studied the effect of reducing the PAPR of an OFDM signal in the time-domain before compression in a C-RAN architecture. The analytical distributions of the PAPR-reduced OFDM IQ samples were derived and validated by simulation analysis for both clipping and TR-QCQP PAPR reduction techniques. The obtained IQ distributions were used to optimize the codebook levels of a non-uniform scalar quantizer and the average number of bits allocated by the entropy encoder, in addition to the analysis of the MER, modeling the joint effect of PAPR reduction and quantization operations.

First, by analyzing the theoretical Shannon bound for the obtained distributions of the clipped and tone-reserved signals, we show how the compression techniques take advantage of changing the distribution of the signal to be compressed. Simulation results have shown that the Gaussian-optimized quantizer is not very robust when applied to a clipped signal. Moreover, the gain of using an optimized quantizer considering the clipping operation is highlighted, especially in the low-rate quantization region. However, the gain decreases significantly in the high-rate region. In the case of the TR-QCQP algorithm, the Gaussian quantizer works properly when applied to a tone-reserved signal, since the distribution of the latter is close to a Gaussian one. However, we show that the optimized quantizer based on the distribution of the tone-reserved IQ samples fulfils a significant gain in the high-rate quantization region, especially when the number of tones reserved for TR is high and the power constraint is high.

Thus, the importance of optimizing the compression techniques based on the true distribution of the input signal has been shown. Moreover, it is shown that it is not always straightforward to derive the PDF that characterizes the input signal. Therefore, the next chapter explores how the learning-based techniques can help to jointly obtain the optimized PAPR-reduced and compressed IQ samples without the need to know the distribution of the PAPR-reduced signal for compression.

Chapter 4

Joint PAPR reduction and signal compression based on deep neural networks

4.1 Introduction

In the last two chapters, we have studied the compression of time-domain OFDM IQ before transmission over the fronthaul link in the C-RAN architecture. In addition, we have studied the impact of reducing the PAPR of the OFDM signal before compression. It has been shown that knowing the true distribution of the input signal is important for optimising compression techniques.

In recent years, DNN has led to many breakthrough results in machine learning and has been widely used in industry. Advances in neural network training have helped to improve performance over conventional methods in numerous domains, including wireless communications.

In particular, the idea of end-to-end learning of communication systems by DNN and especially the Auto-Encoder (AE) architecture has recently shown great promises. Unlike conventional communication systems, such a design enables joint optimization of different blocks for any differentiable specific task without being limited to component-wise optimization. This approach breaks the limitations that exist in conventional block-based signal processing by moving from hand-crafted, carefully optimized sub-blocks to adaptive and flexible Neural Network (NN).

Therefore, in this chapter, we first propose an auto-encoder-based neural network in Section 4.2 to reduce the PAPR of the OFDM system and further improve the performance of the conventional techniques.

Then, to go further in the analysis, we develop the network to jointly reduce the PAPR and quantize the OFDM signal by using an AE-based neural network deployed in the C-RAN architecture in Section 4.3. The idea of the compression network for the C-RAN architecture is straightforward. The proposed auto-encoder is jointly trained with a multi-objective optimization function to obtain the PAPR-reduced signal, followed by the proposed quantization network to map the features of the PAPR-reduced signal into a finite set and then the quantized features are encoded by an entropy coding method, e.g., Huffman coding. Finally, the decoder reverses the compression process.

4.2 PAPR reduction based on deep neural networks

At the beginning, we examine the ability of the AE-based architecture to reduce the PAPR of the OFDM signal as preliminary results before presenting quantization-based compression for the C-RAN architecture. We first explain the overall structure of the proposed PAPR reduction network and explain each of its modules in detail. We then present the training process for the model to achieve the required performance in reducing the PAPR of the OFDM signal without affecting the system performance.

4.2.1 Structure of the feedforward neural networks

The proposed model for PAPR reduction is based on the Feedforward Neural Network (FFNN), a.k.a., Multi Layer Perceptron (MLP), which is explained in more detail below.

4.2.1.1 Feedforward neural networks

The FFNN is a DNN architecture consisting of an input layer, several hidden layers, and an output layer, as shown in Fig. 4.1. Each layer contains multiple neurons, and the neurons of adjacent layers are interconnected, while the neurons of the same layer are independent. Thus, the architecture of the FFNN, denoted by F, consists of L_{lay} layers and describes a mapping function $F(\mathbf{r}_o;\boldsymbol{\theta}) : \mathbb{R}^{d_0} \mapsto \mathbb{R}^{d_L}$ of an input vector $\mathbf{r}_0$ to an output vector through L_{lay} iterative processing steps:

$$\mathbf{r}_l = f_l(\mathbf{r}_{l-1};\boldsymbol{\theta_l}); \quad l = 1, \cdots, L_{\text{lay}}, \tag{4.1}$$

where $f_l(\mathbf{r}_{l-1};\boldsymbol{\theta_l}) : \mathbb{R}^{n_{l-1}} \mapsto \mathbb{R}^{n_l}$ is the mapping performed by the lth layer, n_l and n_{l-1} are the number of neurons in the lth and $l-1$th layers, respectively. This mapping depends on the output vector of the previous layer and the set of parameters of the layer, i.e., $\boldsymbol{\theta}_l$. We denote the matrix $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_L]$ as the matrix containing all parameters of the network.

The lth layer is called *dense* or *fully connected (FC)* when the basic task of the neuron is to multiply its input by the appropriate weight and add bias, i.e., $f_l(\mathbf{r}_{l-1};\boldsymbol{\theta_l})$ has the form

$$f_l(\mathbf{r}_{l-1};\boldsymbol{\theta_l}) = \sigma(\mathbf{W}_l\mathbf{r}_{l-1} + \mathbf{b}_l), \tag{4.2}$$

where $\mathbf{W}_l \in \mathbb{R}^{n_l \times n_{l-1}}$ is the weight matrix, $\mathbf{b}_l \in \mathbb{R}^{n_l}$ is the bias vector of the FC layer and $\sigma(\cdot)$ is the nonlinear activation function, which processes the output of the neuron and introduces nonlinearity for the NN, such as sigmoid, tanh and ReLU, etc. The activation function is applied individually to each element of the input vector. Thus, the parameters for the dense layer is generally $\boldsymbol{\theta}_l = [\mathbf{W}_l, \mathbf{b}_l]$.

Backpropagation has been considered as a valid training algorithm for the deep-based neural networks [69, Ch. 6.5]. It adaptively adjusts the parameters of the network in $\boldsymbol{\Theta}$, using the Gradient Descent (GD) algorithm to minimize the loss function. The parameters of the network are usually learned w.r.t training data $\mathbf{x}_i \in \mathbb{R}^{d_0}, \forall i = 1, \cdots, N_{\text{TS}}$ and labels $\mathbf{y}_i \in \mathbb{R}^{d_l}, \forall i = 1, \cdots, N_{\text{TS}}$ by minimizing a real-valued loss function $\boldsymbol{\mathcal{L}}(\mathbf{x}_i, \mathbf{y}_i; \boldsymbol{\Theta}), \forall i = 1, \cdots, N_{\text{TS}}$, where N_{TS} is the length of

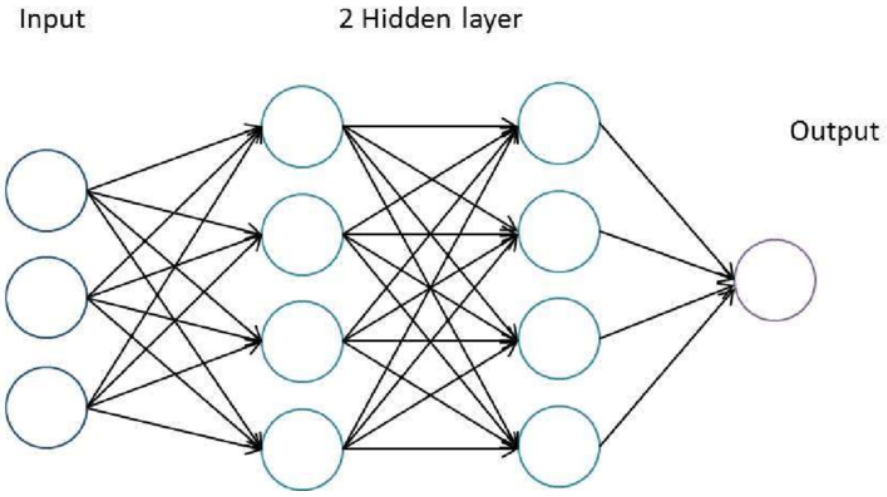


Figure 4.1: An example of a FFN with two hidden layers of 4 neurons each.

the training data. Typically, the loss function can be decomposed as a sum over the training data as follows:

$$\mathcal{L}(\boldsymbol{\Theta}) = \frac{1}{N_{\mathrm{TS}}} \sum_{i=1}^{N_{\mathrm{TS}}} l(F(\mathbf{x}_i; \boldsymbol{\Theta}), \mathbf{y}_i), \tag{4.3}$$

where $l(F(\mathbf{x}; \boldsymbol{\Theta}), \mathbf{y}) : \mathbb{R}^{d_0} \times \mathbb{R}^{d_L} \mapsto \mathbb{R}$ is the samples loss.

The GD algorithm starts with some random initial values for each parameter of the network $\theta = \theta_0$ and then iteratively updates θ as

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta \mathcal{L}(\theta_t), \tag{4.4}$$

where $\eta > 0$ is the learning rate and $\mathcal{L}(\boldsymbol{\theta})$ is the loss function defined in (4.3) considering the whole data set and ∇_θ denotes the gradient operation $\frac{\partial \mathcal{L}(\theta)}{\partial \theta}$.

The most commonly used algorithm for finding a good set of parameters $\boldsymbol{\Theta}$ is the mini-batch GD. The mini-batch GD algorithm is implemented when an approximation of the cost function $\tilde{\mathcal{L}}(\boldsymbol{\Theta})$, given for a random mini-batch of training examples, denoted by Bs, replaces the cost function in (4.3) at each iteration, that is,

$$\boldsymbol{\Theta}_{t+1} = \boldsymbol{\Theta}_t - \eta \nabla \tilde{\mathcal{L}}(\boldsymbol{\Theta}_{\boldsymbol{t}}), \tag{4.5}$$

where

$$\tilde{\mathcal{L}}(\boldsymbol{\Theta}) = \frac{1}{Bs} \sum_{i=1}^{Bs} l(F(\mathbf{x}_i; \boldsymbol{\Theta}), \mathbf{y}_i). \tag{4.6}$$

Thus, Updating the cost function in the mini-batch GD algorithm is significantly less computationally expensive than in the GD algorithm by choosing BS small compared to N_{TS}, but this reduces the variance of updating the network parameters. Several variants of the GD algorithm have been proposed to dynamically adjust the learning rate to improve convergence, such as Adam, Adagrad and RMSProp [69, Ch. 8.5].

4.2.1.2 Auto-encoder feedforward neural network

AE is a neural network architecture that is trained to attempt to reconstruct the original input signal [69, Ch. 14]. The general structure of the AE consists of two main components: the encoder part and the decoder part, which are trained jointly to minimize a certain loss function between the input of the encoder and the output of the decoder.

The encoder and the decoder can be modeled mathematically as two mapping functions: $E(\mathbf{x}; \mathbf{W}_1) : \mathbb{R}^m \rightarrow \mathbb{R}^n$ and $D(E(\mathbf{x}); \mathbf{W}_2) : \mathbb{R}^n \rightarrow \mathbb{R}^m$, where $\mathbf{x}$ and $E(\mathbf{x})$ denote the original and mapped encoded signals, respectively, and $\mathbf{W}_i, i \in \{1, 2\}$ denotes their corresponding weights.

The goal of the AE is to reconstruct the input signal by jointly minimizing a certain loss function between the input of the encoder and the output of the decoder, i.e., $\mathcal{L}(\mathbf{x}, D(E(\mathbf{x}; \mathbf{W}_1); \mathbf{W}_2))$. The choice of the loss function depends on the nature of the features of the input signal $\mathbf{x}$.

An interesting architecture for the AE is to constraint the hidden layer at the output of the encoder $E(\mathbf{x}) \in \mathbb{R}^m$, known as *latent space* representation, to a smaller dimension than $\mathbf{x} \in \mathbb{R}^n$, i.e., $m < n$. This architecture is called *undercomplete AE* .

Thus, the undercomplete AE is a neural network with dimension reduction, where the hidden layer with lower dimension used to represent the input is usually called the *bottleneck* layer. In particular, the number of neurons in the bottleneck layer is smaller than the number of neurons in the input and output layers to reduce the dimension of the data. This forces the AE to capture the most salient features of the training data distribution rather than learning to copy it perfectly.

AE has recently been used in various digital communication tasks. It is further investigated to model the end-to-end communication systems. In particular, the encoder and decoder of an AE are designed to represent the transmitter and receiver of a communication system, respectively [70], [71], [72]. In addition, AE is often used for denoising the corrupted data [73]. Therefore, it may be suitable for dealing with the nonlinear distortions caused by reducing the PAPR of the OFDM signal.

4.2.2 State of the art in PAPR reduction schemes based on deep learning

Traditional PAPR reduction techniques presented in Section 3.3 have experimentally been proven to have certain limitations in PAPR reduction performance [74]. Therefore, in recent years, the ANN-based models have been proposed to break the limitations of the traditional techniques in reducing the PAPR of the OFDM signal.

In recent years, deep Learning based on MLP architecture has shown significant performance improvement over traditional methods. In [75] and [76], the authors consider an ANN, to reduce the complexity of clipping and filtering and ACE conventional PAPR reduction methods, respectively. Specifically, they design a NN that uses the time-domain OFDM signal as the training input signal and the time-domain PAPR-reduced signals as the desired output signal and the goal was to train the network to act as a clipper or an ACE mapper to reduce the PAPR of the OFDM signal with less complexity. For this reason, they cannot outperform the conventional methods.

To break the performance limitation of conventional methods, the authors in [74] propose a PAPR reduction scheme, namely *PRNet*, based on a MLP AE. The model consists of a PAPR reduction module at the transmitter and the inverse decompression module at the receiver with identical structures. The transmitter module consists of five FC layers, each with 2048 neurons and ReLU activation function. They process the signal in the frequency-domain before the IFFT transformation, i.e., the network learns the constellation mapping and demapping of the symbols on each subcarrier in an OFDM system. Therefore, the number of neurons in the input and output layers is set to $M * N_{\text{fft}}$, where N_{fft} is the number of subcarriers for each OFDM symbol and M is the order of the modulation scheme used.

The training objective of the PRNet is to recover the original OFDM signal distorted during PAPR reduction by minimizing the MSE of the system at the output of the decoder and to produce a low PAPR transmission signal at the output of the encoder. However, both the complexity of the model and its efficiency are significantly affected by the number of subcarriers in the OFDM system. Due to the NN model and the split loss function, this method is superior to the conventional methods. Moreover, compared to other NN-based methods discussed later, the PRNet shows better performance in PAPR.

The authors in [77] propose to move the PAPR reduction to the time-domain after the IFFT transformation, using a simpler ANN model based on the AE structure. The module for PAPR reduction consists of two FC layers with 10 and 5 neurons, respectively. The input and output layers then contain only 2 nodes corresponding to the IQ components of each complex OFDM sample. The training objective of this module is to reduce the PAPR of the OFDM signal and minimize the out-of-band spectrum.

The decompression module on the receiver also consists of an identical structure to the PAPR reduction module. The training objective of the decompression module is to recover the original OFDM signal that was distorted during PAPR reduction to minimize the MSE of the system. Finally, the two modules are trained together, combining both objectives to achieve good PAPR reduction and BER performance.

Thus, the idea of their proposed model is to reduce the processing rate of the model by using multiple parallel processing links to reduce the PAPR of each complex sample of the OFDM symbol. Moreover, the proposed method can be executed with negligible overhead after the training is completed, so that real-time operation is possible.

In [78], the authors propose a novel tone reservation network, namely *TRNet*, based on deep learning to improve the PAPR reduction performance of the conventional TR technique. As shown in Section 3.3.2, there are suboptimal solutions for the TR algorithm with lower complexity than the QCQP optimization algorithm.

Therefore, the authors propose a feedforward neural network that adaptively generates the PRTs of the peak cancellation signal of the TR algorithm according to the characteristics of the input signal, without signal distortion. Compared to the AE-based PAPR reduction methods, their proposed network is used only at the transmitter, which reduces the cost of the system. The training objective of the network is to find the peak cancellation signal that reduces the PAPR of the transmitted tone-reserved signal. The network consists of five layers, i.e., an input layer, three hidden layers and an output layer. This method outperforms the suboptimal tone reservation

techniques.

Finally, in [79], the authors propose a tone reservation scheme based on the online sequential Extreme Learning Machine (ELM) [80] with a single-hidden layer feedforward neural network to improve the PAPR reduction performance of the suboptimal TR schemes. The ELM is a feedforward machine learning model based on the topological structure of MLP. Its distinctive feature is that it has only one layer of hidden nodes, where the weights between the input layer and the hidden layer are randomly assigned and never tuned, while the weights between the hidden and the output layer are usually learned only once in a single step with the matrix inverse using least square methods, without the need for a repetitive gradient-based backpropagation algorithm [80].

The ELM has the advantages of simple structure, fast learning speed, and strong generalization ability. They used a suboptimal Clipping Control Tone Reservation (CC-TR)algorithm [81] to generate the peak cancellation signals for the OFDM training symbols. These peak cancellation signals are then used as targets to train the sequential ELM network online until a certain training accuracy is achieved. Their method outperforms traditional suboptimal TR schemes. The PAPR performance is lower than that of the deep AE-based PAPR reduction technique, i.e., PRNet, proposed in [74], but at the gain of much lower complexity.

4.2.3 Structure of the proposed network

An overview of the OFDM system structure, including the proposed neural network for PAPR reduction, is given in Fig. 4.2. Specifically, the proposed model includes the OFDM system, the PAPR reduction module to reduce the PAPR of the transmitted signal and the decompression module at reception to reconstruct the original signal.

In the OFDM system, the OFDM signal $\mathbf{x} \in \mathbb{C}^{N_{\text{fft}}}$ is generated by an IFFT, which maps the M-QAM data symbols $\tilde{\mathbf{x}} \in \mathbb{C}^{N_{\text{fft}}}$ to N_{fft} subcarriers, as stated in (1.2). Current deep learning libraries do not support direct handling of complex numbers [82]. Therefore, we alternatively consider a mapping of the complex-valued OFDM signal $\mathbf{x} \in \mathbb{C}^{N_{\text{fft}}}$ into a real-valued vector $\mathbf{x} \in \mathbb{R}^{2N_{\text{fft}}}$, which can be interpreted as a concatenation of the in-phase and Quadrature components of the OFDM signal.

The encoder and decoder have identical architectures, as shown in Fig. 4.2. They consist of four layers, i.e., an input layer, two hidden layers and an output layer. The number of neurons in the input layer is set to twice the number of subcarriers $2N_{\text{fft}}$ corresponding to the IQ components of each OFDM symbol. The output layer is set to a smaller dimension N_b than the input layer, i.e., $N_{\text{b}} < N_{\text{fft}}$. Finally, each hidden layer consists of a FC layer, a Batch Normalization (BN) layer and the Scaled Exponential Linear Unit (SELU) function, as shown in Fig. 4.2.

The core of the FC layer is the matrix multiplication between the input to the FC layer and the parameters of the layer, as explained in (4.2). Then, the BN layer normalizes the output of the FC layer, to train the network more efficiently by avoiding the vanishing gradient problem [1] and increasing the rate of training convergence.

[1]When the NN learns with the gradient-based backpropagation algorithm. The network parameters are updated in the directions of the gradients. Backpropagation can cause the gradients to become too small, leading to a problem called *Vanishing gradients*. This causes the weights to not adjust and the learning to stop [83].

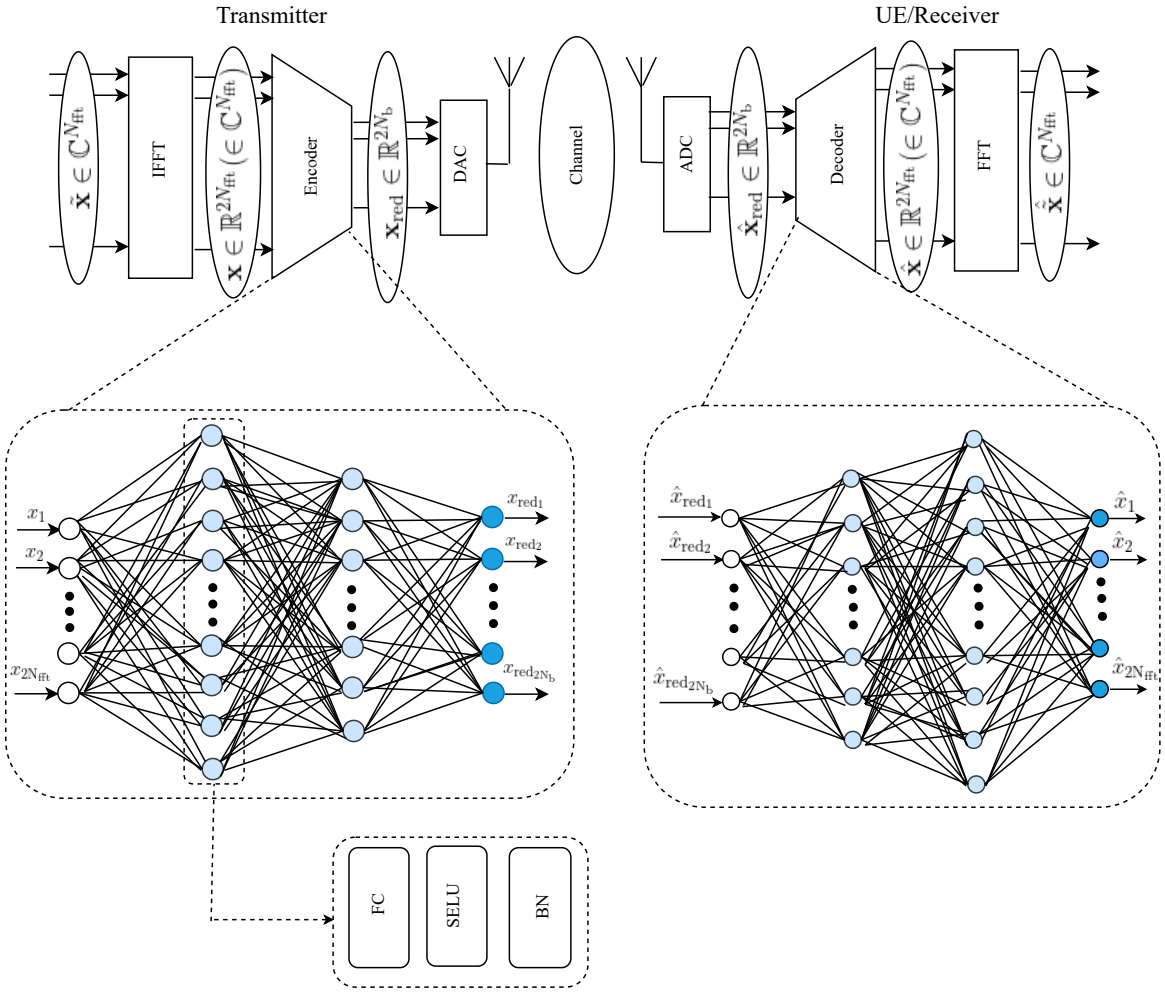


Figure 4.2: The architecture of the AE-based PAPR reduction module.

More precisely, BN can be expressed mathematically as:

$$\hat{\mathbf{y}}_{\mathrm{FC}}^{(k)} = \gamma \frac{\mathbf{y}_{\mathrm{FC}}^{(k)} - \mathbb{E}[\mathbf{y}_{\mathrm{FC}}^{(k)}]}{\sqrt{\mathrm{var}(\mathbf{y}_{\mathrm{FC}}^{(k)}) + \epsilon}} + \beta, \tag{4.7}$$

where $\mathbf{y}_{\mathrm{FC}}$ denotes the output signal of the FC layer, var(.) is the variance of the signal, ϵ is a constant small value that avoids division by zero [69] and γ and β are adjustable learnable scaling and shift parameters whose values are found through training and increase the expressiveness of the neural network. This helps the neural network adaptively normalize the distribution during training, which accelerates convergence.

The number of neurons in the two FC layers is set to 2048 and 2000, respectively. The SELU is used as the activation function, which can be expressed as follows:

$$f(x) = \begin{cases} \lambda x; & if\ x \geq 0 \\ \lambda\alpha(e^x - 1); & if\ x < 0, \end{cases} \tag{4.8}$$

with $\alpha \sim 1.6733$ and $\lambda \sim 1.0507$ known as scaling constants and helps to bring the output normalization close to the normal distribution [84].

SELU is an activation function that induces *self-normalization*, i.e., SELU output activation automatically converges to a normal distribution with zero mean and unit variance. Thus, the SELU, in conjunction with batch normalization, is characterized by very good performance, which is higher than that of other activation functions [85]. Moreover, it has no problem with vanishing gradient in deep neural networks.

To achieve the best performance of the SELU, the hidden layer weights must be initialized with the leCun normal initialization, which is defined as [84]

$$\mathbf{W}_l \sim \mathcal{N}(\mu = 0, \sigma = \sqrt{1/n_{\mathrm{par}}^{(l)}}), \tag{4.9}$$

where $n_{\mathrm{par}}^{(l)}$ is the number of trainable parameters of the hidden layer l.

The generated PAPR-reduced signal $\mathbf{x}_{\mathrm{red}} \in \mathbb{R}^{2N_{\mathrm{fft}}}$ is transmitted over the wireless channel. It is worth mentioning that a normalization layer is added to the PAPR-reduced signal to force the average transmitted power of the signal to the value of one. Thus, the network does not try to increase the power, but learns the features of the channel to be able to reconstruct the signal.

Finally, after down-conversion and analogue-to-digital conversion, the received signal is denoted as $\hat{\mathbf{x}}_{\mathrm{red}}$, as shown in Fig. 4.2, passed to the decoder of the AE to reconstruct the original signal $\hat{\mathbf{x}} \in \mathbb{R}^{2N_{\mathrm{fft}}}$.

4.2.4 Complexity analysis

The prediction complexity of NN-based algorithms depends on the number of layers and the number of neurons in each layer. The complexity analysis in terms of the big $\mathcal{O}$ notation of the proposed network can be illustrated as follows.

The training complexity of the FFNN based on the MLP NN with L_{lay} hidden layers can be computed as $2 \times N_{\mathrm{TS}} \times \mathcal{O}(n_i n_1 + n_1 n_2 + \cdots + n_{L_{\mathrm{lay}}} n_o)$, where N_{TS} is the number of training samples, n_i is the number of neurons in the input layer,

$n_m,\ m = 1, \cdots, L_{\text{lay}}$, is the number of neurons in the m-th hidden layer and n_o is the number of neurons in the output layer.

In the case of adding a BN layer in the MLP NN, the trainable parameters of the network include not only the weights and biases, but also the scaling and shift parameters in the batch normalization. Consequently, the training complexity of the MLP network with BN can be calculated as follows: $2 \times N_{\text{TS}} \times \mathcal{O}((n_i + 2)n_1 + (n_1 + 2)n_2 + \cdots + (n_{L_{\text{lay}}} + 2)n_o)$. Finally, the complexity of the prediction process of a MLP network is $\mathcal{O}(n_i n_1 + n_1 n_2 + \cdots + n_{L_{\text{lay}}} n_o)$. Note that the constant terms are omitted in these analyzes.

In summary, the complexity of the training depends mainly on the number of training samples and the number of trainable parameters. In particular, the number of trainable parameters depends on the network structure, the number of layers and the number of neurons in each layer.

Implementing the proposed AE in the time-domain rather than in the frequency-domain, as proposed in the PRNet, helps to significantly reduce the complexity of the network. This is because the dimension of the feature space is reduced only to the number of subcarriers per OFDM symbol, rather than the order of the modulation scheme times the number of subcarriers.

On this basis, the proposed MLP scheme has lower computational complexity than the other schemes based on a MLP neural network, except [77], which proposes a very simple network but at the cost of using more processors.

4.2.5 Training of the proposed network

The proposed AE network is to be trained to reduce the PAPR of the transmitted signal while maintaining its quality. The original OFDM signal $\mathbf{x}$ passes through the encoder module to produce the PAPR-reduced signal $\mathbf{x}_{\text{red}}$. Then, the signal passes through the channel to obtain the $\hat{\mathbf{x}}_{\text{red}}$ signal, which is finally passed through the decoder module to reconstruct the original OFDM signal $\hat{\mathbf{x}}$.

Thus, the objective function of the network is twofold. First, the network must reduce the PAPR of the transmitted signal at the output of the encoder $\mathbf{x}_{\text{red}}$. Second, the network must be able to reconstruct the signal at the output of the decoder without degrading the signal quality.

Therefore, the appropriate loss function to jointly achieve the stated objectives can be explained as follows. The loss function used to optimize the reconstruction of the signal is the MSE between the input signal and the reconstructed decoded signal, given by

$$\mathcal{L}_1(\mathbf{x}, \hat{\mathbf{x}}) = \text{MSE}(\mathbf{x}, D(E(\mathbf{x}; \boldsymbol{\Theta}); \boldsymbol{\Phi})) = \frac{1}{N_{\text{fft}}} \sum_{n=0}^{N_{\text{fft}}-1} \|x_n - \hat{x}_n\|^2, \tag{4.10}$$

where $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$ denote the parameters of the encoder and decoder, respectively.

On the other hand, reducing the PAPR of the transmitted signal at the output of the encoder can be achieved as follows, by limiting the PAPR of the output signal of the encoder $\mathbf{x}_{\text{red}}$:

$$\mathcal{L}_2(\mathbf{x}_{\text{red}}) = |\text{PAPR}(E(\mathbf{x}; \Theta)) - \text{PAPR}_t| = |\text{PAPR}\mathbf{x}_{\text{red}} - \text{PAPR}_t|, \tag{4.11}$$

where

$$\text{PAPR}\mathbf{x}_{\text{red}} = \frac{\max\limits_{k=0,\cdots,N_{\text{fft}}-1}\left[\|x_{\text{red}}(k)\|^2\right]}{\mathbb{E}\left[\|x_{\text{red}}(k)\|^2\right]}, \tag{4.12}$$

where PAPR_t is the target PAPR. During training, $\mathcal{L}_2(\mathbf{x}_{\text{red}}) \to 0$, so the PAPR of the signal $\mathbf{x}_{\text{red}}$ is close to the target to achieve the purpose of PAPR reduction.

Finally, the proposed network trains the encoder and decoder parameters, i.e., Θ and Φ jointly, by considering the joint loss function $\mathcal{L}(\mathbf{x}, \hat{\mathbf{x}})$, which combines both $\mathcal{L}_1$ and $\mathcal{L}_2$ such that both MSE and PAPR can be minimized, expressed as follows:

$$\Theta^*, \Phi^* = \arg\min_{\Theta,\Phi} \frac{1}{N_{\text{TS}}} \sum_{i=1}^{N_{\text{TS}}} \mathcal{L}_1(\mathbf{x}, D(\mathbb{H}(E(\mathbf{x};\Theta));\Phi)) + \alpha\mathcal{L}_2(E(\mathbf{x};\Theta)), \tag{4.13}$$

where $\mathbb{H}$ denotes the impact of the wireless channel and α denotes the weighted hyperparameter that balances the two different loss components, i.e., $\mathcal{L}_1$ and $\mathcal{L}_2$.

4.2.6 Performance analysis

In this section, simulation results are presented to demonstrate the performance of the AE proposed scheme in terms of the CCDF of PAPR and BER. In addition, comparisons are made with conventional techniques including clipping and QCQP-based TR technique.

4.2.6.1 Simulation parameters and discussion of network hyperparameters

We consider an OFDM frame structure with $N_{\text{fft}} = 64$ subcarriers and 64-QAM modulation scheme. During the training of the network, 100,000 OFDM symbols are used for training, 50,000 symbols are used for validation and 50,000 symbols are used for testing to generate all simulation results.

The configuration of the network is summarized in Table 4.1. The network is modeled in Python software using the TensorFlow deep learning framework module [86]. Based on the joint cost function in (4.13), the parameters of the network are determined using the Adaptive Moment Estimation (ADAM) optimization algorithm [87].

We consider AWGN for the wireless channel, i.e., $\mathbb{H}$ when training the network according to (4.13). The value chosen for training for the energy-per-bit to noise ratio E_b/N_0 is 15 dB, which gives the optimal performance when the network is trained for E_b/N_0 in the range of 5 to 30 dB with a step of 5 dB. In the inference phase, the trained AE is then tested for the entire range of the E_b/N_0.

Moreover, training E_b/N_0 is an important factor for convergence to the optimal solutions. In particular, when training with low E_b/N_0, neural networks have difficulty learning the data structure due to the high noise power. On the other hand, if the training value E_b/N_0 is too high, neural networks lose the ability to generalize noisy data.

The key parameters of the neural network, including the number of layers and neurons used for AE simulations, are determined by numerical optimization to find the best possible training accuracy. Moreover, the way the bottleneck dimension N_b is chosen will be made clear in the upcoming compression analysis in the next sections.

The batch size (Bs) is the number of training samples used for each learning iteration to update the parameters of the network in the mini-batch GD algorithm. The number of epochs indicates how many times the neural network sees the entire training samples again and again to improve the precision of the network.

Table 4.1: Configuration of the proposed PAPR reduction model

Parameter	Value
Framework	Tensorflow
Programming language	Python 3.9
Modulation scheme	64-QAM
Number of subcarriers per OFDM symbol (N_{fft})	64
Number of neurons in the input and output layers	128, 112
Number of hidden layers	Two layers
Number of neurons per hidden layer	2048 and 2000, respectively
Activation function of the hidden layer	SELU
Number of training OFDM symbols	100,000
Number of test OFDM symbols	50,000
Optimization algorithm	ADAM algorithm
Learning rate (η)	1e-4
Batch size (Bs)	16000
Number of epochs	4000
Weight parameter (α)	0.001

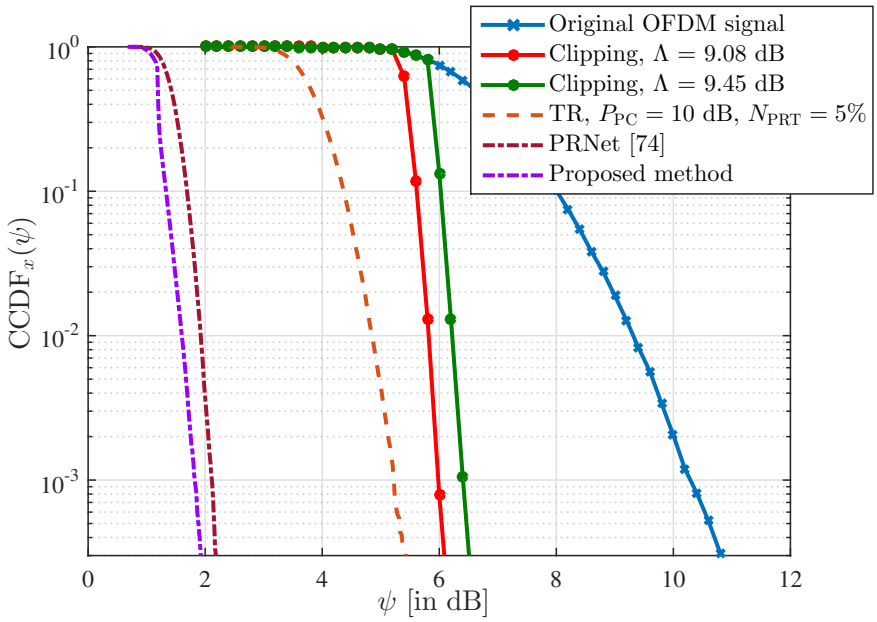
4.2.6.2 CCDF for PAPR reduction analysis

The CCDF of the PAPR of the encoder output signal of the proposed model is used to test the performance of the PAPR reduction defined according to (1.6). Experimental results show that CCDF and BER can achieve good results when α is at the order of magnitude of 0.001. For comparison, we consider the conventional clipping and TR-QCQP PAPR reduction schemes, described in Sections 3.3.1 and 3.3.2, respectively, in addition to the neural network based approach proposed in [74], i.e., PRNet.

Fig. 4.3 shows the CCDF of the PAPR of the PAPR-reduced signal $\mathbf{x}_{\text{red}}$ of the proposed scheme. It can be seen that at CCDF $= 1 \times 10^{-2}$, the network can achieve a PAPR reduction of 7.7 dB compared to the OFDM signal. Moreover, the proposed AE-based network outperforms both clipping and TR-QCQP conventional schemes without the complexity and convergence time of the TR algorithm or the distortion of the clipping algorithm, which will be clearly seen in the BER analysis in the next section. Finally, it has slightly better PAPR performance than the ANN-based PRNet algorithm, but with lower computational complexity, as explained in Section 4.2.4.

4.2.6.3 Bit error rate analysis

Fig. 4.4 shows the BER of the proposed architecture for PAPR reduction compared to the conventional techniques in AWGN channel. Considering the tradeoff between



)

Figure 4.3: CCDF of PAPR for proposed and conventional PAPR reduction schemes.

the PAPR reduction and BER performance, the target PAPR for the proposed network is set to $\text{PAPR}_t = 1.2$ dB in the joint loss function according to (4.11).

Although the clipping scheme can reduce the PAPR of the OFDM signal, as the reduction increases by increasing the clipping ratio Λ, the BER of the system is degraded due to the distortion of the signal by the PAPR reduction process, which is the disadvantage of the clipping scheme. TR-QCQP is a distortionless scheme due to the orthogonality between the data and the tone-reserved signals in the frequency-domain. While, the power loss observed with the TR algorithm is due to the additional power allocated to the PRts for PAPR reduction. At constant transmit power, this means that less power is allocated to the data subcarriers, resulting in higher sensitivity to the AWGN. The power loss in dB for a given percentage of PRTs N_{PRT} and P_{PC} is given by:

$$10\log_{10}\frac{(100 - N_{\text{PRT}}) + N_{\text{PRT}} \times 10^{P_{\text{PC}}/10}}{100}. \tag{4.14}$$

From Fig. 4.4, it can be seen that compared to clipping and PRNet, the BER of the proposed network is reduced by 7 and 10 times, respectively, with the same E_b/N_0 of 10 dB. This is because in the proposed AE-based network, the signal reconstruction function of the decoder allows the PAPR of the transmitted signal to be reduced to a greater extent without distortion.

Finally, Fig. 4.5 shows the BER of the proposed network during the training of the network at different corruption levels E_b/N_0. It can be seen that training the network at $E_b/N_0 = 15$ dB over the whole SNR range leads to a lower BER than in other cases, which justifies the choice of this value for our analysis.

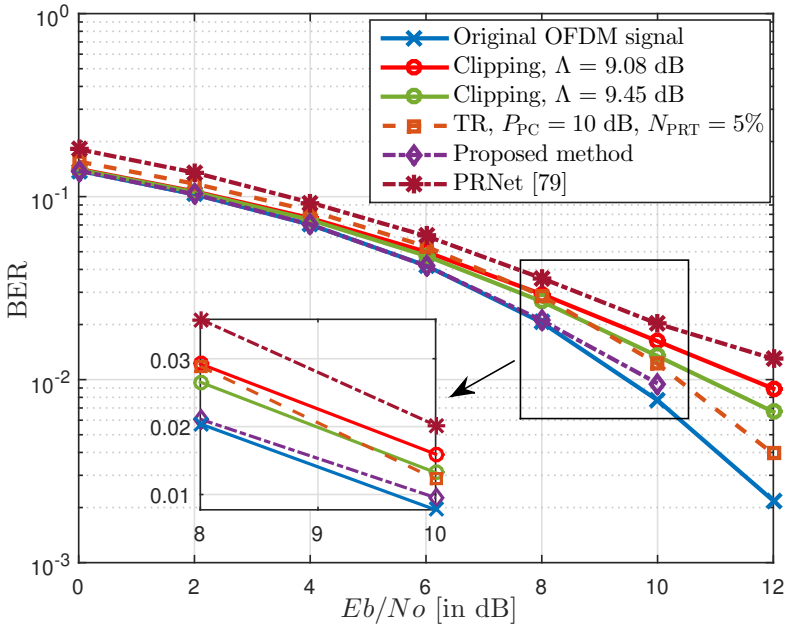


Figure 4.4: The BER performance of the different algorithms over AWGN channel for 64-QAM modulation.

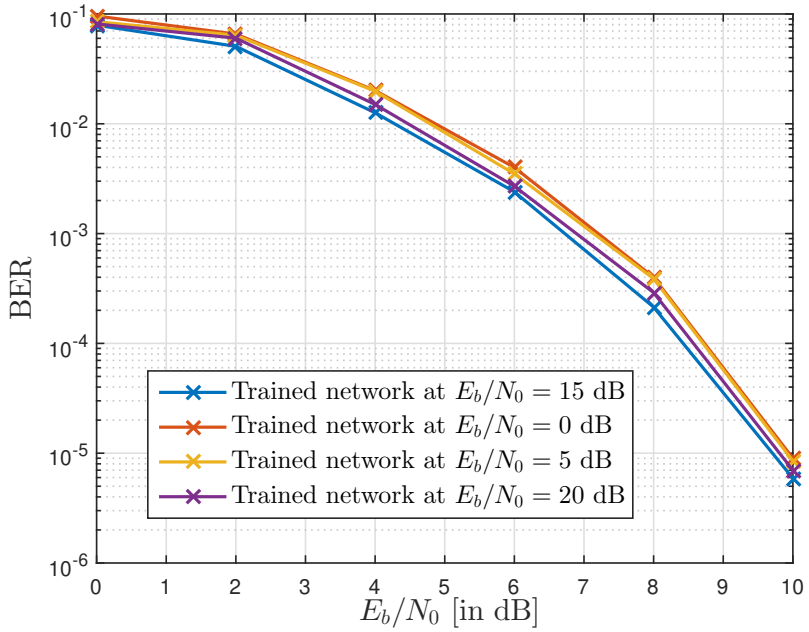


Figure 4.5: BER of the proposed NN trained at different E_b/N_0.

4.3 Proposed joint PAPR reduction and compression network for C-RAN architecture

In this section, we extend our previous work in the perspective of developing a joint PAPR reduction and compression scheme for the context of the C-RAN architecture. We first present the general AE architecture of the proposed joint model and then describe the proposed quantizer, followed by an illustration of the joint training.

4.3.1 Challenges

End-to-end AE-based neural network optimization in the context of compression has recently been investigated for lossy image compression. The quantizer is considered an integral part of the compression pipeline. However, it is inherently non-differentiable function, i.e., it generates zero gradients almost everywhere. Therefore, despite the effectiveness of deep learning studies in this area, the optimization of the quantizer along with the auto-encoder becomes a challenge for joint end-to-end training.

This generally renders the optimization algorithms ineffective, e.g., gradient descent, which requires almost-everywhere differentiability of the loss function with respect to the trainable parameters. To solve this problem, the studies propose to apply an approximately differentiable continuous relaxation of the quantization operation.

In [88], the authors propose to approximate the quantization function by a continuously differentiable function. Their assumption is valid only for uniform scalar quantization. Given a sufficiently long training sequence, they assume without loss of generality that the size of the quantization bins is always one and the quantization values lie in the middle of the bins, i.e., the quantization is simply a rounding function to the nearest integer $x_q = \text{round}(x)$.

The derivatives of the rounding function $x_q = \text{round}(x)$ w.r.t the network parameters are zero. To allow optimization over SGD, the deterministic quantizer is replaced by an additive i.i.d uniform noise source (Δx) that has the same width as the quantization bins (one), i.e., $x_q = x + \Delta x$, which simulates the quantization operation. This idea is based on the fact that in information theory the uniform noise is usually used as a model for the quantization error [89].

The authors in [90] propose to replace the derivative of the rounding function in the backward pass of backpropagation by 1. In other words, they do not completely replace the rounding function with a smooth approximation, but only its derivative in the backpropagation, which means that the quantization in the forward pass is still performed as usual.

It should be noted that the gradient can be calculated with respect to the parameters of the decoder without resorting to approximations. Thus, the idea proposed in [90] has the advantage that it does not change the decoder's gradients, since the forward pass remains the same, unlike [88].

In the previous deep compression methods, the quantizer is often fixed during training and the regular quantization is used during inference because the quantization is not differentiable. As a result, the quantization intervals and quantized levels are not optimized. Furthermore, it is unclear to what extent this discrepancy between

training and testing phases affects performance.

4.3.2 The structure of the proposed joint model

4.3.2.1 Auto-encoder structure

The architecture of our AE-quantizer network for the C-RAN is shown in Fig. 4.6. For a given training dataset $\mathcal{X} = \mathbf{x}_{i=1}^{N_{\text{TS}}}$, the goal is to learn a joint PAPR reduction and compression AE module for the C-RAN architecture consisting of an undercomplete encoder, an NN-based quantizer and a decoder.

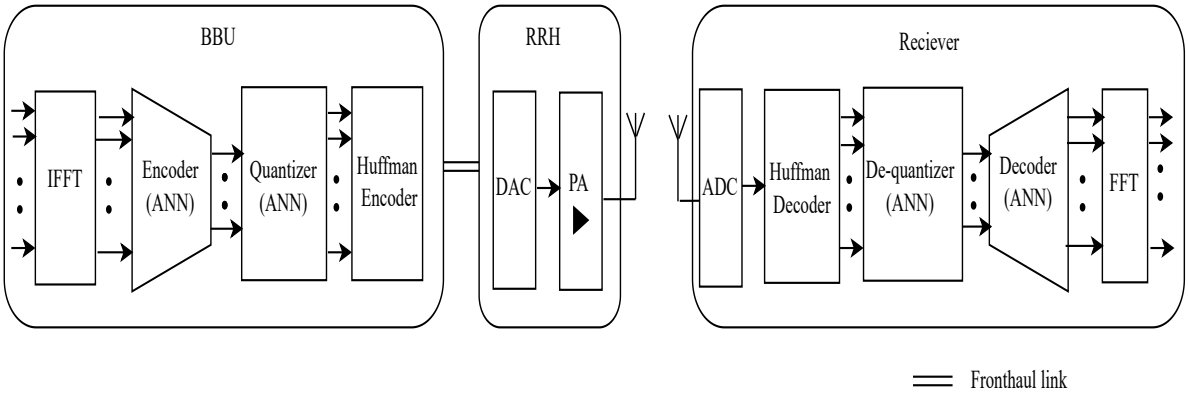


Figure 4.6: The architecture of the proposed joint PAPR reduction and compression network for the C-RAN architecture.

The encoder part of the AE is located in the BBU and generates the latent representation of the time-domain OFDM signal $\mathbf{x}$, denoted by $\mathbf{x}_{\text{red}} = E(\mathbf{x})$. The encoder is responsible for reducing the PAPR of its output signal which is also compressed to a lower dimension. The encoder is followed by the quantization network Q, which discretizes the PAPR-reduced signal, i.e., it represents each PAPR-reduced sample/vector with one of N quantization levels. Then, the index of the representative quantization level is losslessly entropy coded.

The compressed PAPR-reduced signal is then transmitted over the fronthaul link to the RRHs, where it passes through the power amplifier and is prepared for wireless transmission. Finally, the deocer part of the AE is placed at the end user to reconstruct the original OFDM signal from the dequantized signal $\hat{\mathbf{x}} = D(\hat{\mathbf{x}}_{\text{red}})$, where $\hat{\mathbf{x}}_{\text{red}}$ denotes the recovered PAPR-reduced signal decoded from the bitstream.

The generation of the OFDM signal and its preprocessing before applying it to AE has already been explained in Section 4.2.3. We use the same setting previously used in the PAPR reduction network. The structure of the encoder and decoder is summarized in Table 4.2.

4.3.2.2 ANN-based quantizer and de-quantizer models

The algorithm is developed by implementing the iterative conventional Lloyd algorithm as layers of the deep learning network. The quantizer $Q : \mathbb{R}^{2N_b} \mapsto [N]^J$ begins by

Table 4.2: Proposed deep-based neural network for the C-RAN architecture

Layer Name	Type	Output Shape
Encoder input	Input	$(Bs, 2N_{\text{fft}})$
Encoder 1	FC, BN, activation: SELU	$(Bs, 2048)$
Encoder 2	FC, BN, activation: SELU	$(Bs, 2000)$
Encoder output	FC, BN, activation: Linear	$(Bs, 2N_{\text{b}})$
Quantizer input	Reshape	(Bs, k, J)
Quantizer, layer 1	Time Distributed, Euclidean distance	(Bs, J, N)
Quantizer, layer 2	Time Distributed, softmax	(Bs, J, N)
Quantizer, layer 3	Time Distributed, argmin	(Bs, J, N)
De-quantizer, layer 1	FC, activation: Linear	(Bs, k, J)
De-quantizer output	Reshape	$(Bs, 2N_{\text{b}})$
Huffman encoding	Non-trainable	
Decoder input	Input	$(Bs, 2N_{\text{b}})$
Decoder 1	FC, BN, activation: SELU	$(Bs, 2000)$
Decoder 2	FC, BN, activation: SELU	$(Bs, 2048)$
Decoder output	FC, activation: Linear	$(Bs, 2N_{\text{fft}})$

reshaping the PAPR-reduced signal $\mathbf{x}_{\text{red}} \in \mathbb{R}^{2N_b}$ into a matrix $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_J] \in \mathbb{R}^{k \times J}$, where k is the dimension of the vector to be quantized, i.e., $k = 1$ refers to scalar quantization and $J = \frac{2N_b}{k}$ is the number of vectors per OFDM symbol.

Then, the quantizer tries to quantize each column $\mathbf{z}_j; j = [1, \ldots, J]$ to the nearest quantization level of the N-levels codebook $\mathcal{C} = \{\mathbf{y}_1, \cdots, \mathbf{y}_N\} \subset \mathbb{R}^k$, denoted as $\mathbf{y}_{n_j}, n = [1, \cdots, N]$.

Thus, we can say that the PAPR-reduced vector $\mathbf{x}_{\text{red}} \in \mathbb{R}^{2N_b}$ can be modeled as a sequence of J vectors partitioned into the Voronoi tessellation over the centers $\mathcal{C}$. More precisely, a given vector $\mathbf{z}_j \in \mathbb{R}^k$ is quantized to the index of its nearest neighbor, i.e., $Q(\mathbf{z}_j) = n_j$. Moreover, the index n_j can be represented with a fixed codeword length of $R = \frac{\log_2(N)}{k}$ bits per scalar sample.

Finally, the dequantizer $Q^{-1} : [N]^J \mapsto \mathbb{R}^{2N_b}$, first dequantizes each vector by selecting its corresponding quantization level from its index, i.e., $\hat{\mathbf{z}}_j = Q^{-1}(n_j) = \mathbf{y}_{n_j}$. Then it constructs the PAPR-reduced quantized signal $\hat{\mathbf{x}}_{\text{red}}$ by transforming $\hat{\mathbf{Z}} \in \mathbb{R}^{k \times J}$ back to $\mathbb{R}^{2N_b}$.

NN-based quantizer structure: The structure of the proposed NN-based quantizer is shown in Fig. 4.7. We recall that the quantizer assigns a given vector $\mathbf{z}_j$ to the index of the quantization level satisfying the nearest neighbor condition, stated as follows

$$n_j = \arg\min_{n} \quad \|\mathbf{z}_j - \mathbf{y}_n\|_2^2; \quad n = 1, 2, \cdots, N, \tag{4.15}$$

where $\|\cdot\|$ denotes the Euclidean distance between $\mathbf{z}_j$ and $\mathbf{y}_n \in \mathbb{R}^k$, expressed as $d_{j,n} = \sum_{i=0}^{k-1} |z_{j,i} - y_{n,i}|^2$.

On this basis, the first layer of the quantizer is designed to consist of N neurons, each representing one of the quantization levels. The parameters of the layer are set

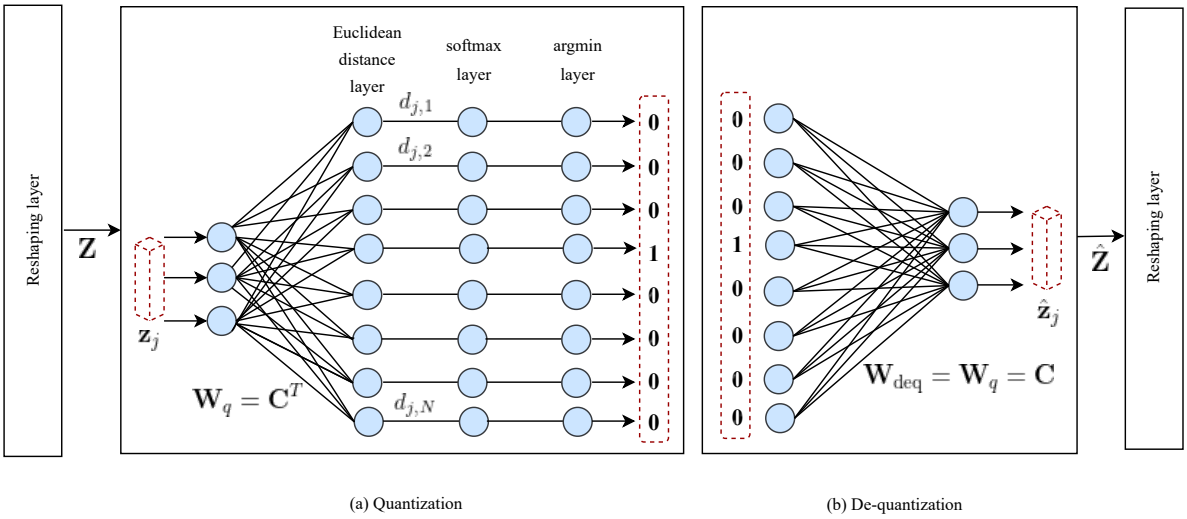


Figure 4.7: The architecture of the quantizer and de-quantizer.

to the codebook, i.e., $\mathbf{W}_q = \mathbf{C}^T \in \mathbb{R}^{N \times k}$ and the activation function is the Euclidean distances between the given vector $\mathbf{z}_j$ and the N quantization levels, denoted as $\mathbf{d}_j = [d_{j,1}, d_{j,2}, \cdots, d_{j,N}]$. The weights $\mathbf{W}_q$ are initialized with the random coding of the PAPR-reduced encoder output $\mathbf{x}_{\text{red}}$, refer to Section 1.4.4.2, which was beneficial for the convergence of the codebook $\mathbf{C}$ during the backpropagation process.

Then we used the *softmax* operator defined as $\text{softmax}(d_{j,1}, d_{j,2}, \cdots, d_{j,N})_l = \frac{e^{d_{j,l}}}{\sum_{n=1}^{N} e^{d_{j,n}}}$, to convert the distances $\mathbf{d}_j$ into probabilities. The result is thus a probability vector of size N. The chosen quantization vector corresponds to the index of the lowest probability in the output vector, i.e., the one with the smallest Euclidean distance (nearest neighbor condition).

The activation function $\min(\cdot)$ is used to obtain the index of the chosen quantization level, which produces a one-hot encoding vector $\mathbf{h}_j = [0, \cdots, 1, \cdots, 0] = \arg\min(\text{softmax}(\mathbf{d}_j))$, in which the index of the chosen quantization level n_j is the only non-zero element in the vector and is equal to one.

From an implementation perspective, end-to-end training requires the availability of the gradient of each layer in the computational graph in order to apply gradient descent training. Since the $\min(\cdot)$ operator is not differentiable w.r.t $\mathbf{d}_j$ or the parameters of the layer that feed into $\mathbf{d}_j$, we set the derivative of the $\min(\cdot)$ operator to 1 during backpropagation, similar to other methods facing the same problem in [90], [91].

In contrast to the implementation of the hard mapping function as adapted in our quantization network, the authors in [92] propose the implementation of a soft-to-hard scheme to implement a vector quantization model where the quantization levels are optimized along with the network parameters.

Their idea is based on soft assignments of the given scalar or vector to be quantized, rather than the hard nearest neighbor condition with a parameter that controls the "hardness" of the assignments. This introduced parameter is annealed during training and allows the gradual transition from soft to hard assignments. Unlike rounding-

based or stochastic quantization schemes commonly used in the literature, this scheme is differentiable and thus end-to-end trainable.

NN-based de-quantizer structure: The input to the de-quantizer Q^{-1} is the one-hot encoding vector $\mathbf{h}_j$, where the true position represents the index of the chosen quantization level $\mathbf{y}_{n_j}$. The de-quantizer can simply recover the quantized value with a fully-connected layer by setting the matrix of layer parameters as follows: $\mathbf{W}_{\text{deq}} = \mathbf{W}_q = \mathbf{C} \in \mathbb{R}^{k \times N}$, i.e., $\hat{\mathbf{Z}}_j = \mathbf{C}\mathbf{h}_j^T = \mathbf{y}_{n_j}$ [2].

Finally, after de-quantization, the reconstructed vectors $\hat{\mathbf{Z}}_j$ are reshaped back into the PAPR-reduced quantized vector $\hat{\mathbf{x}}_{\text{red}} \in \mathbb{R}^{2N_b}$ which is then fed into the decoder to reconstruct the OFDM signals.

Based on this structure, the quantization levels $\mathbf{C}$ can be optimized jointly with other AE parameters. The detailed settings of the AE-quantization network for the C-RAN architecture shown in Fig. 4.6 are summarized in Table 4.2. Here Bs in the first dimension of the output shape is the mini-batch size determined by the GD algorithm and N_b is the lower dimension of the bottleneck layer.

During inference, the parameters of the encoder, quanatizer and decoder are set after optimization. In addition, Huffman coding is used to losslessly encode the quantized discrete values into a bitstream based on their probabilities.

4.3.3 Training of the proposed joint model

The proposed AE network is deployed in a C-RAN architecture to optimize end-to-end performance. Our objective is to optimize the parameters of the encoder, quantizer and decoder to achieve good PAPR reduction and compression performance.

Thus, the joint objective function of the network can be stated as follows. First, the MSE distortion measure is used to minimize the distortion between the original signal and the reconstructed signal at the output of the decoder, as specified in the loss function $\mathcal{L}_1(\mathbf{x}, \hat{\mathbf{x}})$ defined in (4.10).

In addition, the PAPR of the transmitted signal $\mathbf{x}_{\text{red}}$ must be minimized at the output of the encoder, as specified in the loss function $\mathcal{L}_2(\mathbf{x}_{\text{red}})$ defined in (4.11). To achieve good compression performance, the bit rate of the quantized signal $\hat{\mathbf{x}}_{\text{red}}$ must be minimized.

In general, the rate of the quantized signal can be measured by the entropy, which is calculated based on the population distribution of the quantized bins as follows:

$$H(\hat{X}_{\text{red}}) = -\sum_{i=1}^{N} \mathbb{P}[Y = y_i] \log_2 \mathbb{P}[Y = y_i], \tag{4.16}$$

where $\mathbb{P}[Y = y_i]$ denotes the probability of the ith quantization level. $\mathbb{P}[Y = y_i]$ can be estimated by counting the number of samples following cluster i over the entire sample set, i.e.,

$$p_i = \mathbb{P}[Y = y_i] = \frac{|\mathcal{S}_i|}{JN_{\text{TS}}}, \tag{4.17}$$

[2]The constraint that $\mathbf{W}_{\text{deq}} = \mathbf{W}_{\text{q}}$ can be removed to allow the network to learn the quantization levels for reconstruction at the de-quantizer.

where $\mathcal{S}_i$ is the set of vectors assigned to the ith quantization level and $|\mathcal{S}_i|$ is its cardinality.

End-to-end training of the proposed network requires a differentiable loss function w.r.t the arguments of the network. It is evident that p_i is discrete and consequently it is impossible to minimize the entropy defined in (4.16) w.r.t the network parameters. Therefore, we define a soft differentiable histogram that depends on the network parameters. This can be achieved by summing the partial assignments to each quantization level instead of counting as defined in (4.17).

We used the Euclidean distance between each training vector and the set of quantization levels during training to represent the partial assignment to each quantization level. Then, the negative exponential function is used to map this vector of N mappings into a valid PMF: $\mathbb{R}^N \rightarrow (0,1)^N$ by applying it to each element of the N mappings and normalizing these values by the sum of all exponentials. On this basis, the soft histogram is defined as follows:

$$q_i = \mathbb{Q}[Y = y_i] = \frac{\sum_{j=1}^{J_i} e^{-d_{j,i}}}{\sum_{n=1}^{N}\sum_{j=1}^{J_n} e^{-d_{j,n}}}, \tag{4.18}$$

where J_i denotes the number of training vectors assigned to the ith quantization level, i.e. $\sum_{i=1}^{N} J_i = JN_{\mathrm{TS}}$.

Rate loss can now be redefined as the cross entropy between the true PMF $\mathbb{P}$ defined in (4.17) and the estimated PMF $\mathbb{Q}$ in (4.18) as follows:

$$\mathcal{L}_3(\hat{\mathbf{x}}_{\mathrm{red}}) = \tilde{H}(\hat{X}_{\mathrm{red}}) = H(p,q) = -\sum_{i=1}^{N} p_i \log_2 q_i = H(p) + D_{\mathrm{KL}}(p||q), \tag{4.19}$$

where $D_{\mathrm{KL}}(p||q) = \sum_i p_i \log(p_i/q_i)$ denotes the Kullback-Leibler divergence. Since $D_{\mathrm{KL}}(p||q) \geq 0$, the cross entropy $H(p,q)$ is considered an upper bound for the marginal entropy $H(p)$ with equality when $p = q$. Therefore, $H(p)$ can be minimized indirectly by minimizing the differentiable loss function $H(p,q)$ w.r.t. q, while treating p as a constant for gradient calculations.

In summary, to optimize the encoder $E(\cdot)$, the quantizer $Q(\cdot)$ and the decoder $D(\cdot)$ so that both the PAPR reduction and the bit rate are minimized, the joint objective loss function can be formulated as follows

$$\begin{aligned}\Theta^*, \mathbf{W}_q^*, \Phi^* = \underset{\Theta, \mathbf{W}_q, \Phi}{\arg\min} \frac{1}{N_{\mathrm{TS}}} \sum_{i=1}^{N_{\mathrm{TS}}} & \mathcal{L}_1(\mathbf{x}, D(\tilde{Q}(E(\mathbf{x};\Theta));\Phi)) + \alpha\mathcal{L}_2(E(\mathbf{x};\Theta)) \\ & + \beta\mathcal{L}_3(\tilde{Q}(E(\mathbf{x};\Theta))),\end{aligned} \tag{4.20}$$

where $\tilde{Q} = Q^{-1}(Q(.))$, Θ and Φ denote the parameters of the encoder $E(\cdot)$ and the decoder $D(\cdot)$, respectively, α and β are the hyperparameters that balance the trade-off between the reduction of the PAPR of the transmitted signal and its bit rate, respectively and the quality of the reconstructed OFDM signal.

4.3.4 Performance analysis

4.3.4.1 Simulation parameters and discussion of network hyperparameters

We consider an OFDM frame structure with $N_{\text{fft}} = 64$ subcarriers and 128-QAM modulation scheme. In addition, we place empty subcarriers in about $\sim 4/25$ of the spectrum to have almost the same signal redundancy that characterizes the broadcasting technologies in our analysis. The parameters of the network are determined using ADAM algorithm based on the joint cost function given in (4.20).

The key parameters of the neural network, including the number of layers and neurons used for AE simulations, are determined by numerical optimization to find the best possible training accuracy and illustrated in the configuration of the network summarized in Table 4.3.

Table 4.3: Configuration of the proposed joint PAPR reduction and compression model

Parameter	Value
Framework	Tensorflow
Programming language	Python 3.9
Modulation scheme	128-QAM
Number of subcarriers per OFDM symbol (N_{fft})	64
Number of neurons in the input and output layers	128, 112
Number of hidden layers of AE	Two
Number of neurons per hidden layer	2048 and 2000, respectively
Activation function of the hidden layer	SELU
Number of training OFDM symbols	1000,000
Number of test OFDM symbols	50,000
Optimization algorithm	ADAM algorithm
Learning rate (η)	5e-6
Batch size (Bs)	512
Number of epochs	4000
PAPR weight parameter (α)	0.005
Entropy weight parameter (β)	0.7

4.3.4.2 Rate-distortion and PAPR analysis

Our analysis were performed for a resolution range between 2 and 7 bits per IQ sample. To begin, we typically varied the regularization hyperparameter β of the joint loss function in (4.20) to generate different bit rates. However, in this case, the network tends to reduce the rate of the quantized signal without making any effort to achieve good MER performance with the rate reduction.

Therefore, better and more stable performance was obtained by slightly modifying the rate constraint calculated with the cross entropy defined in the joint loss function in (4.20), denoted $\mathcal{L}_3(\hat{\mathbf{x}}_{\text{red}})$. Instead of minimizing the rate $\tilde{H}(\hat{\mathbf{x}}_{\text{red}})$, we limit the rate of the quantized signal to a target rate R_t, i.e., the rate loss function can be redefined

as follows:

$$\mathcal{L}_3(\hat{\mathbf{x}}_{\text{red}}) = |\tilde{H}(\hat{X}_{\text{red}}) - R_t|,$$
$$\text{where,} \quad \tilde{H}(\hat{X}_{\text{red}}) = -\sum_{i=1}^{N} p_i \log_2 q_i. \tag{4.21}$$

Thus, to generate different bit rates, we vary the number of quantization levels of the quantizer shown in Fig. 4.7 in the range of $N = [4, 128]$ levels. Moreover, we set the regularization parameter β to a higher value of 0.5. Therefore, during training $\mathcal{L}_3(\hat{\mathbf{x}}_{\text{red}}) \to 0$, so that the rate of the quantized signal is close to the target entropy and the network tries to achieve the minimum MSE distortion at this target rate.

For all experiments, the batch size and learning rate were set to $Bs = 512$ and $\eta = 7e - 6$, respectively, which yields the most stable numerical results for all resolutions. For simplicity, the generated quantization levels were encoded using the Huffman algorithm, although other advanced entropy encoding schemes can also be used.

As illustrated in Section 4.3.2.1, the architecture of the encoder is constrained at the bottleneck, which may be forced to discard certain information. The dimension of the feature space at the bottleneck layer is crucial for the tradeoff between MSE, PAPR of the transmitted signal and its rate.

To show how much information is lost by dimensionality reduction, Fig. 4.8 shows the MER achieved by the proposed AE-quantization network for different bottleneck dimensions. These analyses were simulated for a target PAPR of $\text{PAPR}_t = 3.3$ dB and a target rate of $R_t = 2.8$ bits per sample with a one-dimensional quantization structure of $N = 8$ levels.

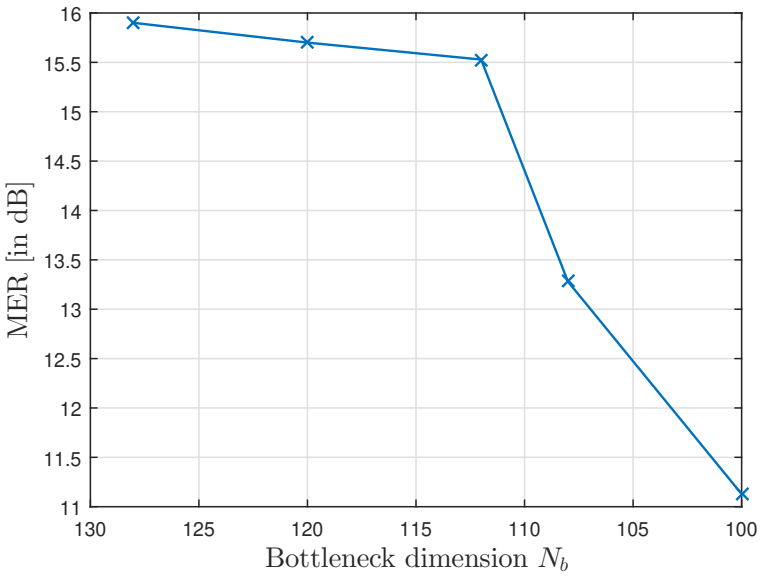


Figure 4.8: MER of the proposed network at different bottleneck dimensions N_b, for $PAPR_t = 3.3$ dB, $N = 8$-levels 1-dimensional quantizer and $R_t = 2.7$ bits/sample.

It can be seen that the MER performance of the proposed network differs significantly at different bottleneck dimensions. More specifically, we can see that a lower dimension than $N_b = 112$ significantly reduces the system performance, so we consider $N_b = 112$ as the bottleneck dimension for all upcoming analysis. It must be clearly stated that reducing the dimension of the bottleneck layer of the AE leads to a reduction in the sampling rate of the transmitted PAPR-reduced signal $\mathbf{x}_{\text{red}}$.

Therefore, the undercomplete structure of the AE assumed for our analysis leads to signal compression at the beginning by reducing the sampling rate of the transmitted PAPR-reduced OFDM signal. The degree of compression can be calculated using the compression ratio already defined in (1.10), i.e.,

$$\begin{aligned} \text{CR} &= \frac{N_{\text{fft}}}{N_b} \quad \text{samples/OFDMsymbol} \\ &= \frac{N_{\text{fft}} R_0}{N_b R_q}, \quad \text{bits/sample} \end{aligned} \tag{4.22}$$

where R_0 and R_q are the original number of bits of the IQ OFDM samples and the entropy of the quantized output signal of the proposed NN $\hat{\mathbf{x}}_{\text{red}}$, respectively. Thus, the AE structure achieves CR = 1.1429.

Fig. 4.9 shows the MER as a function of resolution R for the proposed AE-quantization network implemented for scalar quantization, i.e., the quantization model described in Section 4.3.2.2, is explored with scalar samples $z_j \in \mathbb{R}$ and quantized into 1-dimensional scalar clusters. Moreover, results are shown for different PAPR target values as defined in the joint loss function (4.20).

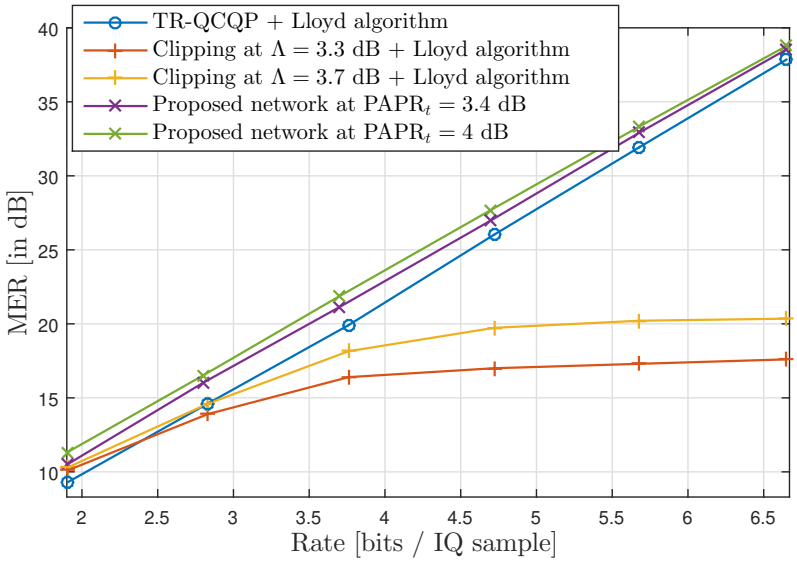


Figure 4.9: MER vs resolution R for the proposed AE-quantization network applied on the OFDM signal for different target PAPR values PAPR$_t$.

According to the objective function of the proposed method, there is a tradeoff between the PAPR reduction performance and the quality of the reconstructed signal.

Therefore, we set the target PAPR values to $\text{PAPR}_t = 3.4$ and 4 dB to evaluate the MER performance of the proposed method under different target PAPR.

For comparison, we used an OFDM system where a PAPR reduction algorithm is first used to reduce the PAPR of the OFDM signal. We consider the clipping and TR-QCQP algorithms. Considering that the simple clipping technique causes signal distortion while the TR-QCQP algorithm is distortionless, in addition the QCQP-based algorithm for TR is considered as an upper bound for the PAPR reduction performance of the TR scheme, but at the cost of its high computational complexity. Then, the conventional Lloyd algorithm is used to quantize the PAPR-reduced OFDM signal, followed by a Huffman entropy coder.

From Fig. 4.9, we can first see that the MER performance of the TR-QCQP signal outperforms the performance of the clipped signal, which is due to the distortions added to the system by clipping. Subsequently, we can see that the proposed AE-quantization network consistently outperforms the compression performance of the conventional Lloyd algorithm applied to the PAPR-reduced signal.

The proposed network at $\text{PAPR}_t = 3.4$ dB, improves the MER by about 1.1 dB compared to the conventional Lloyd algorithm when an average of 3 bits per IQ sample is used. At higher bit rates of 7 bits per IQ sample, the improvement in the MER performance reduced to 0.7 dB. The reason could be that with the proposed network architecture, it is difficult to effectively learn the fine-grained quantization centers for higher bit rates to improve the compression performance.

Finally, increasing the PAPR target value to $\text{PAPR}_t = 4$ dB improves the performance of the MER by 1.90 dB at $R = 3$ bits per sample compared to the conventional Lloyd algorithm and by 0.92 dB at a higher rate of 7 bits per sample. Therefore, the PAPR target value is set according to the system performance requirements of the MER. If the system can tolerate lower MER, we can set a smaller PAPR_t to improve the energy efficiency of the power amplifier.

For a fair and meaningful comparison, Fig. 4.10 shows the CCDF of the PAPR of the encoder's output transmitted signal $\mathbf{x}_{\text{red}}$ at $\text{PAPR}_t = 3.4$ and 4 dB. We have presented the performance of the conventional clipping and TR-QCQP PAPR reduction schemes for comparison.

It can be seen that at $\text{CCDF} = 1 \times 10^{-2}$, the network at $\text{PAPR}_t = 3.4$ dB can achieve a PAPR reduction of 3.2 dB compared to the OFDM signal. It also outperforms clipping and has approximately the same performance as the optimal TR-QCQP algorithm, where the power constraint for PRT subcarriers P_{PC} is set to 10 dB and the number of subcarriers reserved for PAPR reduction relative to the total number of subcarriers N_{PRT} is set to 5%. While, at $P_{\text{PC}} = 10$ dB and a PRT ratio of 1%, as defined in the DVB-T2 specifications, the PAPR performance of the proposed network clearly outperforms the distortionless TR-QCQP algorithm. Moreover, it should be noted that the subcarriers reserved for the TR algorithm do not carry useful information, which leads to a reduction in the throughput of the system, a drawback that is avoided by our proposed network.

When the target PAPR_t is increased to 4 dB to allow more compression performance, as previously shown in Fig. 4.9, the PAPR reduction performance of the network still outperforms clipping and TR-QCQP with a power constraint of $P_{\text{PC}} = 10$ dB and a PRT ratio of 1%.

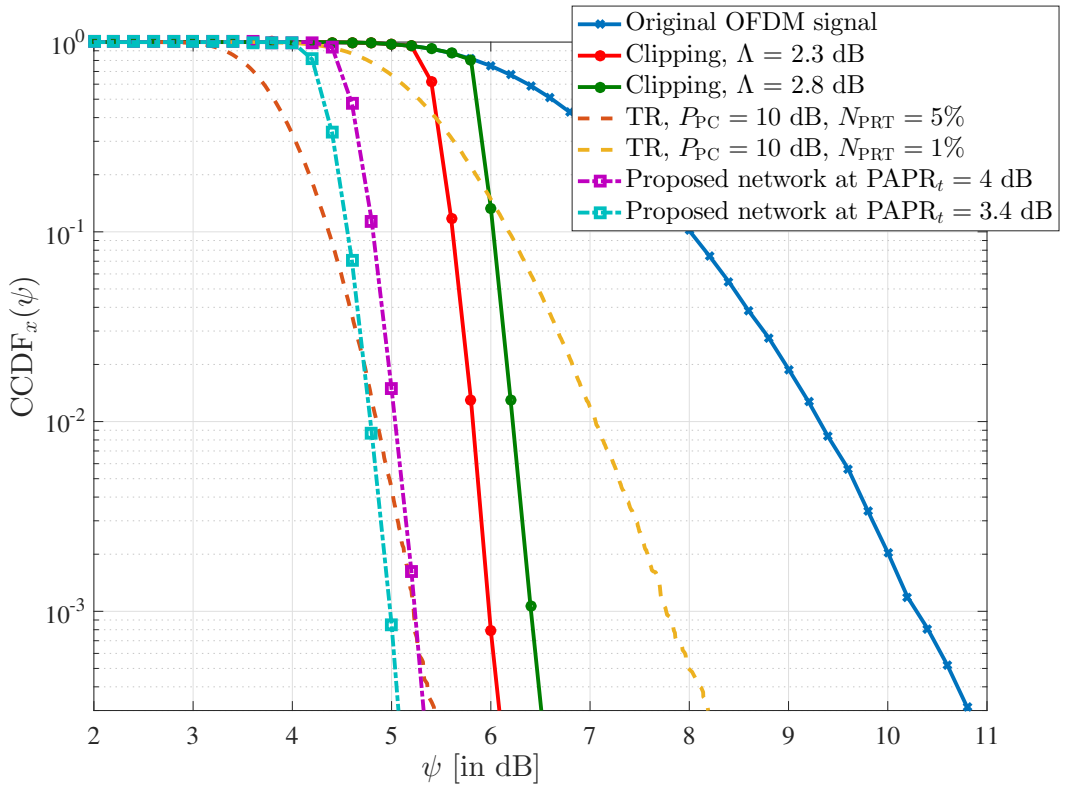


Figure 4.10: CCDF of PAPR for proposed AE-quantization network at different target PAPR values PAPR_t compared with the conventional PAPR reduction schemes.

Finally, for the sake of clarification, the reduction in the sampling rate of the transmitted OFDM signal due to the dimensionality reduction achieved by the under-complete AE structure is equivalent to the decimation process shown in Section 1.4.6, which reduces the sampling rate of the oversampled OFDM signal to approach the Nyquist rate.

In our analysis, we consider spectrum redundancy corresponding to broadcast technologies that have a decimation factor limited to $F \leq \frac{f_s}{f_m} = 1.142$ to preserve the useful data. Therefore, updating the conventional system with a decimation process applied to the OFDM signal before PAPR reduction achieves an additive compression ratio of $\text{CR} = 1.142$. This compression ratio has already been achieved with the AE structure used in our proposed network, as shown previously.

In summary, our proposed AE-quantization network performs the reduction of the sampling rate of the OFDM signal as well as the reduction of the PAPR for efficient transmission and finally the quantization of the signal for compression before transmission over the fronthaul link within the C-RAN architecture. Then the decoder is implemented at the receiver side, which is responsible for the reconstruction of the original OFDM signal.

4.4 Conclusion

In this chapter, we proposed an AE-based neural network that jointly optimizes the encoder, quantizer and decoder to reduce the PAPR of the OFDM system and quantize the signal for compression before transmitting it over the fronthaul link in the C-RAN architecture.

As a preliminary result, we propose an AE-based NN for PAPR reduction. The encoder is implemented in the transmitter to reduce the PAPR of the transmitted OFDM signal. At the receiving side, there is the decoder which is responsible for the reconstruction of the OFDM signal. The considered objective function of the proposed network allows to reduce the PAPR of the signal to a larger extent while ensuring a good MER performance.

The CCDF performance of the proposed model is superior to the conventional techniques of clipping and distortionless efficient TR-QCQP. Moreover, it has similar performance to the neural network based PRNet approach [74] but with much lower computational complexity. Finally, it almost approaches the BER performance of the OFDM signal without PAPR reduction.

We extend the architecture of the proposed AE to include the quantization process to jointly achieve the PAPR reduction and compression optimization adopted in a C-RAN architecture. In the proposed network, the latent representation of the PAPR-reduced OFDM signal is generated at the output of the encoder with a smaller dimension than the one of the input signal, which reduces the sampling rate of the OFDM signal.

Then, the latent representation of the PAPR-reduced signal is fed to the quantization network. The later is designed by implementing the iterative conventional Lloyd algorithm as layers of the learning network. The proposed quantizer and de-quantizer are generated with fully connected layers and are therefore simple and intuitive. Finally, the decoder is optimized to reconstruct the original OFDM signal. With the

proposed architecture, the quantization levels can be optimized together with the encoder and the decoder.

The proposed AE-quantization model is trained simultaneously based on multi-objective optimization. Due to the powerful nonlinear representation capability and end-to-end training of the network, better PAPR reduction in terms of CCDF is achieved than the conventional clipping and TR-QCQP PAPR reduction technique, which is considered as the upper bound for PAPR reduction without distortion.

Moreover, in terms of network rate-distortion analysis, the proposed model gains about 1 dB in the MER performance of the system compared to the conventional Lloyd algorithm in the low-rate quantization region and with about 0.7 dB in the high-rate region, thanks to the ability of the AE in signal reconstruction.

This analysis allows the implementation of the recently introduced flexible concept of end-to-end learning, which has been recently explored in various domains in the consideration and compensation of several effects, such as PAPR reduction and quantization, in a straightforward way.

Conclusions and Perspectives

Conclusions

In this thesis, we investigated the compression of the OFDM IQ data samples prior to transmission over the capacity-constrained fronthaul link in the context of C-RAN architecture. We focused our study on the performance evaluation of the data rate compression techniques from different perspectives, which are summarized as follows:

- We first analyze the compression algorithms, including uniform, non-uniform, scalar and vector quantization, entropy coding and spectral domain redundancy removal algorithms. Then, we propose a simple optimized scalar uniform quantization algorithm coupled with entropy coding over the resulting scalar codebook that efficiently exploits the non-uniform characteristics of the Gaussian distribution of the OFDM signal. It achieves a compression gain of 2.4 dB below the idealized entropy-constrained scalar quantizer, while the complexity is much lower compared to the non-uniform quantization approach. Moreover, we provide a numerical comparison between the decimation process and vector quantization by exploiting the correlations between the OFDM IQ samples. We have shown the impact of the appropriate practical parameter values on the tradeoff between the required signal quality, compression performance and system complexity.

- Second, the OFDM signal is characterized by high PAPR, which requires the use of a PAPR reduction technique to feed the signal at the optimal point of the nonlinear amplifier. This changes the distribution of the OFDM IQ samples, making the Gaussian-optimized compression techniques suboptimal. Therefore, we study the effect of reducing the PAPR of the signal in the time-domain before compression. We derive the analytical distribution of PAPR-reduced OFDM IQ samples for both clipping and TR-QCQP techniques. The derived IQ distributions were used to optimize the codebook levels of a non-uniform quantizer and the average number of bits allocated by the entropy encoder, in addition to the analytical analysis of the MER, which models the joint effect of PAPR reduction and quantization operations. Finally, simulation results validate the obtained distribution of the PAPR-reduced IQ samples. Furthermore, we use these analysis to quantify the performance gained by optimizing the compression techniques based on the true distribution of the PAPR-reduced IQ samples compared to the Gaussian-based distribution.

- We thought of going a step beyond conventional techniques by introducing an AE-based DNN that can be end-to-end jointly trained to obtain an optimized

PAPR-reduced and compressed OFDM IQ samples. We first provide an AE-based network to reduce the PAPR of the OFDM signal and compare its performance with conventional and current DNN-based PAPR reduction techniques. Then, we further develop this proposed AE network to incorporate a quantization network that maps the features of the PAPR-reduced signal to a finite set. Then, the quantized features are encoded by an entropy coding scheme, followed by a decoder responsible for reconstructing the original OFDM signal after the distortions caused by quantization and PAPR reduction. This AE-quantization network is trained based on multi-objective optimization to produce the PAPR-reduced compressed signal. The numerical analysis of the rate-distortion performance of the network and the CCDF of the PAPR shows that the proposed network is able to improve the MER of the system by about 1 dB compared to the conventional Lloyd algorithm and moreover, has approximately the PAPR reduction performance of the conventional most efficient TR-QCQP algorithm.

Perspectives

There are many directions in which further research can be done. In our opinion, it would be interesting to consider the following aspects:

- In Chapter 4, we were able to propose a DNN to compress the OFDM IQ samples in addition to taking advantage of the ability of the AE network to account for multiple effects in a straightforward way, i.e., the concept of end-to-end learning, to additionally reduce the PAPR of the signal and allow the quantizer to adapt to the distribution of the PAPR-reduced signal without analytical analysis. Here, we design the quantization network by unfolding the iterative conventional Lloyd algorithm, in which we solved the problem of undifferentiability of the mapping layer by setting the derivative of the mapping operator to 1 during backpropagation. Thus, one of the challenges that could be addressed in future work is to investigate how to implement a smooth differentiable mapping in the context of end-to-end leaning of the joint model.

- In addition, it may be interesting to investigate whether the MER performance of the system is improved by investigating the vector-based quantizer based on the degree of redundancy generated between the IQ samples of the PAPR-reduced signal in the bottleneck layer at the output of the encoder.

- Moreover, it would be beneficial to consider the nonlinear high power amplifier when training the network. On this basis, the decoder is trained to compensate for the distortions caused by the nonlinearity of the high power amplifier.

- To make our proposed AE-quantization network more efficient for realistic scenarios, increasing the FFT mode is crucial. In such a case, increasing the number of subcarriers per OFDM signal leads to a significant increase in the dimesnion of the feature space, especially for the larger modes, e.g., $16K$ and $32K$. Therefore, the network architecture and complexity must be greatly extended and evaluated to achieve good PAPR reduction and compression performance.

- Self organizing maps are a popular unsupervised machine learning technique used for clustering and visualizing high-dimensional data. It is used as a variant of vector quantization that performs better than the traditional Lloyd algorithm, i.e., the k-mean algorithm, on high-dimensional data [93]. Therefore, we can envision evaluating the usefulness of applying the Self organizing maps to PAPR-reduced signals, especially for higher dimensional analysis.

Appendix

A Huffman algorithm

In 1952, D. A. Huffman [35] developed a scheme that is quite close to Shannon entropy in (1.9). When the input probabilities are powers of $\frac{1}{2}$, the bound is indeed reached. Suppose that the input symbols are ordered by probability, e.g., $\mathbb{P}[e_1] \geq \mathbb{P}[e_2] \geq \cdots \geq \mathbb{P}[e_M]$ and represented as a list $\mathcal{L}_{\text{sym.}}$ as listed in Table A.1. Based on the construction of a binary code tree, the Huffman algorithm for a given source E with a known marginal PMF can be summarized as follows:

1. First, select the two symbols e_i and e_j with the smallest probabilities and create a parent node for all nodes representing these two symbols in the binary code tree.

2. Replace these two symbols e_i and e_j with a new symbol with an associated probability equals to the addition of the probabilities of their parent node, i.e., $\mathbb{P}[e_i] + \mathbb{P}[e_j]$.

3. The previous steps are repeated for the remaining symbols and then iteratively repeated until only one symbol remains.

4. Finally, the constructed tree is converted into a binary code tree. This is done by labeling the branch from the parent to one of the child nodes 1 and the branch from the parent to the other child 0 to produce a uniquely decodable code for a particular PMF for each symbol.

Table A.1: List $\mathcal{L}_{\text{symbols}}$.

Symbol	Probability
e_1	$\mathbb{P}[e_1]$
e_2	$\mathbb{P}[e_2]$
.	.
.	.
.	.
e_{M-1}	$\mathbb{P}[e_{M-1}]$
e_M	$\mathbb{P}[e_M]$

A detailed example of the application of the Huffman algorithm is given in Fig. A.1. Huffman developed and proved the optimality of the aforementioned iterative algorithm satisfying the following two properties: (i) For any two symbols e_i, e_j with

probabilities $\mathbb{P}[e_i] > \mathbb{P}[e_j]$, the associated codeword lengths satisfy $l_{e_i} \leq l_{e_j}$. (ii) There are always two codewords that have the maximum codeword length and differ only in the last bit.

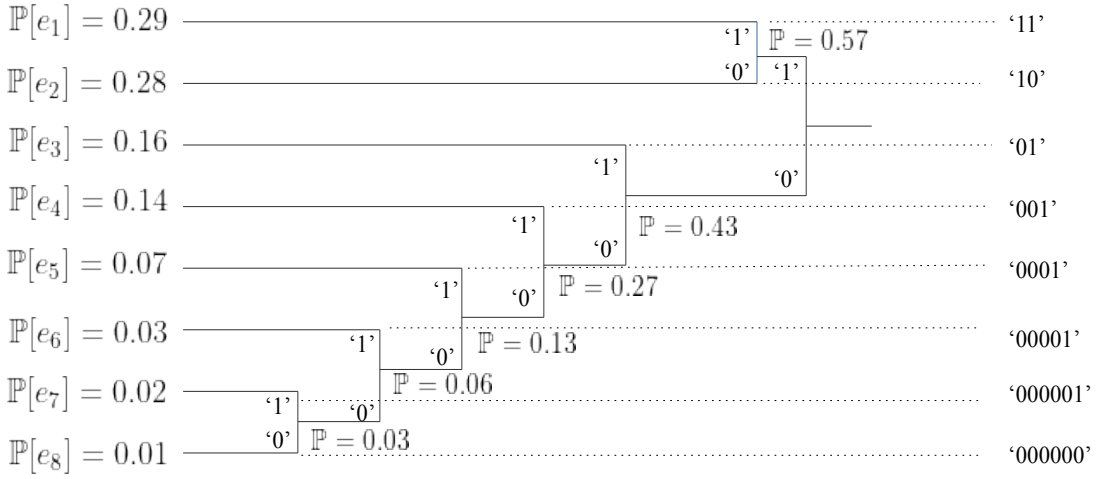


Figure A.1: Example for the design of a Huffman algorithm.

B The distribution of the TR-QCQP PAPR-reduced IQ samples

The multiplication of the Rayleigh distribution by the GEV CDF in the amplitude distribution of the tone-reserved PAPR-reduced signal $\mathrm{f}_{R_{\mathrm{in}}}^{\mathrm{QCQP}}(r)$ in Lemma 3.2 is complex to integrate. Therefore, in [67], the authors propose a suitable approximation that has been numerically tested to ensure a reasonable tradeoff between the complexity of the expression and the accuracy of the computations.

The approximation consists in replacing the GEV CDF in $\mathrm{f}_{R_{\mathrm{in}}}^{\mathrm{QCQP}}(r)$ in Lemma 3.2 by a unit step function truncated at $r_s = \mu_1 + \frac{\sigma_1}{k_1}\left((\ln 2)^{-k_1} - 1\right)$, i.e., $u(r - r_s)$. Thus, $\mathrm{f}_{R_{\mathrm{in1}}}(r)$ in (3.9) in Lemma 3.2 can be expressed by:

$$\mathrm{f}_{R_{\mathrm{in1}}}(r) = \begin{cases} q\, \mathrm{f}_R^{\mathrm{Ray}}(r)(1 - u(r - r_s)); & \textit{if } r \in [0, r_s] \\ 0; & \textit{otherwise.} \end{cases} \tag{4.23}$$

Based on this approximation and remembering that $\mathrm{f}_{R_{\mathrm{in2}}}(r)$ is equal to a GEV distribution over $[0, \mu_2 - \frac{\sigma_2}{k_2}]$ and 0 elsewhere. The respective support of the amplitude distribution of the TR-QCQP signal is redefined over $[0, d^*]$, where $d^* = \max(r_s, \mu_2 - \frac{\sigma_2}{k_2})$. Thus, the evaluation of $\mathrm{f}_{X_1}(x)$ in (3.23) can be developed as follows:

$$\mathrm{f}_{X_1}(x) = \int_{-d^*}^{d^*} \frac{(1-p)q}{2\pi\sigma_{\mathrm{ray}}^2} e^{-\frac{x^2+y^2}{2\sigma_{\mathrm{ray}}^2}}\, u\left(r_s - \sqrt{x^2 + y^2}\right)\, \mathrm{d}y. \tag{4.24}$$

The unit step function redefines the integral over $\left[-\sqrt{r_s^2-x^2}, \sqrt{r_s^2-x^2}\right]$. Then, using the integral identity in [68, eq. (1), Sec. 3.321, p. 336], $\mathrm{f}_{X_1}(x)$ can be expressed as follows:

$$\mathrm{f}_{X_1}(x) = \frac{(1-p)q}{\sqrt{2\pi}\sigma_{\text{ray}}} e^{-\frac{x^2}{2\sigma_{\text{ray}}^2}} \operatorname{erf}\left(\frac{\sqrt{r_s^2-x^2}}{\sqrt{2\sigma_{\text{ray}}^2}}\right), \quad \text{with} \quad r_s = \mu_1 + \frac{\sigma_1}{k_1}\left((\ln 2)^{-k_1} - 1\right). \tag{4.25}$$

In the same way, $\mathrm{f}_{X_2}(x)$ in (3.23) is evaluated from

$$\mathrm{f}_{X_2}(x) = \frac{p}{\pi\sigma_2} \underbrace{\int_0^{d^*} \frac{1}{\sqrt{x^2+y^2}} G(x,y)^{k_2+1} e^{-G(x,y)}\, \mathrm{d}y}_{I}, \tag{4.26}$$

Consider the change of variable $\psi = G(x,y)$ in I, where $G(x,y)$ is in (3.25). Thus,

$$\begin{aligned} \frac{d\psi}{dy} &= \frac{-y}{\sigma_2\sqrt{x^2+y^2}}\left[G(x,y)^{k_2+1}\right], \\ y &= \sqrt{\left(\mu_2 + \frac{\sigma_2}{k_2}\left(\psi^{-k_2}-1\right)\right)^2 - x^2}, \end{aligned} \tag{4.27}$$

and the integral limits are stated as

$$\psi_1 = \left[1 + \frac{k_2}{\sigma_2}(x-\mu_2)\right]^{-\frac{1}{k_2}}, \quad \text{and} \quad \psi_2 = \left[1 + \frac{k_2}{\sigma_2}\left(\sqrt{x^2+d^{*2}} - \mu_2\right)\right]^{-\frac{1}{k_2}}. \tag{4.28}$$

Hence, we get:

$$I = -\sigma_2 \int_{\psi_1}^{\psi_2} \frac{1}{\sqrt{\left(\mu_2 + \frac{\sigma_2}{k_2}\left(\psi^{-k_2}-1\right)\right)^2 - x^2}} e^{-\psi}\, \mathrm{d}\psi. \tag{4.29}$$

Let us assume the following notations

$$c_1 = \frac{\sigma_2^2}{k_2^2}, \quad c_2 = \frac{2\mu_2\sigma_2}{k_2} - 2c_1, \quad \text{and} \quad c_3 = \mu_2^2 - c_1 - c_2 - x^2, \tag{4.30}$$

and with adequately rearranging the terms, I is stated as follows:

$$I = -\sigma_2 \int_{\psi_1}^{\psi_2} \underbrace{\left(c_1\psi^{-2k_2} + c_2\psi^{-k_2} + c_3\right)^{-\frac{1}{2}}}_{I_1} e^{-\psi}\, \mathrm{d}\psi. \tag{4.31}$$

Thus, using the generalized Newton multinomial theorem [94], I_1 can be developed as

$$I_1 = \sum_{p_1=0}^{\infty} \sum_{p_2=0}^{p_1} \binom{-\frac{1}{2}}{p_1} \binom{p_1}{p_2} c_1^{\frac{-(1+2p_1)}{2}} c_2^{p_1-p_2} c_3^{p_2} \psi^{k_2(p_1+p_2+1)}, \tag{4.32}$$

substituting I_1 into I and using the integral property in [68, eq. (2), Sec. 8.35, p. 899], together with substituting I into (4.26) and further simplifying the expression, $\mathrm{f}_{X_2}(x)$ can be expressed as in Theorem 3.2.

C Asymptotic MER and SQNR of a compressed clipped OFDM signal

MER in (1.8) can be shown to be equal to

$$\mathrm{MER} = \frac{\mathbb{E}\left[\|S\|^2\right]}{\mathbb{E}\left[\|S - X_{\mathrm{out}}\|^2\right]} = \frac{P_s}{D_C + D_{\mathrm{LQ}}^{\mathrm{clip}}}. \tag{4.33}$$

For the case of PAPR reduction by clipping, P_s is the power of a complex Gaussian-distributed signal with standard deviation σ_{ray}, i.e., $P_s = 2\sigma_{\mathrm{ray}}^2$. D_C and $D_{\mathrm{LQ}}^{\mathrm{clip}}$ are the powers of the clipping and quantization distortions, respectively, which are considered statistically independent [95]. Let us first derive the clipping distortion D_C, which can be calculated as follows:

$$D_C = \mathbb{E}\left[\|S - X_{\mathrm{in}}^{\mathrm{clip}}\|^2\right] = \mathbb{E}\left[(R_s - R_{\mathrm{in}})^2\right]. \tag{4.34}$$

Expanding the expectation operator and taking the clipping definition in (3.1) into consideration leads to

$$D_C = \mathbb{E}\left[(R_s - R_{\mathrm{in}})^2\right] = \int_{V_{\max}}^{\infty} (r - V_{\max})^2 \mathrm{f}_{R_s}(r)\,\mathrm{d}r. \tag{4.35}$$

After expanding the squared term and using the integral identity in [68, eq. (9), Sec. 3.381, p. 346], D_c can be expressed as

$$D_C = 2\sigma_{\mathrm{ray}}^2 \Gamma\left(2, \frac{V_{\max}^2}{2\sigma_{\mathrm{ray}}^2}\right) - 2^{\frac{3}{2}}\sigma_{\mathrm{ray}} V_{\max} \Gamma\left(\frac{3}{2}, \frac{V_{\max}^2}{2\sigma_{\mathrm{ray}}^2}\right) + V_{\max}^2 e^{-\frac{V_{\max}^2}{2\sigma_{\mathrm{ray}}^2}}. \tag{4.36}$$

The asymptotic scalar quantization distortion $D_{\mathrm{LQ}}^{\mathrm{clip}}$ is derived from the "*Panter and Dite formula*" in Proposition 1.2. The in-phase and Quadrature-phase of a PAPR-reduced complex signal are identical at the input of the quantizer. Therefore, the quantization distortion is the same for both components and can be derived for complex samples as follows:

$$D_{\mathrm{LQ}}^{\mathrm{clip}} = \frac{2^{-2R}}{6}\Bigg(\underbrace{\int_{-\infty}^{\infty} \sqrt[3]{\mathrm{f}_{X_{\mathrm{in}}}^{clip}(x)}\,\mathrm{d}x}_{I}\Bigg)^3. \tag{4.37}$$

The evaluation of (4.37) using the PDF of the clipped IQ samples given in Theorem 3.1 leads to the following integral

$$I = \int_{-V_{\max}}^{V_{\max}} \left(\frac{1}{\sqrt{2\pi\sigma_{\mathrm{ray}}^2}} e^{-\frac{-x^2}{2\sigma_{\mathrm{ray}}^2}} \mathrm{erf}\left(\frac{V_{\max}}{\sqrt{2\sigma_{\mathrm{ray}}^2}}\right) + \frac{e^{-\Lambda}}{2\pi} \frac{1}{\sqrt{{V_{\max}}^2 - x^2}}\right)^{\frac{1}{3}} \mathrm{d}x. \tag{4.38}$$

Let us assume the following notations

$$a = \frac{1}{\sqrt{2\pi\sigma_{\mathrm{ray}}^2}} \mathrm{erf}\left(\frac{V_{\max}}{\sqrt{2\sigma_{\mathrm{ray}}^2}}\right), \qquad b = \frac{1}{2\sigma_{\mathrm{ray}}^2}, \qquad \text{and} \qquad c = \frac{e^{-\Lambda}}{2\pi}. \tag{4.39}$$

Thus, using the generalized Newton binomial theorem [94], I can be developed as

$$\begin{aligned} I &= \int_{-V_{\max}}^{V_{\max}} \sum_{k=0}^{\infty} \binom{\frac{1}{3}}{k} \left(ae^{-bx^2}\right)^{\frac{1}{3}-k} \left(\frac{c}{\sqrt{V_{\max}^2 - x^2}}\right)^k \mathrm{d}x \\ &= \sum_{k=0}^{\infty} \binom{\frac{1}{3}}{k} \int_{-V_{\max}}^{V_{\max}} \frac{\left(ae^{-bx^2}\right)^{\frac{1}{3}-k} c^k}{\left(\sqrt{V_{\max}^2 - x^2}\right)^k} \mathrm{d}x. \end{aligned} \tag{4.40}$$

Rearranging the terms in I leads to

$$I = \sum_{k=0}^{\infty} \binom{\frac{1}{3}}{k} c^k a^{\frac{1-3k}{3}} \underbrace{\int_{-V_{\max}}^{V_{\max}} \frac{e^{-\frac{b(1-3k)}{3}x^2}}{\left(\sqrt{V_{\max}^2 - x^2}\right)^k} \mathrm{d}x}_{I_1}. \tag{4.41}$$

Thus, with proper variable substitution along with the integral property in [68, eq. (1), Sec. 3.383, p. 347], I_1 is solved as

$$I_1 = \frac{1}{b^{\frac{1-k}{2}}} \left[B\left(\frac{2-k}{2}, \frac{1}{2}\right) \left(dV_{\max}^2\right)^{\frac{1-k}{2}} \times {}_1F_1\left(\frac{1}{2}; \frac{3-k}{2}; -dV_{\max}^2\right)\right] \quad \text{with } k \leq 1, \tag{4.42}$$

where $B(u,v) = 2\int_0^1 t^{2u-1}(1-t^2)^{v-1}dt$ is the Beta function, $d = \frac{b(1-3k)}{6\sigma^2}$ and ${}_1F_1$ is the confluent hypergeometric function [68, Sec. 9.18, p. 1019]. Finally, substituting I_1 into I, substituting I into (4.37), and further simplifying the expression, $D_{\mathrm{LQ}}^{\mathrm{clip}}$ can be expressed as

$$D_{\mathrm{LQ}}^{\mathrm{clip}} = \frac{2^{-2R}}{6} \left[\frac{2\sqrt[3]{a}V_{\max}}{\sqrt{6b\sigma_{\mathrm{ray}}^2}} {}_1F_1\left(\frac{1}{2}; \frac{3}{2}; \frac{-V_{\max}^2}{6\sigma_{\mathrm{ray}}^2}\right) - \frac{26cV_{\max}}{25\sqrt{3\sigma_{\mathrm{ray}}^2}\sqrt[3]{a^2}} {}_1F_1\left(\frac{1}{2}; 1; \frac{V_{\max}^2}{3\sigma_{\mathrm{ray}}^2}\right) \right]^3. \tag{4.43}$$

Finally, $P_{z_{\mathrm{in}}}^{\mathrm{clip}}$ is the power of a complex clipped IQ sample, computed with the distribution in Theorem 3.1, i.e.,

$$P_{z_{\mathrm{in}}}^{\mathrm{clip}} = 2\int_{-V_{\max}}^{V_{\max}} x^2 \mathrm{f}_{X_{\mathrm{in}}}^{clip}(x)\,\mathrm{d}x. \tag{4.44}$$

Using the integral identity in [68, eq. (5), Sec. 3.321, p. 336] for the exponential function, $P_{z_{\mathrm{in}}}^{\mathrm{clip}}$ as shown in Lemma 3.3.

List of Acronyms

ACE	Active Constellation Extension
ADAM	Adaptive Moment Estimation
AE	Auto-Encoder
ATSC 3.0	Advanced Television Systems Committee version 3.0
BBU	Baseband Unit
BN	Batch Normalization
BTS	Base Transceiver Station
CAPEX	CAPital EXpenditure
CCDF	Complementary Cumulative Distribution Function
CC-TR	Clipping Control Tone Reservation
CDF	Cumulative Distribution Function
CLT	Central Limit Theorem
CPRI	Common Public Radio Interface
CR	Compression Ratio
C-RAN	Centralized, Cooperative, Cloud Radio Access Network
DCT	Discrete Cosine Transform
DNN	Deep Neural Network
DVB-T2	Digital Video Broadcasting - Terrestrial - second generation
EC	Entropy Coding
ECQ	Entropy Constrained Quantization
ECSQ	Entropy Constrained Scalar Quantization
ECVQ	Entropy Constrained Vector quantization
ELM	Extreme Learning Machine
eNodeB	Evolved NodeB
FC	fully connected
FFNN	Feedforward Neural Network
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
GD	Gradient Descent
GEV	Generalized Extreme Value
GSM	Global System for Mobile Communications
HRA	High-Rate Approximation
ICI	Inter-Carrier Interference
IFFT	Inverse Fast Fourier Transform
i.i.d	Independent and Identically Distributed
IQ	In-phase/Quadrature

ISI	Inter-Symbol Interference
LA	Lloyd Algorithm
LBG	Linde-Buzo-Gray Algorithm
LTE	Long Term Evolution
LTE-A	Long Term Evolution-Advanced
MAC	Media Access Control
MER	Modulation Error Ratio
MLP	Multi Layer Perceptron
MSE	Mean Square Error
NN	Neural Network
NUQ	Non-Uniform Quantizer
OFDM	Orthogonal Frequency Division Multiplexing
OPEX	OPerating EXpenditure
PAPR	Peak-to-Average Power Ratio
PDF	Probability Density Function
PMF	Probability Mass Function
PRTs	Peak Reserved Tones
QCQP	Quadratically Constrained Quadratic Problem
RF	Radio Frequency
RRH	Remote Radio Head
r.v.	Random Variable
SELU	Scaled Exponential Linear Unit
SLB	Shannon Lower Bound
SQ	Scalar Quantization
SQNR	Signal to Quantization and Noise Ratio
TR	Tone Reservation
TS	Training Sequence
UQ	Uniform Quantization
VQ	Vector Quantization

Mathematical Notations

General notations

Random variables are shown in capital font while their realizations in small case. Vectors and matrices are denoted in bold font and in bold capital font, respectively, e.g. $\mathbf{x}$ and $\mathbf{X}$. The signal in the time-domain is denoted by x and its Fourier transform by $\tilde{x}$.. The expectation and probability operators taken under the underlying distribution of the r.v. X are denoted by $\mathbb{E}_X[\cdot]$ and $\mathbb{P}_X[\cdot]$, respectively. If the underlying distribution is obvious from the context, it can be removed.

Basic operations, functions and sets

$\Re(\cdot)$	Real part
$\Im(\cdot)$	Imaginary part
$\angle(\cdot)$	Phase
$\mathbb{E}[\cdot]$	Expectation
$\Gamma(n)$	Gamma function of order n
$\Gamma(n,\phi)$	Upper incomplete Gamma function of order n at a limit ϕ
${}_1F_1$	Confluent hypergeometric function
$\binom{r}{k}$	Binomial coefficient
$\log_2(\cdot)$	The base 2 logarithm
$\log_{10}(\cdot)$	The base 10 logarithm
$\mathbb{R}$	Set of real numbers
$\mathbb{C}$	Set of complex numbers
$\mathbb{Z}$	Set of integer numbers
$\|\cdot\|$	Euclidean norm
$\|\cdot\|_\infty$	Infinite norm, i.e., the maximum value
$\mathbf{0}_{n\times p}$	Null matrix of dimension $n \times p$
$\mathbf{I}_n$	Square identity matrix of size n
$\mathcal{O}(\cdot)$	Big O notation
$\mathcal{F}$	Fourier transform operation
$\mathcal{F}^{-1}$	Inverse Fourier transform operation
$u(t)$	Unit step function
$\delta(t)$	Dirac distribution where $t \in \mathbb{R}$
$J(\cdot,\cdot)$	Jacobian determinant
ζ	Polar coordinates domain
Υ	Cartesian coordinates domain
$\|\cdot\|$	Cardinality of the set

Parameter of the OFDM modulation

N_{fft}	Number of subcarriers per OFDM symbol
N_{active}	Number of loaded subcarriers per OFDM symbol
N_{null}	Number of guard subcarriers per OFDM symbol
$\mathbf{F}$	DFT matrix

Statistical distribution functions

$\mathcal{N}(0, \sigma^2)$	Gaussian distribution with mean μ and variance σ^2
$\mathrm{f}_R^{\text{Ray}}(r)$	Rayleigh distribution
σ_{ray}	Standard deviation of the Rayleigh distribution
$\mathrm{f}_R^{\text{GEV}}(r)$	GEV distribution
$\mathrm{F}_R^{\text{GEV}}(r)$	GEV CDF

Parameters of the quantization process

R	Rate, i.e., number of bits per each sample
N	Number of quantization levels
$Q(\cdot)$	Quantization operation
$\mathcal{C}$	Quantizer codebook
$\mathcal{R}_i$	ith quantization interval
Δ	Quantization step size
B	Quantization range
D	Quantization distortion
$D_{\text{gran.}}$	Granular distortion
D_{overload}	Overload distortion
γ	Loading factor
$\gamma_{\text{opt.}}$	Optimal loading factor
$s_{\text{sc.}}$	Scaling factor for optimal uniform quantizer

Parameters of vector quantization

k	Number of components per vector to be quantized
n	Length of the training data
$\mathcal{S}_i$	ith k-dimensional quantization region
V_k	volume of the k-dimensional unit sphere

Parameters of the decimation process

L	Oversampling factor
f_s	Sampling rate
f_m	System bandwidth

f_{ds}	Reduced sampling rate
m	Upsampling factor
p	Downsampling factor
F	Decimation factor
N_w	Length of FIR filter

Parameters of the clipping technique

V_{max}	Clipping threshold
Λ	Clipping ratio defined as the ratio between the power of the threshold V_{max}^2 and the signal power

Parameters of the tone reservation technique

$\mathcal{B}$	Set of peak reserved tones
υ	Total number of peak reserved tones
N_{PRT}	Percentage of PRTs out of the total N_{fft} subcarriers
P_{data}	Power of data subcarriers
ς	Power control defined as ratio between the maximum power allocated to each PRT and the average power of a data carrier
P_{PC}	Power control in dB
P_{PRT}	Maximum power permitted to PRTs
N_{data}	Number of data subcarriers out of the N_{active} subcarriers when υ subcarriers reserved for PRTs
τ	Maximum power of the time-domain signal

Parameters of the neural network

L_{lay}	Number of layers of the feedforward neural network
n_l	Number of neurons in the lth layer of the feedforward neural network
n_i	Number of neurons in the input layer of the feedforward neural network
n_o	Number of neurons in the output layer of the feedforward neural network
d_o	Dimension of the input vector to the neural network
d_l	Dimension of the output vector of the neural network
$\boldsymbol{\Theta}$	Matrix with all parameters of the feedforward neural network
$\sigma(\cdot)$	Nonlinear activation function of the output of the neuron
$\mathbf{W}_l$	Weight matrix of the lth fully connected layer
$\mathbf{b}_l$	Bias vector of the lth fully connected layer
N_{TS}	Length of the training data
η	Learning rate of the GD algorithm
Bs	Batch size of the mini-batch GD algorithm
$E(\cdot)$	Encoder mapping function for the AE neural network
$D(\cdot)$	Decoder mapping function for the AE neural network
N_b	Dimension of the bottleneck layer of an AE feedforward neural network

$n_{\text{par}}^{(l)}$	Number of trainable parameters of the lth hidden layer
$\mathbb{H}$	Impact of wireless channel
E_b/N_0	Energy-per-bit to noise ratio
J	Number of vectors per the latent representation of the OFDM signal in the AE-quantization network

Bibliography

[1] A. W. Dawson, M. K. Marina, and F. J. Garcia, "On the benefits of RAN virtualisation in C-RAN based mobile networks," in *2014 Third European Workshop on Software Defined Networks*. IEEE, 2014, pp. 103–108.

[2] "5G; procedures for the 5G system," Mar. 2019.

[3] "Frame Structure Channel Coding and Modulation for a Second Generation Digital Terrestrial Television Broadcasting System (DVB-T2)," Sep. 2009.

[4] "ATSC Proposed Standard: Physical Layer Protocol," 2016.

[5] A. CHECKO, "Cloud Radio Access Network architecture. Towards 5G mobile networks." Ph.D. dissertation, PHD thesis. Technical University of Denmark, 2016.

[6] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. Springer Science & Business Media, 2012, vol. 159.

[7] D. Samardzija, J. Pastalan, M. MacDonald, S. Walker, and R. Valenzuela, "Compressed transport of baseband signals in radio access networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 9, pp. 3216–3225, 2012.

[8] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for mobile networks—A technology overview," *IEEE Communications surveys & tutorials*, vol. 17, no. 1, pp. 405–426, 2014.

[9] "Digital cellular telecommunications system (Phase 2+) (GSM); GSM/EDGE Radio subsystem synchronization (3GPP TS 45.010 version 14.2.0 Release 14)," 2000.

[10] "Universal Mobile Telecommunications System (UMTS); General Universal Mobile Telecommunications System (UMTS) architecture (3GPP TS 23.101 version 14.0.0 Release 14)," 2017.

[11] G. Fernández-Getino, "Technical specification group radio access network; evolved universal terrestrial radio access (UTRA): Physical channels and modulation layer," *TS 36.211*, 2009.

[12] "Technical Specification Group Radio Access Network ; Requirements for further advancements for Evolved Universal Terrestrial Radio Access (E-UTRA) (LTEAdvanced)," 2020.

[13] I. Hwang, B. Song, and S. S. Soliman, "A holistic view on hyper-dense heterogeneous and small cell networks," *IEEE communications magazine*, vol. 51, no. 6, pp. 20–27, 2013.

[14] D. Gesbert, M. Kountouris, R. W. Heath, C.-B. Chae, and T. Salzer, "Shifting the MIMO paradigm," *IEEE signal processing magazine*, vol. 24, no. 5, pp. 36–46, 2007.

[15] J. Hoydis, S. Ten Brink, and M. Debbah, "Massive MIMO: How many antennas do we need?" in *2011 49th Annual Allerton conference on communication, control, and computing (Allerton)*. IEEE, 2011, pp. 545–550.

[16] Y. Lin, L. Shao, Z. Zhu, Q. Wang, and R. K. Sabhikhi, "Wireless network cloud: Architecture and system requirements," *IBM Journal of Research and Development*, vol. 54, no. 1, pp. 4–1, 2010.

[17] G. Kardaras and C. Lanzani, "Advanced multimode radio for wireless & mobile broadband communication," in *2009 European wireless technology conference*. IEEE, 2009, pp. 132–135.

[18] A. Ericsson *et al.*, "Common public radio interface (CPRI); interface specification v7. 0," *Huawei Technologies Co. Ltd, NEC Corporation, Alcatel Lucent, and Nokia Networks*, 2015.

[19] "Open Base Station Architecture Initiative (OBSAI) BTS System Reference Document Version 2.0." 2006.

[20] G. ETSI, "001:"Open Radio Equipment Interface (ORI)," *ORI Interface Specification*, vol. 10, 2014.

[21] S. Zhou, M. Zhao, X. Xu, J. Wang, and Y. Yao, "Distributed wireless communication system: a new architecture for future public wireless access," *IEEE Communications Magazine*, vol. 41, no. 3, pp. 108–113, 2003.

[22] M. F. Hossain, A. U. Mahin, T. Debnath, F. B. Mosharrof, and K. Z. Islam, "Recent research in cloud radio access network (C-RAN) for 5G cellular systems-A survey," *Journal of Network and Computer Applications*, vol. 139, pp. 31–48, 2019.

[23] H. Holma and A. Toskala, *LTE advanced: 3GPP solution for IMT-Advanced*. John Wiley & Sons, 2012.

[24] J. Segel and M. Weldon, "Lightradio portfolio-technical overview," *technology white paper*, vol. 1, 2011.

[25] T. Salman, "Cloud RAN: Basics, advances and challenges," *A Surv. C-RAN Basics Virtualization Resour. Alloc. Chall*, pp. 1–16, 2016.

[26] S. Weinstein and P. Ebert, "Data transmission by frequency-division multiplexing using the discrete Fourier transform," *IEEE transactions on Communication Technology*, vol. 19, no. 5, pp. 628–634, 1971.

[27] R. DISNEY and A. Clarke, "Probability and random processes: A first course with applications 2nd ed," *WILEY SERIES IN PROBABILITY AND MATHEMATICAL*, 1985.

[28] T. Jiang and Y. Wu, "An overview: Peak-to-average power ratio reduction techniques for OFDM signals," *IEEE Transactions on broadcasting*, vol. 54, no. 2, pp. 257–268, 2008.

[29] R. v. Nee and R. Prasad, *OFDM for wireless multimedia communications*. Artech House, Inc., 2000.

[30] M. Cover Thomas and A. Thomas Joy, "Elements of information theory," *New York: Wiley*, vol. 3, pp. 37–38, 1991.

[31] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Springer US, 1992.

[32] S. Lloyd, "Least squares quantization in PCM," *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.

[33] J. Max, "Quantizing for minimum distortion," *IRE Transactions on Information Theory*, vol. 6, no. 1, pp. 7–12, 1960.

[34] P. Fleisher, "Sufficient conditions for achieving minimum distortion in a quantizer," *IEEE Int. Conv. Rec.*, pp. 104–111, 1964.

[35] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, no. 9, pp. 1098–1101, 1952.

[36] A. Moffat, R. M. Neal, and I. H. Witten, "Arithmetic coding revisited," *ACM Transactions on Information Systems (TOIS)*, vol. 16, no. 3, pp. 256–294, 1998.

[37] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy-constrained vector quantization," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 1, pp. 31–42, 1989.

[38] P. Panter and W. Dite, "Quantization distortion in pulse-count modulation with nonuniform spacing of levels," *Proceedings of the IRE*, vol. 39, no. 1, pp. 44–48, 1951.

[39] H. Gish and J. Pierce, "Asymptotically efficient quantizing," *IEEE Transactions on Information Theory*, vol. 14, no. 5, pp. 676–683, 1968.

[40] R. Gray and A. Gray, "Asymptotically optimal quantizers (corresp.)," *IEEE Transactions on Information Theory*, vol. 23, no. 1, pp. 143–144, 1977.

[41] T. D. Lookabaugh and R. M. Gray, "High-resolution quantization theory and the vector quantizer advantage," *IEEE Transactions on Information Theory*, vol. 35, no. 5, pp. 1020–1033, 1989.

[42] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on communications*, vol. 28, no. 1, pp. 84–95, 1980.

[43] R. Gray, J. Kieffer, and Y. Linde, "Locally optimal block quantization for sources without a statistical model," *Inform. Syst. Lab., Stanford Univ., Stanford, CA, Tech. Rep. L-904-1*, 1979.

[44] D. Lee, S. Baek, and K. Sung, "Modified k-means algorithm for vector quantizer design," *IEEE Signal Processing Letters*, vol. 4, no. 1, pp. 2–4, 1997.

[45] T. Koch and G. Vazquez-Vilar, "A rigorous approach to high-resolution entropy-constrained vector quantization," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2609–2625, 2018.

[46] P. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 139–149, 1982.

[47] A. Gersho, "Asymptotically optimal block quantization," *IEEE Transactions on information theory*, vol. 25, no. 4, pp. 373–380, 1979.

[48] J. Makhoul, S. Roucos, and H. Gish, "Vector quantization in speech coding," *Proceedings of the IEEE*, vol. 73, no. 11, pp. 1551–1588, 1985.

[49] H. Si, B. L. Ng, M. S. Rahman, and J. Zhang, "A novel and efficient vector quantization based CPRI compression algorithm," *IEEE Transactions on vehicular technology*, vol. 66, no. 8, pp. 7061–7071, 2017.

[50] Y. Su, M. LiWang, L. Huang, X. Du, and N. Guizani, "Green communications for future vehicular networks: Data compression approaches, opportunities, and challenges," *IEEE Network*, vol. 34, no. 6, pp. 184–190, 2020.

[51] L. Ramalho, M. N. Fonseca, A. Klautau, C. Lu, M. Berg, E. Trojer, and S. Höst, "An LPC-based fronthaul compression scheme," *IEEE Communications Letters*, vol. 21, no. 2, pp. 318–321, 2016.

[52] L. Ramalho, I. Freire, C. Lu, M. Berg, and A. Klautau, "Improved LPC-based fronthaul compression with high rate adaptation resolution," *IEEE Communications Letters*, vol. 22, no. 3, pp. 458–461, 2018.

[53] Y. Su, X. Lu, L. Huang, X. Du, and M. Guizani, "A novel DCT-Based compression scheme for 5G vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 10 872–10 881, 2019.

[54] J. Terry and J. Heiskala, *OFDM wireless LANs: A theoretical and practical guide*. Sams publishing, 2002.

[55] S.-E. Lee, K.-J. Bang, M.-H. Park, Y.-S. Lee, and D.-S. Hong, "A New Peak-Windowing Algorithm with Window-length Adaptation for PAPR Reduction of OFDM Systems," in *Proceedings of the IEEK Conference*. The Institute of Electronics and Information Engineers, 2005, pp. 185–188.

[56] X. Huang, J. Lu, J. Zheng, J. Chuang, and J. Gu, "Reduction of peak-to-average power ratio of OFDM signals with companding transform," *Electronics letters*, vol. 37, no. 8, p. 1, 2001.

[57] S. H. Müller and J. B. Huber, "OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequences," *Electronics letters*, vol. 33, no. 5, pp. 368–369, 1997.

[58] J. Tellado, "Peak-to-average power reduction for multicarrier modulation," *Ph. D. thesis*, 1999.

[59] J. Hou, C. Tellambura, and J. Ge, "Tone injection for PAPR reduction using parallel tabu search algorithm in OFDM systems," in *2012 IEEE global communications conference (GLOBECOM)*. IEEE, 2012, pp. 4899–4904.

[60] B. S. Krongold and D. L. Jones, "PAPR reduction in OFDM via active constellation extension," *IEEE Transactions on broadcasting*, vol. 49, no. 3, pp. 258–268, 2003.

[61] J. Armstrong, "Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering," *Electronics letters*, vol. 38, no. 5, p. 1, 2002.

[62] J. Tellado and J. M. Cioffi, "Peak power reduction for multicarrier transmission," in *IEEE GLOBECOM*, vol. 99, 1998, pp. 5–9.

[63] R. Mounzer, "New tone reservation PAPR reduction techniques for multicarrier systems," Ph.D. dissertation, INSA de Rennes, 2015.

[64] Y. Rahmatallah and S. Mohan, "Peak-to-average power ratio reduction in OFDM systems: A survey and taxonomy," *IEEE communications surveys & tutorials*, vol. 15, no. 4, pp. 1567–1592, 2013.

[65] P. Banelli, G. Leus, and G. B. Giannakis, "Bayesian estimation of clipped Gaussian processes with application to OFDM," in *2002 11th European Signal Processing Conference*. IEEE, 2002, pp. 1–4.

[66] A. Cheaito, M. Crussière, J.-F. Hélard, and Y. Louët, "Quantifying the memory effects of power amplifiers: EVM closed-form derivations of multicarrier signals," *IEEE Wireless Communications Letters*, vol. 6, no. 1, pp. 34–37, 2016.

[67] M. El Hassan, M. Crussière, J.-F. Hélard, Y. Nasser, and O. Bazzi, "EVM closed-form expression for OFDM signals with tone reservation-based PAPR reduction," *IEEE Transactions on Wireless Communications*, vol. 19, no. 4, pp. 2352–2366, 2020.

[68] A. Jeffrey and D. Zwillinger, *Table of integrals, series, and products*. Elsevier, 2007.

[69] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[70] T. O'shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[71] M. Ibnkahla, "Applications of neural networks to digital communications–a survey," *Signal processing*, vol. 80, no. 7, pp. 1185–1215, 2000.

[72] S. Dörner, S. Cammerer, J. Hoydis, and S. Ten Brink, "Deep learning based communication over the air," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 132–143, 2017.

[73] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion." *Journal of machine learning research*, vol. 11, no. 12, 2010.

[74] M. Kim, W. Lee, and D.-H. Cho, "A novel PAPR reduction scheme for OFDM system based on deep learning," *IEEE Communications Letters*, vol. 22, no. 3, pp. 510–513, 2017.

[75] I. Sohn, "A low complexity PAPR reduction scheme for OFDM systems via neural networks," *IEEE Communications Letters*, vol. 18, no. 2, pp. 225–228, 2014.

[76] I. Sohn and S. C. Kim, "Neural network based simplified clipping and filtering technique for PAPR reduction of OFDM signals," *IEEE Communications Letters*, vol. 19, no. 8, pp. 1438–1441, 2015.

[77] Z. Liu, X. Hu, K. Han, S. Zhang, L. Sun, L. Xu, W. Wang, and F. M. Ghannouchi, "Low-complexity PAPR reduction method for OFDM systems based on real-valued neural networks," *IEEE Wireless Communications Letters*, vol. 9, no. 11, pp. 1840–1844, 2020.

[78] B. Wang, Q. Si, and M. Jin, "A novel tone reservation scheme based on deep learning for PAPR reduction in OFDM systems," *IEEE Communications Letters*, vol. 24, no. 6, pp. 1271–1274, 2020.

[79] Z. Li, N. Jin, X. Wang, and J. Wei, "Extreme learning machine-based tone reservation scheme for OFDM systems," *IEEE Wireless Communications Letters*, vol. 10, no. 1, pp. 30–33, 2020.

[80] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "A fast and accurate misc sequential learning algorithm for feedforward networks," *IEEE Transactions on neural networks*, vol. 17, no. 6, pp. 1411–1423, 2006.

[81] A. Gatherer and M. Polley, "Controlling clipping probability in DMT transmission," in *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers (Cat. No. 97CB36136)*, vol. 1. IEEE, 1997, pp. 578–584.

[82] A. Hirose, *Complex-valued neural networks*. Springer Science & Business Media, 2012, vol. 400.

[83] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.

[84] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks," *Advances in neural information processing systems*, vol. 30, 2017.

[85] A. Nguyen, K. Pham, D. Ngo, T. Ngo, and L. Pham, "An analysis of state-of-the-art activation functions for supervised deep neural network," in *2021 International Conference on System Science and Engineering (ICSSE)*. IEEE, 2021, pp. 215–220.

[86] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

[87] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[88] J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end optimized image compression," *arXiv preprint arXiv:1611.01704*, 2016.

[89] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE transactions on information theory*, vol. 44, no. 6, pp. 2325–2383, 1998.

[90] L. Theis, W. Shi, A. Cunningham, and F. Huszár, "Lossy image compression with compressive autoencoders," *arXiv preprint arXiv:1703.00395*, 2017.

[91] I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio, "Quantized neural networks: Training neural networks with low precision weights and activations," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6869–6898, 2017.

[92] E. Agustsson, F. Mentzer, M. Tschannen, L. Cavigelli, R. Timofte, L. Benini, and L. V. Gool, "Soft-to-hard vector quantization for end-to-end learning compressible representations," *Advances in neural information processing systems*, vol. 30, 2017.

[93] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.

[94] C.-s. Liu, "The essence of the generalized Newton binomial theorem," *Communications in Nmiscar Science and Numerical Simulation*, vol. 15, no. 10, pp. 2766–2768, 2010.

[95] M. Bernhard, D. Rörich, T. Handte, and J. Speidel, "Analytical and numerical studies of quantization effects in coherent optical OFDM transmission with 100 Gbit/s and beyond," *ITG-Fachtagung Photonische Netze*, pp. 34–40, 2012.

COLLEGE DOCTORAL BRETAGNE
MATHS, TELECOMS
INFORMATIQUE, SIGNAL
SYSTEMES, ELECTRONIQUE

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES RENNES

Titre : Compression du débit de données des signaux OFDM en bande de base pour le traitement virtualisé dans les réseaux d'accès radio en nuage d'accès en nuage.

Mots clés : C-RAN, OFDM, Quantification, Réduction du PAPR, L'apprentissage profond

Résumé : La compression des échantillons IQ dans le domaine temporel avant la transmission sur la liaison fronthaul est un moyen efficace de réduire les données transmises dans le réseau d'accès radio en nuage virtualisé (C-RAN). Pour une forme d'onde multiporteuse telle que l'OFDM, dont les échantillons IQ suivent une distribution quasi-gaussienne, le quantificateur gaussien conventionnel et le codeur entropique peuvent être utilisés comme solution optimale au problème de compression. Cependant, diverses techniques peuvent être appliquées au signal OFDM pour réduire son rapport de puissance crête-moyenne (PAPR) élevé, ce qui entraîne une modification de sa distribution. Ce fait rend le quantificateur gaussien sous-optimal. Dans ce contexte, cette thèse vise à étudier l'impact de la réduction du PAPR du signal OFDM avant la compression dans l'architecture C-RAN. Nous considérons les algorithmes de réduction du PAPR par écrêtage et par réservation de tonalité.

Le premier est l'approche la plus simple, tandis que le second est l'un des algorithmes les plus efficaces utilisés dans les normes de diffusion telles que DVB-T2 et ATSC 3.0. Nous évaluons analytiquement l'optimisation du quantificateur et du codeur entropique pour les signaux OFDM traités avec des techniques de réduction du PAPR. Ces résultats sont d'un grand intérêt pour quantifier le gain de performance lors de l'utilisation d'un quantificateur optimisé par rapport à la vraie distribution des échantillons IQ à PAPR réduit. En outre, nous allons plus loin que les approches conventionnelles et introduisons un réseau neuronal profond basé sur un autoencodeur qui est entraîné de bout en bout conjointement pour obtenir des échantillons de QI OFDM optimisés, réduits en fonction du PAPR et quantifiés dans le contexte de l'architecture C-RAN.

Title : Data rate compression of baseband OFDM signals for virtualized processing in cloud radio access networks.

Keywords : C-RAN, OFDM, Quantization, Entropy coding, PAPR reduction, Deep learning

Abstract : Compressing the time-domain IQ samples prior transmission over the fronthaul link is an effective way to reduce the transmitted data in the virtualized cloud radio access network (C-RAN). For a multicarrier waveform such as the OFDM, whose IQ samples follow a quasi-Gaussian distribution, the conventional Gaussian quantizer and entropy encoder can be used as the optimal solution to the compression problem.
However, various techniques can be applied to the OFDM signal to reduce its high peak-to-average power ratio (PAPR), resulting in a change in its distribution.
This fact makes the Gaussian quantizer suboptimal. In this context, this thesis aims to study the impact of reducing the PAPR of the OFDM signal before compression in the C-RAN. We consider clipping and tone reservation PAPR reduction algorithms.

The former is the simplest approach, while the latter is one of the most effective algorithms used in broadcast standards such as DVB-T2 and ATSC 3.0. We analytically evaluate the optimization of the quantizer and entropy coder for OFDM signals processed with PAPR reduction techniques. These results are of high interest in quantifying the performance gain when using a quantizer optimized with respect to the true distribution of PAPR-reduced IQ samples. Moreover, we go a step beyond conventional approaches and introduce an autoencoder-based deep neural network that is trained end-to-end jointly to obtain an optimized PAPR-reduced and quantized OFDM IQ samples in the context of the C-RAN architecture.