



**HAL**  
open science

# Analysis of Baseband IQ Data Compression Methods for Centralized RAN

Aya Shehata, Matthieu Crussière, Philippe Mary

► **To cite this version:**

Aya Shehata, Matthieu Crussière, Philippe Mary. Analysis of Baseband IQ Data Compression Methods for Centralized RAN. 28th European Signal Processing Conference (EUSIPCO 2020), Jan 2021, Amsterdam, Netherlands. hal-02941426

**HAL Id: hal-02941426**

**<https://hal.science/hal-02941426>**

Submitted on 17 Sep 2020

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

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

# Analysis of Baseband IQ Data Compression Methods for Centralized RAN

Aya Shehata, Matthieu Crussière and Philippe Mary

Univ Rennes, INSA Rennes, CNRS, IETR-UMR 6164, F-35000 Rennes

**Abstract**—Through recent wireless technologies, such as Centralized Radio Access Network, baseband unit and remote radio heads are physically separated and connected using fronthaul links. Compressing complex baseband signal samples prior transmission over fronthaul link is an effective way to satisfy the pressing need to decrease the huge required transported data rates. In this paper, we analyze the existing IQ data compression schemes exploiting time and spectral signal characteristics. We consider compression system evaluation parameters to have a smooth trade-off between required signal quality and complexity performance while achieving an acceptable compression gain. We propose an optimized uniform quantization technique combined with entropy coding achieving non-uniform quantization performance in exploiting signal temporal statistical characteristics with much less computational complexity. We also present a comparison between simulation results analyzing the trade-off between the removal of the signal spectral redundancies and vector quantization in terms of performance and complexity.

**Index Terms**—CPRI, Uniform and Non-Uniform Quantization, Decimation, Vector Quantization, Compression Ratio.

## I. INTRODUCTION

The Centralized, Cooperative, Cloud Radio Access Network (C-RAN) relying on centralized processing, collaborative radio and real time cloud infrastructure, define the next generation wireless network architecture. Through C-RAN, processing is done in a Baseband Unit (BBU) located centrally in a pool configuration and connected to many Remote Radio Head (RRHs), where solely radio frequency (RF) units remain [1]. Fronthaul links, which are implemented over electrical or optical based interfaces, are then responsible for the transmission of digitized complex IQ baseband signals [2]. Consequently, the C-RAN strategy leads to an increased amount of traffic on the Fronthaul links which hereby become the bottleneck of such an architecture. A natural solution to face this issue is to accommodate the increasing data traffic demand through the installation of more higher bandwidth optical fibers. A less expensive approach is to find ways to decrease the required transmitted data rate over the fronthaul before being transmitted. Hence, baseband IQ data rate compression has gained increased interest through the last years. Various strategies of data rate compression have recently been investigated, namely taking basis on time-domain samples streams resulting from Orthogonal Frequency Division Multiplexing (OFDM), as used in many of the recent technologies. Therefore, fronthaul link compression techniques commonly found in literature are based on exploiting the temporal and spectral characteristics of the OFDM signal.

In [3] and [4], uniform scalar quantization was used with oversampling to remove the inherent redundancy in oversampled OFDM signals and block scaling has been carried out to compensate for the redundant information appeared due to the large dynamic range of signal power. This is especially useful for uplink signals facing large and small scale propagation effects. In [5], non-uniform quantization based on an iterative gradient algorithm was proposed to take

advantage of the signal statistical structure over uniform quantization, and dithering is performed across parallel links to reduce compression error. In [6], relatively low complexity compression methods were used, where compression is done by encoding the difference between the current and the previous sample and the algorithm was tested using high oversampling factors. Vector quantization combined with decimation, block scaling before quantization and entropy coding for the quantizer output was adopted in [7]. The aim was to exploit time correlated samples and led to improved performance however at the expense of substantial increase of computational complexity. In [8] and [9] authors explored the correlation between samples by the well known linear predictive coding, hereby trying to solve the complexity issue introduced by vector quantization. This approach remains however restricted to uplink signal.

Throughout this paper, we explore the well known lossy and lossless compression methods exploiting scalar-based and vector-based characteristics of the signal samples to satisfy the goal of finding a solution which balances the compression performance, complexity and system end to end performance. After a reminder of the main ingredients which can be used to build a global compression scheme, we explore two main strategies at the scalar and vector levels.

We first propose to leverage the signal statistical characteristics by using a simple optimized uniform quantizer combined with entropy coding instead of designing a more complex non-uniform quantizer as usually done in the literature. Our proposed approach yields a substantial compression gain and lower computational complexity.

Secondly, we investigate in which extend correlation between signal samples can be efficiently exploited from a compression goal perspective. Thus, according to wireless system specifications, we present an illustration in attaining optimum performance in terms of compression, signal quality and complexity between removing OFDM signal oversampling overhead using decimation or exploiting this time correlated IQ baseband signals using vector quantization. Differently to what is proposed in [7] where decimation is followed by vector quantization, we rather compare both strategies since the former eliminates the main memory advantage of the latter due to diminishing the signal correlation. A performance/complexity analysis is then led and discussed.

The remaining of this paper is organised as follows. Section II presents the system model. In Section III, compression algorithms are described in details. Numerical analysis are presented in Section IV. Finally, Section V concludes the paper.

## II. SYSTEM MODEL

Fig. 1 depicts the basic functional blocks of the system used where time domain baseband signal is subject to compression before being transmitted through fronthaul link. Throughout this study, the considered modules for compression are decimation, quantization

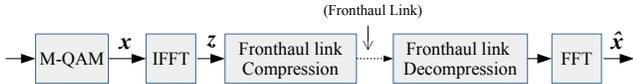


Figure 1: Fronthaul link compression algorithm framework.

and entropy encoding at the transmitter side and reverse operations are performed at decompression in the receiver side.

Let the signal  $\mathbf{x} \in \mathbb{C}^{N_c \times 1}$  be the M-QAM frequency domain symbol vector where  $N_c$  is the number of loaded sub-carriers.  $\mathbf{z} \in \mathbb{C}^{N_f \times 1}$  is the corresponding time domain symbol vector loaded with  $N_f - N_c$  empty sub-carriers, i.e. known as guard sub-carriers and  $N_f$  is the FFT size.  $\mathbf{z}$  can be expressed as

$$\mathbf{z} = \mathbf{F}^H \mathbf{M} \mathbf{x} \quad (1)$$

where  $\mathbf{F} \in \mathbb{C}^{N_f \times N_f}$  is the DFT matrix and  $\mathbf{M} \in \mathbb{C}^{N_f \times N_c}$  is a mapping matrix decomposed into  $\frac{N_f - N_c}{2} \times N_c$  null matrices and identity matrix of size  $N_c$  as follows

$$\mathbf{M} = \begin{bmatrix} \mathbf{0}_{\frac{N_f - N_c}{2} \times N_c} \\ \mathbf{I}_{N_c} \\ \mathbf{0}_{\frac{N_f - N_c}{2} \times N_c} \end{bmatrix}$$

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,  $\hat{\mathbf{x}}$  is the demodulating vector at the receiver. According to the Central Limit Theorem (CLT), for a sufficiently large IFFT length  $N_f$ , the resulting amplitude statistical distribution of the real and imaginary components converge to a zero-mean Gaussian distribution. Thus, exploiting OFDM signal statistical distribution is an important consideration in the quantizer design. Practically, the sampling value of the OFDM signal is usually higher than the minimum required by Nyquist theorem ( $N_f > N_c$ ). Thus, time correlated baseband signals are expected and compression techniques exploiting this correlation must be considered.

On this basis, Compression Ratio (CR) and Modulation Error Ratio (MER) are the metrics used to evaluate the performance of the compression system, as introduced hereafter.

#### A. Compression Ratio

The compression ratio is defined as the ratio between the uncompressed size and the compressed one, defined as following

$$\text{CR} = \frac{R_o}{R} \quad (2)$$

$R_o$  is the uncompressed bit width of I or Q component of each sample set, e.g. 15 (bits/sample) according to [2], and  $R$  is the compressed I or Q component bit width.

#### B. Performance

MER quantifies the ratio between the power of the original signal over the distortion introduced by the quantization in the log domain:

$$\text{MER}(\text{dB}) = 10 \log_{10} \left( \frac{\mathbb{E}\{|x|^2\}}{\mathbb{E}\{|x - \hat{x}|^2\}} \right) \quad (3)$$

### III. COMPRESSION SCHEMES

#### A. Scalar based Compression Techniques

1) *Scalar Uniform Quantization*: IQ samples are quantized sample by sample using a quantizer with  $R_{sq}$  bit resolution per each complex component. Uniform quantization (UQ) is optimum only for a uniform distributed signal [10].  $N$  quantization levels are simply uniformly distributed between  $[-2^{R_{sq}-1}, \dots, 2^{R_{sq}-1}-1]$ , centered at zero, with total number of  $N = 2^{R_{sq}}$  levels. Quantizer performance is evaluated by quantization distortion measured by the Mean Square Error (MSE) between the signal and the chosen quantization level expressed as

$$D = \sum_{i=1}^N \int_{t_{i-1}}^{t_i} (x - y_i)^2 p_X(x) dx \quad (4)$$

where  $y_i$  is the  $i^{\text{th}}$  quantization level,  $t_{i-1}$  and  $t_i$  are the decision thresholds of the  $i^{\text{th}}$  level. The first and last threshold levels  $t_0$  and  $t_N$  are set to  $-\infty$  and  $\infty$  respectively. MSE in (4) could be decomposed into two effects as following

$$D = \underbrace{\sum_{i=2}^{N-1} \int_{t_{i-1}}^{t_i} (x - y_i)^2 p_X(x) dx}_{D_1} + \underbrace{\int_{-\infty}^{t_1} (x - y_1)^2 p_X(x) dx + \int_{t_{N-1}}^{\infty} (x - y_N)^2 p_X(x) dx}_{D_2} \quad (5)$$

where  $D_1$  is the distortion occurred when an input value lies in the bounded intervals (granular distortion) and  $D_2$  is the distortion at the first and last unbounded intervals a.k.a. overload distortion.

UQ performance could be enhanced by dynamically adapting the input signal level to have a trade-off between granular and overload distortions [9]. The loading factor  $\gamma$  is defined as

$$\gamma = \frac{y_N}{\sigma_x s_x} \quad (6)$$

where  $y_N$  is the maximum quantizer's output amplitude,  $\sigma_x$  is the original random input signal standard deviation and both are fixed attributes, while  $s_x$  is the adjustable factor scaling the input signal before quantization to have optimum quantizer performance. Hence, a convenient value of  $\gamma$  controlling this trade-off is required. When  $R_{sq} \rightarrow \infty$ ,  $D_1$  can be expressed as follows [10]

$$D_1 = \frac{\gamma^2 \sigma^2}{3N^2} \quad (7)$$

and we derive a closed form of  $D_2$  for a zero-mean Gaussian distributed signal in terms of  $\gamma$  as

$$D_2 = \left[ \sigma^2 (1 + \gamma^2) Q \left( \gamma \left( 1 - \frac{1}{N} \right) \right) \right] - \left[ \frac{\gamma \sigma^2}{\sqrt{2\pi}} \left( 1 + \frac{1}{N} \right) e^{-\frac{\gamma^2}{2} \left( 1 - \frac{1}{N} \right)^2} \right] \quad (8)$$

where  $Q(u) = (1/\sqrt{2\pi}) \int_u^{\infty} \exp(-v^2/2) dv$  is the Q-function. According to (8) which consists of a constant and monotonically decreasing functions, overload distortion is decreasing by increasing  $\gamma$ . while, in contrast, granular distortion increases with  $\gamma$ . Thus, as well known, explicit solution for  $\gamma$  trading between granular and overload distortions satisfying

$$\gamma_{opt} = \arg \min_{\gamma} D \quad (9)$$

is impossible and numerical search is used to find the optimum depending on the input signal variance and the number of quantization levels.

2) *Scalar Non-uniform Quantization*: Optimum quantizer for a non-uniformly distributed data samples requires more quantization levels in the range of samples with high probability of occurrence. Thus, non-uniform quantizer (NUQ) has non-equal distances between quantization levels according to input signal statistical amplitude distribution. According to [11], for minimum mean square error (MMSE) optimality, decision thresholds for each quantization level are derived by setting the partial derivative of (4) with respect to  $t_i$  to zero, which generates

$$t_i = \frac{y_i + y_{i+1}}{2} \quad (10)$$

and by setting the partial derivative of (4) with respect to  $y_i$  to zero, each quantization level is optimally the centroid of its decision region

$$y_i = \frac{\int_{t_{i-1}}^{t_i} xp_X(x)dx}{\int_{t_{i-1}}^{t_i} p_X(x)dx}. \quad (11)$$

For a zero-mean Gaussian distributed signal, we have

$$y_i = \frac{\sigma}{\sqrt{2\pi}} \frac{\left( e^{-\frac{t_{i-1}^2}{2\sigma^2}} - e^{-\frac{t_i^2}{2\sigma^2}} \right)}{Q\left(\frac{t_{i-1}}{\sigma}\right) - Q\left(\frac{t_i}{\sigma}\right)} \quad (12)$$

3) *Entropy Coding*: Entropy coding is a lossless data compression scheme that utilizes the probability mass function (PMF) of the quantization levels. More bits for levels with lower frequency and lower bits for the higher frequency levels. Huffman coding [13] is the most common practically used entropy coding technique approaching the Shannon's lossless source coding theorem. Thus the average codeword length assigned to the  $i^{th}$  quantization level is

$$L_{\text{Huff},i} = -\log_2 \int_{t_{i-1}}^{t_i} p_X(x)dx \quad (13)$$

For a zero-mean Gaussian distributed signal, we derive it in a closed form as

$$L_{\text{Huff},i} = -\log_2 \left( Q\left(\frac{t_{i-1}}{\sigma}\right) - Q\left(\frac{t_i}{\sigma}\right) \right) \quad (14)$$

Hence, additional compression is gained by eliminating the potential redundancy from the distribution of quantization levels.

4) *Entropy Constrained Quantization*: it merges quantization with entropy coding [14]. Through ECQ, distortion and codeword lengths are optimized simultaneously by embedding entropy coding into the Lloyd's algorithm. The optimal quantizer is then obtained by minimizing the distortion subject to a constraint on the maximum transmission rate as following

$$\min_{t_i, y_i} D = \sum_{i=1}^N \int_{t_{i-1}}^{t_i} (x - y_i)^2 p_X(x)dx \quad (15)$$

$$\text{Subject to } -\sum_{i=1}^N p(y_i) L_i \leq R_{max}$$

Quantization level are still as in (11) but decision thresholds are updated to

$$t_i = \frac{y_i + y_{i+1}}{2} + \frac{\lambda}{2} \frac{L_{i+1} - L_i}{y_{i+1} - y_i} \quad (16)$$

where  $\lambda$  is the Lagrange multiplier of the rate constraint. This method has a remarkable complexity enhancement compared to using quantization and entropy coding functions separately.

## B. Vector based Compression Techniques

Additional gained compression could be achieved by considering the fact that the amplitude of the OFDM baseband signals are oversampled. Two solutions will be used to exploit this spectral redundancy.

1) *Spectral Domain Redundancies Removal*: OFDM signal is transmitted in broader spectrum than required by the system bandwidth, i.e. oversampled by factor  $L$  defined as

$$L = \frac{f_s}{f_m} \quad (17)$$

where,  $f_s$  is the original signal sampling rate and  $f_m$  is the system bandwidth. This can be exploited to further increase the compression without degrading signal quality by removing the redundant spectrum leaving a signal with lower sampling rate  $f_{ds}$  before being transmitted through fronthaul link.

Decimation is used in the redundancy removal process, first by upsampling the input signal by factor  $m$ . The upsampled signal is then filtered with an FIR low-pass filter with a bandwidth limited to  $[-f_{ds}/2, f_{ds}/2]$ . Finally, filtered signal is downsampled by factor  $k$ . the decimation factor  $F$  is the key parameter for performance evaluation and is defined as

$$F = \frac{f_s}{f_{ds}} = \frac{k}{m} \geq 1 \quad (18)$$

which is limited to the oversampling factor  $L$ , i.e. that depends on the amount of redundancy spectrum to be removed while retaining useful data  $F \leq L$ , and the filter length which must be chosen to have the best trade-off between ensuring good amplitude and phase response in the passband of the filter and the filter complexity.

2) *Vector Quantization*: Vector quantization (VQ) differs from the scalar one in exploiting time correlation between samples. VQ is a vector codebook based quantizer. The input is a  $k$ -dimensional vector whose components are the source IQ samples  $\mathbf{x} = [x_1, x_2, \dots, x_k]$  and mapped to one of the  $k$ -dimensional vectors  $\mathbf{y}_i$  in the codebook, where  $i \in \{1, 2, \dots, 2^{k \cdot R_{vq}}\}$  and  $R_{vq}$  is the number of bits per each scalar sample. According to [12], Lloyd's optimality conditions can be extended to the vector quantizer (VQ) design.

By having a set of training samples extracted from the source, vector quantizer codebook is designed offline by partitioning the  $k$ -dimensional space of random vectors  $\mathbf{x}$  into  $N$  convex quantization cells, which seeds are the output vector  $\mathbf{y}$  occurrences. Each input is mapped to a certain cell which minimizes its Euclidean distance with the codeword representing it. VQ is supposed to outperform the performance of scalar quantization by exploiting the correlation occurred by oversampling. Hence, to investigate the time correlated samples, consecutive IQ samples are vector grouped as an input to the quantizer.

## IV. ANALYSIS AND PROPOSAL

In this section we first numerically investigate the scalar quantization performance described in Section III-A. A 64-QAM signal is considered, which modulates  $N_c = 27260$  sub-carriers out of  $N_f = 32768$  IFFT entries per each OFDM symbol, that corresponds to a 8-MHz DVB-T2 frame structure. In order to assess the performance and compare with the idealized quantizer, we consider independent identically distributed (i.i.d) input samples, i.e. no redundancy in the spectral domain. MER of an ideal scalar quantization compression system is  $\text{MER}_{ub} = 2^{2R_{sq}}$  [15] and it is considered as an upper bound.

### A. Scalar based analysis

In Fig. 2, MER is presented as a function of the resolution  $R_{sq}$ . First, the dashed curves present the proposed optimized uniform quantizer and Lloyd-based non-uniform quantizer. UQ at each resolution is optimized at certain  $\gamma_{opt}$  according to the number of quantization levels and signal variance as shown in Fig. 3 which presents the MER as a function of the loading factor  $\gamma$  from 6 to 10 quantization bits [9]. Non-uniform quantizer has a clear advantage over the optimized

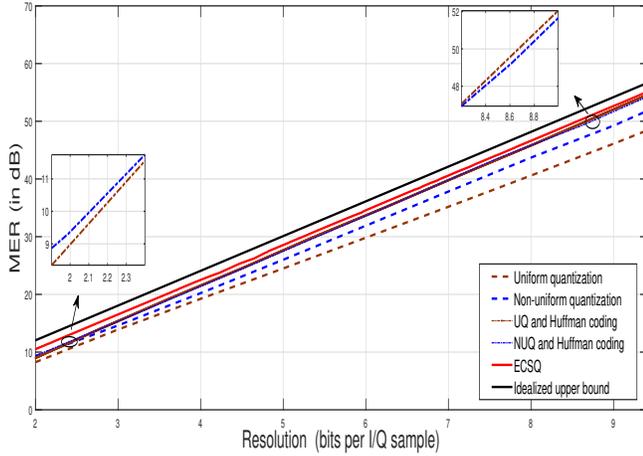


Figure 2: MER as a function of the resolution  $R_{sq}$  for *i.i.d.* complex Gaussian distribution.

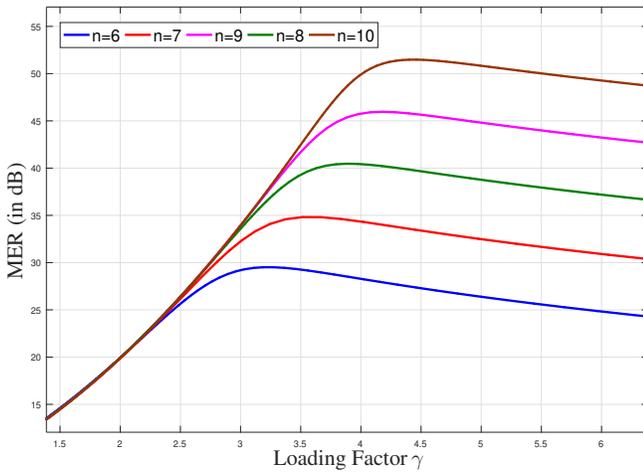


Figure 3: MER versus loading factor  $\gamma$  for different resolution bits.

uniform one and the difference is increasing at higher resolutions. Using the Huffman coding at the output of both uniform and non-uniform quantizers allow more compression by assigning variable-length codewords to different quantization levels according to their probability. As shown in the dashed-dotted lines, the difference between the two algorithms with Huffman coding is approximately indistinguishable. As illustrated in the sub-figures, for lower resolution NUQ is slightly better than UQ by approximately 0.4 dB at 2 bits resolution, while, UQ becomes better by increasing resolution per sample to almost 0.45 dB at 9 bits.

This performance happens due to the increasing of optimum loading factor  $\gamma_{opt}$  with increasing resolution as shown in (6) and also illustrated in Fig. 3. Higher  $\gamma$  leads to a lower scaling factor  $s_k$  according to (6), thus, the scaled signal amplitudes at the input of the quantizer diminishes and quantized to levels near zero. Huffman encoded these levels with shorter codeword lengths due to their higher probability of occurrence under Gaussian distribution. Hence, The loading factor based optimized uniform quantizer combined with entropy coding is having the role of exploiting signal statistical distribution like non-uniform quantization with lower computational complexity. This can be assessed analytically by analyzing the asymptotic gap

Table I: Illustration simulation results

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

between the entropy of uniform and non-uniform quantizer outputs derived from the asymptotic distortion-rate functions for uniform and non-uniform quantization derived in [10] and [16] respectively. This gap is expressed as following

$$\lim_{R \rightarrow \infty} R_{UQ} - R_{NUQ} = \frac{1}{2} \log_2 \left( \frac{3\sqrt{3}\pi}{2\gamma^2} \right) \quad (19)$$

Substituting in (19) by the simulated  $\gamma_{opt}$  obtained from Fig. 3 for different resolutions is shown in Table I, which shows the improvement of UQ over NUQ by increasing number of resolution bits, i.e. more compression gain can be achieved by UQ at the same MER. Finally, ECSQ performance can be considered as an upper limit to the compression gain obtained by a scalar quantization scheme and it is approximately 1.6 dB lower than the upper bound  $MER_{ub}$ . While, UQ followed by Huffman coding performance is approximately 2.4 dB lower than the upper bound. Thus, according to system requirements, we can decide whether approximately 1 dB gain obtained by ECSQ over using UQ and Huffman code worth the additional complexity of ECSQ implementation.

### B. Vector based analysis

We now consider the fact that the amplitude of the OFDM baseband signals are correlated due to oversampling as described in Section III-B. We will present a comparison between exploiting this correlation in gaining additional compression for different oversampling factors, i.e the attribute affecting the compression performance of the two solutions: i) decimation followed by scalar uniform quantization and ii) vector quantization as shown in Fig. 4a and 4b respectively.

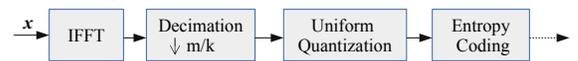


Figure 4a: Decimation followed by scalar quantization.



Figure 4b: Vector quantization.

As the oversampling factor  $L$  increases, the number of guard zero-padded sub-carriers  $N_{Null} = N_f - N_c$  within OFDM symbol increases. Thus, the ratio  $\frac{N_{Null}}{N_f}$  will be our oversampling parameter, i.e increases by increasing the oversampling factor. In Fig. 5, compression ratio is presented as a function of oversampled factor represented by  $N_{Null}/N_f$ . Compression ratio is calculated for the first system of decimation followed by SQ as  $CR = \frac{R_0}{\frac{m}{k} R_{sq}}$  and for VQ as  $CR = \frac{R_0}{R_{vq}}$ . Thus, in general compression ratio increases by increasing the oversampling factor. For the first system due to increasing the decimation factor ( $F$ ) degree of freedom maximally limited to the oversampling factor as mentioned in Section III-B1.

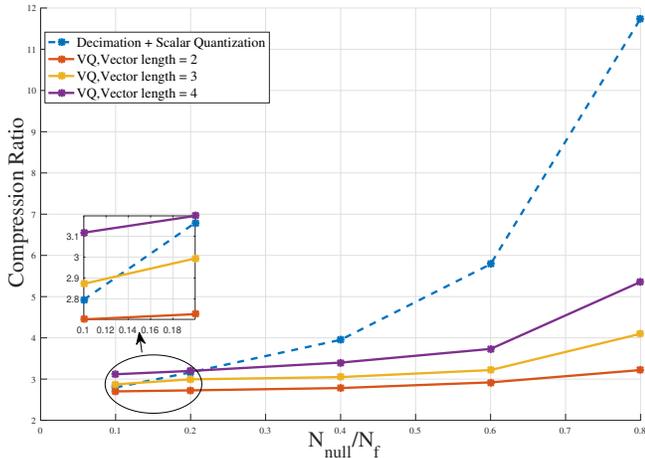


Figure 5: Decimation versus vector quantization.

While for vector quantization implemented using Lloyd algorithm, CR increases by increasing the quantizer memory advantage of exploring the dependency between input vector components, which increases by increasing vector dimension  $k$ .

In Fig. 5, we assume fixed MER = 30 dB to compare the compression achieved by the two systems. Hamming window based FIR filter is implemented in the decimation process and Table II shows the selected decimation parameters  $(m, k)$  and filter length  $N_w$  that achieve an acceptable signal quality and complexity performance for different oversampling factors. Simulations results show that the gain obtained by decimation at low oversampling factor is very small due to the small decimation factor degree of freedom and VQ with higher vector lengths can achieve a better performance. While the decimation gain increases significantly by increasing the oversampling factors, achieving about 6 times compression gain higher than vector quantizer with vector length 3 at 0.8 oversampling factor and 30 dB MER. Nevertheless, computational complexity of the two algorithms is an important parameter to be taken into consideration in deciding the preferable used algorithm beside the compression performance. The timing computational complexity of the first algorithm is  $\mathcal{O}(N_w \log N_w + 2^{R_{sq}})$ , which is the FIR filter complexity and SQ codebook size and for VQ is  $\mathcal{O}(2^{kR_{vq}})$ . VQ has higher computational complexity compared to the decimation and scalar uniform quantization due to the exponential increase of the VQ codebook size by increasing the vector length compared to the logarithmic increasing rate of the filter complexity with the filter length, which leads to a longer codebook training time, search process and larger codebook storage. Hence, the above numerical analysis may be used to decide on the application of the most suitable compression scheme. As an illustration, without loss of generality, DVB-T2 or ATSC3.0 broadcasting technologies are having empty sub-carriers in approximately  $\sim 4/25$  of their spectrum. Thus, for having more compression, VQ with higher vector lengths is preferable, taking into consideration its complexity. While, for an LTE like system has an approximately  $\sim 2/5$  empty sub-carriers, decimation with UQ can achieve an acceptable compression gain at 30 dB MER and with nearly acceptable complexity according to Table II.

Table II: 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

## V. CONCLUSION

In this paper, we have analyzed various IQ data compression strategies for the data rate limitation on fronthaul links of a C-RAN architecture. Through our analysis, two main conclusions can be driven. First, we have showed that an optimized scalar uniform quantization algorithm coupled by entropy coding over the resulting scalar codebook efficiently exploits the non-uniform OFDM signal statistical distribution. This proposed strategy achieves a substantial compression gain, only 2.4 dB lower than 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 in exploiting time-domain correlations. We have showed that a smooth trade-off exists between the required signal quality, compression performance and system complexity based on operator choice of suitable parameter values.

## ACKNOWLEDGEMENT

This work is funded by bpifrance for FUI25Cloudcast project.

## REFERENCES

- [1] Chen, Kulin, and Run Duan. "C-RAN the road towards green RAN." China Mobile Research Institute, white paper 2 (2011).
- [2] Interface, Common Public Radio. "CPRI Specification V7. 0." Standard Document Specification 1 (2015).
- [3] B. Guo, W. Cao, A. Tao and D. Samardzija, "LTE/LTE-A signal compression on the CPRI interface," in Bell Labs Technical Journal, vol. 18, no. 2, pp. 117-133, Sept. 2013.
- [4] Peng-ren, Ding, and Zhao Can. "Compressed transport of baseband signals in cloud radio access networks." 9th International Conference on Communications and Networking in China. IEEE, 2014.
- [5] D. Samardzija, J. Pastalan, M. MacDonald, S. Walker and R. Valenzuela, "Compressed Transport of Baseband Signals in Radio Access Networks," in IEEE Transactions on Wireless Communications, vol. 11,
- [6] A. Vosoughi, M. Wu and J. R. Cavallaro, "Baseband signal compression in wireless base stations," 2012 IEEE Global Communications Conference (GLOBECOM), Anaheim, CA, 2012, pp. 4505-4511.
- [7] H. Si, B. L. Ng, M. S. Rahman and J. Zhang, "A Novel and Efficient Vector Quantization Based CPRI Compression Algorithm," in IEEE Transactions on Vehicular Technology, vol. 66, no. 8, pp. 7061-7071, Aug. 2017.
- [8] L. Ramalho et al., "An LPC-Based Fronthaul Compression Scheme," in IEEE Communications Letters, vol. 21, no. 2, pp. 318-321, Feb. 2017.
- [9] L. Ramalho, I. Freire, C. Lu, M. Berg and A. Klautau, "Improved LPC-Based Fronthaul Compression With High Rate Adaptation Resolution," in IEEE Communications Letters, vol. 22, no. 3, pp. 458-461, March 2018.
- [10] Gersho, Allen, and Robert M. Gray. Vector quantization and signal compression. Vol. 159. Springer Science and Business Media, 2012.
- [11] S. Lloyd, "Least squares quantization in PCM," in IEEE Transactions on Information Theory, vol. 28, no. 2, pp. 129-137, March 1982.
- [12] Y. Linde, A. Buzo and R. Gray, "An Algorithm for Vector Quantizer Design," in IEEE Transactions on Communications, vol. 28, no. 1, pp. 84-95, January 1980.
- [13] D. A. Huffman, "A Method for the Construction of Minimum-Redundancy Codes," in Proceedings of the IRE, vol. 40, no. 9, pp. 1098-1101, Sept. 1952.
- [14] P. A. Chou, T. Lookabaugh and R. M. Gray, "Entropy-constrained vector quantization," in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 37, no. 1, pp. 31-42, Jan. 1989.
- [15] Berger, Toby. "Rate-distortion theory." Wiley Encyclopedia of Telecommunications (2003).
- [16] H. Gish and J. Pierce, "Asymptotically efficient quantizing," in IEEE Transactions on Information Theory, vol. 14, no. 5, pp. 676-683, September 1968.