CPRI compression transport for LTE and LTE-A signal in C-RAN

Bin Guo; Wei Cao; An Tao
Alcatel-Lucent Shanghai Bell Company Limited
Shanghai, China
bin.d.guo@alcatel-sbell.com.cn
Dragan Samardzija
Alcatel-Lucent Bell Laboratories
Holmdel, USA
Dragan.Samardzija@alcatel-lucent.com

***Abstract*—C-RAN is the next-generation clean wireless access network architecture, which is based on centralized processing, the Collaborative Radio and Real-time Cloud Infrastructure. In C-RAN architecture, different access technology (eg.GSM / TD-SCDMA / WCDMA / LTE) can be support on the same hardware platform in baseband pool system; C-RAN can be applied to reduce the consumption of operators. LTE (Long Term Evolution) and LTE-A, which are based on OFDM and MIMO technologies, are regarded as the main wireless access technology in the evolution from 3G to 4G. LTE introduces variety of novel technologies to improve the system performance, especially in C-RAN architecture, such as Multi antennas MIMO, Carrier Aggregation, CoMP. However, C -RAN architecture also brings about lager data transmission in the optical fiber. In this paper a low-latency baseband signal compression algorithm used in ALU is introduced to reduce the data rate. Based on the Characteristics of LTE signal data, the algorithm removal redundancies in the spectral domain firstly, and then the block scaling, which is combined with non-linear (non-uniform) quantizer, is designed to minimize quantization error. This algorithm can effectively reduce the amount of data transferred between BBU and RUU in LTE system, to facilitate the deployment of LTE in the C-Ran architecture. Simulation analysis is given out in the paper; meanwhile TD-LTE system demo verification is also implemented. The results indicate that the scheme introduced can get good performance under a certain compression rate in the real system. Further research is also in progress.**

Keywords-component; CPRI; compression; LTE; C-RAN

I. INTRODUCTION

C-RAN [1] is the next-generation clean wireless access network architecture, which is based on centralized processing, the Collaborative Radio and Real-time Cloud Infrastructure). In C-RAN architecture, the baseband units (BBU) are centrally located to constitute a pooling of BBU, which are connected to the Remote RF unit (RRU) via optical fiber. The BBU Pooling can use common architecture CPU and share baseband processor resources. Meanwhile the cloud computing technology can be also used. The performance of wireless networks can be significantly enhanced through collaborative radio in C-RAN.

C-RAN's baseband centralization brings about processing resources sharing, reduce energy consumption, and improve the utilization of infrastructure. However, we should note that the C-RAN consumes more fiber resources compared with traditional network.

If C-RAN relies on the the transmission of untreated wireless signal between BBUs and RRUs, much high transmission rate, transmission delay and jitter requirements will be required. Therefore, how to reduce the fiber resources consumption becomes a core issue in the C-RAN architecture implementation.

Based on OFDM and MIMO technology, LTE (Long Term Evolution) is regarded as the main technology in the evolution from 3G to 4G, which has been adopted by many operators. In Order to improve the cell throughput, multi-antenna technology has become an important direction of the wireless access technology. Multi-antenna MIMO and beamforming technology can be used to improve system performance [4]. LTE-advanced introduces novel technology, such as CA: Carrier Aggregation, multi-antenna enhanced MIMO (Enhanced MIMO) and Coordinated Multi-point transmission [5]. Meanwhile LTE-Advanced baseband bandwidth is five times of LTE [3] [6]. LTE is designed to improve system performance; however LTE also greatly increase the amount of data transmission between BBU and RRU, namely, need more fiber optic links. In this paper a low-latency baseband signal compression scheme of baseband signals in Radio Access Networks in ALU is introduced to solve above problem in LTE.

According to LTE system architecture, the eNB contain two main parts: BBU and RRU. For downlink, LTE baseband signal is generated in the BBU, and then the baseband signal is transferred to RRU. In the the RRU, the digital-analog and analog-digital conversion is done. Finally, the signal is send to the radio interface. Between BBU and RRU, the CPRI [7] standard interface is used to transfer data in the optical fiber. The signals are samples with a certain bit-width complex value, which are transferred between BBU and

Identify applicable sponsor/s here. *(sponsors)*

RRU, such as 15bit I-part and 15 bits Q-part. In C-RAN architecture, many RRUs are connected to BBU pooling, which requires a large number of fiber-optic links. The purpose of algorithm in this paper is to compress the number of digital signal and the bit width before CPRI framing. Meanwhile this compression should not cause obvious injury to the signal. If this purpose can be achieved, the number of optic fiber link can be reduced.

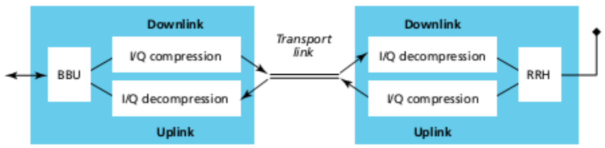


Figure 1. BBU-RRH system with I/Q compression

The basic functional blocks of the compression solution are shown in Figure 1, which is introduced in ALU light-radio white paper [2]. Between the BBU and RRU, one pair optic fiber link is connected, one is for downlink transmission from BBU to RRU, and the other is for uplink transmission from RRU to BBU. The compression and decompression processing module are added respectively inside BBU and RRU for the uplink and downlink transmission. For downlink, the BBU generate the baseband uncompressed signal, and IQ samples are compressed through compression module, then the compressed IQ samples are transfer through CPRI link to RRU side. In the RRU, the compressed IQ samples are firstly decompressed through decompression module, and then do the other normal processing of RRU. For uplink, the process is similar with downlink; just the samples are output of analog-to-digital converter (ADC). The input of ADC is the received analog radio signal from antennas. In the next section the algorithm details is described.

II. ALGORITHM DETAIL

The Compression module consists three mainly process steps as shown in Fig2, while in the deCompression module the process is opposite.

- Removal of redundancies in the spectral domain
- Block scaling
- Quantization

The first step is removal of redundancies in the spectral domain. This module is used to remove redundant spectral data through passing the low-pass filter. This module is designed to achieve that only the data on valid bandwidth is transferred. So the number of samples can be reduced significantly.

Then the data are segmented block by block. In each block, one scaling factor is found to do the data scaling. The piecewise scaling is to meet the fluctuations of the data in the time domain, especially for large value and small value. Block scaling is used to reduce the quantization error.

The third step is to quantify the pre-processed IQ samples from Q_S bit-width to Q_q bit-width, normally $Q_q < Q_S$. According to the data characteristics and the bits-width Q_q and Q_S, the quantizer will be designed or selected. After the quantization, the bits width of one IQ will be reduced to Q_q. This function is performed sample by sample.

After above treatment, the amount of data samples will be reduced significantly compared with original data. Meanwhile the bit-width of one sample will be also reduced. Decompression side will use the same method to recover the original data. In following subsection, each module will be described in detail.

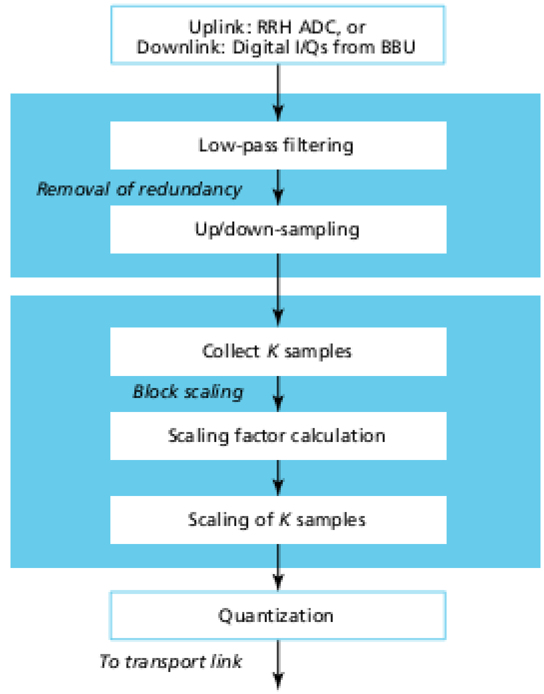


Figure 2. I/Q samples compression block diagram

A. *Removal of redundancies in the spectral domain s*

Based on the current practice, the sampling rate of the ADC, DAC and BBU processing is higher than the minimum required according to the Nyquist sampling theorem in most wireless access technology. In LTE system, the sampling rate also exceeds the signal bandwidth.

In LTE downlink, IFFT is used to generate the OFDM signal in implementation. Usually, the smallest appropriate IFFT size was chosen according to the Nyquist sampling theorem. However, considering the implementation complexity, general IFFT size based on power of two is willing to use. Take 20 MHz bandwidth of TD-LTE system for example, according to 3GPP 36.101[9], 100 PRB are used. So useful sub-carrier number is 100*12 (the number of subcarriers within each RB is 12), subcarrier width is $\Delta f = 15\,\text{kHz}$, the width of signal bandwidth is 1200 * 15 kHz = 18 MHz [8]. While the smallest size power of two, which is larger than1200, is 2048, the signal is mapped on 1200 subcarriers in the middle of 2048 subcarriers; the remaining subcarriers are filled with zeros. So 2048*15k=30.72M bandwidth is used in practice. Taking into account the 2M bandwidth is reserved to the filter roll-off. There are about 10M of bandwidth is redundancies in the spectral domain, That is, in the uncompressed form, a spectrally broader than necessary signal is transmitted in the CPRI frames.

TABLE I. TRANSMISSION BANDWIDTH CONFIGURATION N_{RB} IN E-UTRA CHANNEL BANDWIDTHS

Channel bandwidth BWChannel [MHz]	1.4	3	5	10	15	20
Transmission bandwidth configuration N_{RB}	6	15	25	50	75	100

For cutting these redundancies, the first process is upsampling the input signal; zeros are inserted, and then let these signal with zeros pass the low-pass filter. Finally downsampling is done on the output signal of filter. The sampling rate of the original is f_s. These processes will downsample the input signal to a lower sampling rate f_{ds}.

The selection of f_{ds} depends on how much redundant spectrum need to cut. It should be noted that the adequate protection bandwidth beside the useful signal bandwidth should be reserved. The protection bandwidth insures that the filter doesn't damage the useful information. The system bandwidth is limited to $[-f_{ds}, f_{ds}]$ after downsampling.

The downsampling factor F is a rational number, where L is the upsampling times, K is the downsampling times. L and K are positive integers.

$$\mathrm{F} = \frac{f_s}{f_{ds}} = \frac{L}{K} \geq 1 \tag{1}$$

The filter coefficients designing target is to make the filter have good amplitude and phase response in the pass band, the figure 3 shows amplitude and phase response of the filter used in the simulation in third section.

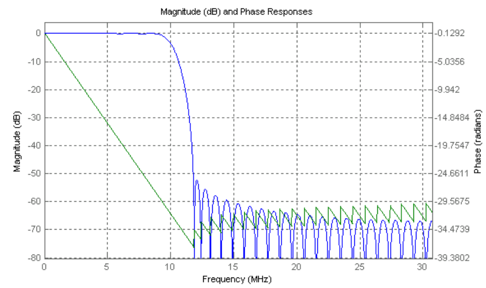
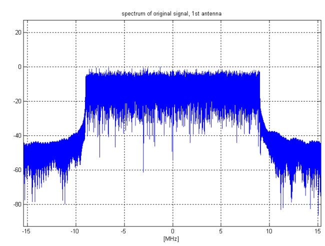
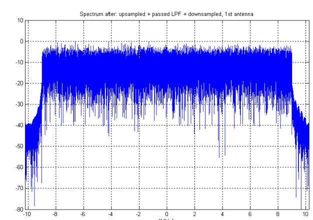


Figure 3. amplitude and phase response of filter

In the figure 4a, it is the frequency spectrum of a typical LTE downlink 20M system bandwidth baseband signal in time domain. As shown, the useful spectrum in the center and redundancies are at the both sides. We use the filter in Fig3 to treat this original time signal. The spectrum of filter output is shown in Fig4b. The most of redundant spectrum is filtered out.



a. The original signal spectrum



b. the signal spectrum pass filter

Figure 4. The fequency spectrum of original signal and the signal psss filter

Particular parameters, such as downsampling factor and filter coefficients, should be specified for a given system bandwidth. Furthermore, the parameters should be selected to achieve the trade-off between complexity and performance, eg.t he number of filter taps.

B. *Block scaling*

In LTE communications system, OFDM is used in downlink transmission. In OFDM transmission technology, multi orthogonal carriers are used to improve the performance, which also causes large PAPR. A typical LTE radio signal has a large dynamic range. LTE uplink uses the SC-FDMA transmission technology to reduce the PAPR, and generally the AGC block is also used to reduce power difference between the different users. However, the signal amplitude is still with greater volatility in large range. Typically, time domain signal samples are transported using Q_s bits per complex component in fixed-point solution. If these bits are compresses to Q_{q} bits directly without any other processing, the saturation will take place frequently. While the small signal will also loss more precision. To solve this problem, the block scaling is introduced. In block scaling, the output of processing 'Removal of redundancies in the spectral domain' is segmented to many small blocks with $\mathrm{N_S}$ samples. In each block with block scaling, a scaling factor is determined for a block of $\mathrm{N_S}$ I/Q samples, such that the subsequent quantization error is minimized. The scaling factor is also transmitted adding an overhead in each block. However, due to minimized quantization error, a lower quantizer resolution is applied, which results in overall reduction of the transport data rates. This function is performed on a block of $\mathrm{N_S}$ I/Q samples, at the output of the decimator. The scaling factor is transmitted once every $\mathrm{N_S}$ I/Q samples. Lowering the block length $\mathrm{N_S}$ will lower the subsequent quantization error, while increasing the factor transmission overhead. Therefore, the block length $\mathrm{N_S}$ is a design parameter derived from the trade-off analysis between the required signal quality and transport data rates.

In the k th block with $\mathrm{N_S}$ samples, a factor with the largest absolute value is determined as below formula

$$A(k) = \max_{i=\mathrm{N_s}*k,\ldots,\mathrm{N_s}(k+1)-1} \{|\operatorname{Re}(\mathrm{s_d}(i))|,|\operatorname{Im}(\mathrm{s_d}(i))|\} \quad (2)$$

The above scaling factor is an integer and it does not exceed $2^{Q_s}-1$, where Q_s is the bit width used for transferring scaling factor. Each sample in the block is then scaled as formula (4), where the Q_{q} is the bit width used in Quantization

$$\mathrm{S}(k) = \begin{cases} \lceil A(k) \rceil & for \lceil A(k) \rceil \le 2^{Q_s}-1 \\ 2^{Q_s}-1 & for \lceil A(k) \rceil > 2^{Q_s}-1 \end{cases} \quad (3)$$

$$\mathrm{s_s}(i) = s_d(i)\frac{2^{Q_q-1}-1}{S(k)} \quad (4)$$

C. Quantization

After block scaling, I/Q samples are quantized using a quantizer with Q_{q} bits resolution per each complex component. This function is performed sample by sample. A simple linear (uniform) quantizer with resolution Q_{q} may be applied. We note that, if the Q_{q} is close to original bitwidth Q_s, this linear quantizer works well. However, when Q_{q} becomes small, application of a quantizer with optimized distances between quantization levels will result in lower quantization error, along with improved signal quality. The design non-linear (non-uniform) quantizer should consider the data characteristic. Figure5 shows the amplitude distribution of data with scaling of $Q_{\mathrm{q}} = 6$. It can be seen that a larger portion of data is with minor amplitude. In the quantizer design, these features will be considered, and the smaller quantization distance will be used for signal with large distribution. Usually using a long training sequence to train the quantizer is a common method in quantizer designing. Multiple iterations are needed to quantify an appropriate distance with the minimum quantization error.

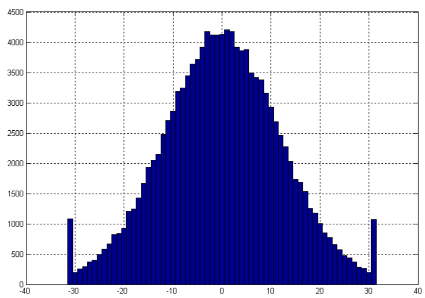


Figure 5. 6bits data amplitude distribution

Alcatel-Lucent Bell Labs has implemented a non-linear (non-uniform) quantizer. In this case, the quantization levels are optimized in conjunction with block scaling. Higher resolution Q_{q} will improve signal quality by lowering quantization noise, while increasing transport data rates. Therefore, the resolution Q_{q} is also a design parameter derived from the trade-off analysis between the required signal quality and data rates. Meanwhile, the custom's requirement is also a main factor in the practice implementation.

The original signal after the processing of the above three steps, the number of symbols and the bit-width are both compressed. In the RRU side the reverse process is used to recovery the original data. Of course, this reverse process can't recovery all the original information in data. The choice of the parameters and the compression ratio will have a significant impact on compression injury. Simulation analysis and testbed evaluation has been done on this compression algorithm in research. The results are in next section.

III. Performance Evaluation

The compression injury on data mainly depend on the compression ratio, the compression ratio in the algorithm is calculated as below formula: Compression ratio formula:

$$C_{\text{ratio}} = \frac{K/L*Q_q + Q_s/N_s}{Q_s} \tag{5}$$

Here we use the error vector magnitude (EVM) to evaluate the error between original data and the data with compression algorithm processing. EVM is defined as follows:

$$EVM\% = \sqrt{\frac{E\left[\| \bar{x} - x|^2\right]}{E\left[\| x|^2\right]}}*100[\%] \tag{6}$$

Where $\bar{x}$ is the output signal (after the compression and decompression have been performed), while x is its idealized noise-free version.

The higher the EVM, the larger injury of the compression algorithm resulted on the data. In order to evaluate the algorithm more carefully, we defined two types EVM: the time domain EVM and frequency-domain EVM. The frequency domain EVM is defined on the used bandwidth of the frequency domain signal, because after all, the changes of the signal on the useful bandwidth are mainly concerned. While the time-domain EVM is defined on entire bandwidth signal.

There are the EVM requirements for different modulation scheme in the 3GPP specification 36.104. The requirements are based on the end-to-end EVM definition between eBN and UE, and it is also a frequency domain EVM definition. EVM loss caused by IQ compression is only a part of end to end EVM.

TABLE II. DOWNLINK LTE EVM REQUIREMENTS

MODULATION SCHEME	MAXIMUM EVM%
QPSK	17.5
16QAM	12.5
64QAM	8

In addition, in order to assess the impact of scaling and quantization, following SQNR is defined.

$$SQNR = \frac{E|s_d(i)|^2}{E|\bar{s}_d(i) - s_d(i)|^2} \tag{7}$$

Where $s_d(i)$ is the block scaling input in (6) and $\bar{s}_d(i)$ the output of block rescaling at decompression side.

A. Simulation results

In simulation the algorithm parameters are configured as follows

TABLE III. PARAMETER CONFIGURATION

Parameter	Value
K	2
L	3
Q_s	11/8/6 bits
Q_q	15 bits
N_s	32 samples

In order to achieve the ergodic of signal, six configurations are used to simulate different downlink user scheduling.

TABLE IV. SIMULATION CASES CONFIGURATION

Case index	RB number	Modulation scheme
1	10	QPSK
2	10	16QAM
3	10	64QAM
4	100	QPSK
5	100	16QAM
6	100	64QAM

- The simulation results are list in Table V-VII. The compression ratios are 0.5201, 0.3868 and 0.2979 respectively.

TABLE V. 11 BITS COMPRESSION RESULTS COMPRESSION RATIO: 0.5201

Case index	Frequency EVM	Time EVM	SQNR
1	[0.4183]	[1.2287]	[62.8566]
2	[0.4005]	[1.1792]	[62.8206]
3	[0.4069]	[1.1982]	[62.8387]
4	[0.4173]	[1.2265]	[62.6868]
5	[0.4605]	[1.3278]	[62.6219]
6	[0.4070]	[1.1823]	[62.5962]

TABLE VI. 8 BITS COMPRESSION RESULTS COMPRESSION RATIO: 0.3868

Case index	Frequency EVM	Time EVM	SQNR

1	[0.6782]	[1.3506]	[44.8722]
2	[0.6668]	[1.3066]	[44.8969]
3	[0.6722]	[1.3253]	[44.8544]
4	[0.6890]	[1.3547]	[44.6512]
5	[0.7166]	[1.4476]	[44.5771]
6	[0.6852]	[1.3166]	[44.6194]

TABLE VII. 6 BITS COMPRESSION RESULTS COMPRESSION RATIO: 0.2979

Case index	Frequency EVM	Time EVM	SQNR
1	[2.0042]	[2.4018]	[33.6422]
2	[2.0027]	[2.3769]	[33.5828]
3	[2.0023]	[2.3870]	[33.5595]
4	[2.0362]	[2.4241]	[33.4597]
5	[2.0461]	[2.4791]	[33.4748]
6	[2.0395]	[2.4097]	[33.4588]

It can be seen from the above results that the EVM of the decompressed signal EVM changes significantly for the different compression ratio. Under 11bits and 8bits compression, the compression ratios are respectively 0.5201, 0.3868. the EVM can be controlled by less than 1%. 1/2 compression ratio means that of the original data transmitted on two optical fiber in the practical application, now can be transmitted on only one optical fiber . Taking into account the further compression ratio, for example, in the 6 bits compression with low SQNR, the EVM is significantly increased, but which is still within 3PGG requirement. Consider achieving system joint optimization, other parts of the system, such as RRU and UE, must be carefully designed to obtain further compression space.

B. *TD-LTE demo test results*

Based on the simulation results, the TD-LTE system demo testbed is also built in lab. Few FPGA resources inside BBU and RRU are used to implement the compression and decompression module. In the demo evaluation, the 11bits 1/2 compression ratio is evaluated.

The demo system is tested in a lab environment. Due to the performance in both compressed and uncompressed cases will be compared, a more stable state should be kept, so the platform should be achieved following status firstly:

(1) The platform is working in normal condition; (2) the equipment has been fully preheated, the performance indicators a stable state.

The test steps are as follows:

(1) Connect the test equipment as shown in figure 6

(2) Configure the D-band carrier, and achieve the above steady state

(3) configure TDL carrier work in the TM3 mode, open the AMC open of HARQ, set the maximum number of retransmissions time 3;

(4) load the SCM channel model; gradually increase the noise to reduce the SNR;

(5) Use the test UE to demodulate the downlink signal and statistical the throughput, SNR values;

(6) Close CPRI compression repeat steps (2) to (4).

The test results are shown as below:

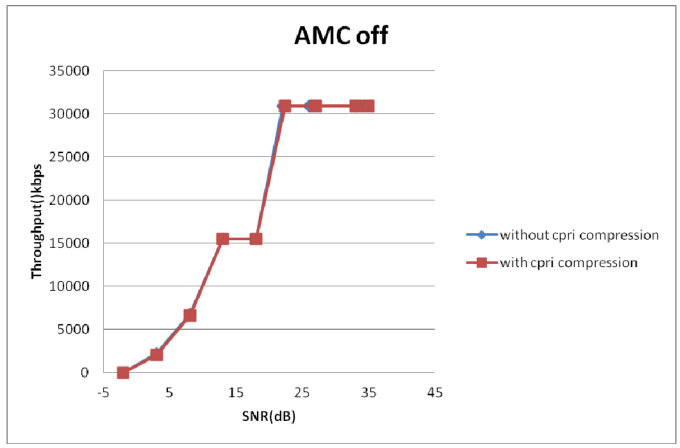


Figure 6. Throughput comparison between case with compression and case without compression under AMC off

Figure6 shows the throughput comparison between case with compression and case without compression under AMC (Adaptive modulation and coding schemes) off. There is no obviously reduction between performance with 11 bits compression and the performance without compression. The 11 bits compression affect on the performance is quite limited.

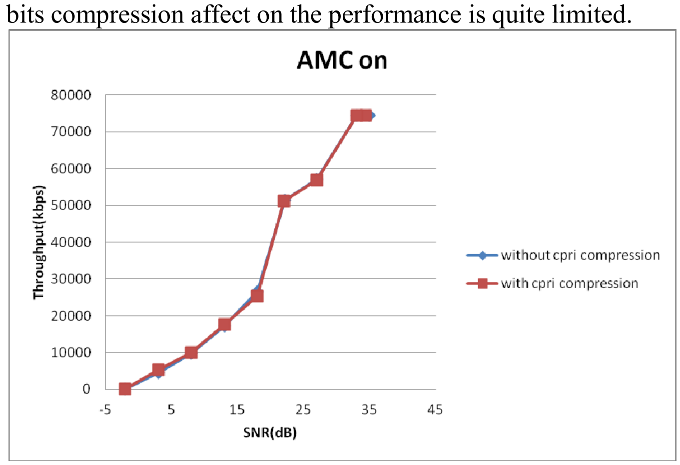


Figure 7. Throughput comparison between case with compression and case without compression under AMC on

Figure 7 is the test results in the case of AMC on. In this case, the user's MCS is scheduled changing with the channel conditions. So the compression algorithm can be measured under different modulation scheme in this case. It can be seen from the figure that throughput with compression and the one without compression are still very close. The compression algorithm works well with limited performance loss.

IV. CONCLUSIONS

The CPRI transmission is the bottleneck in the C-RAN implementation. With the evolution of TLE and LTE-A, new technology introduced also increases the amount of data transmitted on the CPRI interface. A CPRI compression algorithm for LTE system is introduced in this paper to reduce the data rates. Through eliminating redundant spectrum bandwidth and the amount of bit-bit-wide compression, this algorithm can effectively reduce the transmission amount of data. In simulation, algorithm under different compression ratio is analysis. The loss of information is little under low compression ratio; the EVM can be controlled less than 1%. The compression scheme is also verified in LTE lab demo. The performance with 11bits compression is ideal. But we also see that the EVM has shown a significant deterioration in the case of high compression ratio. Therefore, the specific implementation of this kind of compression ratio is a tradeoff between performance and consumption. In the future research, the algorithm configuration optimization for a specific scene will be considered, such as the design of the quantizer.

ACKNOWLEDGMENT

.

REFERENCES

[1] China Mobile Research Institute, “C-RAN: The road towards green RAN,” April 2010.

[2] LightRadioArch6_CompressedIQ_Transport_TechWhitePaper.pdf

[3] “Feasibility Study for Further Advancements for E-UTRA (LTE-Advanced),” 3GPP TR 36.912, 2009.

[4] 8 Antennas Eigen-based Single Stream Beamforming Algorithm for TD-LTE System Bin Guo; Xun Fan; Wei Cao; Yalin Li; Zhining Jiang

[5] M. Sawahashi, Y. Kishiyama, A. Morimoto, and M. T. D. Nishikawa, “Coordinated Multipoint Transmis-sion/Reception Techniques For LTE-Advanced,”IEEE Transactions on Wireless Communications, Jun 2010.

[6] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, “LTE Advanced: Next generation wireless broadband technology,”IEEE Transactions on Wireless Communications, vol. 17, pp. 10–22, Jun 2010.

[7] CPRI Specification, V 4.2, November 2010.

[8] “[illegible]estrial Radio Access (E-UTRA); Physical channels and modulation”, **3GPP TS 36.211** 2010

[9] “Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception”, **3GPP TS 36.101** 2010