◆ LTE/LTE-A Signal Compression on the CPRI Interface

Bin Guo, Wei Cao, An Tao, and Dragan Samardzija

The Centralized, Cooperative, Cloud Radio Access Network (C-RAN) is a next-generation wireless access network architecture based on centralized processing, collaborative radio, and real time cloud infrastructure. In this architecture, different access technologies (Global System for Mobile Communications (GSM)/Time Division Synchronous Code Division Multiple Access (TD-SCDMA)/Wideband Code Division Multiple Access (WCDMA)/Long Term Evolution (LTE)) can be supported on the same hardware platform in a baseband pool system, which can largely reduce system costs. Long Term Evolution (LTE) and Long Term Evolution-Advanced (LTE-A), which are based on Orthogonal Frequency Division Multiplexing (OFDM) and multiple input multiple output (MIMO) technologies, are regarded as the main wireless access technologies in the evolution from 3G to 4G. A variety of novel technologies such as multi-antenna MIMO, carrier aggregation (CA), and coordinated multipoint have been introduced in LTE/LTE-A to improve system performance, especially in the C-RAN architecture. However, one of the technical challenges for the C-RAN architecture is the fiber bandwidth required for data transmission between the remote radio unit (RRU) and the baseband unit (BBU). We propose using a low-latency baseband signal compression algorithm to solve this problem by reducing the fiber data rate. Using the characteristics of the LTE signal data, we remove the redundancy in the spectral domain. We also leverage block scaling in conjunction with using a linear or nonlinear (non-uniform) quantizer to minimize quantization error. This algorithm effectively reduces the amount of data transmitted between the BBU and RRU, and facilitates the deployment of LTE in the C-RAN architecture. We verified the robustness of the algorithm via simulations and lab tests. The proposed algorithm yields good system performance at a 1/2 compression rate and at a 1/3 compression rate in a practical propagation environment. © 2013 Alcatel-Lucent.

Alcatel·Lucent

Panel 1. Abbreviations, Acronyms, and Terms

3G—Third generation
3GPP—3rd Generation Partnership Project
4G—Fourth generation
ADC—Analog-to-digital converter
AGC—Automatic gain control
AMC—Adaptive modulation and coding
BBU—Baseband unit
CA—Carrier aggregation
CoMP—Coordinated multipoint
CPRI—Common public radio interface
CPU—Central processing unit
C-RAN—Centralized, Cooperative, Cloud RAN
DAC—-Digital-to-analog converter
eNodeB—Evolved NodeB
EVM—Error vector magnitude
FDMA—Frequency Division Multiple Access
FPGA—Field programmable gate array
GSM—Global System for Mobile
   Communications
IFFT—Inverse fast Fourier transformation
IQ—In-phase/quadrature-phase

LTE—Long Term Evolution
LTE-A—LTE-Advanced
MCS—Modulation and coding scheme
MIMO—Multiple input multiple output
OFDM—Orthogonal Frequency Division
   Multiplexing
PAPR—Peak-to-average power ratio
PRB—Physical resource block
RAN—Radio access network
RRU—Remote radio unit
SC—Single carrier
SCDMA—Synchronous code division multiple
   access
SNR—Signal-to-noise ratio
SQNR—Signal-to-noise quantization ratio
TD—Time division
UE—User equipment
VSA—Vector signal analysis
WCDMA—Wideband Code Division Multiple
   Access

Introduction

The Centralized, Cooperative, Cloud Radio Access Network (C-RAN) is a wireless access network architecture for next-generation communications which is based on centralized processing, collaborative radio, and a real time cloud infrastructure [4]. In the C-RAN architecture, the baseband units (BBUs) are centrally located in a pool configuration which is connected to the remote radio units (RRUs) via optical fiber. The BBU pool can be built using general purpose central processing units (CPU). Baseband processing resources in the BBU pool can be leveraged by using cloud computing technology. The C-RAN collaborative radio technology can significantly improve wireless network performance, while C-RAN's baseband centralization enables resource sharing, reduces energy consumption, and improves the utilization of infrastructure [10]. C-RAN does however consume more fiber resources than a traditional architecture. The prospect of transmitting untreated signals between BBUs and RRUs in a C-RAN would lead to a much higher transmission rate and a smaller transmission delay. Therefore,

reducing fiber resource consumption becomes a critical issue in implementing a C-RAN architecture.

Based on Orthogonal Frequency Division Multiplexing (OFDM) and multiple input multiple output (MIMO) technology, Long Term Evolution (LTE) is regarded as the technology at the forefront of the evolution from 3G to 4G services, and it has already been adopted by many operators. In order to improve cell throughput, multi-antenna technology has become an important direction in wireless access technology. Multi-antenna MIMO and beamforming technology can be used to improve system performance [7]. LTE-Advanced introduces novel technologies such as carrier aggregation (CA), multi-antenna enhanced MIMO, and coordinated multipoint transmission (CoMP) [8]. However, the system bandwidth required for LTE-Advanced can be five times higher than that required for LTE [1, 6]. LTE is designed to improve system performance, but it also increases the amount of data transmitted between the BBU and RRU, and therefore the need for more costly fiber optic links.

In this paper, we introduce a low-latency baseband signal compression algorithm. In the LTE
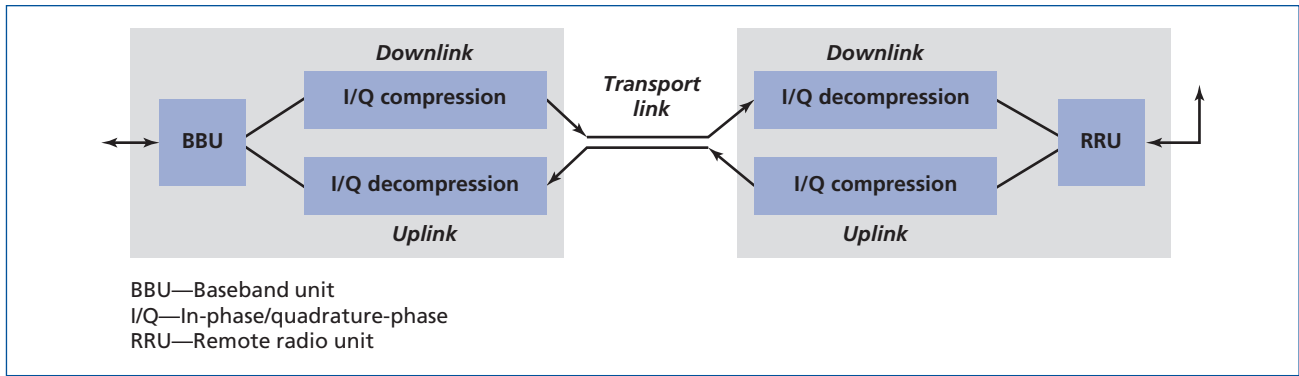
Figure 1.
BBU-RRU system with J/Q compression.

system architecture, the evolved NodeB (eNodeB) includes two main components, the BBU and the RRU. The LTE baseband signal is generated in the BBU for the downlink, and then the baseband signal is transferred to the RRU. Digital-to-analog and analog-to-digital conversion is done in the RRU. Finally, the signal is sent to the radio interface. A standard common public radio interface (CPRI) [5] is used to transfer data via optical fiber between the BBU and the RRU. The signals are a volume of samples with complex value of a certain bit width, such as an I-part of 15 bits and a Q-part of 15 bits. In the C-RAN architecture, a large number of RRUs are connected to the BBU pool, which thus requires a large number of fiber optic links. Our algorithm is designed to compress the digital signals and the bit width before CPRI framing and to do so without any unacceptable degradation in system performance. If we can achieve this goal, we can reduce the costs associated with optic fiber link resources quite significantly.

The basic functional blocks of our compression scheme were introduced in Alcatel-Lucent's lightRadio™ solution, and are shown in **Figure 1** [9]. An optical fiber pair is used between the BBU and RRU, one for the downlink and the other for uplink. Compression and decompression processing modules are respectively added inside the BBU for uplink and inside the RRU for downlink transmission. For downlink, the BBU generates an uncompressed baseband signal, and then in-phase/quadrature-phase (I/Q) samples are compressed by the compression module.

Finally the compressed I/Q samples are sent to the RRU through the CPRI link. In the RRU, the compressed I/Q samples are decompressed by the decompression module, and then undergo normal processing inside the RRU. The uplink process is similar to the downlink, except that the original samples come from the analog-to-digital converter (ADC). The input to the ADC is the analog radio signal received from the antennas. Algorithm details are described in the next section.

Algorithm Details

The compression process consists of three steps as shown in **Figure 2**. In the decompression module, the processes are reversed. They include:

- Removal of redundancy in the spectral domain,
- Block scaling, and
- Quantization.

The first step is removing the redundancy in the spectral domain. The redundancy is caused by over-sampling on signals. The oversampling is used to prevent out-of-band radiation from interfering with useful signals in the OFDM system. This module is designed to remove redundant spectral data by means of a low-pass filter, which ensures that only the data on valid bandwidth is sent. The number of samples can be reduced significantly in this block.

The second step is to segment the samples into blocks. A scaling factor is selected for each block. Piecewise scaling is used to meet the fluctuations of data in the time domain, especially for large and
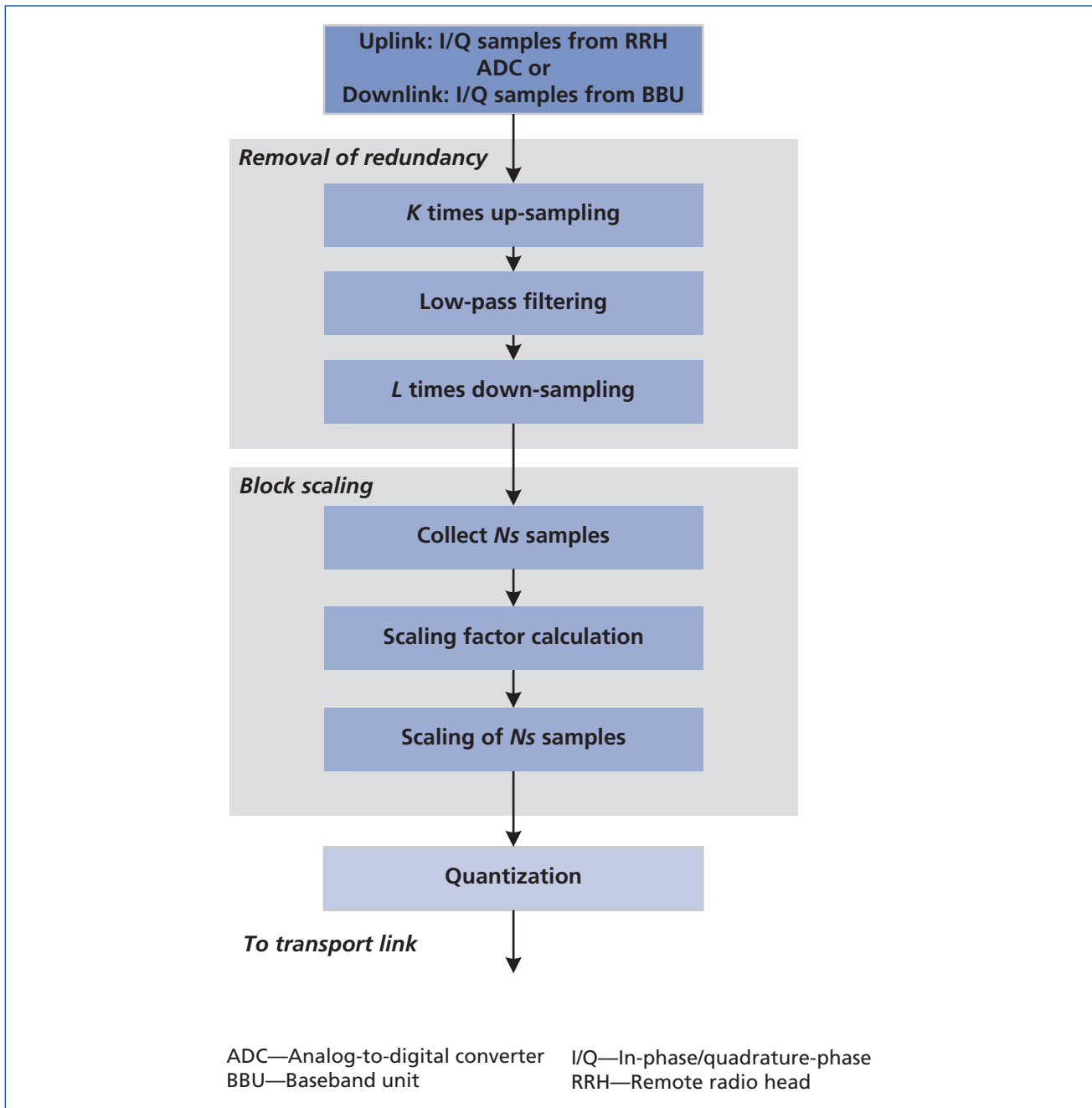
Figure 2.
I/Q samples compression block diagram.

small values. Block scaling is used to reduce the quantization error.

The third step is to quantize the pre-processed I/Q samples from the original bit width of Q_s to a Q_q bit width, normally $Q_q < Q_s$. The quantizer was specifically designed according to the data characteristics assuming a bit width of Q_q and Q_s. After the quantization, the bit width of one I/Q sample will be reduced to Q_q. This function is performed sample by sample.

After this processing is complete, the number of data samples is reduced significantly from that in the original data. Meanwhile the bit width of each sample

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

E-UTRA—Evolved UTRA
UTRA—Universal terrestrial radio access

is also reduced. The decompression module will use the same method to recover the original data. In the subsection following, each module will be described in detail.

Removal of Redundancy in Spectral Domain

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

In LTE implementations, inverse fast Fourier transformation (IFFT) is used to generate OFDM signals. Typically, the Nyquist sampling theorem would then be used to select the smallest appropriate IFFT size. However, due to the complexity of this implementation, in practice, we use an IFFT size based on a power of two. For example, consider 20 MHz of spectrum for a Time Division (TD)-LTE system. As shown in **Table I**, we used a 100 physical resource block (PRB) configuration from 3rd Generation Partnership Project (3GPP) specification 36.101 [2]. So the useful subcarrier number is 100*12 = 1200 (the number of subcarriers within each PRB is 12), the subcarrier width is $\Delta f = 15$ kHz, the width of signal bandwidth is 1200*15 kHz = 18 MHz [3]. The smallest size power of two which is larger than 1200 is 2048, so the signal is mapped on 1200 subcarriers in the middle of 2048 subcarriers. The remaining subcarriers are filled with zeros, so in practice, 2048*15k = 30.72 MHz bandwidth is used. Taking into account the 2 MHz bandwidth reserved for filter edge roll-off, there is about 10 MHz of bandwidth redundancy in the spectral domain. Thus, in the uncompressed

form, a signal that is spectrally broader than necessary is being transmitted in the CPRI frames.

The first step in reducing redundancy is upsampling on the input signal. In this process, zeros are inserted into the original signal. We then let this signal with zeros pass through a low-pass filter. Finally, downsampling is performed on the filter's output signal. The sampling rate of the original signal is f_s. After these processes, the sampling rate of the output signal becomes a lower sampling rate f_{ds}.

The value of f_{ds} depends on the amount of redundant spectrum to be removed. It is important to note that in addition to the useful signal bandwidth, adequate protection bandwidth should also be reserved. Adequate protection bandwidth ensures that the filter doesn't damage useful information in the central band. The system bandwidth is limited to $[-f_{ds}/2, f_{ds}/2]$ after downsampling.

The downsampling factor F is a rational number used to evaluate the efficiency of the process, where K is the upsampling time and L is the downsampling time. K and L are positive integers.

$$F = \frac{f_{ds}}{f_s} = \frac{K}{L} \leq 1 \qquad (1)$$

Our aim in designing the filter coefficients was to ensure good amplitude and phase response in the passband. Meanwhile, to the greatest extent possible, signal in the block band should also be removed, and this becomes a tradeoff in the filter design. Our algorithm uses a Kaiser window to generate the filter coefficient, which shows desirable amplitude/phase response in its passband.

Figure 3 shows the amplitude and phase response for the filter used in the simulation we described in the performance evaluation section.

Figure 4 compares the frequency spectrum from the original signal and the signal pass "redundancy removal" block. Figure 4a shows the frequency spectrum of a typical LTE downlink 20 MHz system bandwidth baseband signal with a sample rate of 30.72 MHz. As shown, the useful spectrum is in the center and the redundancy is seen on both sides. We use the filter in Figure 3 to deal with the original time-domain signal. The output spectrum of the "redundancy
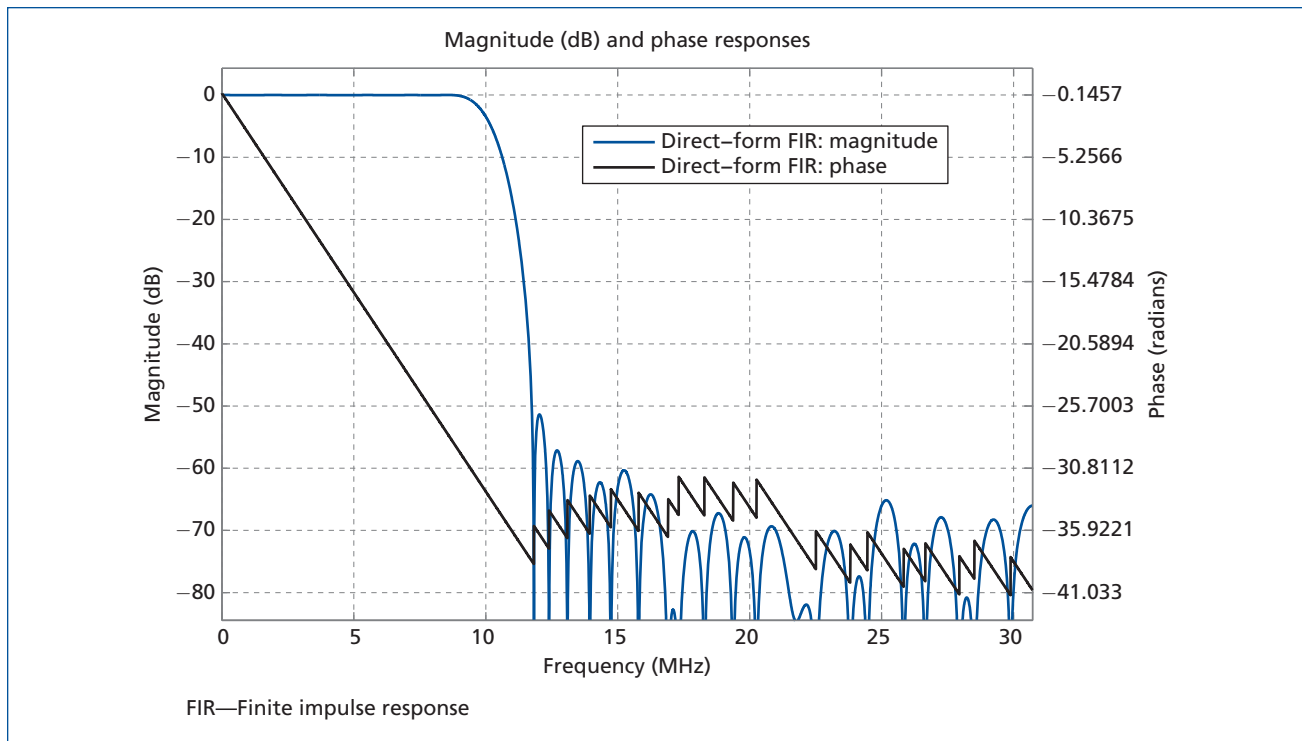
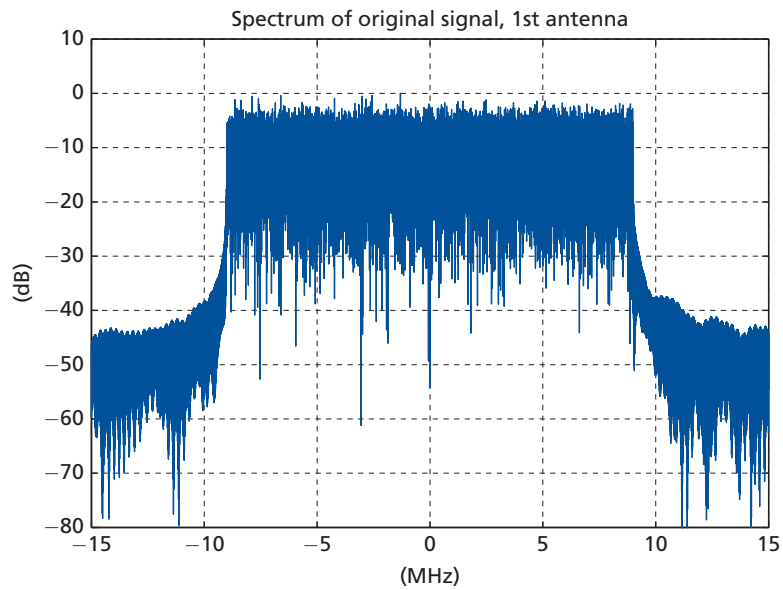**Figure 3.**
Amplitude and phase response of filter.

removal" block is shown in Figure 4b. We demonstrate that most of the redundant spectrum has been filtered out, while the useful band is retained.

Specific parameters, such as the downsampling factor and filter coefficient, will vary given different given system bandwidths. In practical implementations, key parameters (e.g., the number of filter coefficient taps) should be chosen carefully to achieve the best tradeoff possible between complexity and performance.

Block Scaling

In LTE, OFDM is used in downlink transmission. Orthogonal subcarriers are used to implement multiple narrowband systems, which introduces large peak-to-average power ratio (PAPR). A typical LTE radio signal has a large dynamic amplitude range. A Single Carrier Frequency Division Multiple Access (SC-FDMA) transmission technology is used in the LTE uplink to generate a signal with a lower PAPR,

and in general an automatic gain control (AGC) block is also used to reduce the difference in power between user signals. However, the signal amplitude still has great volatility within a large range. Typically, time domain signal samples are transmitted using Q_S bits per complex component in a fixed-point scheme. If all of these signals are compressed to target Q_q bits directly without any other processing, data saturation will frequently occur. At the same time, signal components with low amplitude are being impacted by high quantization noise. We introduce block scaling to solve this problem. In block scaling, the output from previous processing is segmented into a number of small blocks with N_S samples. In each small block, a scaling factor is determined for the N_S I/Q samples within the block. This scaling factor is used to process the amplitude of the samples into the normal range provided by Q_q bits in order to minimize the quantization error. The scaling factor is also transmitted as overhead in each block. However, due to minimized
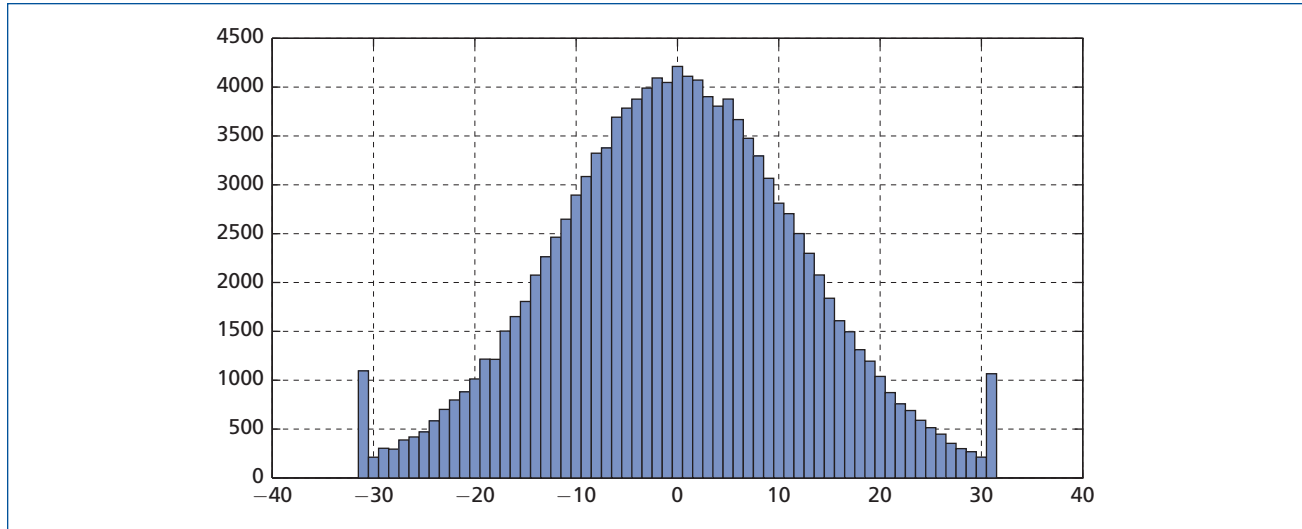
(a) The original signal spectrum for 20 MHz system bandwidth
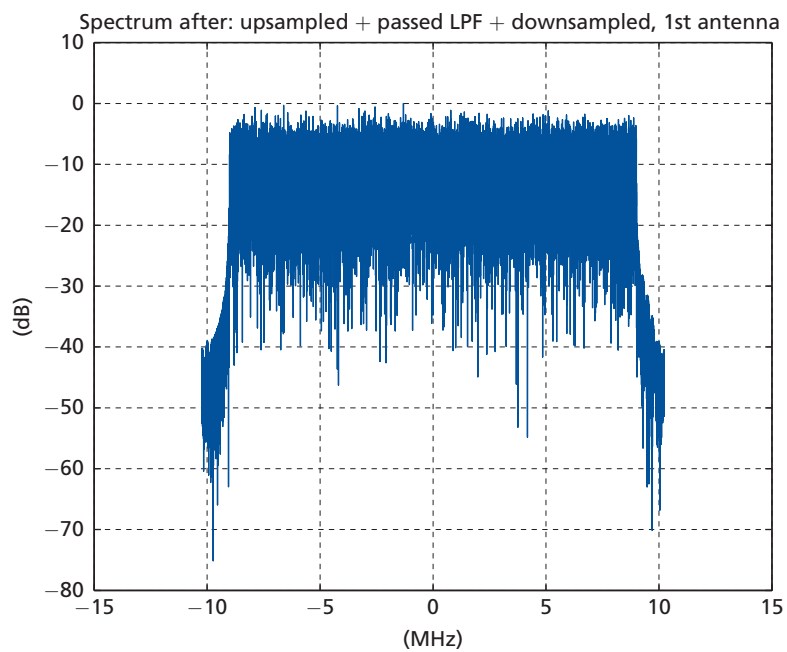
(b) The signal spectrum pass "redundancy removal" block

LPF—Low pass filter

Figure 4.
Frequency spectrum of the original signal and the signal pass "redundancy removal" block.

Figure 5.
Amplitude distribution of a 6 bit downlink signal.

quantization error, a lower quantizer resolution is applied, which results in an overall reduction in the transport data rate. This function is performed on each block with N_S I/Q samples, at the output of the decimator. The scaling factor is transmitted together with each N_S I/Q sample. Minimizing the block length N_S will reduce the subsequent quantization error, however it will also increase the transmission overhead. Therefore, the selection of block length N_S requires a tradeoff between the signal quality and the transport data rate.

In the kth block with N_S samples, the sample with the largest absolute value will be selected as the scaling factor:

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

The scaling factor above is an integer and it does not exceed $2^{Q_s}-1$, where Q_s is the bit width used for transferring the scaling factor. Each sample in the block is then scaled with this scaling factor as shown in equation 3, where the Q_q is the bit width used in quantization. $s_d(i)$ is the signal sampled at the frequency f_{ds}. $S_s(i)$ is the output of black scaling.

$$s_s(i) = s_d(i) \frac{2^{Q_q-1}-1}{S(k)} \quad (3)$$

Where

$$S(k) = \begin{cases} \lceil A(k) \rceil \, for \, \lceil A(k) \rceil \le 2^{Q_s}-1 \\ 2^{Q_s}-1 \, for \, \lceil A(k) \rceil > 2^{Q_s}-1 \end{cases} \quad (4)$$

Quantization

After block scaling, I/Q samples are quantized using a quantizer with a Q_q bit resolution for each component. This function is performed sample by sample. We apply a simple linear (uniform) quantizer with a resolution of Q_q. We found that if Q_q is close to the original bit width Q_s, the linear quantizer works well. However, if Q_q decreases significantly, applying a nonlinear quantizer with optimized distances between quantization levels can result in a lower quantization error. This is an important consideration in the design of the non-linear (non-uniform) quantizer. **Figure 5** shows the amplitude distribution of the downlink signal with scaling of $Q_q=6$. It is clear that a large portion of data has minor amplitude around zero, except for two impulses at the extreme values caused by block scaling saturation. We took these features into account in our quantizer design, and used the smaller quantization distance for signal components when there was high probability of occurrence (here, at low amplitudes). Using a training sequence to tune a quantizer is common practice
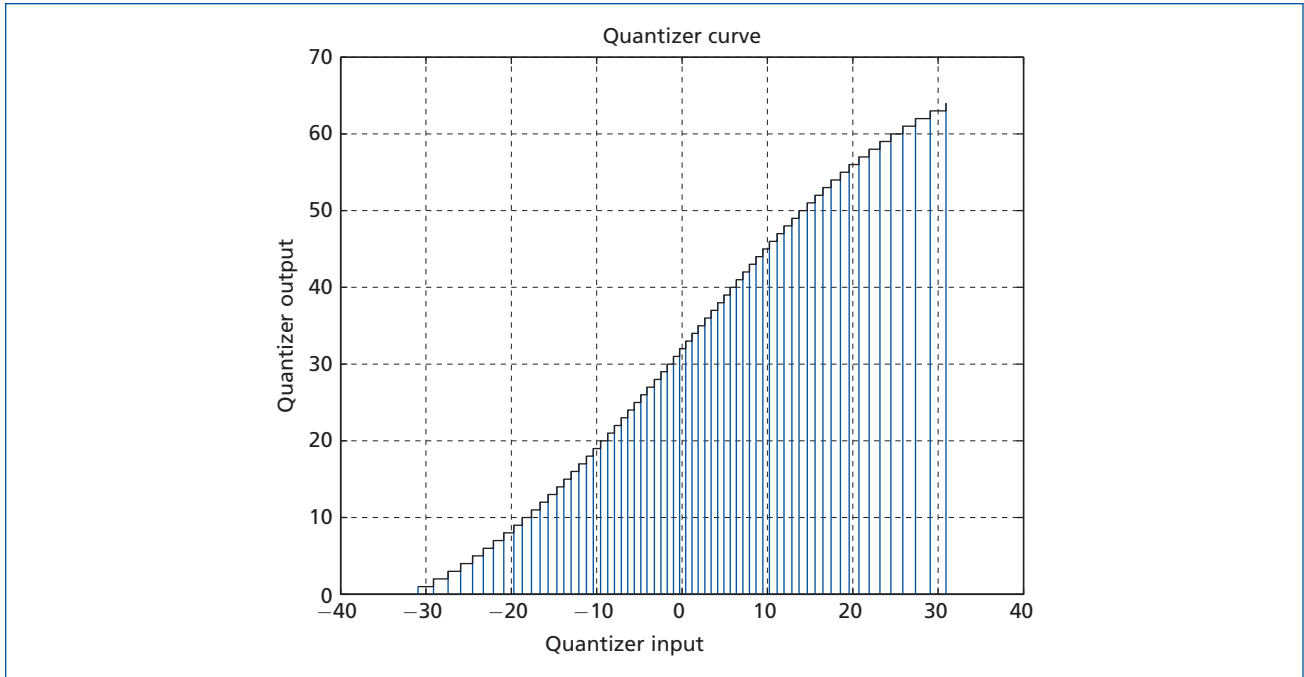
Figure 6.
The trained quantizer curve for 6 bits.

in non-linear quantizer design. Multiple iterations are required to quantify an appropriate distance with a minimum quantization error. According to this rule, one trained nonlinear (non-uniform) quantizer curve of 6 bits is shown in **Figure 6**. We can see that it has a smaller quantizer distance around zero.

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 by analyzing the tradeoff between the required signal quality and data rates. Meanwhile, customer requirements are also a major factor in practical implementation.

After completing the three steps above, the number of original symbols as well as the bit width are both compressed. At the RRU, the process is reversed to recover the original data. Of course this reverse process can't recover all the original information. The choice of the parameters and the compression ratio will have a significant impact on compression injury. We also performed a simulation analysis and lab

evaluation of the compression algorithm. We describe our results in the section below.

Performance Evaluation

Performance degradation depends on the compression ratio and parameter configuration. The compression ratio in the algorithm is calculated as below based on the parameters configured in the algorithm:

$$C_{ratio} = \frac{K/L^*Q_q + Q_s/N_s}{Q_s} \quad (5)$$

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

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

Where $\bar{x}$ is the output signal (after the compression and decompression processes have already been completed), while x is its idealized noise-free version as shown in **Figure 7**. A basic principle of compression algorithm design is to reduce the EVM under a certain compression ratio.
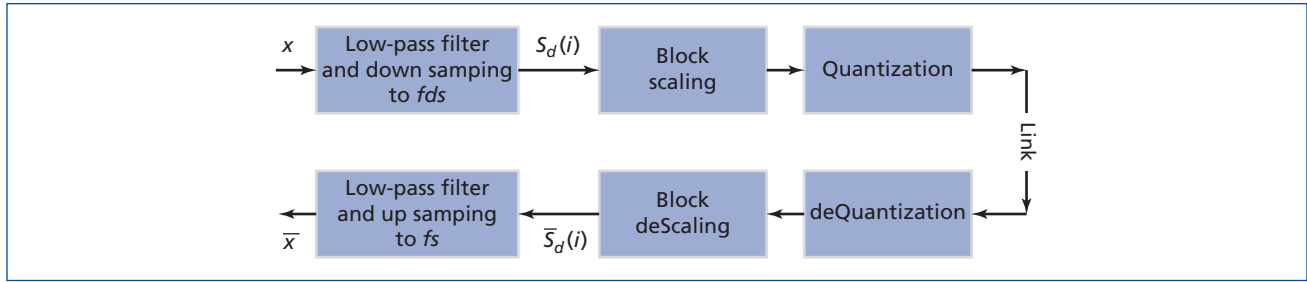
Figure 7.
Block diagram of compression and decompression processing.

Table II. Downlink LTE EVM requirements.

Modulation scheme	Maximum EVM (%)
QPSK	17.5
16 QAM	12.5
64 QAM	8

EVM—Error vector magnitude
LTE—Long Term Evolution
QAM—Quadrature amplitude modulation
QPSK—Quadrature phase shift keying

Table III. Parameter configuration.

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

In order to evaluate the algorithm more carefully, we have defined two types of EVM, the time domain EVM and frequency domain EVM. The frequency domain EVM is defined based on the bandwidth used by the frequency domain signal, while the time domain EVM is defined on the entire bandwidth signal. This is because our primary focus is on signal changes across the entire useful bandwidth.

3GPP has defined EVM requirements for different modulation schemes, as shown in **Table II.** The requirements are based on an end-to-end definition of EVM between the eNodeB and UE, and it is also a frequency domain EVM definition. EVM loss caused by I/Q compression is a part of end-to-end EVM in practical systems.

In addition, in order to assess the impact of scaling and quantization, we define the following SQNR:

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

Where, $s_d(i)$ is the block scaling input in equation 3, $sd(i)$ is the output of block rescaling at decompression side as shown in Figure 7.

Simulation Analysis

The algorithm parameters for compression are shown in **Table III**. In the simulation and implementation, linear quantization is used for 11 bit and 8 bit compression, while nonlinear quantization is used for 6 bit compression.

In order to achieve an ergodic signal, we used the six data configurations (with different PRB numbers and modulation schemes) shown in **Table IV** which were selected to simulate different downlink user scheduling signals.

Our simulation results at compression ratios of 0.5201, 0.3868, and 0.2979 are respectively are detailed in **Table V, Table VI,** and **Table VII**.

The results show that EVM changes markedly at different compression ratios. Under 11 bit and 8 bit compression, the compression ratios are 0.5201 and 0.3868 respectively. The frequency EVM can be controlled to less than one percent. The 1/2 compression ratio means that if the original data was transmitted on two optical fibers in the practical application, it can now be transmitted on only one optical fiber. Looking more closely at the compression ratio, for example the 6 bit compression with low SQNR, the

Table IV. Simulation cases configuration.

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

PRB—Physical resource block
QAM—Quadrature amplitude modulation
QPSK—Quadrature phase shift keying

Table VI. 8 Bits compression results compression ratio: 0.3868.

Index	Frequency domain EVM (%)	Time domain EVM (%)	SQNR (dB)
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)

EVM—Error vector magnitude
SQNR—Signal-to-noise quantization ratio

Table V. 11 Bits compression results compression ratio: 0.5201.

Index	Frequency domain EVM (%)	Time domain EVM (%)	SQNR (dB)
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)

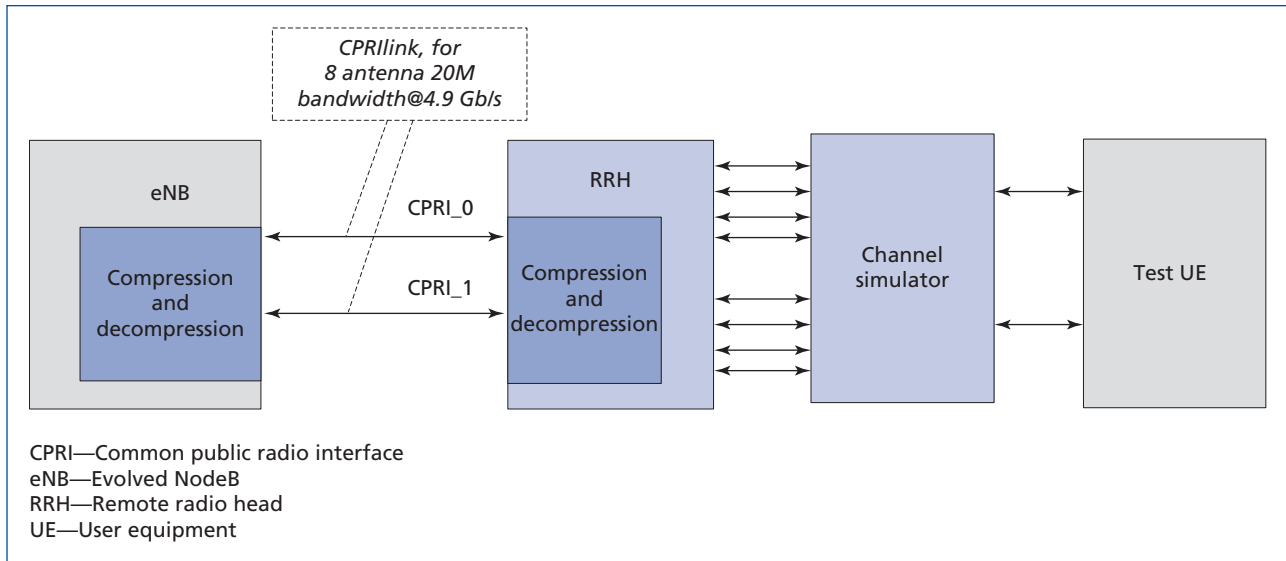EVM—Error vector magnitude
SQNR—Signal-to-noise quantization ratio

Table VII. 6 Bits compression results compression ratio: 0.2979.

Index	Frequency domain EVM (%)	Time domain EVM (%)	SQNR (dB)
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)

EVM—Error vector magnitude
SQNR—Signal-to-noise quantization ratio

EVM has obviously increased, but it still falls within 3PGG requirements. And, in the spirit of joint optimization across the entire ecosystem, other parts of the system, such as the RRU and the UE, can be carefully designed in parallel to obtain further compression space.

TDD-LTE Demo Test Analysis

Based on the simulation results, we built a TD-LTE system demo testbed in our lab. We used field programmable gate array (FPGA) resources inside the eNodeB and RRU to implement the compression and decompression module. Our demo evaluated the 1/2 and 1/3 compression rates. The algorithm parameters were consistent with those used in simulation.

Our test sequence included:

1. Connecting the test equipment as shown in **Figure 8**. The LTE CPRI compression lab test environment is shown in **Figure 9**. The BBU is on the back framework. The black box at left is the test UE. The vector signal analysis (VSA) is on the top of the UE. The RRH and channel simulator are not shown in this figure because of their size. Test performance was monitored via the terminal shown at right.
2. Setting the system to work on the D-band carrier and to achieve a steady state.
3. Setting the system to transmission mode 3.
4. Loading the SCM channel model and gradually increasing the noise to reduce the signal-to-noise ratio (SNR).
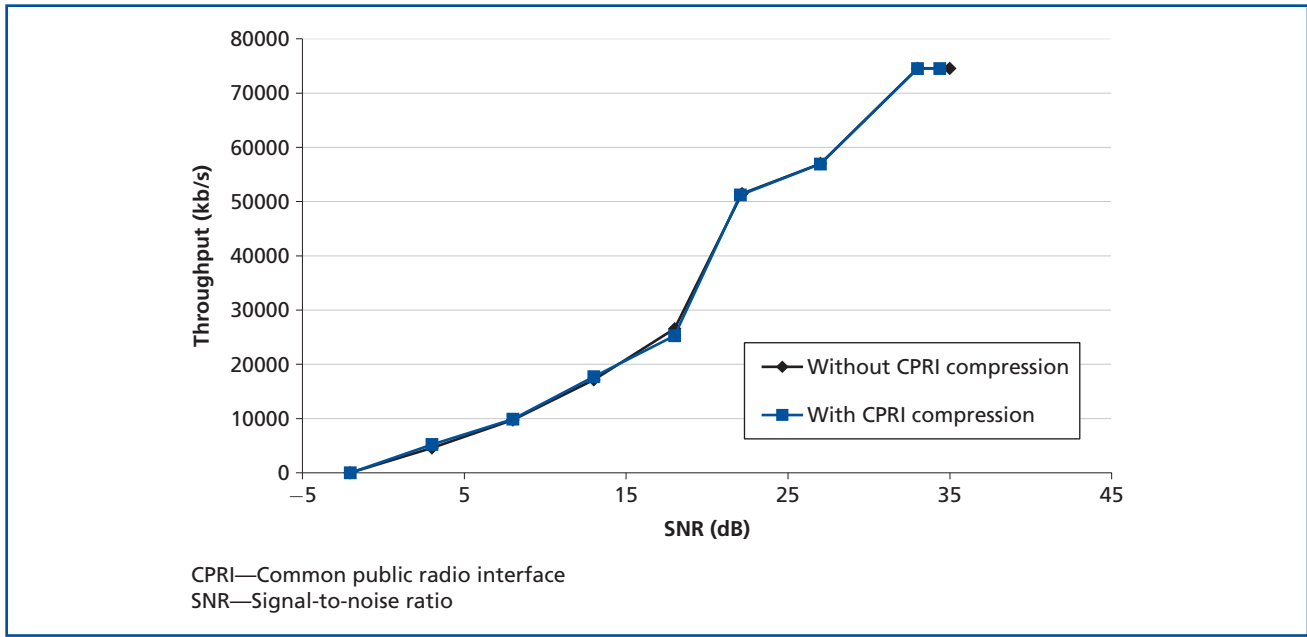
Figure 8.
Eight antenna system CPRI compression lab test equipment connection.

Diagram labels:

CPRIlink, for
8 antenna 20M
bandwidth@4.9 Gb/s

eNB

Compression
and
decompression

CPRI_0

CPRI_1

RRH

Compression
and
decompression

Channel
simulator

Test UE

CPRI—Common public radio interface
eNB—Evolved NodeB
RRH—Remote radio head
UE—User equipment

CPRI—Common public radio interface
LTE—Long Term Evolution
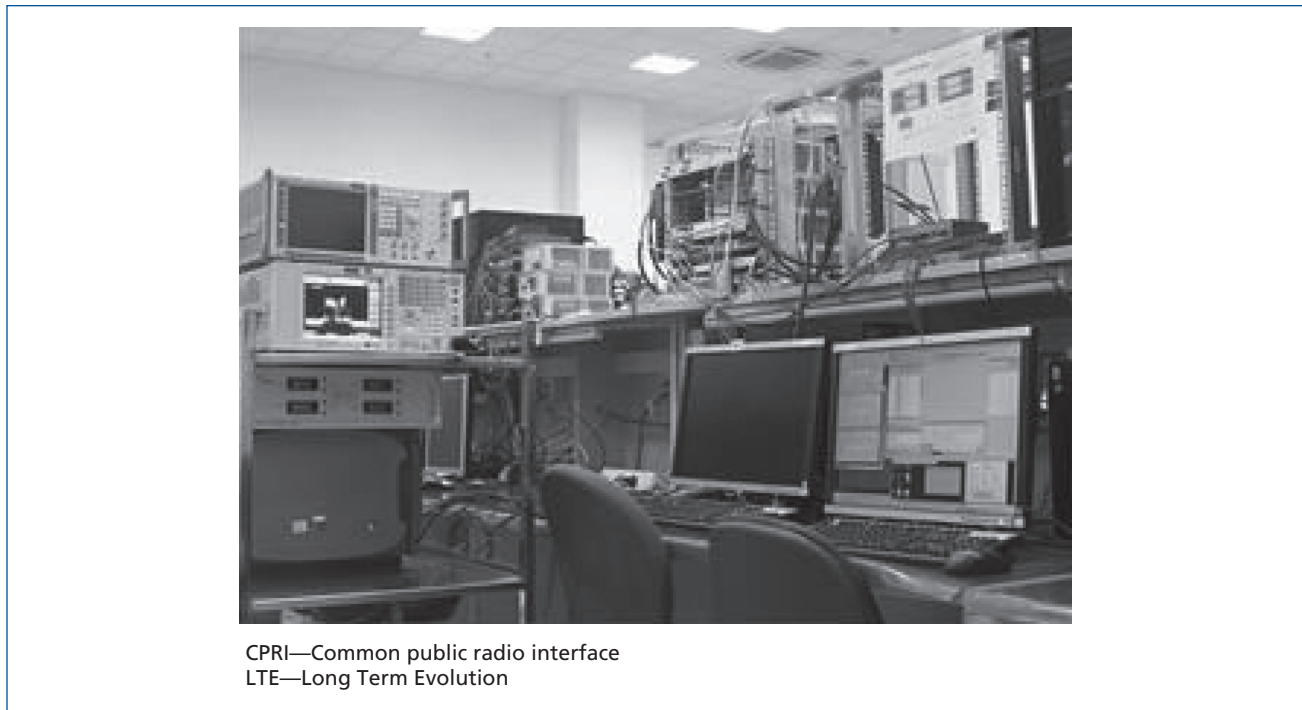
Figure 9.
LTE CPRI compression lab test environment.

5. Using the test UE to receive the downlink signal and obtain an average throughput performance under each SNR.
6. Shutting off the CPRI compression and then repeating steps 2 through 5.

We tested our demo system in a lab environment. Since performance was to be compared for both the compressed and uncompressed transmission schemes, we needed to maintain a stable state. Before moving forward, we needed to ensure that:

Figure 10.
Downlink performance comparison under 1/2 compression.

1. The platform was operating under normal conditions, and
2. The equipment had been fully preheated, and performance indicators were showing stable state. The test results are shown as below.

1/2 CPRI compression rate. In our demo test for downlink, the LTE downlink signal was generated based on a standard 3GPP protocol for a commercial LTE eNodeB, the system bandwidth was 20 MHz, and the uplink/downlink configuration was 1. 100 PRB resources were scheduled for target user equipment (UE), and adaptive modulation and coding (AMC) was used for scheduling. **Figure 10** shows the test results for downlink under 1/2 ratio CPRI compression. The figure shows that the throughputs with and without compression are very close. The compression algorithm works well with limited performance degradation.

In our demo lab test for uplink, the system bandwidth was also 20 MHz, and the uplink/downlink configuration was 1. 10 PRB resources were scheduled for the target UE. We used fixed modulation and coding scheme (MCS) 16. We used an SC-FDMA signal generation scheme for the LTE uplink signal.

The signal PAPR was not as high as the downlink signal. However, since the uplink signal should pass the spatial channel before being received, the signal characteristic will become unstable, which will affect the design of the compression algorithm. The second test case examined the uplink CPRI compression performance for a single UE signal. **Figure 11** shows the uplink throughput performance at a 1/2 compression rate. We observed that the compression algorithm works well with a 1/2 compression rate with a single UE.

Next, we carried out an uplink throughput performance comparison in the presence of interference from multiple UEs. **Figure 12** shows that the performance gap is negligible with a 1/2 compression rate in the presence of strong interfering UEs.

1/3 CPRI compression rate. Figure 13 provides a comparison of downlink throughput performance at a 1/3 compression rate. Although the EVM loss in the simulation analysis is higher with 1/3 compression than with 1/2 compression, our tests indicated that the compression algorithm still worked well at the 1/3 compression rate for downlink.
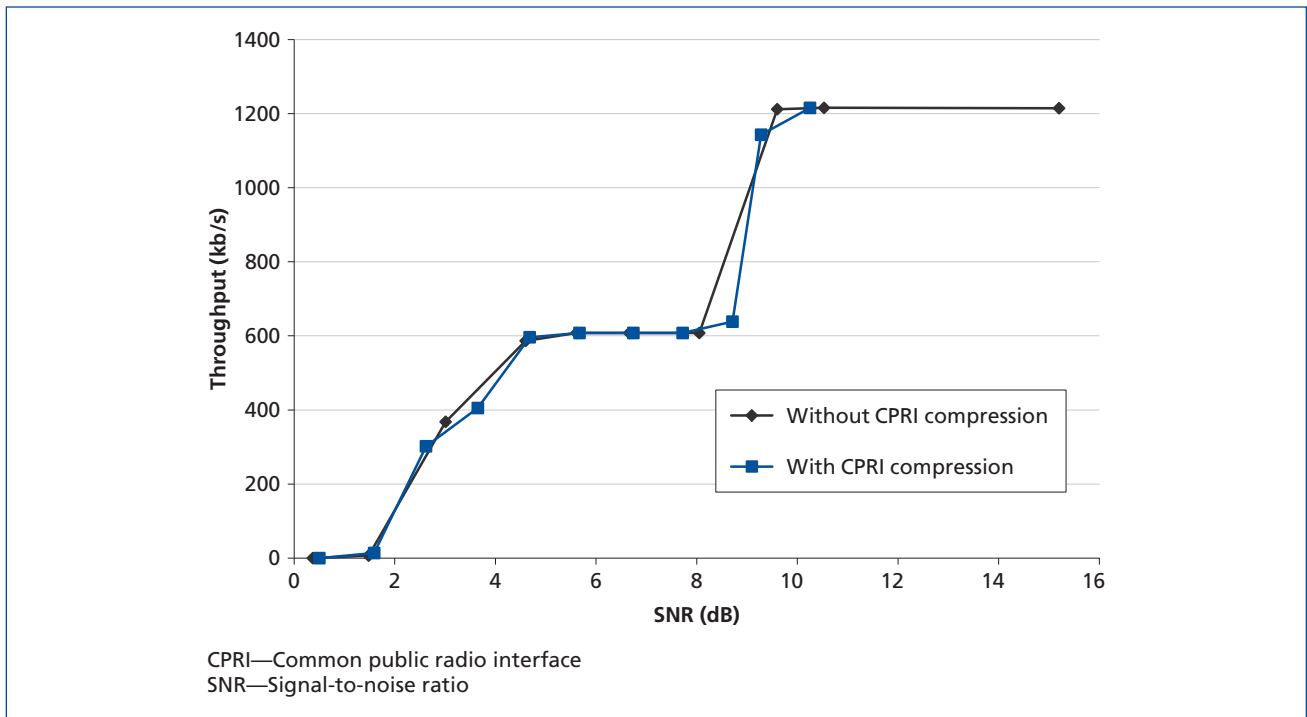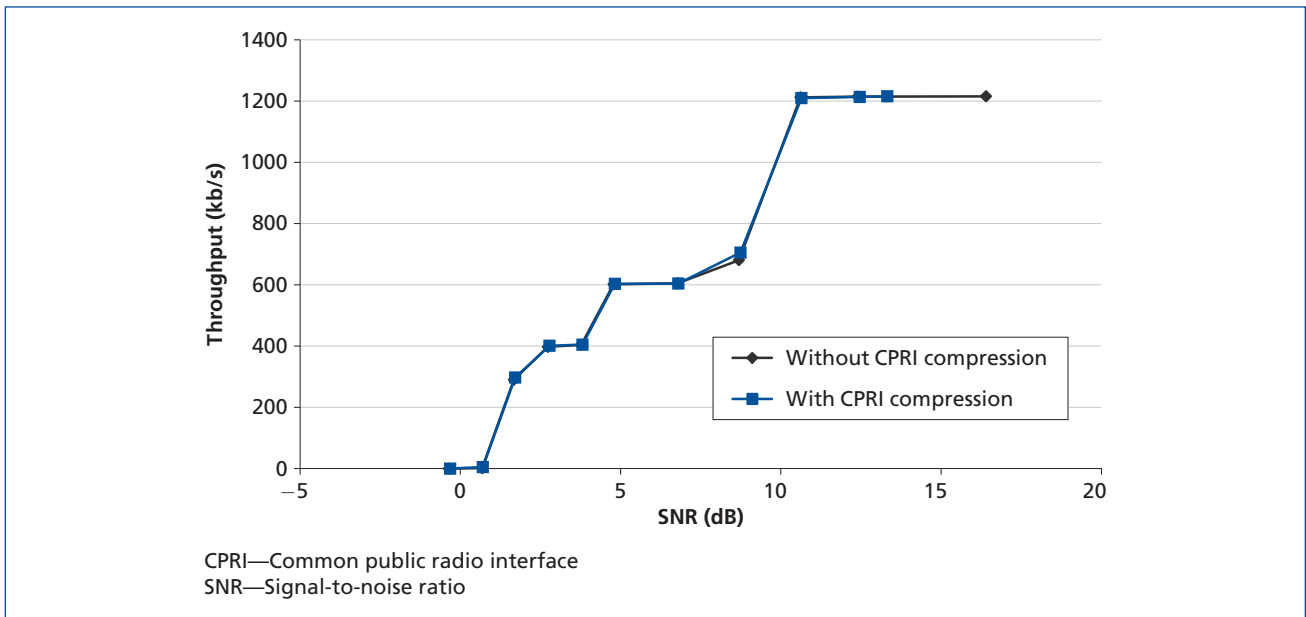
Figure 11.
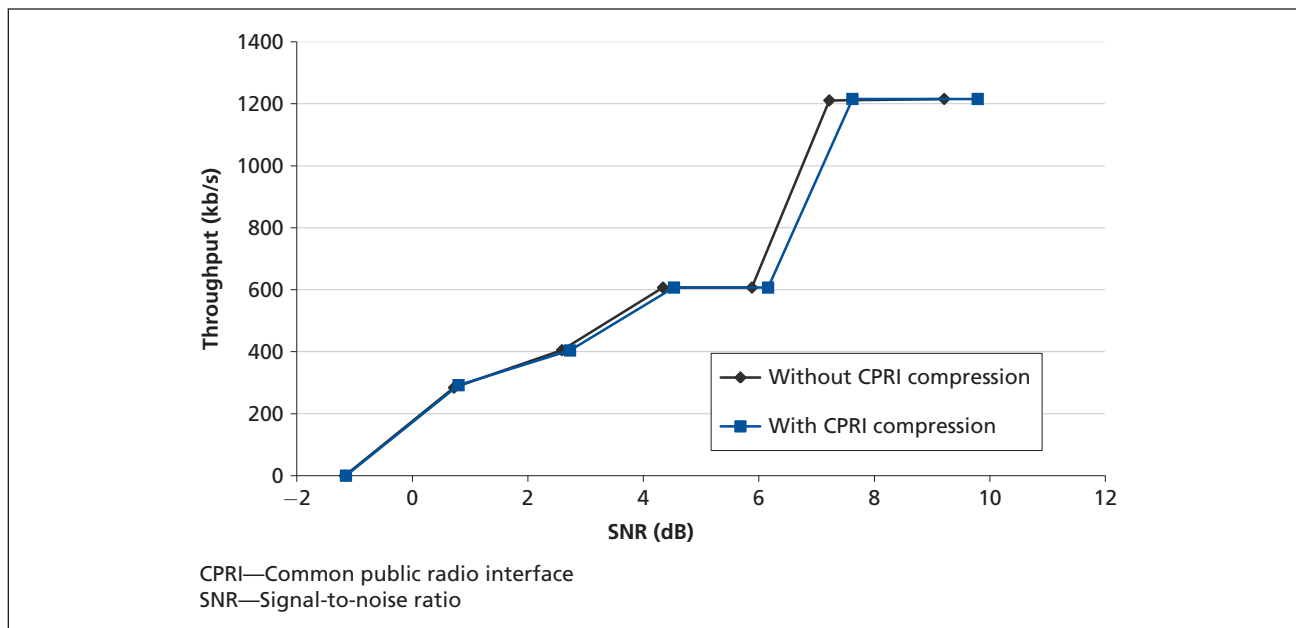**Uplink performance comparison under 1/2 compression.**
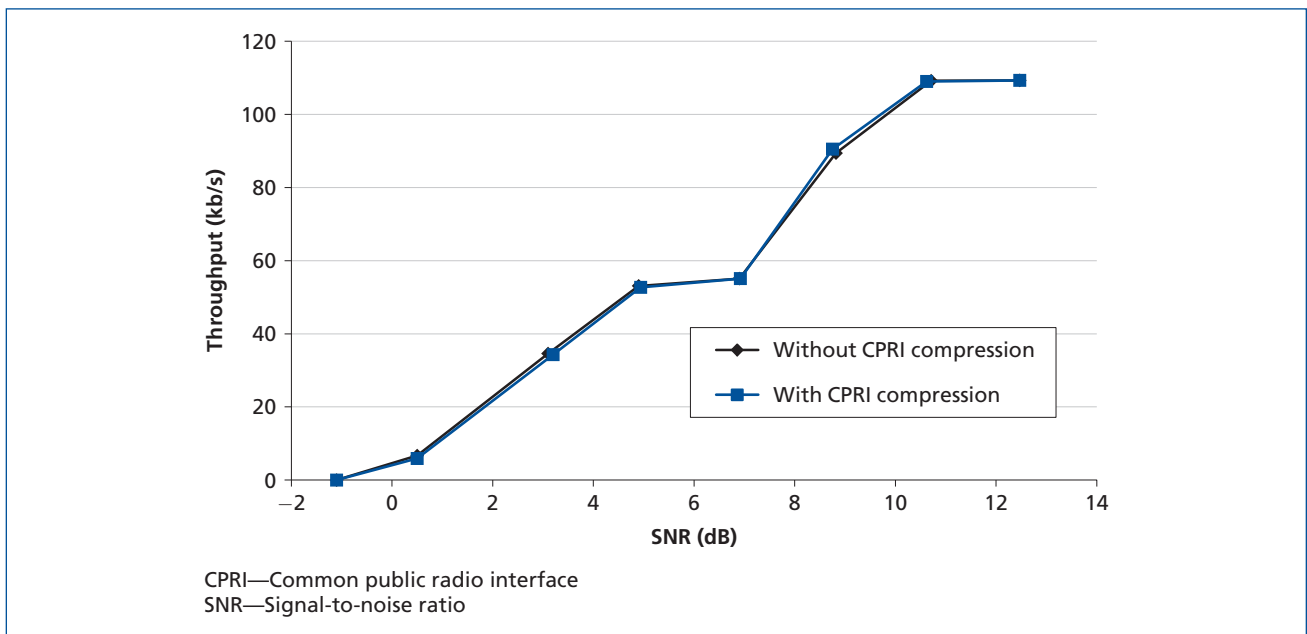
Figure 12.
Uplink performance with interference under 1/2 compression.
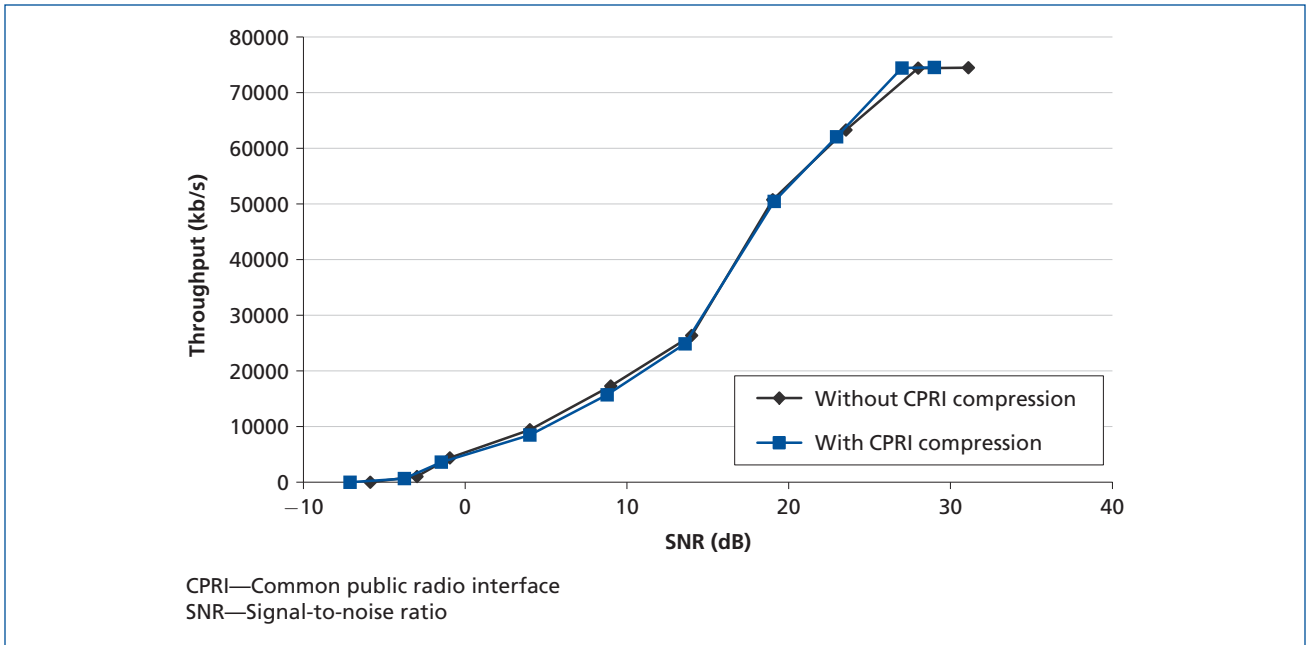
Figure 13.
Downlink performance under 1/3 compression.

Figure 14.
Uplink petformance under 1/3 compression.

Figure 15.
Uplink performance with interference under 1/3 compression.

Figure 14 shows uplink throughput at a 1/3 compression rate when a single UE was deployed. We found that the compression algorithm yielded satisfactory results at the 1/3 compression rate.

Figure 15 shows uplink throughput performance at a 1/3 compression rate in the presence of strong interference from another UE. We observed that with low SNR, the two curves match each other well even when there is strong interference. Under high SNR, the 1/3 compression leads to performance loss, but only 0.3 dB at an SNR of 6 dB to 8 dB.

Conclusions

CPRI transmission is the bottleneck for C-RAN implementation. LTE and LTE Advanced (LTE-A) introduce new and robust technologies, but also increase the amount of data transmitted on the CPRI interface. This paper describes a CPRI compression algorithm for an LTE system which reduces the data rate on the fiber link. By eliminating redundant spectrum bandwidth and compressing the bit width, this algorithm can effectively reduce the amount of data transmitted on the CPRI. Simulation results show that data loss is negligible at a low compression ratio. The EVM can be controlled to less than one percent at the 1/2 compression rate. We verified our compression scheme in an LTE lab demo. Performance with a 1/2 compression rate is ideal; performance with a 1/3 compression rate, a little less so. We also noted that EVM deteriorates significantly at a high compression ratio, which indicates that there may be limited applicability for this specific implementation. In real life applications, there is a tradeoff between performance and resource consumption. In future research, we will study methods for optimizing algorithm configuration in certain scenarios, such as the design of the quantizer.

References

[1] 3rd Generation Partnership Project, "Feasibility Study for Further Advancements for E-UTRA (LTE-Advanced) (Release 9)," 3GPP TR 36.912, v9.3.0, June 2010, <http://www.3gpp.org/ftp/Specs/html-info/36912.htm>.

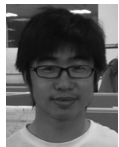[2] 3rd Generation Partnership Project, "Evolved Universal Terrestrial Radio Access (E-UTRA),

User Equipment (UE) Radio Transmission and Reception (Release 10)," 3GPP TS 36.101, v10.4.0, Sept. 2011, <http://www.3gpp.org/ftp/Specs/html-info/36101.htm>.

[3] 3rd Generation Partnership Project, "Technical Specification Group Radio Access Network, Evolved Universal Terrestrial Radio Access (E-UTRA), Physical Channels and Modulation (Release 10)," 3GPP TS 36.211, v10.4.0, Dec. 2011, <http://www.3gpp.org/ftp/Specs/html-info/36211.htm>.

[4] China Mobile Research Institute, "C-RAN: The Road Towards Green RAN," White Paper, Version 1.0.0, Apr. 2010.

[5] Common Public Radio Interface, "CPRI Specification V4.2," Nov. 2010, <http://www.cpri.info/spec.html>.

[6] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-Advanced: Next-Generation Wireless Broadband Technology," IEEE Wireless Commun., 17:3 (2010), 10–22.

[7] B. Guo, X. Fan, W. Cao, Y. Li, and Z. Jiang, "8 Antennas Eigen-Based Single Stream Beamforming Algorithm for TD-LTE System," Telecommun. Eng., 2010:8 (2010), 41–45.

[8] D. Samardzija and T. Sizer, "lightRadio Portfolio: White Paper 6—Compressed I/Q Transport," Alcatel-Lucent Technology White Paper, 2011.

[9] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tanno, "Coordinated Multipoint Transmission/Reception Techniques for LTE-Advanced," IEEE Wireless Commun., 17:3 (2010), 26–34.

[10] C.-H. Tang and C.-E. Wu, "Evaluation of Energy Efficiency for C-RAN Architecture with Centralized BBUs," Proc. 7th Internat. Conf. on Commun. and Networking in China (ChinaCom '12) (Kunming, Chn., 2012), C-RAN Workshop.

(Manuscript approved April 2013)

BIN GUO is a senior systems engineer in the Wireless Research and Development (R&D) Department at Alcatel-Lucent Shanghai Bell. He received his Ph.D. in telecommunication and information systems from Northeastern University, Shenyang, China. Dr. Guo is based in Shanghai, China. His primary research interests include algorithm design in TD-LTE MIMO wireless communication systems and smart antenna theory and techniques.

WEI CAO is a senior systems engineer in the Wireless Research and Development (R&D) Department at Alcatel-Lucent Shanghai Bell. She received her Ph.D. in wireless communications from the National University of Singapore. Dr. Cao is based in Shanghai, China. Her primary research interests include baseband-receiving algorithms in TD-LTE systems and MIMO OFDM technologies.

AN TAO is a systems research and development (R&D) manager in the Wireless business group at Alcatel-Lucent Shanghai Bell. Based in Shanghai, China, his current work focus includes projects related to both Long Term Evolution (LTE) and M&L. His primary research interests include GSM signal processing, code design in WCDMA systems, and LTE system cell deployment.

DRAGAN SAMARDZIJA is a member of technical staff at Alcatel-Lucent Bell Labs in Holmdel, New Jersey. He received his M.S and Ph.D. degrees in electrical engineering from the Wireless Information Network Laboratory (WINLAB), Rutgers University, New Brunswick, New Jersey. Since joining Bell Labs more than a dozen years ago, his primary focus has been on next-generation wireless systems research. Dr. Samardzija also teaches classes at the University of Novi Sad. His research interests include analysis, design, and experimental evaluation of wireless systems. ◆