Interference Mitigation for WCDMA Using QR Decomposition and a CORDIC-based Reconfigurable Systolic Array

Robin SCHEIBLER†, James OKELLO††, Katsutoshi SEKI††, Tomoyoshi KOBORI††, and Masao IKEKAWA††
† Swiss Federal Institute of Technology, Lausanne, Switzerland
†† System IP Core Laboratory, NEC Corporation

Abstract This paper presents implementation and performance of QR Decomposition based Recursive Least-Squares (QRD-RLS) for interference mitigation in Wideband CDMA (WCDMA). The implementation is carried on CORSAEngine which is a new Software-Defined Radio (SDR) processor developed by NEC Corporation and highly optimized for MIMO-OFDM systems. It is shown how QRD-RLS can be mapped on its rectangular CORDIC-based reconfigurable systolic array, hence demonstrating its capability to process WCDMA. In addition, the performance of CORSAEngine is compared to that of other architectures and it is found to achieve at least 91% of the performance of dedicated hardware in terms of computational density.

Key words Software-Defined Radio, QR Decomposition, Wideband CDMA, Interference Mitigation

1. Introduction

In 1991, Mitola [15] introduced the concept of Software-Defined Radio (SDR) that allows operations of different modes of communications systems on a single hardware, dramatically decreasing equipment costs and development time of new technologies. While programmability is attractive to mobile communication equipments manufacturers and operators, it also brings one of the biggest challenges of SDR. The need to maintain high performance while retaining enough flexibility to process as many different standards as possible. This constraint becomes even more difficult to fulfill as modern communication standards require more complex signal processing technology.

In the field of cellular communications, such modern standards are usually referred to as Beyond 3G (B3G) technologies. It has been recognized that B3G systems, already exemplified by WiMAX and 3GPP LTE among others, will heavily rely on Orthogonal Frequency Division Multiplexing (OFDM) and Multiple-Input Multiple-Output (MIMO) technologies [18]. But at the same time, it is important for an SDR to support non-OFDM-based standards like IS-95, CDMA2000 and WCDMA. Firstly, those systems enjoy a very deep market penetration and are likely to remain used for many years. Secondly, in the case of WCDMA, it has the potential to be used in conjunction with an OFDM scheme such as in Multi-Carrier Code Division Multiple Access (MC-CDMA).

Recently, NEC Corporation developed CORSAEngine, a new SDR processor highly optimized for MIMO-OFDM systems [17]. Its rectangular COordinate Rotation DIgital Computer (CORDIC) based reconfigurable systolic array makes it highly suitable to process the computationally intensive baseband algorithms required by those systems, among others QR Decomposition (QRD), Singular Value Decomposition (SVD), least-squares fit or fast Fourier transform. However, performance of efficient interference mitigation algorithms for WCDMA had not been investigated on this processor.

This paper presents the implementation and the performance of QR Decomposition based Recursive Least-Squares (QRD-RLS) for interference mitigation of WCDMA on CORSAEngine. QRD-RLS has been shown to effectively mitigate both Intersymbol Interference (ISI) and Multiple Access Interference (MAI), outperforming the conventional Rake while maintaining reasonable complexity when implemented as a systolic array [14]. In this paper, it is shown how this arbitrarily large systolic array can be split into parts that fit on the reduced size array of CORSAEngine. Through reconfigurability, it is furthermore possible to run successively those different parts on the same hardware structure.

The remainder of this paper is organized as follows. Section 2 gives a brief revision of conventional and QRD-RLS based interference for WCDMA along with simulation results and computational load comparison of those two methods. In Section 3, the architecture of CORSAEngine is described. The mapping of QRD-RLS onto the systolic array
is described in Section 4. Finally in Section 5, the performance of the implementation of WCDMA on CORSAEngine is assessed and a benchmark against other devices is done. Section 6 concludes this paper.

2. Interference Mitigation in WCDMA

2.1 Conventional Interference Mitigation

The conventional interference mitigation for WCDMA is characterized by the RAKE receiver shown in Fig. 1. It uses short-time averaging (typically two slots) of the received pilot symbols to estimate the channel characteristics. Then, long-time averaging (about one frame) is used to get a good power delay profile of the channel. The Path Search uses a threshold-based algorithm to select the paths with a sufficiently large Signal-to-Noise Ratio (SNR). Those paths are despread using a bank of correlators and combined according to the Maximum Ratio Combining (MRC) principle with respect to the channel coefficients. For different algorithm for channel estimation and path search, refer to [8], [9]. For more details about the principles of the RAKE receiver, refer to [16].

2.2 QRD-RLS Interference Mitigation

This section describes QRD-RLS based interference mitigation applied to WCDMA. A block diagram of the receiver considered is shown in Fig. 2. First the Common Pilot Channel (CPICH) and the signal of the user of interest are despread using Matched Filters (MF) corresponding to their respective spreading codes. The despread pilot signal $r_p(n)$ is then sent to the QRD-RLS weight calculation unit which produces the optimal weight vector $\mathbf{w}$. It is then sent to the combiner and used to combine the despread user signal $\tilde{r}_u(n)$.

In the WCDMA system, the despread signal can be written as in [6]:

$$\tilde{r}(n) = \sigma_d(i) + I(n) + \xi(n),$$

where the time index $n = IF + l$ with $i \in \mathbb{N}$ and $l \in \{0, \ldots, F-1\}$, $F$ is the spreading factor, $\sigma_d$ is a multiplicative coefficient introduced by the channel impulse response and the spreading code autocorrelation function and $d(i)$ is the $i^{th}$ symbol sent. $I(n)$ is an interference term created by the ISI and the MAI. $\xi(n)$ is the filtered noise. Let’s define $\mathbf{u}(i) = [\tilde{r}(iF), \ldots, \tilde{r}(iF + M - 1)]^T$, a vector containing the $M$ first chips of the despread signal corresponding to the $i^{th}$ symbol sent. The goal is then to find the optimal weight vector $\mathbf{w}(m) = [w_0(m), \ldots, w_{M-1}(m)]^T$ to combine the elements of $\mathbf{u}(i) = [\tilde{r}_u(iF_0), \ldots, \tilde{r}_u(iF_u + M - 1)]^T$ in order to enhance the symbol $d_u(i)$ and reduce the interference signal $I(n)$, $m$ being the number of symbols received so far. Subscript $p$ and $u$ are used to distinguish between pilot and user signals.

QRD-RLS is a technique borrowed from the adaptive filtering theory [11]. To adaptively calculate $\mathbf{w}(m)$, it attempts to minimize the following error function:

$$E(m) = ||\Lambda(m)(\mathbf{A}(m)\mathbf{w}(m) - d(m))||$$

where $\mathbf{A}(m) = [\mathbf{u}_p(0), \ldots, \mathbf{u}_p(M-1)]^H$ contains the received pilot signal, $\Lambda(m) = \text{diag}(\lambda_0^m, \ldots, \lambda_1^m)$ the exponentially decreasing forgetting factor and $d(m) = [d_p(0), \ldots, d_p(M-1)]^H$ the original pilot symbols.

Minimizing Eq. (2) can be done by multiplying $\Lambda(m)[\mathbf{A}(m)d(m)]$ by a unitary matrix $Q(m)$:

$$Q(m)\Lambda(m)[\mathbf{A}(m)d(m)] = \begin{bmatrix} \mathbf{R}(m) & \mathbf{p}(m) \\ \mathbf{0} & \mathbf{v}(m) \end{bmatrix},$$

where $\mathbf{R}(m)$ is an $M \times M$ upper triangular matrix, $\mathbf{p}(m)$ is a vector of length $M$, $\mathbf{0}$ is an $(m-M) \times M$ null matrix and $\mathbf{v}(m)$ is a vector of length $m - M$. The least-squares estimation of $\mathbf{w}(m)$ is then given by:

$$\hat{\mathbf{w}}(m) = \mathbf{R}^{-1}(m)\mathbf{p}(m).$$

Once $\mathbf{w}(m)$ has been calculated, it is used to combine the signal of the user:

$$\hat{\tilde{r}}_u(i) = \mathbf{w}^H(m)\mathbf{u}_u(i).$$

Using the Extended QRD-RLS algorithm described in [14], the recursion can be done by applying QRD to the following extended $(M+2) \times (2M+2)$ matrix:

$$\begin{bmatrix} \tilde{\mathbf{R}}(m+1) & \tilde{\mathbf{R}}^{-H}(m+1) \\ \mathbf{0} & \mathbf{v}'(m+1) \end{bmatrix} = Q'(m+1) \begin{bmatrix} \Lambda \tilde{\mathbf{R}}(m) & (\Lambda')^{-1}\tilde{\mathbf{R}}^{-H}(m) \\ \hat{\Lambda}^H(m+1) & \mathbf{0} \end{bmatrix},$$

where $\Lambda = \text{diag}(\lambda_0^m, \ldots, \lambda_1^m)$. The QRD-RLS block diagram is shown in Fig. 2.
where $\Lambda' = \text{diag}(\lambda, \ldots, \lambda, 1)$, $\breve{u}_H^H(m+1) = [u_H^H(m+1) d_p^e(m+1)]$, $v'(m+1)$ is an auxiliary vector and
\[
\breve{R}(m) = \begin{bmatrix} R(m) & p(m) \\ 0 & \alpha(m) \end{bmatrix},
\]
(7)
where $\alpha(m + 1)$ is a scalar. After the matrix $Q'$ has zeroed $\breve{u}_H^H(m+1)$, a scaled version of the new weight vector appears in $\breve{R}^{-H}(m+1)$:
\[
\breve{R}^{-H}(m+1) = \begin{bmatrix} R^{-H}(m+1) & 0 \\ -\frac{w_{H(m+1)}}{\alpha(m+1)} & \frac{1}{\alpha(m+1)} \end{bmatrix}.
\]
(8)
This method has the advantage of avoiding back-substitution which can be very time-consuming if it has to be performed frequently.

3. CORSAEngine Architecture

CORSAEngine’s architecture is composed of a 2-dimensional array of processing nodes (PN), a control unit, a memory bank and an address generator which controls the algorithms running on the array. The work presented here was realized on a scaled-down architecture represented in Fig. 3.

This scaled-down version of CORSAEngine has a 2-by-5 array of PNs. Each PN is composed of two CORDIC Processing Elements (CPE) and two Delay Processing Elements (DPE). The CPEs implement the unfolded CORDIC algorithm which allows pipelining. The pipeline is used to implement interleaved threads. Different data sets or even completely unrelated algorithm can be executed in the different threads. The data types supported by the processor are real and complex numbers and rotation angles, which are a subset of real numbers. A complex number is the concatenation of two real numbers. A 20-bit floating point format, consisting of a 16-bit mantissa and a 4-bit exponent is used for real numbers.

The control of the operations on the array is done by a context pointer which is attached to the data by the memory interface when it is sent from the memory to the array. Then, every CPE and DPE possesses an instruction table linking a context pointer to the operation to be done with the incoming data and the destination of the result. The result can be sent to any neighboring PN. PNs have horizontal and diagonal connections. A horizontal connection can hold one complex or two real numbers while a diagonal connection is limited to one real number. As a result, complex data flows can be created in the array, giving an efficient and flexible way to easily implement systolic algorithm.

4. Implementation of QRD-RLS

In this section, the implementation of the Extended QRD-RLS algorithm on the array of CORSAEngine is described. An example of the Extended QRD-RLS systolic array for a 3-tap weight vector is given in Fig. 4. Each non-zero complex-valued coefficient of the matrix $\hat{R}_{ext} = [\hat{R}(m) \hat{R}^{-H}(m)]$ is represented by one cell. This cell holds the coefficient value in its register. Note that the coefficient in the right-bottom corner of the matrix is not needed and hence doesn’t require a cell.

4.1 Cell operations

Two main types of cell can be seen. Border cells are placed on the left diagonal and produce the required Givens rotation to nullify the input. Inner cells apply this rotation to their own input and register value. One more distinction can be made between cells holding the coefficients of $R$, $p$ or $R^{-H}$ and the last row containing the scaling factor $\alpha$ and the scaled weights $-w/\alpha$. The former must multiply the coefficient they hold with the forgetting factor between every two input, while the latter don’t.

Fig. 5 describes how the operations of the cells composing the array can be implemented using CORDIC units in vectoring (VEC), rotation (ROT) and multiplication mode. The two stages of the complex givens rotation are referred to as $\theta$-VEC/ROT and $\phi$-VEC/ROT. In the cells of normal rows, the forgetting factor $\lambda$ must be applied to the register value after every input. However, as the input of the $\phi$-VEC/ROT depends on the output of the CORDIC unit applying the forgetting factor, those two operations cannot be pipelined. As a result, the $\phi$-VEC/ROT can only operate every two cycles. If the same CORDIC is used for both the $\phi$-VEC/ROT and the multiplication by $\lambda$, it is fully utilized. But on the other hand the CORDIC used for the $\theta$-VEC/ROT will only be used every two cycles thus wasting half of this resource. As a solution, the same CORDIC can be time-shared by two adjacent cells for their $\theta$-VEC/ROT as illustrated in Fig. 6. The left cell first receives its input and the angle $\theta$, apply the latter to the former and send the result down to its $\phi$-VEC/ROT unit. However, the angle $\theta$ is not sent further but stored in a register of the CORDIC unit. In the next
cycle, the same CORDIC unit receives only the input of the right cell. It will then reuse the angle stored to rotate the input before sending the result down to the right cell. This time, $\theta$ is not stored but sent to the next cell on the right.

As the cells from the last row don’t apply the forgetting factor, it allows the two CORDIC operations to be fully pipelined. Therefore, successive cells can be connected to each other in a straightforward manner and no time-sharing of CORDIC units is required. And, as the registers of the cells contain a scaled version of the desired weights and the scaling factor $\alpha$, it is possible, by adding one multiplication to each cell, to scale the weights before they are output. The structure of those cells is also illustrated in Fig. 5.

4.2 Partitioning

Now that the cell operations have been mapped to CORDIC units, it is possible to use them to construct a full size array for the production of an $M$-tap weight vector. Such an array has $M^2 + 3M + 1$ cells, each using from 3 to 5 CORDIC units depending on its type. Consequently it has to be divided into smaller partitions that will be successively run on the PN array of CORSAAEngine. Because of the strong vertical dependency in the Extended QRD-RLS array, it is first divided into rows, each row having $M + 2$ cells except the last one with $M + 1$ cells. To make it fit on the PN array, these rows still have to be subdivided into segments of a few cells as shown in Fig. 7. Each of these segments contains 7 cells for a normal row and 3 cells for the last row. Considering a single row there are two types of segments: one with a border cell at the beginning and one containing only inner cells that will be respectively referred to as border and inner segments. In conclusion we have 4 partition types, $T$, $X$, $Y$ and $Z$, with respectively $T$ and $X$ referring to border and inner segments of a normal row and $Y$ and $Z$ to border and inner segments of the last row. Each partition type is implemented on the array as a specific context pointer.

To run the complete algorithm, it is first assumed that the matrix $\hat{R}_{ext}$, as well as the $N$ new pilots received along with their local copies in the form of the matrix:

$$U = \begin{bmatrix} \hat{u}^H(m + 1) \\ \vdots \\ \hat{u}^H(m + N) \end{bmatrix}.$$  

are stored in the memory bank. A flowchart of the algorithm is represented in Fig. 8. The 7 first coefficients of the first row of $\hat{R}_{ext}$ are loaded into the registers of the appropriate CORDIC units. Then the 7 first columns of $U$ are processed through the array configured as partition $T$. The processed columns, the modified coefficients of $\hat{R}_{ext}$ and the angles produced are stored back into memory. The next 7 coefficients of $\hat{R}_{ext}$ are now loaded into the appropriate CORDIC units registers and the next 7 columns of $U$ are processed, this time using a partition type $X$ configuration and the angles produced by the partition $T$. Processed columns and register values are sent back to memory at the end of the execution. This step is repeated until all columns of $U$ have been pro-
matrix where $m > n$ is given by $3n^2(m - n/3)$ [10]. As shown in Table 1, the complexity of the QRD-RLS based method is more than 2.5 times the one of the Rake. The main difference comes from the QRD-RLS algorithm which is computationally intensive compared to the insignificant amount of computation required by the path search in the Rake. However, it is shown in the following sections that using the implementation introduced in Section 4, this complexity can be easily handled by CORSAEngine.

5.3 Resource Usage

In this section, the resource usage of QRD-RLS will be calculated. As shown in Section 5.1, a 16-tap weight vector is sufficient to efficiently mitigate interference. Using the implementation as described in Section 4.2, it is possible to construct an array for the calculation of a 19-tap weight vector which is therefore sufficient to efficiently mitigate the interference. Taking into account the matched filtering of pilot and data channel as well as the combining, QRD-RLS interference mitigation for WCDMA consumes 11.75% of the resources of the scaled-down version of the CORSAEngine. The resource consumptions of the different blocks of the interference mitigation are detailed in Table 2.

5.4 Benchmark

The performance of the implementation of QRD-RLS on CORSAEngine will now be compared to other implementations on different architectures. The architectures considered for comparison are: two dedicated hardwares for QRD-RLS, based on designs conducted on respectively Altera Stratix [5] and Xilinx Virtex-4 [7] FPGAs, and an Application Specific Instruction set Processor (ASIP) for matrix computations (QRD, SVD) using an array of modified CORDIC units [13].

The performance metric used to compare those architectures is the computational density defined as:

$$\rho_{m \times n} = \frac{1}{t_{m \times n}} \sum_i A_i \times u_i,$$

(10)

where $t_{m \times n}$ is the processing time for a complex matrix of size $m \times n$ in seconds [s], $A_i$ and $u_i$ respectively the chip area in [Kgates] and a resources utilization factor. The index $i$ accounts for architectures with totally independent parts. To make a fair comparison, the CORSAEngine implementation is adapted to the matrix sizes that were used for evaluating performance of the referred architectures [5], [7], [13]. The performance of the CORSAEngine is furthermore used to normalize the results.
Fig. 6 An example of a border (vectoring) and an inner (rotation) cell on a normal row sharing a CORDIC respectively for the vectoring and rotation of their inputs. The CORDIC is used in vectoring mode during the odd cycles and in rotation mode during the even cycles. Dashed lines represent values that are kept in a register.

Fig. 8 A run of the algorithm for a 19-tap weight vector. The gray rectangle represents the array of CORSAEngine, the dashed line is for angles and scaling factor that return to memory. $U_{i,j}$ is the matrix composed of the $i$th to the $j$th columns of $U$. The matrix $W$ output by the last row contains all the weight vectors produced by the processing of the matrix $U$ through the Extended QRD-RLS systolic array.

Fig. 9 Performance of QRD-RLS in quarter system load (4 users) with a spreading factor of 16.

Table 3 shows the results of the benchmark. The area estimation of the dedicated hardware was based on the number of lookup tables used in the FPGA design. The corresponding number of gates was estimated according to the available literature [12], [19]. The Altera Stratix design uses two CORDIC blocks for the QRD and the Embedded Nios Soft processor for the back-substitution. The performance of the latest version of the Nios (II) were used [4]. In the case of the ASIP, as it only handles real-valued QRD-RLS, the fact that a $128 \times 20$ real-valued matrix can be used to represent a $64 \times 10$ complex-valued matrix is used. For CORSAEngine, a utilization factor is introduced as an input matrix with 10 columns such as the ones used in the benchmark only use 80.2% of the resource available.

The result of the benchmark shows that CORSAEngine achieves respectively 50% and 80% more computational density than the dedicated hardware II (based on Xilinx design) and the ASIP processor. The dedicated hardware I (based on
Table 3 Performance of the different architectures in terms of the computational density $\rho$. The final result is normalized in terms of the performance of CORSAEngine to give a fair comparison when the matrix sizes used are different.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Clock frequency [MHz]</td>
<td>170</td>
<td>250</td>
<td>300</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>Matrix size</td>
<td>$64 \times 10$</td>
<td>$10 \times 10$</td>
<td>$64 \times 10$</td>
<td>$64 \times 10$</td>
<td>$10 \times 10$</td>
</tr>
<tr>
<td>$l_m \times n$ [µs]</td>
<td>268.67</td>
<td>56.76</td>
<td>7.04</td>
<td>10.63</td>
<td>2.89</td>
</tr>
<tr>
<td>$A$ [gates]</td>
<td>33480</td>
<td>95310</td>
<td>7.0M</td>
<td>1150K</td>
<td>1150K</td>
</tr>
<tr>
<td>Utilization factor</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>80.2%</td>
<td>80.2%</td>
</tr>
<tr>
<td>$\rho$ [update/s/Kgates]</td>
<td>111.22</td>
<td>184.85</td>
<td>20.29</td>
<td>102</td>
<td>375.17</td>
</tr>
<tr>
<td>Normalized to CORSA</td>
<td><strong>109.04%</strong></td>
<td><strong>49.27%</strong></td>
<td><strong>19.8%</strong></td>
<td><strong>100%</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

6. Conclusion

In this paper a new implementation of QRD-RLS interference mitigation for WCDMA on CORSAEngine has been presented. First the necessary complex Givens operations were mapped to the available CORDIC units in a way that maximize the utilization of resources. Then the Extended QRD-RLS systolic array was split into manageable sizes that fit on the PN array of CORSAEngine. Simulations were furthermore used to determine the necessary size of the weight vector to be about 19 taps. Finally, the performance of this implementation was compared to other available architectures for QRD-RLS and it was shown to achieve at least 91% of the dedicated hardware performance in terms of computational density. In conclusion, CORSAEngine was shown to be able to handle computationally intensive but efficient interference mitigation algorithm for WCDMA using only 11.75% of its resources.

Acknowledgments This work was realized between March 2002 and January 2008 while the first author was an internship student at the System IP Core Laboratory, NEC Corporation.

References


— 7 —