

# FPGA implementacija inverzije kompleksnih matrica primenom QR dekompozicije

Bojan Marić, Lazar Saranovac i Miloš S. Trajković

**Sadržaj** — Trendovi u poslednjim istraživanjima odnose se na uvođenje 4G tehnologije i ostvarivanje širokopojsnog pristupa internetu. Protokoli koji imaju tendenciju da postojeću DSL i kablovsku internet strukturu zamene vazduhom jesu WiMAX (*Wireless Interoperability for Microwave Access*) i LTE (*Long Term Evolution*). Primenom MIMO (*Multiple Input Multiple Output*) sistema u okviru ovih standarda moguće je postići veoma visoke protoke na uštrb povećane kompleksnosti. Integracija više antena u sistem na prijemnoj strani stvara matricu estimacije kanala što dodatno uvećava arhitekturu i zahteva primenu matematičkih operacija u cilju uspešne detekcije. Jedna od njih je svakako matrična inverzija čija je hardverska realizacija na FPGA (*Field Programmable Gate Array*) platformi prikazana u ovom radu na primeru matrice dimenzije 4x4. Efikasan dizajn i nov pristup u rešavanju konvencionalnih matematičkih operacija omogućavaju frekvenciju takta od 225MHz na Xilinx Virtex®-5 čipu što donosi vreme računanja jedne inverzne matrice od 285ns.

**Ključne reči** — Bežična komunikacija, FPGA, MIMO sistem, pipeline (protočna obrada), QR dekompozicija, WiMAX.

## I. UVOD

ULTIMATIVNI cilj u savremenim bežičnim komunikacijama oduvek je bio postizanje što većeg protoka podataka uz minimalne gubitke. Jedan od načina za povećanje kapaciteta i performansi prenosa jeste upotreba više antena na prijemu i predaji (MIMO) [1]. Primena MIMO sistema je jedna od glavnih prednosti sistema četvrte generacije i njihova implementacija je detaljno opisana u standardima.

Matrični invertor opisan u ovom radu je dizajniran cilju računanja kompleksnih operacija preuređivanja estimirane matrice kanala po algoritmu koji svojom strukturom nalikuje na V-BLAST algoritam [2], [3]. Sastoji se od više iteracija gde je u svakoj potrebno računati pseudo-inverznu matricu i tražiti normu svake kolone [4]. Na taj način određivan je uticaj (snaga) svake antene ponaosob te se na osnovu toga pristupalo preuređivanju kolona matrice u cilju dobijanja sto bolje detekcije u BER (*Bit Error Ratio*) smislu. Ceo sistem obrađuje signal modulisan po standardu

Bojan Marić, Elektrotehnički fakultet u Beogradu, Srbija (telefon: +381-65-6223059; e-mail: [bozzan@gmail.com](mailto:bozzan@gmail.com)).

Doc. Dr Lazar Saranovac, Elektrotehnički fakultet u Beogradu, Bulevar kralja Aleksandra 73, 11120 Beograd, Srbija; (e-mail: [laza@el.etf.bg.ac.yu](mailto:laza@el.etf.bg.ac.yu)).

Miloš Trajković, Signum Concepts Inc., Kralja Milana 4a, Beograd, Srbija; (telefon: +381-11-2686649; e-mail: [milos.trajkovic@signumconcepts.com](mailto:milos.trajkovic@signumconcepts.com)).

IEEE 802.16e sa specifikacijama prikazanim Tabeli 1.

TABELA 1: WiMAX SPECIFIKACIJE.

<i>FFT size</i>	512
<i>Data Sub-Carriers</i>	360
<i>System Period</i>	102.9 $\mu$ s
<i>Frame Duration</i>	5 ms

U daljim odeljcima će biti detaljno opisan sistem za matričnu inverziju uz pomoć algoritma QR dekompozicije [5] i rešavanjem dobijenog sistema metodom Gauss-ove eliminacije. Nakon kraćeg teorijskog uvoda u problem inverzije i dekompozicije, biće prikazani svi aspekti hardverske realizacije kao i rezultati implementacije.

## II. TEORIJSKO RAZMATRANJE INVERZIJE MATRICE

Neka je  $A$  kompleksna kvadratna matrica dimenzija  $n \times n$ . Problem računanja inverzne matrice se može formulisati na način prikazan u (1)

$$\begin{aligned} A \cdot X &= I, \\ X &= A^{-1}. \end{aligned} \quad (1)$$

Primenom QR dekompozicije na matricu  $A$  dobija se (2), gde je  $R$  gornje-trougona matrica a  $Q$  ortonormalna matrica za koju važi jednačina (3).

$$A = Q \cdot R \quad (2)$$

$$Q \cdot Q^H = I \quad (3)$$

Razlaganjem izraza za računanje inverzije dobija se:

$$A^{-1} = (Q \cdot R)^{-1} = R^{-1} \cdot Q^{-1} = R^{-1} \cdot Q^H, \quad (4)$$

gde je  $X^H$  hermitska matrica matrice  $X$ .

Daljom primenom (4) u (1) dobijamo

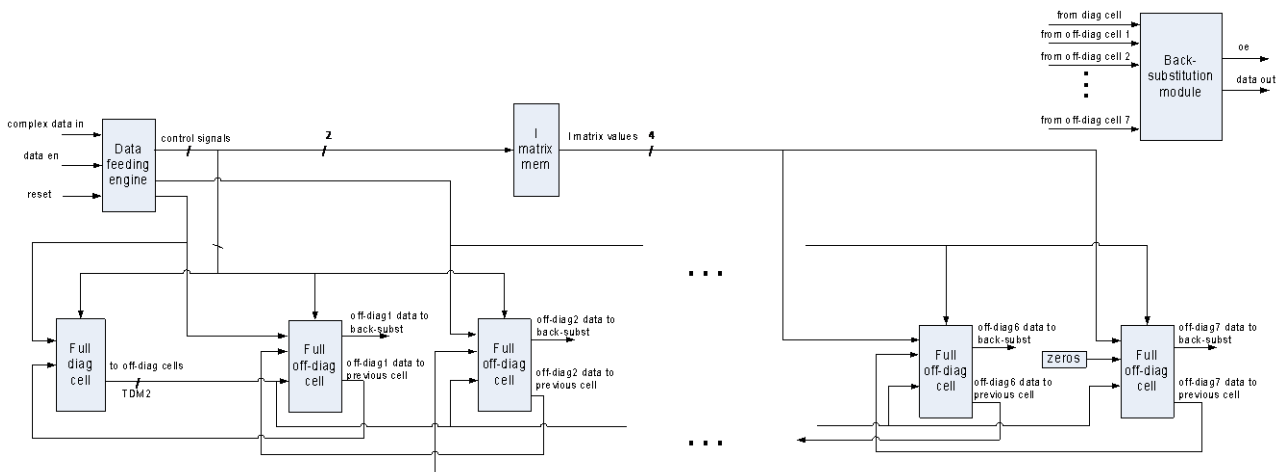
$$R \cdot X = Q^H \cdot I. \quad (5)$$

Na ovaj način se dobija trougaoni Gauss-ov sistem jednačina koji se metodom zamene (*back-substitution*) lako rešava i dobija nepoznata matrica  $X$ .

## III. ARHITEKTURA HARDVERA

### A. QR dekompozicija

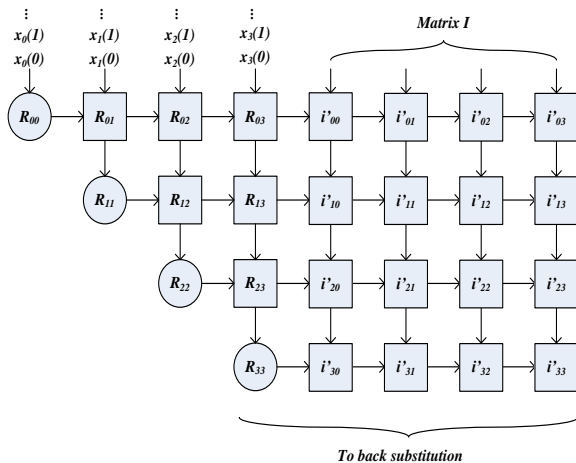
Dekompozicija matrice je veoma čest metod pri rešavanju matričnih problema. U ovom sistemu je izabran metod QR dekompozicije pomoću Givensovih rotacija [6], [7], jer je to standardan metod koji se koristi za QR dekompoziciju. Metoda Givensovih rotacija se zasniva na matematičkim operacijama deljenja i korenovanja, pa je zbog toga jedan od problema bio aproksimacija tih matematičkih operacija. Aproksimacija je neophodna zbog uštede hardverskih resursa FPGA čipa i o tome će biti reči



Sl. 1. Blok dijagram hardvera za matricnu inverziju sa posebnim osvrtom na kompleksnu QR dekompoziciju.

u narednom odeljku.

Na Sl.1 je prikazan blok dijagram sistema za QR dekompoziciju matrice  $4 \times 4$ . Broj dijagonalnih i van-dijagonalnih ćelija implementiranih za matricu  $4 \times 4$  je 1 i 7, respektivno. Na Sl. 2. je prikazana sistolička struktura niza za realizaciju QR dekompozicije [5]. Svaka ćelija se višestruko koristi (*reuse*) za ponovna računanja u zavisnosti koji red strukture, prikazne na Sl. 2, je aktuelan u izračunavanju. U dijagonalnoj ćeliji se vrši estimacija ugla kompleksnog ulaznog vektora i to u dva nivoa koji su potrebni za rotacije oko realne i imaginarne ose u van-dijagonalnim ćelijama. Vektore koji se dovode na ulaz van-dijagonalne ćelije je potrebno rotirati za ugao koji se estimira iz vrednosti dovedene na ulaz u dijagonalnu ćeliju.



Sl. 2. Sistolička struktura niza za QR dekompoziciju.

Estimacija ugla i rotacija su se u dosadašnjim FPGA implementacijama bazirale na upotrebi CORDIC (*COordinate Rotation DIGital Computer*) bloka [8]. Takva realizacija je trošila mnogo hardverskih i logičkih elemenata. Ovde se pribeglo novom metodu realizacije matematičkih operacija upotrebom elemenata za množenje i akumulaciju (DSP48 Slice, videti [9]). Pribegava se što većem korišćenju tih elemenata, jer se na veoma sofisticiran način mogu unaprediti brzina rada sistema, kao i paralelni rad blokova sistema (*pipeline*). Shodno tome, rotacija za odgovarajući ugao je realizovana množenjem

konjugovano-kompleksnim brojem i odgovarajućim post skaliranjem za vrednost modula kompleksnog broja.

### B. Aproksimacija funkcije $1/\sqrt{x}$

Neka je  $z$  obeležena vrednost u dijagonalnoj ćeliji. Tada se množenjem sa konjugovano-kompleksnim brojem  $Z^*$  i skaliranjem sa  $1/\sqrt{|Z|^2}$  realizuje rotacija. Deljenje (u ovom slučaju i korenovanje) je izvršeno kao množenje sa recipročnom vrednošću pomoću polinomske aproksimacije na definisanom intervalu. Analizom funkcije definisane u (6) je određen opseg gde je funkcija najviše linearna. Najmanja greška aproksimacije funkcije polinomom prvog reda može biti ostvarena u opsegu  $[c, +\infty)$ , gde je konstanta  $c$  izabrana sa najvećom mogućom vrednošću. Opseg ulaznih podataka je određen sa  $0 < |Z|^2 < 2$ ;  $\Re(Z), \Im(Z) \in [-1, 1]$ ; odakle sledi da vrednosti koje su manje od  $c$  moraju biti skalirane da bi bile unutar opsega  $[c, 2]$ . Skaliranje ulaznih podataka je realizovano pomeranjem podataka za broj bitova znaka broja, a faktor pomeranja je određen metodom binarnog pretraživanja vodećih bita. Faktor pomeranja je izabran kao paran broj, pa se tako izlazno skaliranje (po završenoj *sqr*t aproksimaciji), radi vraćanja prave vrednosti, izvodi kao obično pomeranje. Uzimajući u obzir sve ove zahteve, proračunom (upotrebom MATLAB-a [10]) je dobijena vrednost konstante  $c = 0.25$  [3]. Aproksimacija prvog reda je:

$$f(x + \Delta x) = f(x) + \Delta x \cdot f'(x); \quad f(x) = \frac{1}{\sqrt{x}} \quad (6)$$

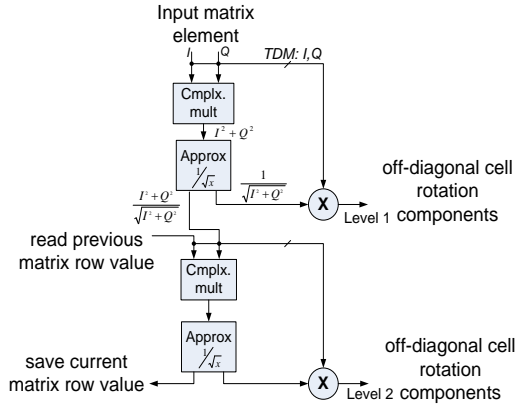
Vrednosti  $f(x)$  and  $f'(x)$  su smeštene u memorije (LUTs), dok je množenje izvršeno sa DSP48 Slice elementima.

### C. Arhitektura dijagonalne i van-dijagonalne ćelije

Arhitektura dijagonalne ćelije je prikazana na Sl.3. Podaci u dijagonalnoj ćelini su već pripremljeni za obradu u van-dijagonalnim ćelijama i u njima treba izvršiti samo kompleksno množenje, tj. rotacije po obema osama. Formula (7) prikazuje matematičku operaciju koja se

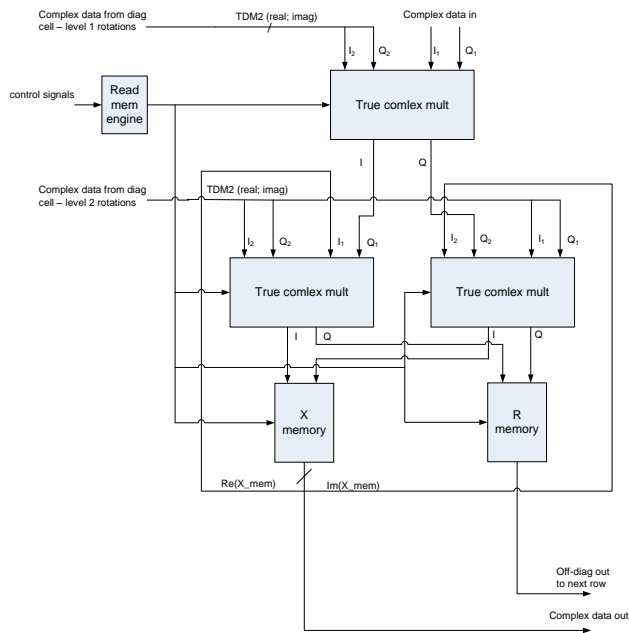
izvodi u van-dijagonalnoj ćeliji.

$$\begin{aligned} & \left[ (I + jQ) * (I_{rot} - jQ_{rot}) \right] \frac{1}{\sqrt{I_{rot}^2 + Q_{rot}^2}} = \\ & = (I + jQ) \left( \frac{I_{rot}}{\sqrt{I_{rot}^2 + Q_{rot}^2}} - j \frac{Q_{rot}}{\sqrt{I_{rot}^2 + Q_{rot}^2}} \right) \end{aligned} \quad (7)$$



Sl. 3. Blok dijagram dijagonalne ćelije.

Blok dijagram van-dijagonalne ćelije je prikazan na Sl. 4.



Sl. 4. Blok dijagram van-dijagonalne ćelije.

Shodno specifikacijama datim u Tabeli 1, kao i radnoj učestanosti, potreban i dovoljan broj taktnih intervala za obradu jedne matrice je:

$$num\_cycles = \frac{(102.9us / 360)}{(1 / 225MHz)} \cong 64. \quad (8)$$

Kašnjenje, koje je nastalo zbog uvođenja modula za aproksimaciju funkcije i kompleksnog množenja, prevaziđeno je uvođenjem TDM (*Time Division Multiplexing*) signala u kome se protočno obrađuju 5 uzastopnih podataka. Da bi se postigli vremenski zahtevi i uspešno implementirao TDM signal, ulazni podaci su stabilni 3 taktna intervala. Podaci se prosleđuju dijagonalnoj i van-dijagonalnim ćelijama na svaka 3 taktna intervala iz ulazne bafer memorije, koja je realizovana kao ping-pong struktura. Vreme dekompozicije jedne matrice

4x4 je  $3 \times 4 \times 4 = 48$  taktnih intervala (od mogućih 64 prema formuli (8)). Preostalo vreme ( $64 - 48 = 12$  taktnih intervala) se koristi za IO i inicijalizaciju memorija za sledeći TDM blok podataka. Svaka ćelija sadrži internu distribuiranu memoriju za smeštanje rezultata obrade i to po redosledu kako odgovarajući podaci matrice 4x4 dolaze u TDM formatu. Van-dijagonalna ćelija sadrži i još jednu memoriju u kojoj se smeštaju obrađeni podaci potrebni za rad sledećeg nivoa sistoličke strukture prikazane na Sl.2.

#### D. Blok za rešavanje sistema jednačina

Blok za rešavanje sistema jednačina je prikazan na Sl. 5. Po završenoj QR dekompoziciji, podaci raspoloživi u internim memorijama ćelija su stabilni još 12 taktnih intervala. Za to vreme ih je potrebno proslediti bloku za rešavanje sistema jednačina. Podaci se i dalje obrađuju u TDM formatu dužine 5.

Formula (9) predstavlja matrični oblik formule (5) izvedene u drugom poglavlju. Trougaoni Gauss-ov sistem jednačina se metodom zamene (*back-substitution*) lako rešava i dobija nepoznata matrica X.

$$\begin{bmatrix} R_{11} & R_{12} & R_{13} & R_{14} \\ 0 & R_{22} & R_{23} & R_{24} \\ 0 & 0 & R_{33} & R_{34} \\ 0 & 0 & 0 & R_{44} \end{bmatrix} \begin{bmatrix} X_{11} & X_{12} & X_{13} & X_{14} \\ X_{21} & X_{22} & X_{23} & X_{24} \\ X_{31} & X_{32} & X_{33} & X_{34} \\ X_{41} & X_{42} & X_{43} & X_{44} \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} & Y_{13} & Y_{14} \\ Y_{21} & Y_{22} & Y_{23} & Y_{24} \\ Y_{31} & Y_{32} & Y_{33} & Y_{34} \\ Y_{41} & Y_{42} & Y_{43} & Y_{44} \end{bmatrix} \quad (9)$$

Obezbeđeno je da podaci dolaze takvim redosledom da se jednačine (10)-(13) izvršavaju jedna za drugom za redom, jer je za svaku sledeću jednačinu potreban prethodno izračunati podatak. Izračunate vrednosti se smeštaju u internu blok RAM memoriju. Za potrebe izračunavanja novih vrednosti se očitavaju prethodno izračunate vrednosti, koje su sačuvane u memoriji.

$$X_{4i} = \frac{Y_{4i}}{R_{44}} \quad (10)$$

$$X_{3i} = \frac{Y_{3i} - R_{34} \cdot X_{4i}}{R_{33}} \quad (11)$$

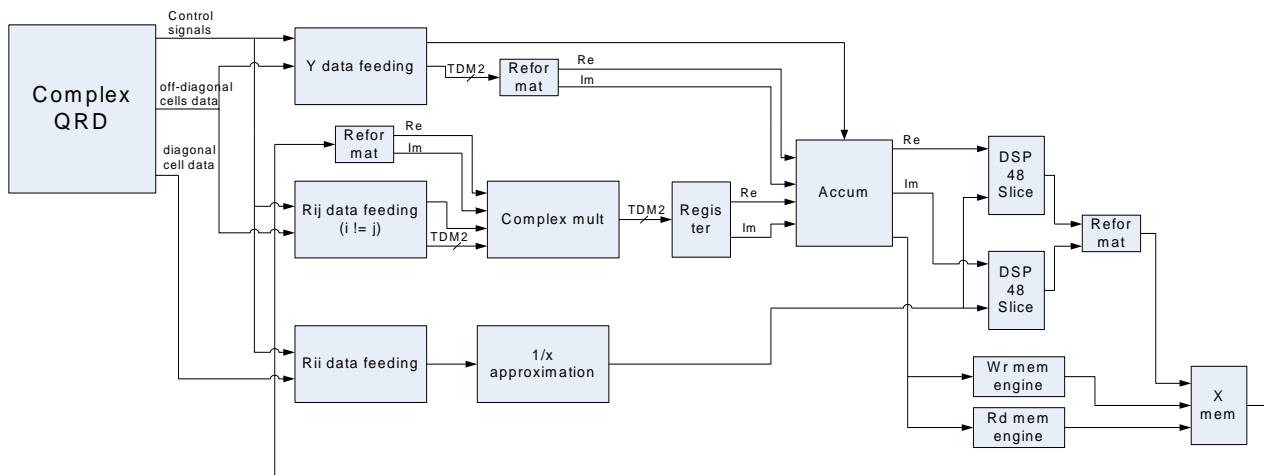
$$X_{2i} = \frac{Y_{2i} - R_{23} \cdot X_{3i} - R_{24} \cdot X_{4i}}{R_{22}} \quad (12)$$

$$X_{1i} = \frac{Y_{1i} - R_{12} \cdot X_{2i} - R_{13} \cdot X_{3i} - R_{14} \cdot X_{4i}}{R_{11}} \quad (13)$$

Svaki od blokova za prosleđivanje odgovarajućih podataka vrše baferisanje ulaznih podataka i logičko oblikovanje izlaznog signala u zavisnosti od jednačine koja se rešava. Deljenje sa odgovarajućom vrednošću je realizovano množenjem sa  $1/x$ . Funkcija  $1/x$  je aproksimirana na sličan način kao i prethodno opisana aproksimacija funkcije  $1/\sqrt{x}$  pomoću polinomske funkcije. Množenje je izvršeno upotrebom DSP48 Slice elementima.

#### IV. FPGA IMPLEMENTACIJA

Arhitektura opisana u trećem poglavlju je realizovana pomoću alata Xilinx System Generator za DSP [11] projektovanje i implementaciju. Ciljna platforma je Virtex@-5 XC5VFX130T-2FF1738 FPGA čip [9]. Frekvencija hardverske platforme je 225MHz. Za ceo sistem je napisana referentna *Hardware-matched* simulacija u MATLAB-u. Greška koja se unosi aproksimacijom funkcija  $1/\sqrt{x}$  i  $1/x$  je reda  $10^{-5}$ , što je



Sl. 5. Blok dijagram hardvera za matričnu inverziju sa posebnim akcentom na blok za rešavanje sistema jednačina.

izuzetan rezultat. U Tabeli 2 je prikazano zauzeće hardverskih resursa na ciljnoj hardverskoj platformi.

TABELA 2: ZAUZEĆE RESURSA NA FPGA ČIPU.

Function	Slices	LUTs/FFs	DSP48	Block RAM
Matrix	3,415	8,142/10,369	57	16
Inversion	(16%)	(9%)	(17%)	(3%)

## V. ZAKLJUČAK

U radu je prikazana hardverska implementacija matematičkog algoritma koji se veoma često koristi u komunikacionim sistemima. Postignuta je veoma visoka fleksibilnost i skalabilnost sistema za računanje inverzije kvadratnih matrica. Takođe, postignut je izrazito visok nivo protočne obrade i paralelnog rada hardverskih blokova na visokim učestanostima, upotrebom specifičnih tehnika digitalne obrade signala. Efikasan dizajn sistema i realizacije matematičkih operacija je razlog relativno malog zauzeća hardverskih resursa čipa. Unapređenja su moguća u pogledu još veće hardverske optimizacije pojedinih blokova sistema.

## LITERATURA

- [1] M. Karkooti, J.R. Cavallaro, C. Dick, "FPGA Implementation of Matrix Inversion Using QRD-RLS Algorithm", *Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers*, page(s): 1625-1629, 2005.
- [2] J. Yue, K. J. Kim, J. Gibson and R. A. Iltis, "Channel Estimation and Data Detection for MIMO OFDM Systems", in *Proceedings of IEEE Global Telecommunication Conference*, vol. 2, page(s): 581-585, 1-5 Dec 2003.
- [3] C. Dick, M. Trajković, S. Denić, D. Vuletić, Raghu Rao, F. Harris, K. Amiri, "FPGA Implementation of a Near-ML Sphere Detector for 802.16e Broadband Wireless Systems", *Proceedings of the SDR '09 Technical Conference and Product Exposition*, December 2009, submitted for publication.
- [4] N. D. Gohar, Z. Rafique, "V-Blast: A Space-Division Multiplexing Technique Providing a Spectral Efficiency Necessary for High Data Rate Wireless Network", *2<sup>nd</sup> International Bhurban Conference on Applied Science and Technology*, Bhurban, Pakistan, June 16-23, 2003, page(s): 238-248, 2003.
- [5] D. Cescato, M. Borgmann, H. Bölcskei, J. Hansen, and A. Burg, "Interpolation-based QR decomposition in MIMO-OFDM systems," *presented at the IEEE Workshop on Signal Processing*

*Advances in Wireless Communications (SPAWC)*, New York, NY, Jun. 2005.

- [6] J. Gotze, U. Schwiegelshohn, "A Square Root and Division Free Givens Rotations for Solving Least Squares Problems on Systolic Arrays", *Siam J. Sci. Stat. Comput.*, vol. 12, No. 4, page(s): 800-907, July 1991.
- [7] Lei Ma, Kevin Dickson, John McAllister, John McCanny, "Modified Givens Rotations And Their Application To Matrix Inversion", *Acoustic, Speech and Signal Processing, 2008. ICASSP '08. IEEE International Conference*, page(s): 1437-1440, 2008.
- [8] M. Myllyla, J. Hintikka, J. Cavallaro, M. Juntti, M. Limingjoja and A. Byman, "Complexity Analysis of MMSE Detector Architecture for MIMO OFDM Systems", in *Proceedings of the 2005 Asilomar Conference, Pacific Grove, CA, Oct 30-Nov2 2005*.
- [9] Xilinx, "Virtex-5 FPGA – User Guide", UG190 (v4.5), January 2009.
- [10] The Mathworks, Simulink – Simulation and Model Based Design, <http://www.mathworks.com/products/simulink>.
- [11] Xilinx, "System Generator for DSP – User Guide", Release 10.1, March 2008.

## ABSTRACT

The last exploration has been related with leading in 4G technology and achieving broadband internet access. Protocols which has tendency to replace existing DSL and cable internet structure with air are WiMAX (*Wireless Interoperability for Microwave Access*) and LTE (*Long Term Evolution*). Using the MIMO (*Multiple Input Multiple Output*) systems within these standards can provide high data rates the prejudice of the complexity increase. The multiple antenna integration into the system makes channel matrix estimation at the receive side what enlarge architecture additionally and require using mathematical operations to achieve a successful detection. Anyway, one of them is the matrix inversion which hardware realization on the FPGA (*Field Programmable Gate Array*) platform is presented in this paper for the example on a 4x4 matrix. An efficient design and a new method in solving of the conventional mathematical operations makes possible clock frequency of 225 MHz at the Xilinx Virtex®-5 chip what afford that a time for single matrix calculation is 285ns.

## FPGA IMPLEMENTATION OF THE COMPLEX MATRIX INVERSION USING QR DECOMPOSITION

Bojan Marić, Lazar Saranovac, Miloš S. Trajković