

54. konferencija za
elektroniku, telekomunikacije,
računarstvo, automatiku i
nuklearnu tehniku

2010

ETRAN

Zbornik apstrakta
i program



Organizator
Društvo za ETRAN

Donji Milanovac
7-10. juna 2010.

PARCIJALNO VISOKOPOUZDANO 1D SISTOLIČKO POLJE ZA MNOŽENJE MATRICE I VEKTORA

Vladimir Simić, Vladimir Ćirić, Ivan Milentijević, Elektronski fakultet, Univerzitet u Nišu
{vladimir.simic, vladimir.ciric, milentijevic}@elfak.ni.ac.rs

Sadržaj – Od samog početka razvoja računarskih arhitektura razmatrala su se rešenja koja mogu prevazići ili tolerisati otkaze u radu. To je dovelo do razvoja oblasti visoke pouzdanosti koja podrazumeva posebne tehnike projektovanja arhitektura. Cilj projektovanja je da se sistem u celini učini otpornim na privremene, ponavljajuće ili trajne defekte i otkaze. Parcijalna visoka pouzdanost analizira i označava oblasti arhitekture koje su značajne za dobijanje ispravnih ili prihvatljivih izlaznih rezultata obrade i na njih primenjuje neki od metoda visoke pouzdanosti. U radu je predložena primena metoda parcijalne otpornosti na defekte na 1D arhitekturi sistoličkog polja za množenje matrice i vektora. Dat je opis matematičkog metoda mapiranja arhitekture na graf i analiza prostiranja grešaka kroz arhitekturu. Dat je poseban osvrt na primenu kod izračunavanja DCT kod H.264/AVC standarda video kodiranja.

1. UVOD

Cilj otpornosti na otkaze (eng. *Fault Tolerance*) je kreiranje sistema koji mogu da kompenzuju otkaze pojedinih delova i nastave sa radom. Počev od sedamdesetih godina prošlog veka do danas, razvijeno je mnogo metoda projektovanja ovakvih sistema. Projektovan je veliki broj različitih arhitektura [1].

Moderni nanotehnološki procesi sve su češće praćeni varijacijama u proizvodnom procesu, sve većim brojem fabrički defektnih čipova, kao i sve većom osetljivošću na spoljne i unutrašnje smetnje [2]. Postizanje 100% ispravnosti u proizvodnji tranzistora i puteva podataka je sve teže i skuplje postići [2].

Kao odgovor na izazove u projektovanju modernih VLSI sistema uvode se kvalitativne izmene u tradicionalnom pristupu. Korektnost rezultata obrade najčešće se tumači nepostojanjem grešaka na bilo kom izlazu iz sistema. Druge definicije uvode pojmove zadovoljavajuće ili prihvatljivo u tumačenju korektnosti rada komponenti [3]. Ovo je naročito važno za DSP sisteme za multimediju koji koriste nedostatak savršenosti ljudskih čula kako bi tolerisali korektnost i preciznost izračunavanja. Upravo se na ovom zasniva princip parcijalne otpornosti na defekte (eng. *PDT – Partial Defect Tolerance*) [3]. Cilj metoda je obezbediti da se postigne visoka pouzdanost za delove sistema za koje se utvrdi da za datu primenu mogu defektom uticati na zahtevanu tačnost u obradi. Kada se u obzir uzme tipičan primer standarda video kodiranja H.264/AVC i niz tehnika i metoda postizanja efikasnosti može se doći do zaključka o važnosti razmatranja stepena korektnosti rezultata obrade na račun cene implementacije.

U radu je predložena primena metoda parcijalne otpornosti na defekte na 1D arhitekturi sistoličkog polja za množenje matrice i vektora. Dat je opis matematičkog metoda mapiranja arhitekture na graf i analiza prostiranja grešaka kroz arhitekturu. Dat je poseban osvrt na primenu kod izračunavanja DCT kod H.264/AVC standarda video kodiranja. Izvršena je implementacija i data analiza

isplativosti primene metoda parcijalne visoke pouzdanosti u odnosu na verovatnoću pojave defekata.

2. DCT NA 1D SISTOLIČKOM POLJU

U ovom poglavlju ćemo dati kratak opis aplikacije DCT na 1D sistoličkom polju za množenje matrice i vektora u odnosu na koju predlažemo primenu parcijalne visoke pouzdanosti.

Jedna od upotreba DCT je priprema za kompresiju pravougaonih delova slike. Ideja je da se nakon primenjene transformacije iz prostornog dobije frekventni domen, iz kojeg je onda moguće ukloniti komponente visoke frekvencije [4, str. 57.]. Na taj način, postaje moguće približno rekonstruisati izvorni signal uzimajući komponente nižih frekvencija. Diskretna kosinusna transformacija (eng. *Forward Discrete Cosine Transform - FDCT*) nad blokom dimenzija $N \times N$ definisana je uvođenjem transformacione matrice A i definiše se proizvodom (1).

$$Y = A \cdot X \cdot A^T \quad (1)$$

Inverzna diskretna kosinusna transformacija (*IDCT – Inverse Discrete Cosine Transform*) data je sa

$$X = A^T \cdot Y \cdot A,$$

gde je A matrica DCT koeficijenata [4, str. 46].

Na slici 1a prikazan je deo slike dimenzija 4×4 , a odgovarajući skup DCT koeficijenata nakon primene diskretne kosinusne transformacije dat je na slici 1b.

126	159	178	181	537.2	-76.0	-54.8	-7.8
98	151	181	181	106.1	-35.0	-12.7	6.1
80	137	176	156	-42.7	46.5	10.3	-9.8
75	114	88	68	20.2	-12.9	-3.9	8.5

a)

b)

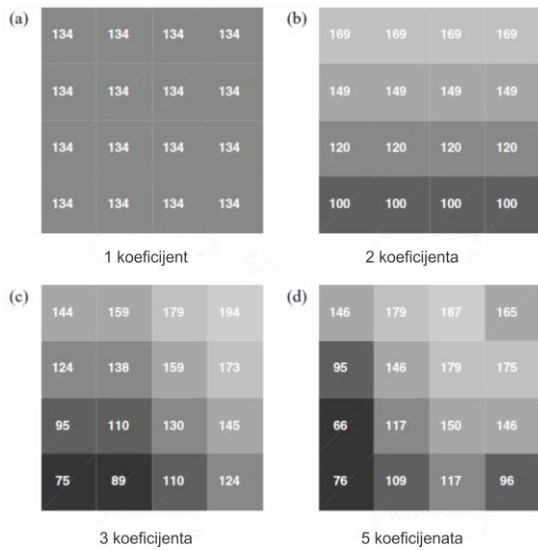
Slika 1. a) Originalni blok i b) DCT koeficijenti nakon transformacije

Ako za postupak inverzne kosinusne transformacije izaberemo najznačajnije cifre iz DCT matrice, tj. cifre $a_{0,0}$, $a_{0,1}$ itd., a za ostale pretpostavimo da su narušene usled defekta, tada broj izabranih cifara određuje stepen tačnosti rekonstruisanog dela slike. Izbor značajnih cifara najčešće se ostvaruje cik-cak obilaskom matrice krećući se od gornjeg levog ugla [4, str. 57.].

Ova činjenica opravdava uvođenje DCT-a jer se smanjuje broj koeficijenata koje treba preneti kako bi se dobio zadovoljavajući kvalitet na strani prijemnika. Broj

izabranih koeficijenata određuje kvalitet rekreiranog dela slike.

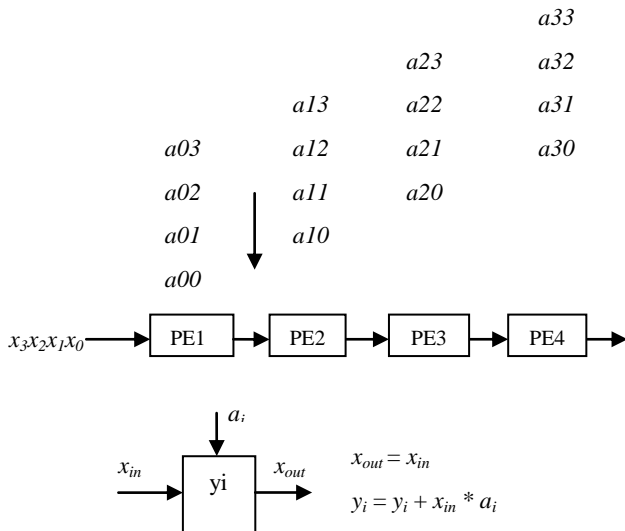
Na slici 2 prikazane su vrednosti nakon rekonstrukcije dela slike na slici 1a kada su iz DCT matrice izabrani redom 1, 2, 3 i 5 najznačajnijih cifara.



Slika 2. Blokovi rekonstruisani iz a) 1 b) 2 c) 3 d) 5 koeficijenata DCT matrice

Za implementaciju DCT-a izabrali smo arhitekturu linearnog 1D sistoličkog polja.

Unidirekciono linearno polje (eng. *Unidirectional Linear Array - ULA*) - linearno polje sastavljeno od N procesnih elemenata kod kojih veze između elemenata postoje samo u jednom smeru. Procesne jedinice u polju su povezani jednim tokom podataka.



Slika 3. Tok podataka ULA implementacije algoritma za množenje matrice vektorom za $n=4$ i funkcionalni opis ćelije

Operacija koja se odvija na ULA polju predstavlja operaciju množenja matrice vektorom. Neka je ulazna matrica A dimenzija $m \times n$ a vektor X dimenzija $n \times 1$. Rezultujući proizvod predstavlja vektor Y dimenzija $m \times 1$, sa elementima $\{y_i\}$, $i=0,1,\dots,m-1$, takvim da važi

$$Y = A \cdot X.$$

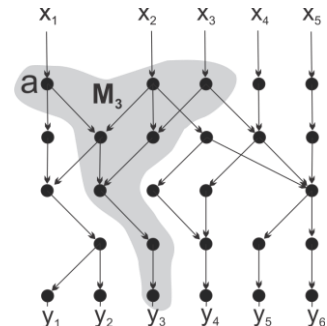
Množenje matrica u izrazu (1) obavljamo u dve faze. U prvoj fazi računa se $Y' = A \cdot X$ što predstavlja računanje jedno-dimenzionalne DCT transformacije za svaku kolonu matrice X . U drugoj fazi računa se $Y = Y' \cdot A^T$, što odgovara računanju jedno dimenzionalne DCT transformacije za svaku red matrice Y' . Za obavljanje prve faze u polje dovodimo matricu A i redom kolone matrice X . Kao izlaz dobijamo redom kolone matrice Y' . U drugoj fazi u polje uvodimo matricu A i redom vrste matrice Y' , kako bismo na 1D polju izračunali redom rezultujuće vrste DCT matrice. Pri tome se vodi računa da se u prvoj fazi na polju generišu rezultujuće kolone, a u drugoj fazi rezultujuće vrste matrice.

Gubitke u procesu proizvodnje prilikom implementacije arhitekture sa slike 3, nastale usled pojave defekata, moguće je smanjiti primenom neke od metoda visoke pouzdanosti. Dalju uštedu moguće je postići analizom osobina IDCT transformacije. Naime, ukoliko se defekt javi na krajnje desnoj ćeliji (slika 3), ovakav defekt neće značajno uticati na rekonstruisani deo slike (slika 2). Ovakva analiza nameće zaključak da je moguće parcijalno primeniti neki od metoda visoke pouzdanosti samo na ćelije polja koje vrše obradu najznačajnijih DCT koeficijenata i time povećati efikasnost fabrikacije ove arhitekture.

3. PARCIJALNA VISOKA POUZDANOST

U ovom poglavlju dat je kratak pregled i opis metode parcijalne otpornosti na defekte koji je predložen i formalno opisan u [3]. U drugom delu poglavlja data je analiza dobiti uvođenja pouzdanosti u sistem [3].

Metod ćemo ilustrovati na primeru sa slike 4.



Slika 4. Primer arhitekture i putevi prostiranja grešaka

Na slici 4 predstavljen je graf arhitekture sistema čije komponente vrše obradu nad ulaznim podacima $\{x_i\}$ i na izlazima generišu izlazne rezultate $\{y_i\}$. Vezama između elemenata označavamo prenos rezultata obrade.

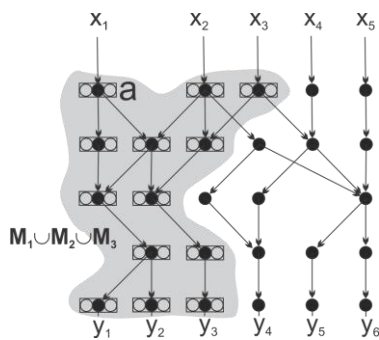
Metod opisan u [3] predlaže analizu prostiranja grešaka kroz arhitekturu. Pojava defekta u čvoru a sa slike 4 prostire se kroz arhitekturu sve do izlaza y_1 , y_2 i y_3 . Metod iz [3] formalno uvodi skup čvorova M_n za svaki od izlaznih bitova y_n , koji mogu propagirati greške do posmatranog izlaza y_n . Skup čvorova M_3 koji odgovaraju izlaznom čvoru y_3 označen je na slici 4. Da bi se računski podržala metoda, skup M_n se predstavlja matricom M_n koja elementima sa vrednošću 1 označava odgovarajuće čvorove u arhitekturi. Ako se za datu primenu ustanovi da je korektnost izlaza y_1 bitna, skup čvorova M_1 projektuje se visokopouzdan. Skup čvorova koji pripada skupu M_n označava se kao netolerantna oblast (eng. *NTA - Non Tolerant Area*) [3] za izlaz y_n . Da li čvor pripada NTA oblasti određeno je primenom arhitekture.

Označimo sa α broj najznačajnijih izlaznih bitova arhitekture $\{y_i\}$ $i=1,\dots,\alpha$. Parametar koji je određen

arhitekturom i geometrijom veza $\Gamma(\alpha)$ označava procenat polja zahvaćenog skupom $\mathbf{M}_1 \cup \mathbf{M}_2 \cup \dots \cup \mathbf{M}_\alpha$. Sa T ćemo označiti ukupan broj elemenata u sistemu.

Kako je propagacija grešaka po prirodi tranzitivna, mapa uticaja defekata može se dobiti iz matrice tranzitivnog zatvaranja grafa pridruženog arhitekturi G u oznaci A^* . Skup čvorova koji defekt propagiraju do izlaznog bita y_n , nalaze su u koloni matrice A^* pridruženoj izlazu y_n [3].

Drugi deo analize obuhvata analizu isplativosti fabrikacije arhitekture za primenu parcijalne otpornosti na defekte [3]. Na slici 5 prikazana je TMR metoda parcijalne visoke pouzdanosti na delu grafa koji pripada skupu $\mathbf{M}_1 \cup \mathbf{M}_2 \cup \mathbf{M}_3$. Isplativost metode moguće je proveriti određivanjem verovatnoće pojave defekata na arhitekturi za koje je cena uvedene parcijalne visoke pouzdanosti manja od cene utrošene na fabrički defektno čipove implementirane bez visoke pouzdanosti.



Slika 5. Parcijalna visoka pouzdanost

Neka je p verovatnoća da sistem ima defektan podsistem. Označimo sa DT sisteme koji su u potpunosti i sa PDT sisteme koji su delimično implementirani otporni na defekte, respektivno. Cena izrade jednog nedefektnog primerka DT i PDT sistema data je sa [3]

$$U^{DT}(p, \alpha) = \frac{C}{(1-p)^{\Gamma(\alpha)^{*T}}}, U^{PDT}(p, \alpha) = \frac{C_{PDT}}{R^{\Gamma(\alpha)^{*T}}},$$

gde je sa R označena verovatnoća otkaza visokopouzdanosti, i zavisi od izabrane metode visoke pouzdanosti.

Dobit DT i PDT sistema (Y) dobija se korišćenjem prethodne definicije i dat je sa [3]

$$Y_{DT}(p, \alpha) = \frac{C}{U^{DT}(p, \alpha)} = (1-p)^{\Gamma(\alpha)^{*T}}, Y_{PDT}(p, \alpha) = \frac{C}{C_{PDT}} R^{\Gamma(\alpha)^{*T}}. \quad (2)$$

Kod parcijalno visoko pouzdanih arhitektura podsisteme u netolerantnoj oblasti implementiramo primenjujući neku od tehnika za postizanje visoke pouzdanosti, tako da to povećava cenu sa C na C_{PDT} , jer se sistemu dodaju novi hardverski elementi. Uvođenjem n istovetnih kopija procesnog elementa cena implementacije ovih podsistema uvećala se n puta dok je cena ostatka sistema ostala ista tj.

$$C_{PDT} = \Gamma(\alpha) \cdot n \cdot C + (1 - \Gamma(\alpha)) \cdot C = (1 + (n-1)\Gamma(\alpha))C$$

Poznate tehnike visoke pouzdanosti su redundantni moduli (eng. *Module Redundancy*) i rezervne komponente (eng. *Spare Components*). Za potrebe implementacije izabrana je SC metoda kod koje je rezultujuća verovatnoća postojanja bar jedne ispravne komponente od ukupno n data sa

$$R = \sum_{k=1}^n \binom{n}{k} (1-p)^k \cdot p^{n-k}.$$

4. PROJEKTOVANJE 1D SISTOLIČKOG POLJA

Analiza sistema kod primene delimične otpornosti na defekte svodi se na nekoliko koraka: izbor stepena tačnosti izlazne reči uzimajući u obzir domen primene, opisivanje hardverskog sistema grafom toka podataka, izračunavanje tranzitivnog zatvaranja nad definisanim grafom, mapiranje tranzitivnog zatvaranja na hardversku arhitekturu preko mapa uticaja koje označavaju kritične podsisteme [3]. Analiza dobiti pokazuje kada će cena parcijalno visoko pouzdane arhitekture biti manja od implementacije bez uvođenja visoke pouzdanosti za datu verovatnoću nastanka defekta.

Poznato je da se matrica tranzitivnog zatvaranja može odrediti iz matrice susedstva [5]. Matrica susedstva grafa koji odgovara ULA polju za množenje matrice dimenzija $m \times n$ i vektora dimenzija $n \times 1$ sa slike 3 prikazana je na slici 6.

$$A = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & 1 & 0 & & \\ 0 & 0 & 1 & 1 & 0 & \\ \vdots & & & & & \\ \vdots & & & & & \\ \vdots & & & & & \\ 0 & 0 & \dots & \dots & 1 & 1 \\ 0 & 0 & \dots & \dots & & 1 \end{bmatrix}_{m \times m}$$

Slika 6. Matrica susedstva pridružena ULA polju

Matrica tranzitivnog zatvaranja A^* je oblika:

$$A^* = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ 0 & 1 & & & & 1 \\ 0 & & 1 & & & \vdots \\ \vdots & & & 1 & & \vdots \\ \vdots & & & & & \vdots \\ 0 & & & & & 1 \\ 0 & 0 & \dots & \dots & & 1 \end{bmatrix}_{m \times m} \quad (3)$$

Mape uticaja defekata mogu se dobiti iz odgovarajuće matrice tranzitivnog zatvaranja izborom odgovarajuće kolone. Na osnovu matrice A^* date sa (3), matrice M_η za $\eta \in [0, m-1]$ su:

$$M_0 = [1 \ 0 \ 0 \ \dots \ 0 \ 0]_{1 \times m}$$

$$M_1 = [1 \ 1 \ 0 \ 0 \ \dots \ 0]_{1 \times m}$$

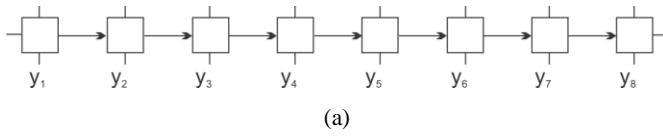
⋮

⋮

⋮

$$M_{m-1} = [1 \ 1 \ 1 \ \dots \ 1 \ 1]_{1 \times m}$$

Na slici 7a prikazano je osnovno polje. Osenčene ćelije na slici 7c predstavljaju skup $\mathbf{M}_1 \cup \mathbf{M}_2 \cup \mathbf{M}_3$ i posebno su označeni skupovi \mathbf{M}_1 , \mathbf{M}_2 i \mathbf{M}_3 . Kolone matrice A^* , koje odgovaraju skupovima \mathbf{M}_1 , \mathbf{M}_2 i \mathbf{M}_3 prikazane su na slici 7b i odgovaraju mapama uticaja za odgovarajuće elemente y_1 , y_2 i y_3 , redom.

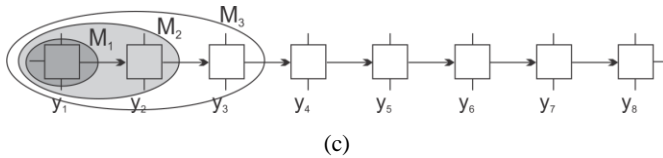


$$M_0 = [1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]_{1 \times m}$$

$$M_1 = [1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]_{1 \times m}$$

$$M_2 = [1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0]_{1 \times m}$$

$$M_3 = [1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0]_{1 \times m}$$



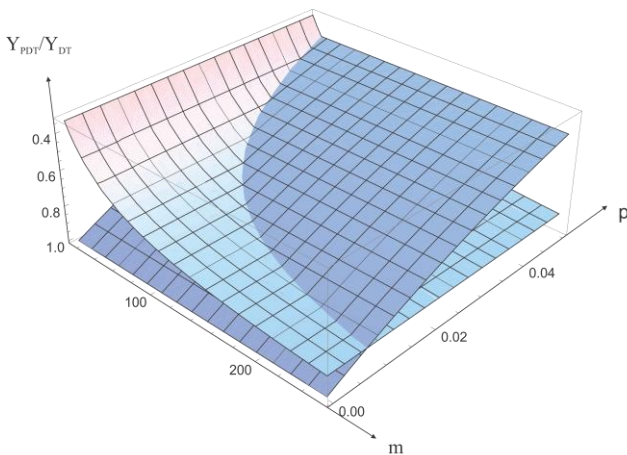
Slika 7. Primer ULA polja: a) Usmereni graf arhitekture ULA polja za $m = 8$ b) skup ćelija M_2 vezanih za izlaz y_2 c) mape uticaja M_0 do M_3

Da bi se sprovela analiza o međusobnom odnosu DT i PDT pristupa, izračunavamo i $\Gamma(\alpha)$, tj. procenat zahvaćenosti arhitekture čvorovima iz skupa $M_1 \cup M_2 \cup \dots \cup M_\alpha$. Koristeći uvedene definicije iz (2) dolazimo do sledećih izraza za dobiti kod ULA polja:

$$Y_{DT}(p, \alpha) = (1-p)^{\Gamma(\alpha)^T} = (1-p)^{\frac{\alpha T}{m}} = (1-p)^\alpha,$$

$$Y_{PDT}(p, \alpha) = \frac{C}{C_{PFT}} \cdot R^{\Gamma(\alpha)^T} = \frac{m}{m+2 \cdot \alpha} \cdot [(1-p)^3 + 3 \cdot (1-p)^2 \cdot p + 3 \cdot (1-p) \cdot p^2]^\alpha.$$

Na slici 8 grafički su prikazani izrazi dati sa (4). Sa slike se može uočiti granična oblast preseka dva grafika, za koji važi jednakost dveju dobiti. Oblast od interesa za sprovedenu analizu je oblast iznad preseka, za koju važi da je dobit Y_{PDT} veća od dobiti Y_{DT} .



Slika 8. Prikaz odnosa PDT i DT dobiti iz (2) za ULA polje za $\alpha=32$

5. REZULTATI IMPLEMENTACIJE

Da bi se potvrdila metoda i stekao uvid u obim hardvera potrebnog za implementaciju parcijalne visoke pouzdanosti, implementacija unidirekcionog 1D sistoličkog polja obavljena je u Xilinx ISE Web pack okruženju za familiju čipa Virtex 4. U tabeli 5.1 prikazani su uporedni rezultati

implementacije 1D sistoličkog polja za različite dimenzije vektora i matrice, $m \leq 32$, i različit procenat parcijalne pouzdanosti izražen preko α . Iz implementacije su izostavljeni memorisjski elementi i upravljačka logika.

Tabela 5.1. Rezultati implementacije N bitnog ULA polja sa trostrukim SC ćelijama

Dimenzija vektora tj. matrice m	Osnovna arhitektura (0%)	PDT na arhitekturi		DT na arhitekturi (100%)
		25%	50%	
$m=4$	3.2	5.3	7.3	11.4
$m=8$	7.3	12.0	16.7	26.0
$m=16$	18.3	29.8	41.4	64.6
$m=32$	50.9	83.1	115.4	180.0

U tabeli 5.1 prikazani su resursi utrošeni za implementaciju ULA polja u ekvivalentnom broju gejtova kG. Kolona koja odgovara PDT arhitekturi pokazuje utrošene resurse za realizaciju 25% i 50% parcijalne pouzdanosti.

6. ZAKLJUČAK

U radu je predložena primena metoda parcijalne otpornosti na defekte na 1D arhitekturi sistoličkog polja za množenje matrice i vektora. Dat je opis matematičkog metoda mapiranja arhitekture na graf i analiza prostiranja grešaka kroz arhitekturu. Dat je poseban osvrt na primenu kod izračunavanja DCT kod H.264/AVC standarda video kodiranja. Prikazani su rezultati implementacije.

LITERATURA

- [1] D.A. Rennels, „Fault-Tolerant Computing, Encyclopedia of Computer Science“, ed., Anthony Ralston, Edwin Reilly, and David Hemmendinger, London; Nature Publishing Group, 2000
- [2] International Technology Roadmap for semiconductors, Recommendations, <http://public.itrs.net/>, 2001.
- [3] Vladimir Ćirić, Aleksandar Cvetković, Ivan Milentijević, "Yield analysis of partial defect tolerant bit-plane array", Elsevier, Computers and Mathematics with Applications (2009), January 2010
- [4] Iain E. G. Richardson, "H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia", John Wiley & Sons, 2003.
- [5] T. H. Cormen, C. E. Leiserson, R. L. Rivest, "Introduction to Algorithms", MIT Press, 2003 (2nd edition)

Abstract – In this paper the application of partial defect tolerance on the architecture of 1D matrix-vector multiplication systolic array is proposed. Error propagation through the architecture is obtained by method of partial defect tolerance application on architecture's data flow graph. Partial defect tolerance is applied in respect to DCT H.264/AVC video coding application.

PARTIAL TOLLERANT DESIGN ON 1D SYSTOLIC ARRAYS FOR MATRIX VECTOR MULTIPLICATION

Vladimir Simić, Vladimir Ćirić, Ivan Milentijević