Analiza high-speed I/O tehnologija iz ugla low-cost peer-to-peer komunikacionog sistema

Mirjana Ž. Stojilović

***Sadržaj* — Prilikom projektovanja sistema u kojima se koristi *high-speed board-to-board* komunikacija se uvek javlja potreba za analizom aktuelnih tehnologija i izborom one koja najbolje odgovara tom sistemu. Zadatak ovog rada je izbor tehnologije za sistem u kome je potrebno realizovati *high-speed peer-to-peer* komunikaciju između nekoliko štampanih ploča međusobno povezanih pomoću *mezzanine* konektora. Da bi se odabrala odgovarajuća tehnologija pošlo se od danas najčešće korišćenih, kao što su RapidIO, HyperTransport i PCI Express. Na osnovu specifikacija tih tehnologija dat je pregled njihovih karakteristika od značaja za sistem. Potom su prikazana neka od konkretnih rešenja za realizaciju komunikacije u sistemu. Na kraju je izvršen izbor najpogodnije tehnologije na osnovu specifikacije sistema, karakteristika tehnologija i cena prikazanih rešenja.**

***Ključne reči* — RapidIO, HyperTransport, PCI Express, Peer-to-peer, FPGA.**

I. UVOD

Na tržištu danas postoji puno rešenja za komunikaciju između čipova, bilo da su oni istog tipa (npr. dva procesora) ili različitog tipa (npr. procesor i memorija), bilo da su smešteni na istoj ili na različitim štampanim pločama. Mnoga od tih rešenja su standardizovana i osnovane su organizacije za njihovo održavanje i unapređivanje. Često se javlja potreba da se u mnoštvu mogućih rešenja pronađe najpogodnije sa stanovišta specifikacije sistema koji je potrebno realizovati. Koje rešenje će biti najbolje zavisi od prednosti koje ono pruža nad ostalim rešenjima. Idealno bi bilo kada bi jedno rešenje objedinilo više prednosti, kao što su npr. brzina komunikacije, pouzdanost, jednostavnost realizacije i prihvatljiva cena. Takođe, važno je da odabrano rešenje bude aktuelno i perspektivno. Time sama njegova implementacija doprinosi vrednosti sistema. Međutim, obično je prilikom izbora potrebno napraviti kompromis jer ni jedno od rešenja ne objedinjuje sve navedene prednosti.

Cilj ovog rada je da se pronađe optimalno rešenje za jedan konkretan test sistem i da se obrazlože kriterijumi koji su doveli do tog izbora. Motiv za to je temeljno upoznavanje sa modernim tehnološkim rešenjima komunikacije između čipova i primena odabranog rešenja u realnom hardverskom sistemu.

Sistem za koji treba odabrati tehnologiju ima ulogu kontrolera više elektronskih uređaja istovremeno. Zbog toga sadrži puno interfejsa za komunikaciju sa tim uređajima. Interfejsi su raspoređeni na više štampanih ploča. Ploče su međusobno povezane pomoću *mezzanine* konektora (naslagane jedna na drugu). Na svakoj ploči se nalazi po jedan FPGA čip čija je funkcija da obavlja komunikaciju sa bilo kojom drugom pločom u sistemu (*peer-to-peer* veza) i sa uređajima povezanim na interfejse implementirane na toj ploči. Ukoliko je potrebno, isti čip bi vršio obradu primljenih podataka. Zahtevani maksimalni protok između dva FPGA čipa u sistemu je 1 GBps.

U mnoštvu opcija analizirane su sledeće tri: RapidIO, HyperTransport i PCI Express. U poglavlju II je dat kratak osvrt na istoriju razvoja navedene tri tehnologije. Na osnovu literature [1]-[3] izdvojene su najvažnije karakteristike svake od tehnologija i predstavljene u poglavlju III. Nakon toga su predložena neka od konkretnih rešenja za realizaciju komunikacije između ploča. Na osnovu karakteristika tehnologija i cena predloženih rešenja, u poglavlju IV je opisan postupak izbora najoptimalnijeg rešenja za dati sistem.

II. ISTORIJA RAZVOJA RAPIDIO, HYPERTRANSPORT I PCI EXPRESS TEHNOLOGIJA

Sa jedne strane, efikasne kompjuterske sisteme karakteriše balans između performansi procesora, memorija i magistrala koje ih spajaju. Sa druge strane, *Moore*-ov zakon [4] je precizno predvideo da će se broj tranzistora po integrisanom kolu udvostručavati na svake dve godine. Stoga se može zaključiti da se *input/output* (I/O) magistrale moraju radikalno menjati svakih par godina kako bi se održale performanse sistema. Međutim, performanse I/O magistrala su diktirane mehaničkim i drugim ograničenjima električnih vodova i konektora na štampanim pločama, koja su mnogo strožija od ograničenja za povećanje kapaciteta integrisanih kola. Pored toga, razvoj arhitektura I/O magistrala odlikuje velika inercija jer od nje zavise arhitekture interfejsa ka magistrali unutar ASIC, DSP ili FPGA čipova. Iz tih razloga, uspešne arhitekture magistrala se obično ne menjaju i ne unapređuju po desetak godina i više. Ranije su arhitekture sistema bile usko povezane sa arhitekturom procesora na kome je sistem bio zasnovan, a često su se koristile *proprietary* magistrale razvijene za tačno određene sisteme.

Peripheral Component Interconnect (PCI) magistrala je bila najčešći izbor među *custom* i *semi-proprietary*

Mirjana Ž. Stojilović, Institut Mihajlo Pupin, Volgina 15, 11050 Beograd, Srbija; (e-mail: mirjana.stojilovic@institutepupin.com).

magistralama, i postala je veoma korišćena u sistemima kao što su PC računari, serveri, pa čak i *embedded* sistemi, iako je frekvencija takta na PCI magistrali od 66 MHz bila daleko ispod gigahercnih frekvencija modernih procesora. PCI je paralelna, *multi-drop* magistrala sa multipleksiranim adresnim i signalima podataka, uz određeni broj kontrolnih signala. Bila je zamišljena da podrži kako *board-to-board* tako i *chip-to-chip* komunikaciju [3].

Kada je uočeno da tadašnje I/O magistrale nisu bile u mogućnosti da podrže gigahercne i brže procesorske sisteme, formirale su se dve struje za projektovanje sadašnjih I/O tehnologija. Prva je bila usmerena ka proširenju i unapređenju PCI magistrale, i rezultovala je stvaranjem specifikacije za PCI-X magistralu kod koje je brzina rada povećana na 533 MHz. Druga struja je težila kreiranju potpuno nove tehnologije u kojoj bi prenos podataka na fizičkom sloju bio *point-to-point* tipa. Pored toga, koristio bi se *Low Voltage Differential Signalling* (LVDS) ili diferencijalni *Current Mode Logic* (CML) metod prenosa podataka. Unutar druge struje izdvojila su se tri nezavisna pravca predvođena od strane poznatih proizvođača procesora – AMD, Intela i Motorole, respektivno. AMD i njihovi partneri su 1997. godine razvili HyperTransport tehnologiju za *chip-to-chip* komunikaciju. Motorola je razvila RapidIO tehnologiju 2001. godine. Prva verzija je bila namenjena za *chip-to-chip* paralelnu komunikaciju, a kasnije je evoluirala u *backplane* orijentisanu serijsku komunikaciju nazvanu Serial RapidIO. Intel je razvio PCI Express, serijski metod prenosa podataka namenjen pružanju *chip-to-chip*, *board-to-board* i *system-to-system* komunikacije. PCI Express 1.0 je odobren krajem 2003. godine, a PCI Express Advanced Switching (AS) u 2004. godini.

III. KARAKTERISTIKE RAPIDIO, HYPERTRANSPORT I PCI EXPRESS TEHNOLOGIJA

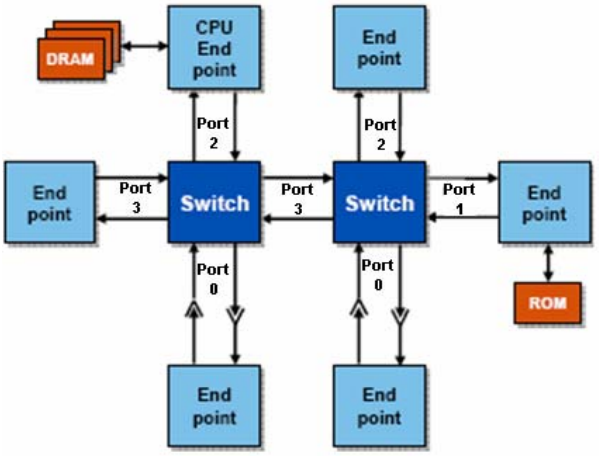
U ovom poglavlju je na osnovu literature [1]-[3] dat prikaz osnovnih karakteristika RapidIO, HyperTransport i PCI Express tehnologija. Navedene su one karakteristike koje su od značaja za realizaciju komunikacije između štampanih ploča u test sistemu, čiji je opis dat u uvodnom poglavlju.

A. RapidIO

RapidIO je tehnologija za *chip-to-chip* i *board-to-board* komunikaciju u *embedded* sistemima srednje i velike kompleksnosti. Ova tehnologija pruža veliku brzinu uz paketski prenos i *peer-to-peer* komunikaciju između čipova kao što su ASIC, DSP, FPGA, mikroprocesori, itd. RapidIO koristi diferencijalne *current steering* drajvere definisane IEEE 802.3 XAUI specifikacijom. Ti drajveri su projektovani da omoguće prenos signala na veće udaljenosti u *backplane* sistemima.

Fizički sloj na kome se implementira RapidIO protokol može biti paralelni ili serijski komunikacioni link. U slučaju paralelnog prenosa koristi se 8 ili 16 linija, uz *double data rate* (DDR) signalizaciju i frekvencije takta od 250 MHz do 1 GHz. Prenos se vrši korišćenjem *source-synchronous* metoda i nekoliko kontrolnih signala. Komunikacija je paketskog tipa. Paralelni prenos se najčešće koristi u *low-cost embedded* sistemima gde je potrebno postići veliki protok u *chip-to-chip* komunikaciji. U slučaju serijskog prenosa koriste se od jedan do četiri diferencijalna para uz primenu 8b/10b kodiranja. Pri tome se postiže maksimalni protok 12.5 Gbps. Serijski prenos se najčešće primenjuje u telekomunikacionim sistemima koji zahtevaju brzu i pouzdanu *board-to-board* komunikaciju. Za razliku od paralelnog prenosa gde se za razlikovanje podataka od kontrolnih simbola koristi poseban *FRAME* signal, u slučaju serijskog prenosa koriste se rezervisane 10-bitne reči iz skupa komandnih reči 8b/10b koda. Za zaštitu podataka primenjuje se *Cyclic Redundancy Check* (CRC). Ukupan *overhead* po paketu se kreće od 16B do 28B.

Centralni deo RapidIO sistema su tzv. *switch*-evi koji usmeravaju pakete od jedne do druge krajnje tačke (*end point*) u sistemu, kao što je prikazano na Sl. 1.



Sl. 1. Blok arhitektura RapidIO sistema.

Vodeći proizvođači integrisanih kola, kao što su npr. Texas Instruments i Xilinx, implementirali su RapidIO u svoje DSP procesore i FPGA čipove. Vodeći proizvođači telekomunikacione opreme, kao što su Alcatel, Ericsson i Motorola, implementirali su RapidIO u svoje uređaje. RapidIO pruža *Quality of Service* (QoS) mehanizme, uključujući i mogućnost hardverske detekcije i korigovanja grešaka, bez reakcije softvera. RapidIO podržava razne topologije povezivanja *end point*-a u sistemu, kao što su zvezda, *daisy-chain*, itd. Moguće je koristiti redundantne linkove u sistemu, kako bi se povećala njegova pouzdanost. Protok po linku i širina linka se mogu menjati. RapidIO sadrži i tzv. *bridging* funkcije koje mu omogućavaju vezu sa drugim tehnologijama, kao što su PCI, PCI Express, InfiniBand. Pored toga podržava i enkapsulaciju podataka, tako da se RapidIO mrežom mogu prenositi i Ethernet paketi. RapidIO *pear-to-pear* arhitektura je vrlo robusna, pošto sistem nikad nije prepušten na milost i nemilost jednom jedinom uređaju.

TABELA 1: PREGLED KARAKTERISTIKA RAPIDIO, HYPERTRANSPORT I PCI EXPRESS I/O TEHNOLOGIJA.

	Serial RapidIO	HyperTransport	PCI Express
Physical layer	• serijski *current mode logic* (CML) prenos (1.25, 2.5, 3.125 GHz) • 1 ili 4 *lane*-a • ekstrakcija takta, kompenzacija • *deskewing* za više *lane*-ova • 8b/10b kodiranje • *framing*, korišćenje *start* simbola	• paralelni LVDS prenos (0.4 – 1.6 GHz) • 2-32 linije podataka, 1 kontrolna linija, 1-4 linije takta • kompenzacija signala takta • *deskewing* za više *lane*-ova • bez dodatnog kodiranja • Nema *framing*-a	• serijski CML prenos (2.5 GHz) • 1, 2, 4, 8, 16 ili 32 *lane*-a • ekstrakcija takta, kompenzacija • *deskewing* za više *lane*-ova • 8b/10b kodiranje • *framing*, *start/end* simboli • skremblovanje
Link layer	• 16b CRC zaštite za svaki paket, 5b zaštite za kontrolne simbole • *Ack/Nack* protokol po linku • Povratna informacija o grešci • klasifikacija i priprema paketa • 3 B *data link layer* paketi	• 32b CRC zaštite na svakih 512 B • bez *hardware error recovery* mehanizma • klasifikacija paketa • bez data link layer paketa	• 16b CRC zaštite za svaki paket, i opcionalno 32b CRC zaštite *end-to-end* • *Ack/Nack* protokol po linku • klasifikacija i priprema paketa • 6 B *data link layer* paketi
Transport layer	• bez posebnog tipa saobraćaja • 4 prioriteta u saobraćaju • *data payload* do 256 B, uz tipičnu veličinu *header*-a 6B	• tri tipa saobraćaja: posted, non-posted i completion • opcionalno se može definisati poseban nivo prioriteta • različiti paketi za podatke (do 64 B) i komande (4 ili 8 B)	• tri tipa saobraćaja: posted, non-posted i completion • 8 prioriteta u saobraćaju • *data payload* do 4096 B, uz tipičnu veličinu *header*-a 12-16B

B. HyperTransport

AMD je zajedno sa svojim partnerima projektovao i razvio HyperTransport metod *chip-to-chip* komunikacije krajem 90-tih godina. U ovoj tehnologiji se za prenos podataka koristi *point-to-point* paralelna magistrala i 1.2 V LVDS DDR signalizacija [2]. Signali komande, adrese i podataka (CAD signali) su multipleksirani i može ih biti 2, 4, 8, 16 ili 32. Svaki bajt CAD signala ima svoj kontrolni (CTL) signal. Svaki bajt CAD signala i odgovarajući CTL signal imaju svoj signal takta. Pored tih signala magistrala sadrži i dva dodatna signala – PWROK i RESET. HyperTransport komunikacioni link se sastoji od dva jednosmerna linka – *upstream* i *downstream*. Komunikacija je paketskog tipa. Postoje dva tipa paketa – paket podataka i kontrolni paket. Oni sadrže 4-64B podataka i 8-12B *overhead*-a. Razmena paketa između uređaja se vrši u *daisy-chain* konfiguraciji sistema. Na početku lanca se nalazi tzv. *host* uređaj, a na kraju tzv. *cave* uređaj. Između *host* i *cave* uređaja su neki od sledeća dva tipa uređaja:

- *bridge* – uređaj koji omogućava komunikaciju između dva uređaja od kojih jedan koristi HyperTransport protokol, dok drugi uređaj koristi drugačiji protokol, npr. PCI Express;
- *tunnel* – uređaj koji omogućava komunikaciju između dva uređaja koji koriste HyperTransport protokol.

Host uređaj je obično procesor koji može da komunicira sa maksimalno 32 uređaja u jednom lancu. Uređaji se mogu dogovarati oko širine linka, frekvencije takta i nivoa protokola koji će koristiti. Maksimalna frekvencija takta je 1.4 GHz [2], čime je omogućen prenos do 22.4 GBps u punoj konfiguraciji. Pošto se signal takta šalje odvojeno od signala podataka, nema potrebe za dodatnim kašnjenjem signala usled serijalizacije, paralelizacije i dekodovanja, što je neophodno u sistemima koji koriste RapidIO ili PCI Express tehnologiju. Za zaštitu podataka se koristi CRC kodovanje.

C. PCI Express

PCI Express tehnologija je nastala na osnovu PCI tehnologije, i stoga pruža punu PC kompatibilnost. Koristi se u PC računarima za komunikaciju između modula, u mrežnim serverima, u telekomunikacionim uređajima, praktično svuda gde je potrebno uspostaviti komunikaciju između modula povezanih pomoću PCI konektora.

PCI Express tehnologija primenjuje *point-to-point* serijsku vezu, pri čemu se dvosmerna komunikacija postiže pomoću dva jednosmerna linka [3]. Serijska veza funkcioniše slično kao Serial RapidIO, s tim što se koriste posebni diferencijalni *current steering* drajveri. Ti drajveri se razlikuju od onih definisanih u IEEE 802.3 XAUI speficikaciji. Linkova može biti 1, 2, 4, 8, 16 ili 32. Takt se šalje zajedno sa podacima, tako da se u prijemniku mora vršiti ekstrakcija takta. Komunikacija je paketskog tipa. Struktura paketa se sastoji od sledećih slojeva: *Transaction Layer* (unosi 12B ili 16B *overhead*-a), *Data Link Layer* (unosi 8B *overhead*-a) i *Physical Layer* (unosi 20% *overhead*-a zbog korišćenja 8b/10b kodovanja). Maksimalna veličina paketa je 4096B. Za očuvanje integriteta podataka koristi se CRC kodovanje. PCI Express pruža QoS mehanizme. Maksimalni efektivni protok na PCI Express magistrali u punoj konfiguraciji (32 dvosmerna linka) je 32 GBps.

U Tabeli 1. su sumirane najbitnije karakteristike opisanih magistrala.

IV. IZBOR TEHNOLOGIJE

U test sistemu opisanom u uvodu, ploče su povezane pomoću *mezzanine* konektora čije karakteristike moraju biti takve da se omogući očuvanje integriteta *high-speed* signala. Pored toga, poželjno je da FPGA čipovi budu povezani u *daisy-chain* konfiguraciju jer na taj način svaki signal prolazi samo kroz jedan konektorski par i trpi manja izobličenja nego u slučaju *multidrop* konfiguracije.

Ukoliko bi se implementirala u test sistemu, svaka od opisanih tehnologija bi mogla pružiti zahtevani protok. Međutim, efektivni protok podataka u sistemu zavisi od veličine paketa koje bi FPGA čipovi razmenjivali, jer paketi u svakoj od tehnologija imaju određeni *overhead*. Za slučaj prenosa malih paketa, RapidIO i PCI Express unose veliki *overhead*. Nasuprot tome, HyperTransport protokol unosi manji *overhead* i manje kašnjenje, ali u tom slučaju uređaji moraju biti povezani u *daisy-chain* konfiguraciju u kojoj jedan uređaj mora vršiti ulogu *host* uređaja. I pored toga, moguće je implementirati *peer-to-peer* komunikaciju koristeći HyperTransport protokol. Međutim, za raliku od npr. RapidIO protokola, HyperTransport nije zamišljen sa ciljem da podrži veliki broj *peer-to-peer* veza. Zbog toga, ukoliko zanemarimo *overhead* u paketskoj komunikaciji, rešenje treba potražiti u primeni RapidIO ili PCI Express tehnologije.

Bitan kriterijum za izbor rešenja je cena njegove implementacije. Da bi se realizovale veze između štampanih ploča u test sistemu, potrebno je u FPGA čipu instancirati *Intellectual Property* (IP) *core* module koji obavljaju komunikaciju primenom neke od tehnologija. Pored toga, moguće je uz FPGA čip koristiti dodatno *transceiver* kolo. U tom slučaju ulogu IP *core* modula obavljaju sama *transceiver* kola. Neka od mogućih konkretnih rešenja su:

- RapidIO tehnologija – jedno od rešenja se sastoji od *low-cost* Xilinx Virtex II FPGA čipa za koji je potrebno kupiti RapidIO IP *core* modul. Cena tih FPGA čipova je $150-200 [5], dok je cena *core* modula $15000 [6]. Drugo rešenje se sastoji od Xilinx FPGA čipa iz Virtex-4 ili Virtex-5 familije ($1000-2000 [5]), za koje se RapidIO *core* modul iz Xilinx Core Generator paketa može besplatno koristiti [6].
- HyperTransport tehnologija – jedno od rešenja se sastoji od *low-cost* Xilinx Virtex II FPGA čipa za koji je potrebno kupiti HyperTransport IP *core* modul. Cena IP *core* modula je $25000 [6]. Međutim, krajem 2007. godine se pojavio *open-source* HyperTransport *core* modul [7] za Xilinx Virtex-4 FPGA familiju ($1000-2000), za čije je korišćenje potrebno postati član *HyperTransport Technology Consortium* organizacije.
- PCI Express tehnologija – jedno rešenje se sastoji od Altera *low-cost* Cyclone II FPGA čipa ($150-250 [5]) i TI (Texas Instruments) XIO1100 ($10 [5]) ili NXP PX1011A PCI Express *transceiver* čipa ($10 [5]). Drugo, skuplje rešenje, dobija se korišćenjem Altera Stratix IV GX ili Stratix II GX FPGA čipova ($2000-3000 [5]) koji imaju integrisana *transceiver* kola [8]. Treće rešenje se dobija korišćenjem Xilinx FPGA čipa iz Virtex-4 ili Virtex-5 familije ($1000-2000 [5]), za koje se PCI Express *core* modul iz Xilinx Core Generator paketa može besplatno koristiti [6].

Cene svakog od rešenja su, za primenu na malom broju uređaja, dosta visoke. Najmanje se isplati koristiti skupe FPGA čipove, jer je na svaku ploču potrebno staviti po jedan od njih, a da pritom sama primena ne zahteva potpuno iskorišćenje njihovih performansi. Mnogo je isplativije kupiti IP *core* modul i implementirati ga na *low-cost* FPGA čipovima. Izuzetak, u smislu visoke cene realizacije, predstavlja PCI Express rešenje u kome se koriste Altera *low-cost* Cyclone II FPGA čip ($150-250) i TI XIO1100 ($10) ili NXP PX1011A PCI Express *transceiver* čip ($10).

Na osnovu izloženih karakteristika modernih *high-speed* tehnologija kao što su RapidIO, HyperTransport i PCI Express, može se zaključiti da svaka od njih zadovoljava zahteve specifikacije komunikacionog dela test sistema date u uvodu. Međutim, uzimajući cenu implementacije za glavni kriterijum izbora, PCI Express tehnologija se nameće kao najbolje rešenje.

LITERATURA

[1] RapidIO™ Interconnect Specification, RapidIO Trade Association, Available: www.rapidio.org
[2] HyperTransport™ I/O Link Specification Revision 2.0, Available: http://www.HyperTransport.org
[3] PCIe Base 2.0 Specification, Available: http://www.pcisig.com
[4] Moore's law, Available: www.wikipedia.org
[5] www.avnet.com
[6] www.xilinx.com
[7] David Slogsnat, Alexander Giese and Ulrich Brüning, "A Versatile, Low Latency HyperTransport Core", University of Mannheim, Available: http://www.ra.informatik.uni-mannheim.de
[8] www.altera.com

ABSTRACT

The most popular I/O technologies for high-speed board-to-board communication today are RapidIO, HyperTransport and PCI Express. This paper investigates which one of them is the best option for the peer-to-peer communication in the target system made of several printed circuit boards (PCBs) connected using mezzanine connectors. The most important characteristics of each technology are presented. Afterward are presented some hardware solutions for communication subsystem implementation. Finally, the most suitable solution is selected based on the specification of the system, the characteristics of each technology and the prices of presented hardware solutions.

ANALYSIS OF HIGH-SPEED I/O TECHNOLOGIES WHEN IMPLEMENTED IN THE LOW-COST PEER-TO-PEER COMMUNICATION SYSTEM

Mirjana Ž. Stojilović