

- PCI Express -

1. Úvod

Periferní zařízení počítačů zaznamenávají v posledních letech nebývalý rozvoj. Ten je hnán neustálým tlakem na zvyšování a zkvalitňování toku dat v uživatelských aplikacích. Dnešní aplikace nepožadují pouze vysokou přenosovou rychlost, ale také tzv. quality of service, což znamená, že data musí být v určitý čas přístupná pro zpracování. Příkladem může být typické trháni videa při přehrávání, kdy počítač má dostatek systémových prostředků, ale nemá přístup k právě požadovaným datům. Pro splnění těchto náročných požadavků se v roce 1999 začala formovat skupina firem IBM, HP a Compaq pod vedením firmy Intel. V té době byla vytvořena specifikace "3GIO (Third Generation Input and Output) ARCHITECTURE" tedy třetí generace vstup/výstupních zařízení. Dnes je specifikace známa pod obchodním názvem PCI Express. Specifikaci zastřešuje skupina PCI-SIG, tak jako všechny předcházející specifikace PCI a PCI-X.

2. Požadavky na architekturu 3GIO

Uživatelé požadují po moderním elektronickém průmyslu v podstatě dvě protichůdná řešení a to maximalizaci výkonu a minimalizaci ceny. To byl jeden z důvodů tvorby nové specifikace. Dále byla specifikace doplněna o nové požadavky, funkce a vlastnosti:

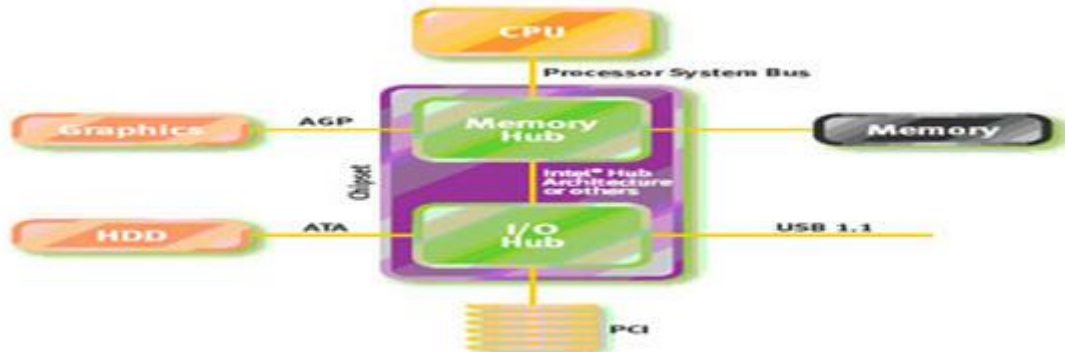
- Sjednocení přístupu pro více platform (desktop, mobilní zařízení, servery, pracovní stanice)
- Podpora různých druhů připojení (čip-čip, karta-karta přes konektor nebo kabel)
 - Nové mechanické uspořádání
 - Zachování současného softwarového modelu
- Detekce a konfigurace PCI Express zařízení pomocí současných mechanismů
 - Bootování existujících operačních systémů bez nutnosti modifikovat SW
 - Podpora existujících driverů zařízení
 - Podpora nových funkcí - vyžaduje úpravu konfiguračního mechanismu
 - Výkon
 - Nízká režie na vlastní komunikaci a rychlá odezva
- Vysoká přenosová rychlost na pin, z toho plynoucí omezení počtu pinů a nožů konektoru
 - Škálovatelná šířka přenosové trasy
 - Speciální vlastnosti
 - Podpora různých datových typů a pravidel pro přenosy dat
 - PowerManagement a úsporné režimy práce
 - Identifikovat podporované režimy pro dané zařízení
 - Přechod zařízení do daného stavu
 - Identifikace současného stavu zařízení
 - Generování wakeup sekvencí pro zdroje
 - Postupný náběh napájecích napětí
 - Podpora QoS (Quality of Service)
- Izochronní datový tok (je požadován nezávislý datový tok v závislosti na čase)
 - HotPlug a HotSwap
 - Zachování podpory současných zařízení typu HotPlug a HotSwap
 - Zavedení ryziho standardu HotPlug a HotSwap bez pomocných signálů
 - Jeden softwarový model pro všechny platformy
 - Datová integrita
 - Podpora detekce chyb na linkové vrstvě jak v datech tak i v paketech
 - Oprava chyb
 - Podpora současné verze PCI
 - Rozšířené generování chyb pro jejich lepší lokalizaci
 - Nezávislost na typu přenosového média a napěťových úrovních
 - Podpora jednoduchého testovacího mechanismu

3. Základní modely sběrnic PCI, PCI-X a PCI Express

3.1 Architektura PCI - PC do roku 2004

Sběrnice PCI Express se významně odlišuje od svých předchůdců, protože mění celkovou filozofii pohledu na sběrnice, tak, jak je známe do dnešních dnů. Obvyklá představa sběrnice vycházela z propojení několika zařízení sdílející stejnou skupinou vodičů

(paralelní nebo sériového). Některé typy sběrnic byly doplněny o vybrané vodiče sloužící pro dekódování daného zařízení, informaci o jeho stavu a synchronizaci hodinovým signálem. Pro lepší představu vyjděme ze sběrnice PCI nebo PCI-X, jak je známe z dnešních počítačů PC (obr. 1).



Tento typ sběrnice dovoluje paralelní řazení zařízení do slotu nebo přímé propojení mezi zařízeními na desce plošného spoje. Počet zařízení na sběrnici je omezen elektricky dovolenou zátěží sběrnice, logicky může být sběrnice rozšířena na dalších 256 segmentů oddělených bridgem. Za bridgem může následovat dalších 32 zařízení. Pracovní frekvence sběrnice PCI je obvykle pevná a to 33 nebo 66 MHz. Sběrnice typu PCI-X řídí pracovní frekvenci hodin podle počtu slotů nebo připojených zařízení.

Tabulka 1 udává přehled jednotlivých typů PCI sběrnic a počet potřebných pinů pro sestavení komunikace mezi zařízeními na sběrnici.

Typ sběrnice PCI - 33 PCI - 66 PCI-X 66 PCI-X 133 PCI-X 266 PCI-X 533

Počet datových bitů 32 64 32 64 32 64 32 64 16 32 64 16 32 64

Počet pinů 49 81 49 81 50 82 50 82 36 50 82 36 50 82

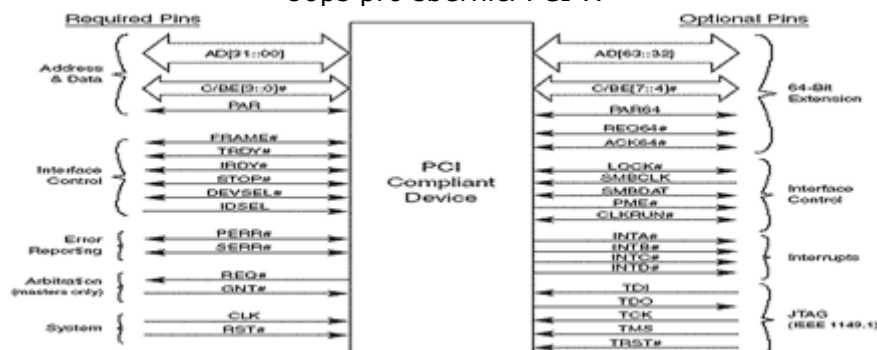
Přenosová rychlost MB/s 133 266 266 533 266 533 533 1066 533 1066 2133 1066 2133 4266

Napájecí napětí 5/3,3V 3,3V 3,3V 3,3V 1,5V a 3,3V 1,5V a 3,3V

Tabulka 1: Přehled konfigurace jednotlivých typů PCI sběrnic

Tento přístup klade vysoké nároky na kvalitu signálů, prostor na desce plošného spoje, mechanickou konstrukci konektorů a odolnost proti změnám parametrů sběrnice. Pokud vyjdeme z typického zapojení PCI nebo PCI-X slotu (obr. 2) je nutné zohlednit několik zásad při návrhu zapojení:

- Zatížení sběrnice se musí pohybovat v rozsahu 8-10pf na jedno zařízení, a to podle typu signálu
- Impedance vedení pro klasickou verzi PCI musí být v rozsahu 60-100Ω a pro PCI-X 57Ω ±10%
 - Délka skupin vedení je omezena v rozsahu viz. tab. 2
- Minimalizovat rozdíl délek mezi datovými a strobovacími signály, musí být menší než 80ps pro sběrnici PCI-X



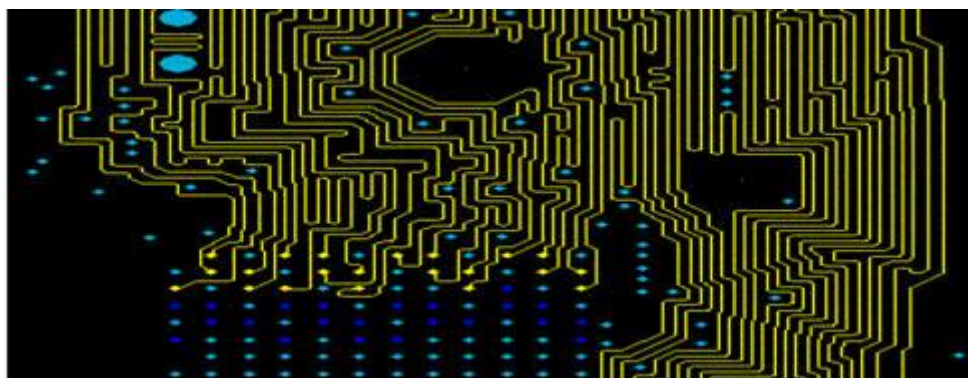
Dále jsou tyto parametry omezeny elektrickými vlastnostmi materiálů plošných spojů a konektorů, kde se používají obvykle materiály FR4. Pro klasické PCI stačí 4 vrstvy a pro PCI-X je doporučeno vrstev 6. Takovéto omezení není jednoduché dodržet, protože se

jedná o vodiče, které mají přesné umístění na nožích konektoru a jsou doplněny o napájecí a další pomocné signály.

Délka signálů PCI-X PCI Jednotka					
min.	max.	min.	max.		
CLK vedení	2,4	2,6	2,4	2,6	palce
CLK zpoždění	360	468	-	-	ps
AD[31:0] vedení	0.75	1.5	-	1.5	palce
AD[31:0] zpoždění	113	270	-	-	ps
AD[63:32] vedení	1.75	2.75	-	2.0	palce
AD[63:32] zpoždění	263	495	-	-	ps

Tabulka 2: Maximální délky vedení a zpoždění sběrnic typu PCI a PCI-X

Jako příklad uvedme část plošného spoje sběrnice PCI-X 533 viz obr. 3. Jak je vidět, konstrukce plošného spoje je poměrně komplikovaná a klade vysoké nároky na návrhový systém. Je nutno podotknout, že takto navrženou sběrnicí bude možné používat maximálně pro jeden slot. V případě potřeby dalších slotů pracujících na této frekvenci je nutné použít komplikovanější HOST-PCI bridge se dvěma oddělenými sběrnicemi PCI-X. Počet vodičů se zdvojnásobí a systém přechází na architekturu typu peer-to-peer.



Obrázek 3: Část plošného spoje sběrnice PCI-X 533

3.1.1 Rekapitulace

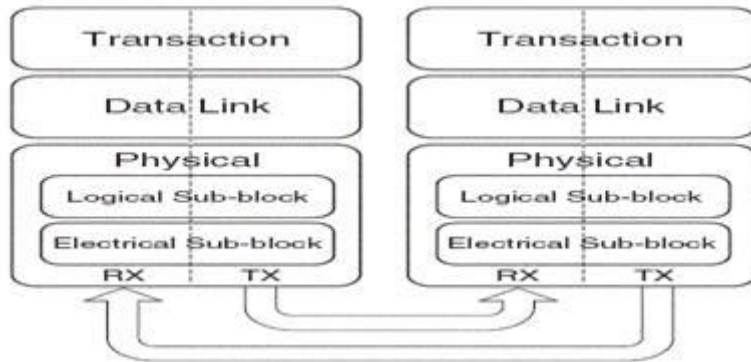
Z výše popsaného výčtu jsou zřejmé nevýhody dnešních PCI zařízení a pochopitelný tlak výrobců na změnu koncepce, hlavně v oblasti zvýšení přenosové rychlosti, zkvalitnění datového přenosu, úspory místa na plošných spojích, snížení elektromagnetického vyzařování (EMI), zjednodušení návrhu plošných spojů a sjednocení přístupu ke sběrnicím pro všechna zařízení.

Omezení architektury PCI:

- Velikost přenášeného bloku dat není omezená, díky tomu není možné jednoznačně definovat velikost bufferu, čímž může docházet k častému odpojování zařízení od sběrnice
 - PCI zařízení může vkládat velký počet wait-state
- Paměťově orientované operace vyžadují spolupráci s cache procesoru, aby byla zajištěna koherence dat - snížení výkonu CPU a PCI
 - Přenos dat je omezen pouze na transakce v délce 32/64 bitů
- Na sběrnici může probíhat pouze jedna transakce a to, jen jedním směrem
 - PCI master neví, jestli cílové zařízení může transakci přijmout - retry

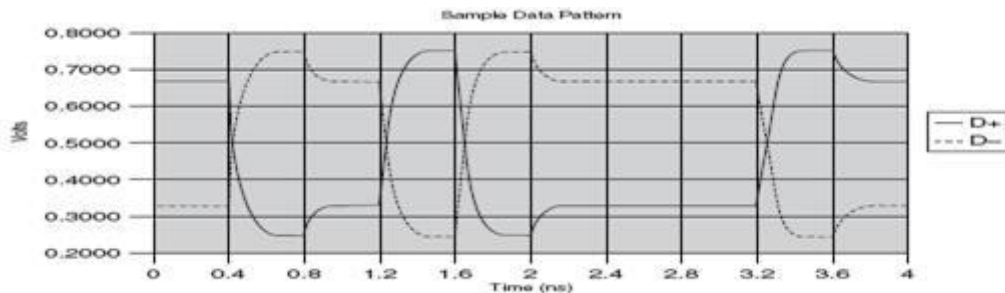
Sběrnice PCI Express

Model fyzické vrstvy sběrnice PCI Express vychází spíše ze sítě typu peer-to-peer než z architektury PCI nebo PCI-X. Jistá podobnost architektury PCI Express je i v dělení vrstev viz obr. 4 se síťovým modelem ISO-OSI. Architektura typu peer-to-peer umožňuje nezávislou komunikaci mezi jednotlivými zařízeními, kdy jedno zařízení nemusí čekat na uvolnění sběrnice při vzniku požadavku na komunikaci s jiným zařízením, jak tomu bylo u architektury PCI. Pochopitelně komunikace neprobíhá pouze jedním směrem, ale oběma - rozhraní je plně duplexní.



Obrázek 4: Vrstvy sběrnice PCI Express

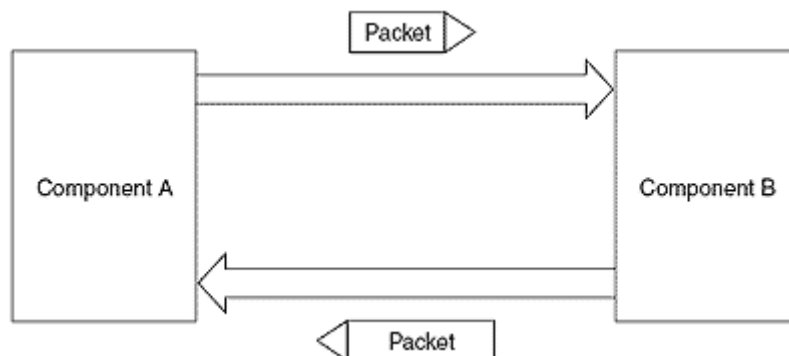
Další významnou změnou je způsob přenosu signálu po vedení. Sběrnice PCI využívaly k přenosu jeden vodič s dvoustavovou modulací (logická úroveň 1 nebo 0) naproti tomu sběrnice PCI Express využívá dva vodiče v diferenciálním zapojení. Modulace vodičů je obvykle vícecestavová - používají se obvody preemfáze sloužící k úpravě signálů na vedení (obr. 5). Pozn. tento typ víceúrovňové modulace neslouží k zvýšení přenosové rychlosti, ale ke korekci ztrát vznikajících na vedení při vysokých frekvencích (skinefekt, vyzářování).



Obrázek 5: Průběh signálu na Lanu sběrnice PCI Express

3.2.1 PCI Express Link

PCI Express Link reprezentuje komunikační kanál mezi dvěma zařízeními sběrnice PCI Express (obr. 6). Základní PCI Express Link je sestaven ze dvou nízkonapěťových diferenciálních párů a to přijímacího a vysílacího komunikačního páru označovaného jako Lane. Činnost vysílače i přijímače je na sobě nezávislá a Link tvoří plně duplexní komunikační kanál.



Obrázek 6: PCI Express Link

Základní vlastnosti komunikačního kanálu Link:

- Základní link se skládá ze dvou jednosměrných diferenciálních párů v každém směru, reprezentující přijímací a vysílací pár. Hodinový signál je kódovaný do datového toku, aby mohlo být dosaženo maximální přenosové rychlosti. Samostatně vedené hodiny a data na vysokých frekvencích jsou náchylné k fázovému posunu a jitteru.
- Každý link může pracovat s příslušnými signálovými úrovněmi pro které byl navržen. Přenosová rychlost dle současné specifikace dosahuje 2,5Gbitu/s na jeden Lane v jednom směru. Zvýšení pracovní frekvence se předpokládá v dalších verzích specifikace.

- Každý Link musí podporovat alespoň jeden Lane. Pro zvýšení přenosové rychlosti je možné využít sdružování Lanes do Linků v povolené šířce. Obvykle se jedná o hodnoty x1, x2, x4, x8, x12, x16 a x32. Stejná šířka musí být dodržena jak pro přijímací, tak vysílací část.

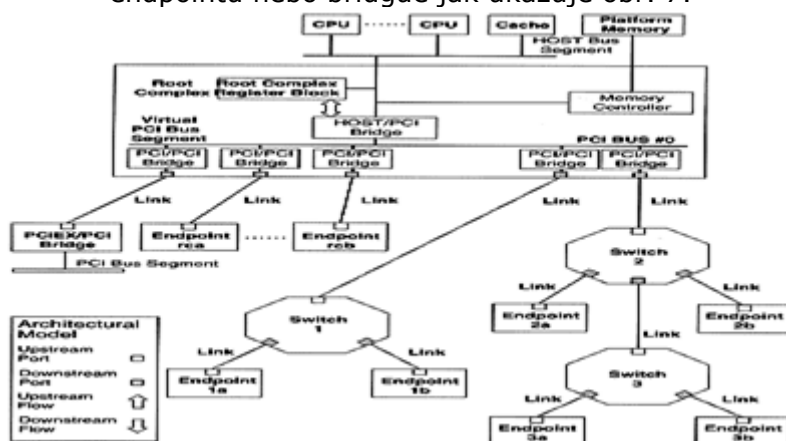
- Během hardwarové inicializace Linku se vyjedná pracovní frekvence a počet Lanes sestavujících Link. Obdoba vyjednávání pracovní frekvence sítí typu Ethernet.

3.2.2 Zařízení sběrnice PCI Express

Podobně jako sběrnice PCI, tak i sběrnice PCI Express je sestavena ze zařízení, která jsou vzájemně propojena a zajišťují nezbytné funkce sběrnice. Jedná se o zařízení: root complex, switches, endpoints a bridges. Jednotlivá zařízení a jejich funkce budou popsány v následujícím textu. Jednotlivé typy zařízení je možné propojovat a kombinovat libovolně při dodržení 4 základních pravidel:

- Na sběrnici může existovat pouze jedno zařízení typu root complex, které zajišťuje propojení mezi PCI Express zařízením, hostitelským systémem a pamětí.
- Každý switch má pouze jeden upstream port, počet downstream portů je omezen na 256.
- Každý endpoint a bridge má pouze jeden upstream port.
- Každý link může propojovat pouze jeden upstream a downstream port. Vyjimku tvoří propojení typu Advanced Peer-to-peer Link (nebude popisováno - využití v multiprocessorových systémech, nebo při propojení více PCI Express sběrnic s několikeroot complexy).

Tyto požadavky neomezují vzájemné propojení jednotlivých zařízení na sběrnici PCI Express. Například na root complex zařízení může být připojeno několik switchu, endpointů nebo bridgů jak ukazuje obr. 7.



Obrázek 7: Architektura sběrnice PCI Express - Root Complex

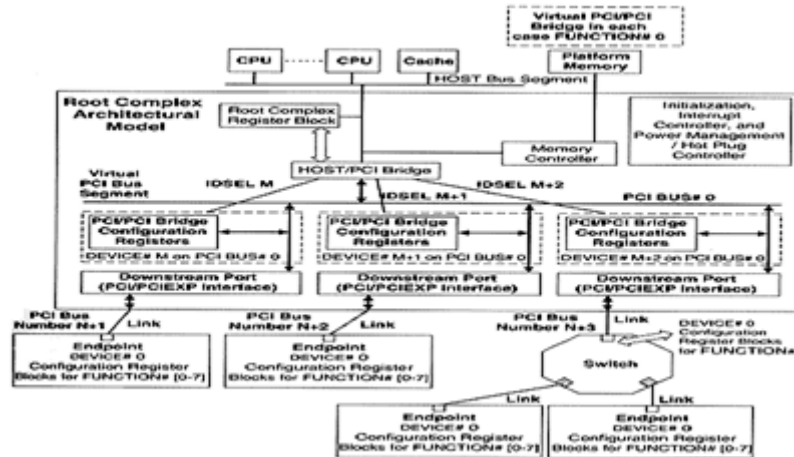
Je obdobou HOST-PCI Bridge u sběrnice PCI, který vytvářel propojení mezi hostitelským procesorem a systémovou pamětí. V současnosti zastává Root Complex následující funkce:

- Propojení sběrnice hostitelského procesoru se systémovou pamětí (řadič paměti)
 - Propojení sběrnice hostitelského procesoru se sběrnicí PCI Express přes downstream porty a linky
 - Propojení systémové paměti se sběrnicí PCI Express přes downstream porty a linky
- Root complex dále zajišťuje podporu PCI Express a systémových prostředků. Systémové prostředky jsou konfigurovatelné přes I/O prostor hostitelské sběrnice. Obvykle se jedná o řadič přerušení, Power Management a řadič Hot Plug zařízení.

Na obr. 8 je základní model root complexu. Jak je vidět z obrázku, veškerá zařízení sběrnice PCI Express jsou na straně downstreamu (tedy "pod Root Complexem") a zařízení hostitelského procesoru a paměti na straně upstreamu. Jak již bylo zmíněno výše, zařízení musejí být kompatibilní s modelem směrnice PCI, což se odráží i v modelu root complexu - vytváří se virtuální HOST-PCI bridge a PCI segmenty. Klíčové jsou především tyto části:

- Virtuální HOST-PCI Bridge, propojující systémovou část a vlastní sběrnici PCI Express

- Blok root complex registrů (RCRB) zajišťující konfiguraci.
- Virtuální segment sběrnice PCI, zajišťující propojení mezi virtuálními bridgi HOST-PCI a PCI-PCI. Dále jsou generovány signály IDSEL obdobně jako u sběrnice PCI, které zajišťují přístup ke konfiguračním registrům.
- Virtuální PCI-PCI bridge zajišťuje větvení transakcí přicházejících z hostitelského procesoru na sběrnici PCI Express nebo dat přicházejících ze sběrnice do hlavní paměti. Root complex se může chovat také jako virtuální switch, kdy dochází k propojení dvou zařízení (Linků) přímo na fyzické úrovni. Příkladem může být propojené zařízení rca a rcb z obr. 8 Tato funkce nemusí být implementovaná.



Obrázek 8: Architektura Root Complexu
- Switch

Switch zajišťuje větvení a rozšiřování sběrnice PCI Express od downstream portu root complexu, nebo jednoho switche k dalším PCI Express zařízením (endpoints, switche a bridge). Primárně switch zajišťuje přenos paketů mezi upstream a downstream portem. Architektura zachovává členění do PCI segmentů a virtuálních sběrnic (obr. 9). Každý switch má následující vlastnosti:

- Je sestaven nejméně ze dvou nebo více virtuálních PCI-PCI bridgu.
- Používá PCI Bridge mechanismus pro přenos dat založený na cílové adrese.
 - Musí podporovat přenos všech typů transakcí (TLP)
- Nesmí generovat pakety s menší délkou, než do něho vstoupily. Pokud přijde paket o délce 512 bytů, nemůže být rozdělen na dva o délce 256 bytů.
- Pokud vznikne současně několik požadavků na jeden virtuální kanál, budou rozděleny podle priority (kanál s nejvyšší prioritou zvítězí), nebo se využije rotující priorita.

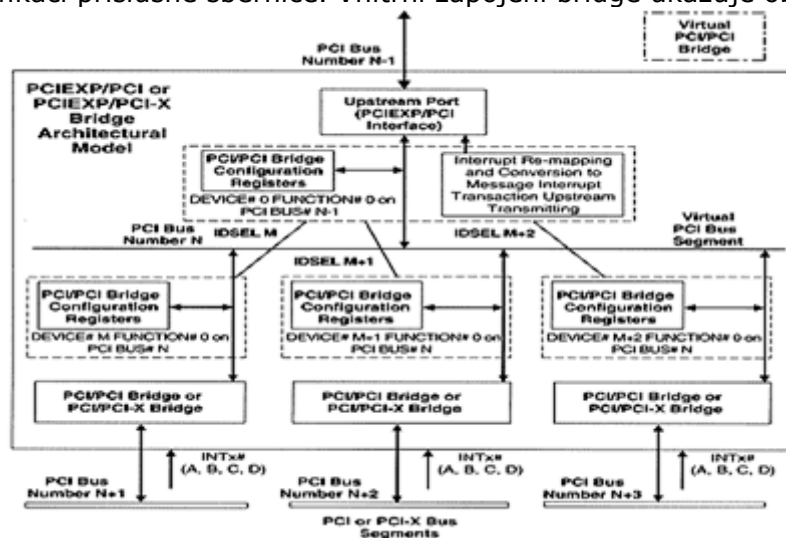


Obrázek 9: Architektura switche

- Bridge

Bridge zajišťuje převod mezi PCI Express a jiným typem sběrnice (PCI, PCI-X, nebo jiným). V podstatě se dá říct, že bridge se chová podobně jako switch, ale musí být

schopen přenášet specifické transakce jiného typu sběrnice a pochopitelně podporovat specifikaci příslušné sběrnice. Vnitřní zapojení bridge ukazuje obr. 10.



Obrázek 10: Architektura bridge
- Endpoint

Pod pojmem Endpoint rozumíme koncové zařízení připojené ke sběrnici PCI Express, které tvoří downstream Root Complexu nebo Switchce. Endpoint obsahuje jedno zařízení s jednou až osmi funkcemi (ekvivalent ke specifikaci PCI). Endpoints můžeme rozdělit do dvou skupin a to Legacy Endpoints a PCI Express Endpoints.

- Legacy Endpoint podporuje následující typy operací:
 - Musí podporovat konfigurační mechanismus typu 0 (ekvivalent s PCI - typ jedna se využívá pro bridge - po dosažení cílového bridge je konvertován na typ 0)
 - Může podporovat žádosti typů I/O
 - Může generovat požadavek na I/O operaci
 - Může podporovat operace typu LOCK (specifikace nedoporučuje jejich používání)
 - Nesmí generovat požadavek na operace typu LOCK
 - Může implementovat rozšířený konfigurační prostor, avšak ten nemusí být využit softwarem (zpětná kompatibilita s PCI)
 - Při vzniku požadavku na transakci do hlavní paměti nemusí generovat adresu větší než 4Gbyty (32bitů)
 - Generování přerušení může podporovat jak 32-bitovou verzi, tak 64-bitovou verzi MSI
 - Má dovoleno používat 32bitové adresování pro báze adresy vyžadující paměťový přístup
 - PCI Express Endpoint
 - Musí podporovat konfigurační mechanismus typu 0
 - Nesmí generovat požadavky na operace typu I/O
 - Nepodporují a negenerují operace typu LOCK
 - Při vzniku požadavku na transakci do hlavní paměti musí podporovat generování adresy větší než 4Gbyty (32bitů)
 - Generování přerušení musí podporovat 64 bitovou verzi MSI4
 - Pro všechny báze adresy, které mají nastavený prefetch bit, musí být podporováno 64-bitové adresování. Adresy nepoužívající prefetch bit, mohou podporovat jen 32-bitovou verzi adresování
 - Minimální velikost požadovaná, pro přidělení báze adresy je 128 bytů
- Z výše zmíněného výčtu a názvu je patrné, že legacy endpoints jsou zařízení, zachovávající zpětnou kompatibilitu se standardem PCI. Především to platí pro SW model a zařízení za bridgem typu PCI Express - PCI.

Vrstvy sběrnice PCI Express

3.2.3 Vrstvy sběrnice PCI Express

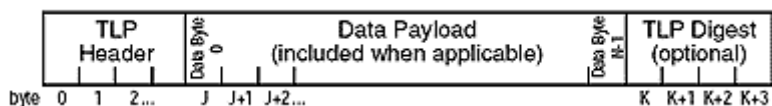
V předcházející části jsme se seznámili se zařízeními, které tvoří základní bloky sběrnice. Jednotlivá zařízení jsou propojena tzv. Linkem, který zajišťuje přenos dat pomocí paketů

mezi jednotlivými komponenty a tvoří část přenosové trasy. Dále je přenosová trasa připojena k vysílači a přijímači fyzické vrstvy. Na tuto vrstvu navazuje linková a transakční vrstva, ve kterých je formován paket a doplňován o potřebné informace. Tyto tři vrstvy tvoří základní komunikační model sběrnice PCI Express definované specifikací 1.a.

Transakční vrstva je obvykle propojena ještě s vrstvou (core logic), která realizuje interface mezi hostitelským procesorem a sběrnicí PCI Express. Tato vrstva není specifikována standardem PCI Express a záleží na konkrétní implementaci použitého procesoru nebo systému.

3.2.4 Transakční vrstva

Nejvyšší vrstvou architektury je transakční vrstva. Tato vrstva je zodpovědná za zpracování (kompozici a dekompozici) paketů transakční vrstvy (Transaction Layer Packet - TLP) obr. 11. Tyto pakety nesou informaci o typu prováděné operace, jako je čtení, zápis, zpráva nebo operace s IO prostorem. Všechny pakety, které vyžadují potvrzení jsou implementovány jako dvě transakce (request/completion) a jsou přijímány nebo posílány do vrstvy core logic. Každý vysílaný paket má svoje identifikační číslo, aby cílové zařízení mohlo poslat completion transakci zdroji. Dále vrstva zajišťuje nastavování atributů, posílání zpráv, které v podstatě realizují virtuální vodiče přerušování, řízení napájení, jak jsme je znali z jiných typů sběrnic.



Obrázek 11: Paket transakční vrstvy

Jak je vidět z obr. 11, paket transakční vrstvy se skládá z hlavičky (TLP Header - obr. 12), vlastních dat, pokud jsou požadována pro příslušný typ operace a nepovinným kontrolním součtem ECRC.

Hlavička se skládá z řídicích informací o typu přenosu. Délka hlavičky může být v rozsahu 3-4 doublewordy podle Fmt (DW = 4 byty) viz. tab. 3. Význam jednotlivých atributů hlavičky je popsán v následujícím textu.

- Fmt udává formát paketu TLP a zda paket obsahuje data:

Fmt[1:0] Formát TLP
 00b 3 DW, bez dat
 01b 4 DW, bez dat
 10b 3 DW a data
 11b 4 DW a data

Tabulka 3: Formát hlavičky paketu TLP

- Type udává typ paketu TLP, respektive typ operace, co se bude provádět. Je vázán také na Fmt. Příslušné kombinace Fmt a typu operace jsou uvedeny v tab. 4.

- TC udává Traffic Class

- TD povoluje připojení ECRC (TLP Digest) ke konci paketu

- EP nese informaci o chybě vzniklé v datech, například při chybě parity při čtení ze sběrnice PCI. Tato chyba nemusí být nutně kritická. Jedná se o informaci pro nadřazený systém, který se s chybou vypořádá. Tato chyba se nesmí vyskytnout v žádné jiné části než datové! Proto se tomuto mechanismu také říká data poisoning.

- Attr atributy udávající informace pro přenos dat a koherenci cache

- Length jde o délku přenášených dat v DW uvnitř paketu. Tedy maximální délka může být $2^{10} \text{ DW} = 4096 \text{ bytů}$

Obrázek 12: Formát hlavičky paketu sběrnice PCI Express

Typ TLP
 Fmt
 Type
 Popis
 MRd
 00/01

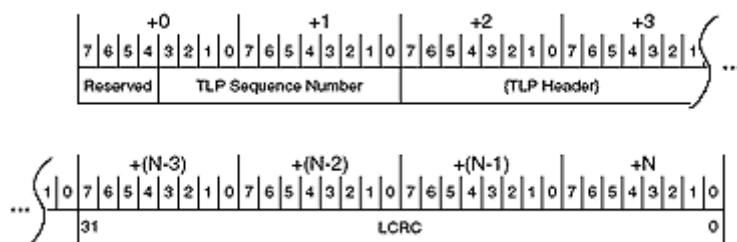
0 0000
Žádost o čtení z paměti
MRdLk
00/01
0 0001
Žádost o čtení z paměti s požadavkem na Lock
MWr
10/11
0 0000
Žádost o zápis do paměti
IORd
00
0 0010
Žádost o I/O čtení
IOWd
10
0 0010
Žádost o I/O zápis
CfgRd0
00
0 0100
Konfigurační čtení typu 0
CfgWr0
10
0 0100
Konfigurační zápis typu 0
CfgRd1
00
0 0101
Konfigurační čtení typu 1
CfgWr1
10
0 0101
Konfigurační zápis typu 1
Msg
01
1 0rrr
Žádost o zprávu - rrr viz tab. 5
MsgD
11
1 0rrr
Žádost o zprávu - rrr viz tab. 5
Cpl
00
0 1010
Dokončení transakce bez dat
CplD
10
0 1010
Dokončení transakce s daty
CplLk
00
0 1011
Dokončení transakce bez dat s Lockem
CplDLk
10
0 1011

Dokončení transakce s daty a s Lockem
 Tabulka 4: Jednotlivé atributy hlavičky paketu TLP

r[2:0]	
Popis	
000	Zpráva je směřována na root complex
001	Směřováno podle adresy
010	Směřováno podle ID
011	Vysíláno root complexem
100	Lokální - ignorováno v přijímači
101	Zpráva je odeslána do RC switchem až po příjmu zprávy od všech zařízení
110 a 111	Nepoužívají se a jsou ignorovány

Tabulka 5: Typy směrování zpráv
 3.2.5 Linková vrstva

Linková vrstva je vložena mezi transakční a fyzickou vrstvu. Jejím úkolem je zajišťování integrity dat - detekce a oprava chyb. Data přijatá z transakční vrstvy jsou opatřena kontrolním kódem, identifikačním číslem a poslána do fyzické vrstvy viz. obr. 13. Naopak data přijatá z fyzické vrstvy jsou otestována, zda neobsahují nějakou chybu a jsou poslána do transakční vrstvy. V případě výskytu chyby, vrstva zajišťuje opakovaný požadavek na data, dokud nejsou požadovaná data přítomná, nebo dokud není linka prohlášena za nefunkční.



Obrázek 13: Paket linkové vrstvy
 3.2.6 Fyzická vrstva

Fyzická vrstva zajišťuje veškeré obvody nutné pro připojení k linku. Jsou to fázové závěsy, buffery, sério-paralelní a paralelně sériové převodníky, impedanční přizpůsobení a v neposlední řadě i logiku pro inicializaci a udržování spojení na linku (vyjednání přenosové rychlosti, formátu přenosu dat). Tato vrstva je zodpovědná za komunikaci mezi linkem a linkovou vrstvou. Paket přijatý z linkové vrstvy LLTP je doplněn o kódy začátku konce paketu podobně, jako je tomu u síťových paketů (Ethernet). Dále je paket zakódován kódem 8 na 10. Tím jsou do paketu doplněny další informace zajišťující synchronizaci. Potom je paket převeden na sériový kód a odvíšlán do příslušného Lanu. Příjímáčí část fyzické vrstvy postupuje opačným způsobem. Dekóduje přijatý paket na řídicí kódy, data a rekonstruuje hodiny. Pokud je rámec paketu v pořádku a odpovídá kontrolní součet, je odeslán do linkové vrstvy, dále je také odesláno potvrzení o přijetí dat zdroji transakce, který vyprázdňuje retry buffer. Pokud přijdou data s chybou, odesílá se do zdrojového portu požadavek na opakování transakce. Je nutno podotknout, že pokud dojde v budoucnu ke změně formátu kódování (64bitů na 66bitů) nebo rychlosti, je třeba změnit pouze fyzickou vrstvu a není nutné měnit žádnou jinou vrstvu modelu PCI Express.

3.2.7 Virtuální kanály - VC a Traffic Class - TC

Každý PCI Express port může být rozdělen až do 8mi virtuálních kanálů, identifikovaných

číslem kanálu. Jak napovídá název, virtuální kanál není fyzickým kanálem, ale mapuje se na fyzické kanály v časovém multiplexu a podle pravidel určených arbitrážní logikou kanálů. Traffic class zajišťuje relativní prioritu mezi jednotlivými přenosy. Pomocí mapování TC na VC je možné zajistit deterministický a izochronní transfer dat tzv. Quality of Service.

Využití virtuálních kanálů a traffic classes umožňuje využít přenosovou trasu z hlediska přenosu velkých objemů dat (paketů). Horší situace nastane při vzniku požadavku na přenos série menších paketů. V takovémto případě je přenosová rychlost ovlivněna délkou paketu, kdy je k aktivním datům připojena řídicí informace a to 5 DW hlavičky + 2 DW DLLP (Data Link Layer Packet). Dále požadavek na čtení je ovlivněn latencí (doba od poslání požadavku na čtení až po příjem dat). U zápisu není latence obvykle kritická, protože se tvoří přirozený pipe-line při průchodu dat směrem od zdroje k cíli. Posledním faktorem, který ovlivňuje rychlost čtení, je nastavení maximální délky požadavku na čtení v registru RCB. Nastavení se může pohybovat v rozsahu 128, 256, 512, 1024, 2048 a 4096 bytů. Přehled využití linky nám ukáže tab. 6.

Délka datové části paketu
 Využití sběrnice PCI Express v[%]

byty

Zápis

Čtení RCB (256B)

Čtení RCB(128B)

4

16

12

12

8

28

22

22

32

62

53

53

64

76

69

69

128

86

82

72

256

92

90

74

512

96

94

75

1024

98

97

75

2048

99

98
75
4096
99
98
76

Tabulka 6: Využití sběrnice PCI Express pro různé délky přenosu a nastavení RCB

3.3 Signály sběrnice PCI Express a mechanické uspořádání

Jak již bylo zmíněno v předešlých kapitolách, tak sběrnice PCI Express pracuje na jiných principech než dosud používané sběrnice v počítačích PC. Základem je Line sestavený ze dvou diferenciálních párů a to přijímacího a vysílacího. Vlastní sběrnice je doplněna o další pomocné a napájecí signály. Přehled signálů pro konektor typu x1 a jejich popis je uveden v tab. 7.

Číslo pinu	Název
Stran A	Popis
Stran B	Název
	Popis
1	+12V
	Napájení +12V
	PRSENT1#
	Detekce přítomnosti karty Hot-Plug
2	+12V
	Napájení +12V
	+12V
	Napájení +12V
3	RSVD
	Vyhrazeno pro budoucí použití
	+12V
	Napájení +12V
4	GND
	Zem
	GND
	Zem
5	SMCLK
	System Management Bus Clock
	JTAG2
	TCK, hodinový signál pro JTAG
6	SMDAT
	System Management Bus Data
	JTAG3
	TDI, vstupní data pro JTAG
7	GND
	Zem
	JTAG4
	TDO, výstupní data od JTAGu

8
 +3,3V
 Napájení +3,3V
 JTAG5
 TMS, výběr modu pro JTAG
 9
 JTAG1
 TRST, reset JTAGu
 +3,3V
 Napájení +3,3V
 10
 3,3Vaux
 Pomocné napájení 3,3V - pro wake-up
 +3,3V
 Napájení +3,3V
 11
 WAKE#
 Signál pro reaktivaci zařízení z PWD
 PERST#
 Reset
 12
 RSVD
 Vyhrazeno pro budoucí použití
 GND
 Zem
 13
 GND
 Zem
 REFCLK-
 Referenční hodinový signál, negativní
 14
 PETp0
 Vysílač - Lane 0 pozitivní
 REFCLK+
 Referenční hodinový signál, pozitivní
 15
 PETn0
 Vysílač - Lane 0 negativní
 GND
 Zem
 16
 GND
 Zem
 PERp0
 Přijímač - Lane 0 pozitivní
 17
 PRSNT2#
 Detekce přítomnosti karty Hot-Plug
 PERn0
 Přijímač - Lane 0 negativní
 18
 GND
 Zem
 GND
 Zem

Tabulka 7: Popis konektoru PCI Express

Konektor pro verze x2, x4, x8 a x16 je rozšířen o další nože konektoru - obvykle o čtyři pozice, které obsahují GND a přijímací a vysílací pár. Je nutno podotknout, že implementace některých signálů není nutná a je pouze doporučena. Detailní popis signálů bude rozebrán v následujícím odstavci, nebo v literatuře [3].

- REFCLK (nutno implementovat)

Referenční hodiny - obdoba signálu CLK jak ho známe ze sběrnice PCI. Vlastní hodinový rozvod se skládá z diferenciálních vodičů REFCLK+ a REFCLK- s rozsahem napěťových úrovní 0 - 0,7V. Pracovní kmitočet je stanoven na 100MHz \pm 300ppm. Zásuvná karta nemusí využívat referenční hodiny z konektoru, ale musí udržovat datový tok v rozsahu 600ppm. Referenční hodiny mohou využívat hodiny s rozprostřeným spektrem.

- PERST (nutno implementovat)

Signál inicializující kartu po zapnutí napájení, nebo může sloužit jako tzv. warm reset.

- WAKE# (nepovinné)

Signál pro reaktivaci zařízení z power down modu, otevřený kolektor. Tato funkce vyžaduje přítomnost napájecího napětí 3,3Vaux. Po detekci signálu WAKE# musí systém zapojit hlavní napájení slotu a obnovit referenční hodiny.

- SMBus System Management Bus (nepovinné)

Sběrnice určená ke komunikaci mezi zařízeními, může zajišťovat doplňkovou zprávu power managementu, identifikovat verze firmwaru a hardwaru. Sběrnice se skládá ze dvou signálů SMCLK a SMDAT a odpovídá standardu I2C.

Napájecí napětí

1x

4x/8x

16x

+3,3V

Tolerance napětí

9%

9%

9%

Proud

3.0A (max)

3.0A (max)

3.0A (max)

Zatížení (kapacita)

1000uF

1000uF

1000uF

+12V

Tolerance napětí

8%

8%

8%

Proud

0,5A(max)

2,1A (max)

4,4A (max)

Zatížení (kapacita)

300uF

1000uF

2000uF

+3,3Vaux

Tolerance napětí

9%

9%

9%

Proud

Povoleny waku-up

375mA (max)

375mA (max)

375mA (max)

Nepovoleny wake-up

20mA (max)

20mA (max)

20mA (max)

Zatížení (kapacita)

150uF

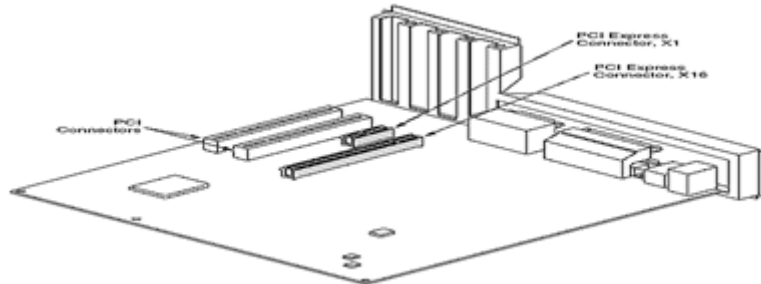
150uF

150uF

Tabulka 8: Maximální proudové zatížení konektoru sběrnice PCI Express

3.3.1 Mechanické uspořádání

Mechanické uspořádání ukazuje obr. 14. Jak je vidět PCI Express konektor s šířkou X16 se využije pro grafickou kartu a nahradí stávající AGP konektor. Další konektory budou pro běžné PC X1 a pro servery X4 nebo X8. Po dobu několika let se bude ještě udržovat standard PCI, který bude postupně nahrazen sběrnicí PCI Express (obdobný přechod se odehrával u sběrnice ISA, která byla postupně vytlačena PCI konektory)



Obrázek 14: Mechanické uspořádání konektorů na základní desce PC v roce 2004
Rozměry přídatných karet by měly splňovat některá omezení, co se týče maximálních rozměrů dle tab. 9. Karty připojované pomocí kabelů mohou mít jiné rozměry a uspořádání.

Šířka linku

Typ karty

Výška karty

Šířka karty

x1

Standardní karty poloviční délky

111,15 mm (max)

167,65 mm (max)

x1,x4,x8,x16

Standardní karty plné délky

111,15 mm (max)

312 mm (max)

x1,x4,x8,x16

Karty s nízkým profilem

68,9 mm (max)

167,65 mm (max)

Tabulka 9: Rozměry přídatných karet sběrnice PCI Express

4 Souhrn

Využití sběrnice PCI Express přináší zvýšení propustnosti dat v oblasti počítačů PC, zmenšení rozměrů karet, zjednodušení návrhu plošných spojů a sjednocení různých platforem, jako jsou síťové karty, grafické karty a komponenty přenosných počítačů. Dále specifikace umožňuje připojovat zařízení pomocí kabelů, což vede k vysoké univerzálnosti standardu (nepředpokládá se nahrazení rozhraní SerialATA).

Jako příklad porovnávající výkonnost sběrnice PCI a PCI Express uvedme síťovou kartu pracující na 1Gbitu/s, kdy máme k dispozici standardní počítač, který má sběrnici PCI v šířce 32 bitů pracující na frekvenci 33MHz. Tedy maximální datový tok, který může vzniknout, je 100Mbytu pro příjem a stejný objem pro vysílání, což není standardní sběrnice PCI schopna přenést (max. propustnost je 133Mbytu/s v jednom směru) a potřebuje 124 nožů konektoru. Sběrnice PCI Express x1 potřebuje 36 nožů konektoru a maximální propustnost je 2,5Gbitů/s v obou směrech (při použití kódování 8/10 asi 250Mbytu surových dat za sekundu). Navíc propojení je typů peer-to-peer, takže datový tok není omezován žádným dalším zařízením na sběrnici. K saturaci může docházet až ve switchi nebo root complexu při přístupu do paměti. Připojení 4 takových síťových karet do systému je bez problému možné, co se týče sběrnice a propustnosti dat mezi pamětí a root complexem. Jako další příklad uvedme grafickou kartu pracující na sběrnici AGP. V současné době je maximální propustnost sběrnice AGP 2Gbyty/s jedním směrem. Při použití sběrnice PCI Express x16 bude propustnost směrem do karty dvojnásobná (viz. tab. 10) a směrem z karty bude možné využít další 4Gbyty/s. Tedy celková propustnost dat mezi grafickou kartou a hlavní pamětí činí 8Gbytu/s. Při využití plné šířky kanálu nebude v současnosti vznikat úzké hrdlo na sběrnici, ale spíše v paměťových modulech.

Zvýšení pracovní frekvence sběrnice ze současných 2,5Gbitu/s na 5Gbitů/s (3,125Gbitů/s) a později na 10Gbitů/s povede k lineárnímu nárůstu výkonu s frekvencí. Nasazení sběrnice s vyšší frekvencí zatím brání relativně vysoká cena nových technologií pracujících na frekvencích do 10Gbitů/s, nedostatečná propustnost dat mezi hlavní pamětí a sběrnici a v neposlední řadě je nutné ověřit, funkčnost a vlastnosti masového nasazení technologie PCI Express.

Počet lanu v linku

x1

x2

x4

x8

x12

x16

x32

Počet vodičů pro jeden směr

2

4

8

16

24

32

64

Přenosová rychlost v Gbitech/s

2,5

5

10

20

30

40

80

Přenosová rychlost v Mbytech/s

250
500
1000
2000
3000
4000
8000

Tabulka 10: Přehled přenosových rychlostí sběrnice PCI Express v závislosti na počtu lanu v linku

Odvrácenou stránkou technologie PCI Express je relativní složitost vlastního protokolu, znemožňující jednoduchý, levný a rychlý návrh řadiče do hradlového pole. Navíc v současné době (rok 2004) neexistují běžně dostupné komponenty převádějící protokol PCI Express na uživatelsky "přívětivou" sběrnici, jak tomu bývalo dřív u výrobců PCI matchmakeru, které převáděly protokol PCI do paměťově mapovaných prostorů doplněných o sadu uživatelských registrů, mailboxu, doorbellu a DMA kanálů. Toto zatím znemožňuje návrh zařízení s touto technologií pro malosériovou výrobu v počtu stovek kusů.

Jako nejschůdnější řešení se jeví využití hradlových polí se sériovým interfacem schopným zpracovávat datový tok fyzické vrstvy kompatibilní s formátem PCI Express a implementovat zbývající vrstvy včetně logic core. Vhodná hradlová pole pro tento účel jsou Virtex II Pro, nebo Virtex II Pro X od firmy Xilinx a Stratix GX od firmy Altera.