

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

PROVOZNĚ EKONOMICKÁ FAKULTA

KATEDRA INFORMAČNÍCH TECHNOLOGIÍ



GRAFICKÉ KARTY NA PC

Bakalářská práce

Praha 2007 ©

Vedoucí práce: Ing. Jiří Vaněk, Ph.D.

Autor práce: Antonín Kříž

Zadání práce

Prohlášení

Prohlašuji, že jsem bakalářskou práci na téma „*Grafické karty na PC*“ vypracoval samostatně, pouze za odborného vedení vedoucího bakalářské práce.

Dále prohlašuji, že veškeré podklady, ze kterých jsem čerpal, jsou uvedeny v seznamu použité literatury.

V Praze, dne 24.4.2007

Podpis:

Poděkování

Rád bych tímto vyjádřil svůj dík vedoucímu bakalářské práce Ing. Jiřímu Vaňkovi, Ph.D. za jeho cenné připomínky, odborné vedení, trpělivost a ochotu při vedení mé bakalářské práce.

Název

Grafické karty pro PC

Souhrn

Práce se zabývá grafickými kartami pro osobní počítače. Popisuje jejich vývoj, architekturu a jednotlivé parametry. Zároveň práce rozděluje karty do tříd a srovnává jejich výkon, který posléze vyhodnocuje.

Klíčová slova

Grafická karta, ATI, nVidia, shader, DirectX, sběrnice

Title

Graphic cards for Personal Computers

Summary

The aim of my BA thesis is to gain a new piece of knowledge in the sphere of graphic cards for personal computers. It describes their advancement, architecture and individual variables. The thesis also divides cards to classes, compares their performance and makes conclusions.

Key words

Graphic card, ATI, nVidia, shader, DirectX, bus

Obsah

1. Úvod	5
2. Cíl práce a metodika	7
3. Historie vývoje grafických karet	8
3.1 Obecný náhled	8
3.1.1 Koncepce osobního počítače IBM PC	9
3.1.1.1 Sběrnice a porty na osobních počítačích	9
3.2 Grafické karty pro IBM PC	13
3.2.1 Čip Motorola MC6845	13
3.2.2 Grafická karta CGA	13
3.2.3 Grafická karta Hercules.....	13
3.2.4 Grafická karta EGA	14
3.2.5 PGC – první grafický akcelerátor pro PC	14
3.2.6 Grafická karta VGA	14
3.2.7 SVGA.....	15
3.2.8 Standard VESA	15
3.3 Grafické akcelerátory	15
3.3.1 Moderní grafická rozhraní:	16
3.3.1.1 OpenGL™.....	16
3.3.1.2 DirectX.....	16
3.3.2 Chronologický vývoj grafických akcelerátorů	16
3.3.3 1995 první grafický akcelerátor	16
3.3.3.1 3Dfx	16
3.3.4 1997.....	17
3.3.4.1 nVidia vystrkuje růžky.....	17
3.3.5 1998.....	17
3.3.6 1999.....	18
3.3.7 2000.....	18
3.3.8 2001 – DirectX 8.0 nastupuje	18
3.3.9 2002 – DirectX 9.0.....	19
3.3.10 2004.....	20
3.3.11 2005, 2006.....	20
4. Hodnocení současného stavu	21
4.1 Architektura grafických karet současnosti	21
4.1.1 Nejnovější čip – nová architektura	21

4.1.1.1	Unifikované shadery.....	21
4.1.1.2	Struktura GPU	22
4.1.1.3	Vyhlazování a filtrování obrazu	24
4.1.2	Parametry grafických karet – pojmy	27
4.1.2.1	Počet tranzistorů.....	28
4.1.2.2	Počet shaderů	28
4.1.2.3	Počet ROP.....	28
4.1.2.4	Šířka paměťové sběrnice	28
4.1.2.5	Propustnost paměti	28
4.1.2.6	Frekvence jádra	29
4.1.2.7	Frekvence shaderů.....	29
4.1.2.8	Frekvence paměti.....	29
4.1.2.9	Velikost paměti	29
4.1.2.10	Fillrate.....	30
4.1.2.11	Verze shaderů	30
4.1.2.12	Výrobní technologie.....	30
4.1.2.13	Zdroj – minimum.....	30
4.2	Porovnání výkonu.....	30
4.2.1	Karty vhodné do kanceláře.....	31
4.2.1.1	Konfigurace	32
4.2.1.2	Parametry grafických karet.....	32
4.2.1.3	Testy	32
4.2.1.4	Závěr	35
4.2.2	Mainstream.....	35
4.2.2.1	Konfigurace	35
4.2.2.2	Parametry grafických karet.....	36
4.2.2.3	Testy.....	36
4.2.2.4	Závěr	39
4.2.3	Nejrychlejší grafické karty současnosti – High-end.....	40
4.2.3.1	Konfigurace počítače.....	40
4.2.3.2	Parametry grafických karet.....	40
4.2.3.3	Testy.....	41
4.2.3.4	Závěr	44
5.	Očekávaný vývoj grafických karet	45
5.1	Intel.....	45
5.2	nVidia.....	45
5.3	ATI.....	45

5.4	Vývoj architektur.....	46
5.4.1	Výpočet fyziky.....	46
5.4.2	Optimalizace ovladačů	46
6.	Závěr	48
7.	Seznam literatury	49
8.	Slovníček	51
Příloha A	Sběrnice PCI, AGP, PCI-E	53
Příloha B	Pasivní a aktivní chlazení grafických karet	54
Příloha C	Rozdíl Min a Max detailů	56
Příloha D	Závislost grafické karty na CPU	58

1. Úvod

V poválečných letech dvacátého století vzbudily celosvětový ohlas první prototypy počítačů. Častým tématem časopisů, zajímající se o vědu a techniku, byl budoucí vývoj výpočetní techniky a míra „robotizace“ ve 21. století. Světové kapacity věřily v plně automatizovaný systém zavedený po celém světě, každá rodina měla mít domácího robota, schopného vykonat všechny práce, auto mělo být dopravní prostředek, kde by se člověk stal pouze o své pohodlí a vše by řídil počítač či robot. Na Měsíc by jezdily zamilované páry na romantický výlet a člověk by byl jen pouhou kontrolou v plně automatizovaném světě. Dnes může tato předpověď vyvolat úsměv, při pohledu na geometricky rostoucí tempo vývoje počítačů ve dvacátém století, se těmto očekáváním nelze divit. Ano, počítače prodělaly obrovský výkonnostní skok, ale stále nemůžeme hovořit o plně automatizované společnosti. Počítače jsou našimi velkými pomocníky, ale hlavní díl práce stále leží na člověku. Co se však objevuje ve většině knih, filmů a komiksů odehrávajících se v blízké budoucnosti, je virtuální svět. Nemají tím na mysli internet, jakožto jistou formu virtuálního světa, ale „schopnost počítače zobrazovat skutečný svět reálně“, tedy pokud možno co nejreálněji. Zobrazovací a výkonové prostředky počítače jsou z tohoto ohledu stále na svém počátku a jistou měrou k tomu přispívá i koncepce dnešních počítačů, dnešních monitorů a grafických karet.

Pokud má práce přiblížit hojně diskutovanou problematiku grafických karet, ať už se jedná o výkon, architekturu či vývojový trend, je nutné začít od prvopočátků celého zařízení, kterého je grafická karta nedílnou součástí, a to od počítačů. Grafická karta bude správně fungovat pouze v dobře sestaveném, odladěném a výkonném počítači. K dnešním výkonným systémům vedla dlouhá a trnitá cesta, kterou se autor pokusí velmi stručně nastínit.

Historie vývoje samočinných počítačů se začíná odvíjet počátkem čtyřicátých let 20. století., první počítače, které byly osazovány přídatnými grafickými kartami, se však objevily až v letech osmdesátých, kdy Intel představil rychlý procesor 4004. Poté již v rychlém sledu následovaly novější a rychlejší typy procesorů, na jejichž rychlosti se odvíjel výkon celého počítače. Procesory byly vyvíjené a produkovány především firmou Intel. V osmdesátých letech byl představen nový procesor 80286, v roce 1985 procesor 80386, který byl již poměrně rozšířen v osobních počítačích (Personal Computer). Od příchodu moderního a výkonného Pentia v roce 1993 se výkon procesorů zvyšuje každým rokem, ne-li měsícem, a tak se z 66 MHz zvyšuje frekvence během necelých patnácti let na 3,6 GHz. Výkon počítače roste díky několika procesorům v počítači (dual-core), zvětšující se kapacitě a rychlosti paměti a zvyšující se rychlosti datových přenosů. Již neplatí „rovnice“ z minulého století, že výkon počítače je dán rychlostí procesoru. Stále má jeho rychlost obrovský vliv, ovšem důležitá se stala rychlost či velikost ostatních zařízení v počítači. Jelikož člověk usiluje o

maximálně reálné zobrazení virtuálního světa, je výkon a možnosti dnešních počítačů z nemalé míry dán výkonností grafické karty.

Je otázkou, zda-li jsou všechny gigaherty, gigabyty a nanosekundy dostačující pro moderní grafické karty, pro moderní systémy, pro nároky a požadavky neustále se rozrůstajícího počtu uživatelů?

2. Cíl práce a metodika

Při pohledu na kterýkoliv obchod zabývající se výpočetní technikou za účelem výběru grafické karty, stránka zobrazí několik desítek druhů grafických karet, lišících se jen nepatrnými změnami v názvu nebo pouze odlišným číselným označením. Může to být poněkud frustrující, jelikož se nedá říci, že čím vyšší číslo název karty obsahuje, tím výkonnější bude. Většiny elektronických obchodů nabízí potenciálnímu kupci mnoho podrobnějších informací v podobě desítek technických parametrů a hodnot. Není výjimkou (natož ostudou), že člověk běžně pracující v oblasti IT, se mezi těmito parametry mnohdy ztratí. Výrobci se často předhánějí v superlativech, a proto nalezení kýžených parametrů, se mnohdy rovná Foglarově Záhadě hlavolamu.

Cílem práce je provést spektrem grafických karet a učinit závěry, které mohou pomoci s výběrem karty pro různé segmenty. Je totiž zřejmé, že uživatel pracující pouze s běžnými kancelářskými programy bude mít jiné preference, než uživatel libující si v nejmodernějších hrách či ve zpracování videa. Práce se tedy bude zabývat jednotlivými technickými parametry karet, především těmi, které nejvíce ovlivňují výkon. Ten bude měřen pomocí testů a speciálních benchmarků, v každém z nich se bude vždy vyskytovat velké spektrum grafických karet. To pomůže odhalit, které komponenty nebo technické parametry mají či nemají vliv na výkonnost grafické karty. Architektura, nebo-li způsob práce grafické karty, se taktéž podílí na výkonu, proto se práce bude současnými architekturami zabývat a poukáže na případná pozitiva či negativa.

3. Historie vývoje grafických karet

3.1 Obecný náhled

Představa, že počítač bude využit i pro generování obrazu nepřišla náhle a nečekaně, první pokusy sahají k prvopočátkům výpočetní techniky. Vývoj speciálních zobrazovacích prostředků začal zhruba v první polovině šedesátých let dvacátého století. V té době se počítače začali naplno uplatňovat ve vědecké sféře a výstup v podobě mnoha čísel byl velmi komplikovaný, málo přehledný a srozumitelný. Platilo to hlavně v případech, když bylo nutné porozumět většímu množství dat, která jsou mezi sebou určitým způsobem provázána.

Bylo zřejmé, že ve většině případů je názornější převést číselné informace na grafické a ty pomocí barevných map, grafů či barevného zvýraznění zobrazit. V jisté míře jsou grafické informace nezávislé na jazyku (přímo nezávislé jsou například u dopravních značek). Grafickým symbolům také porozumí většina národů (velmi jednoduchým případem je šipka, jenž vznikla z letícího šípku).

Základní funkcí grafické karty je tedy převádění číslicových (digitálních) instrukcí do analogové podoby zpracovatelné monitorem. Jedná se tedy o zařízení, jenž se stará o zobrazování informací na monitoru, tzv. počítání polohy pixelů¹⁾ a přiřazování barevného kódu. Počátkem osmdesátých let byla dostačující karta, která podporovala pouze textový (znakový – 80x25 znaků) režim. S nástupem prvních grafických uživatelských rozhraní však tento druh karet nestačil. Výkonový požadavek se zvýšil více než třicetkrát.

Grafické rozhraní začalo být populární zejména na osmibitových domácích počítačích a na herních automatech. S ohledem na dnešní dobu byly možnosti prezentace grafického rozhraní velmi omezené, avšak jedny z prvních animací, grafických triků a efektů učinily z těchto počítačů jedno z prvních, široce dostupných multimediálních zařízení.

Generování grafického obrazu zajišťovaly u osmibitových počítačů speciální grafické čipy. Vzhledem k výkonu „osmibitů“ musela být kapacita obrazové paměti co nejmenší, aby její adresovaný rozsah zbytečně neubíral na již tak skromném výkonu procesorů. I přes tyto nedostatky se daly vytvořit velmi zajímavé a působivé efekty, které uchvátily zejména nadšence počítačových her. Ti do té doby pracovali v plně textovém režimu, s podporou grafického režimu tak hry dostaly nový rozměr.

Téměř všechny grafické čipy u osmibitových počítačů byly schopny pracovat v textovém i grafickém režimu. Velikou předností textového režimu je rychlost, přiřazuje se mu malé množství paměti a výkon tak není příliš ovlivněn. Oproti tomu výhody grafického režimu jsou zcela zřejmé. Uživatel může využít všech možností výstupního grafického zařízení, typicky monitoru či televizoru.

Náročnost grafického rozhraní vyřešili výrobci implementováním mnoha složitých instrukcí do grafického akceleratorů, který zefektivňoval práci mezi procesorem a samotnou grafickou kartou a přebíral práci procesoru na svá bedra. Byl koncipován jako přídatné zařízení ke grafické kartě, jenž se zapojovalo do slotu PCI. Obrovský nárůst výkonu byl znát u trojrozměrné grafiky (3D), kdy výpočet scény bez akceleratoru, by drasticky snížil výkon a znemožnil plynulou práci. Postupem času se přídatné akcelerátory, respektive jejich obvody staly součástí grafických karet, které tak splňují vysoký standard. Vysokou kvalitu zobrazování ve 2D (kancelářské programy, plocha windows, prohlížení obrázků) a vysoký výkon ve 3D (počítačové hry, práce s videem).

3.1.1 Koncepce osobního počítače IBM PC

Vzhledem k tendenci vývoji osobních počítačů a grafických karet se práce bude věnovat počítači, jenž fungoval na téměř stejném principu, jako fungují počítače dnešní. IBM PC.

V době uvedení prvních osobních počítačů nikdo příliš nepředpokládal, že se počítač od firmy IBM dostane před své konkurenty a stane se základním kamenem počítačů konce dvacátého a začátku dvacátého prvního století. IBM PC byl uveden v srpnu 1981, obsahoval procesor Intel 8088, 16kB paměti, disketovou mechaniku o velikosti 160kB a černobílý monitor 11,5". Revoluční nebyly grafické možnosti, které umožňoval, ty byly ve skutečnosti ze začátku velmi strohé, podstatný byl hrubý výkon, který tento počítač nabízel. Bylo to dáno díky nové architektuře, kde procesor komunikuje po sběrnici s přídatnými zařízeními a podstatně tak zefektivňuje práci, jelikož v mnoha případech mu přídatná zařízení práci usnadňují.

Jistým paradoxem bylo, že oproti relativně komplikovaným a propracovaným grafickým adaptérům, jenž se vyráběly k „osmibitům“, videoadaptéry pro PC uměly pouze zobrazit bitovou mapu uloženou ve své videopaměti. Právě oddělení videopaměti od paměti operační a velkou měrou ovlivňuje výkonnost a funkcionalitu grafických karet současnosti.

Tato kapitola převážně čerpá ze zdroje [16].

3.1.1.1 Sběrnice a porty na osobních počítačích

IBM PC je navrženo tak, aby bylo možné použít části počítače od kteréhokoli výrobce a tím byl tedy počítač velmi lehce rozšiřitelný. Otevřenost a přizpůsobitelnost (modulárnost) je dána přenášením informací neboli dat pomocí sběrnice, díky které spolu komunikují jednotlivá zařízení. Modulárnost má za následek, že místo grafického čipu, který by pomáhal operační paměti a procesoru, se objevuje samostatná grafická karta. Ta významnou měrou ulehčuje práci procesoru, a jelikož obsahuje vlastní videopaměť, uvolňuje i paměť operační.

3.1.1.1.1 Sběrnice ISA

Jedna z prvních sběrnic, která se používala v počítačích IBM PC, byla typu ISA (Industry Standard Architecture). Byla navržena v roce 1980 firmou IBM a umožňovala osmibitový, později šestnáctibitový přenos informací.

Díky své konstrukční jednoduchosti se používala relativně dlouho, prakticky do roku 1989, kdy byl vyvinut procesor 80486 a sběrnice ISO již nebyla schopna nabídnout potřebný výkon.

3.1.1.1.2 Sběrnice VESA Local Bus

Společnost VESA navrhla sběrnici VESA Local Bus (VLB) jako rychlejší variantu výkonově již nedostačující sběrnici ISA. Nový typ sběrnice byl vyvinut hlavně pro grafické karty, řadiče pevných disků a síťové karty. Díky tomu byl její výkon mnohonásobně vyšší a tato sběrnice tak byla osazena do nového počítače 486.

Nevýhodou sběrnice VLB byla možnost připojení omezeného množství zařízení (prakticky to byla pouze tři zařízení při frekvenci sběrnice 40Mhz, při maximální rychlosti 50Mhz to byla pouze dvě zařízení) a postupem času, kdy byly vyvinuty rychlejší sběrnice (kterým se práce bude věnovat dále), klesala podpora u výrobců. Dnes se VESA Local Bus vyskytuje pouze u pár průmyslových počítačů.

3.1.1.1.3 Sběrnice PCI – nový věk

V současné době je na platformě²⁾ osobních počítačů (PC) nejvíce rozšířena sběrnice PCI (Peripheral Component Interconnect). Tato sběrnice byla navržena firmou Intel v roce 1992 a stala se standardem pro výrobce periferních zařízení³⁾.

Sběrnice PCI se vyrábí v několika variantách, lišících se taktovací frekvencí a šířkou datové sběrnice. Tyto dva parametry přímo ovlivňují a určují maximální datový tok. U většiny osobních počítačů je nejpomalejší verze sběrnice PCI. Je taktována na 33MHz, šířkou datové sběrnice 32 bitů a datovým tokem $132 \text{ MB} \cdot \text{s}^{-1}$ ($33 \text{ MHz} \times 4 \text{ B} = 132 \text{ MB} \cdot \text{s}^{-1}$). Rychlejší varianty našly své uplatnění ve výkonných specializovaných systémech, kde je nutný maximální datový tok. Jedná se například o práci s videem, či komunikaci několika velmi výkonných pevných disků.

Předností PCI je podpora Plug and Play, což je dynamická detekce a konfigurace zařízení (hardwaru) ihned po startu počítače. Umožňuje tak operačním systémům zjistit, jaký hardware je v počítači a na základě toho může automaticky k danému hardware nainstalovat či přiřadit příslušné ovladače.

3.1.1.1.3.1 Připojení grafických karet pomocí PCI

Pro připojení grafických karet se PCI sběrnice používala velmi často (prakticky od svého uvedení na trh v roce 1992), a to zejména z důvodů jejího progresivního vývoje do budoucna (sběrnice PCI2, PCI-X, PCI-E atd.) a samozřejmě poměrně velké rychlosti přenosu dat. Kvůli zachování zpětné kompatibility však nebylo možné nadále zvyšovat rychlost PCI sběrnice, a tak bylo nutné vyvinout novější a rychlejší sběrnici či port pro grafické karty.

3.1.1.1.4 AGP

Výsledkem snah o výkonnější sběrnici než je PCI, byl nakonec port AGP, představený společností Intel v roce 1997. Nejedná se totiž v pravém slova smyslu o sběrnici, neboť lze k AGP připojit pouze jedno zařízení (takzvané zařízení Point-to-Point). Označení AGP vzniklo zkrácením plného názvu Advanced Graphics Port a jak již tento název naznačuje, jedná se o port určený prakticky výhradně k připojení grafických adaptérů, zejména grafických akceleratorů. Hlavním důvodem k zavedení AGP portu bylo rozšíření trojrozměrné scény, s nímž rostly požadavky na přenos textur a jiných grafických dat. Port AGP vznikl úpravou sběrnice PCI, proto jsou použity velmi podobné řídicí signály. Při návrhu byl kladen důraz na co nejvyšší přenosovou rychlost dat a díky možnosti připojení pouze jednoho zařízení a tím jistého zjednodušení. Změnou je podpora taktovacích frekvencí 33MHz a 66MHz, přičemž na těchto frekvencích musí pracovat všechny karty zapojené do portu AGP.

Novější verze AGP zvýšily původní přenosovou rychlost na čtyřnásobek a později na osminásobek a přinesly snížení napětí z původních 3,3V až na 0,8V. Zároveň zvyšovaly počet přístupů do paměti během jednoho taktu, díky čemuž se zvyšovala přenosová rychlost. Viz. Tabulka 1.

Verze	Frekvence	Počet přístupů do paměti	Max. přenosová rychlost	Napětí
AGP 1x	66 MHz	1	266 MB*s ⁻¹	3,3 V
AGP 2x	66 MHz	2	533 MB*s ⁻¹	3,3 V
AGP 4x	66 MHz	4	1066 MB*s ⁻¹	1,5 V
AGP 8x	66 MHz	8	2133 MB*s ⁻¹	0,8 V

Tabulka 1 – Verze AGP sběrnice

Nejnovější verze (označovaná taktéž jako AGP 3.0) kýžený výkonnostní posun oproti AGP 4x nepřinesla, protože grafické karty, jenž v té době byly nejrychlejší na trhu, se spokojily s datovým tokem 1066 MB*s⁻¹ a rychlejší AGP tak zůstalo de facto nevyužito.

3.1.1.1.4.1 Zvláštní režimy:

Vzhledem k tomu, že port AGP je určen výhradně pro připojení grafických karet a grafických akceleratorů, podporuje AGP některé zvláštní režimy přenosu dat, které nemají u jiné sběrnice či portu obdobu. Tyto režimy jsou určeny pro efektivní přenos případně sdílení grafických dat, zejména textur nebo přenos videa.

3.1.1.1.4.1.1 DMA

V tomto režimu jsou veškerá data textur uložena v paměti grafické karty. Pokud již není v paměti grafického akceleratoru volný prostor pro uložení další textury nebo je textura ve vykreslované scéně použita poprvé, musí se data zkopírovat z operační paměti počítače do paměti integrované na grafickém akceleratoru. Toto kopírování probíhá po velkých blocích (řádově desítky kilobytů) pomocí režimu DMA. Každý přenos tedy znamená velkou zátěž pro celý systém, tato zátěž je však pouze krátkodobá a občasná. Režim DMA se stává výhodným především v případech, kdy je na grafickém akceleratoru nainstalována paměť s dostatečnou kapacitou.

3.1.1.1.4.1.2 Execute

V případě tohoto režimu je paměť grafické karty a operační paměť rovnocenná. To znamená, že se při texturování nemusí na grafický akcelerator přenášet celá textura, ale je umožněn přenos pouze malé části, která je v dané chvíli potřebná.

3.1.1.1.5 PCI-E

Nejnovějším typem sběrnice, je rozšířená sběrnice PCI, označovaná jako PCI Express. V roce 2004 se novinka stává standardem pro osobní počítače a o dva roky později je 99 procent grafických karet vyráběno pouze do PCI-E.

Jedná se o zcela přepracovanou, novou sběrnici, která však používá existující programové návrhy a komunikační standardy. Je založena na mnohem rychlejší sériové komunikaci a tak se její přenosová rychlost pohybuje v desítkách $\text{GB}\cdot\text{s}^{-1}$. V dnešní době se nejčastěji používají měděné spoje, jejichž maximální přenosová rychlost je stanovena na $10 \text{ GB}\cdot\text{s}^{-1}$, čili se dá říci, že sběrnice je nyní brzděna pouze technologií výroby. Mnohem vyšší rychlost ji přímo předurčuje k tomu, aby nahradila veškeré lokální sběrnice, včetně AGP i PCI.

Jak je to u sběrnice běžné, zvyšování přenosové rychlosti se provádí dvěma způsoby. Zvyšováním frekvence a přidáváním dalších signálových párů. Díky tomu se základní rychlost může zvyšovat v násobcích 1x, 2x, 4x, 8x, 12x a 16x. Pro grafické karty se samozřejmě používá nejrychlejší varianta, tedy 16x násobná rychlost, pro ostatní přídatná zařízení většinou postačuje rychlost základní tedy 1x, případně 4x.

3.2 Grafické karty pro IBM PC

Jak již tato práce zmínila, první počítače prezentované společností IBM byly koncipovány jako kancelářské počítače s minimální mírou grafických aplikací. Pozdější vývoj operačního systému a aplikací však ke grafice směřoval, a tak začala firma IBM vyvíjet a produkovat první grafické karty. IBM drželo tříletý cyklus představování nové technologie. První byla představena grafická karta CGA v roce 1981, následovaly EGA, VGA a XGA v tříletém cyklu, tedy do roku 1990. V dalších kapitolách práce tyto grafické karty představí, nejprve se však zmíní o čipu, jenž měl hlavní slovo ve vývoji grafických karet pod hlavičkou IBM. Jedná o čip Motorola MC6845.

3.2.1 Čip Motorola MC6845

Právě tento řadič displeje vybrala firma IBM pro návrh grafických karet určených pro vznikající IBM PC a jeho následovníky. Bylo to z díky velmi dobré podpoře jak textových (hardwarový kurzor), tak i grafických režimů (rozlišení 768x384).

3.2.2 Grafická karta CGA

První dostupnou kartou s podporou grafiky na osobních počítačích typu IBM PC byla grafická karta CGA, což je označení, které vzniklo z celého názvu Color Graphics Adapter. Tato karta, která byla jako oficiální standard uvedena v roce 1981, oficiálně podporuje celkem dva grafické režimy a dva režimy textové. Kromě toho je však pro televizní výstup použit třetí grafický režim a vhodnou manipulací s registry lze vytvořit i další textové, grafické či smíšené režimy.

V prvním grafickém režimu bylo možné zobrazit obraz v rastru 320x200 pixelů ve čtyřech barvách, druhý režim měl rozlišení 640x200 pixelů dvoubarevně.

3.2.3 Grafická karta Hercules

Grafická karta Hercules (celým názvem Hercules Graphics Card – HGC) byla navržena společností Hercules Computer Technology jako konkurenční výrobek ke grafickým kartám CGA od firmy IBM. Hercules byl velmi oblíbený a do svých sestav ho montovali i přední výrobci osobních počítačů. Oblíbenost byla způsobena jak nízkou cenou, tak i poměrně dobrými technickými parametry, například možností práce v kvalitním textovém režimu (to nezvládala karta CGA) či v monochromatickém grafickém režimu s vysokým rozlišením.

3.2.4 Grafická karta EGA

Rozšířením grafické karty CGA vznikla v roce 1984 grafická karta EGA (Enhanced Graphics Adapter). EGA podporovala veliké množství grafických režimů, počínaje základním rozlišením 320x200 pixelů ve 4 barvách až po rozlišení 640x350 pixelů v šestnácti barvách.

3.2.5 PGC – první grafický akcelerátor pro PC

Tato karta působila ve srovnání s videoadaptérem EGA jako zjevení. PGC je zkratka názvu Professional Graphics Controller a karta se dle svého jména i chovala. Akcelerátor umožňoval jeden režim o rozlišení 640x480 pixelů, kde každý pixel mohl nabývat jedné z 256 barev, které byly vybírány z palety 4096 barev. Na svou dobu šlo o vynikající grafické rozhraní, a jak už z názvu vyplývá, bylo to spíše profesionální řešení, jelikož cena za výkon byla vysoká (několik tisíc dolarů). Díky vysokému rozlišení byl nutný i jiný monitor, než pro klasické karty CGA, Hercules či EGA, čili i z tohoto hlediska to nebyla levná záležitost.

PGC nebyla obyčejná karta, ve skutečnosti se skládala ze třech karet, z nichž se dvě zapojovaly do slotů ISA, prostřední se do počítače nezapojovala. Tyto tři desky, byly propojeny kabelem, jenž fungoval jako interní sběrnice a komunikační kanál. Pro zapojení PGC do počítače bylo kvůli její konstrukci nutné, aby se konektory ISA sběrnice nacházely v definované vzdálenosti od sebe, což někteří výrobci základních desek nespĺňovali.

3.2.6 Grafická karta VGA

Rozšířením a technologickým vylepšením karet CGA a EGA vznikla grafická karta Video Graphics Array, neboli VGA. Slovo array v tomto kontextu značí, že celá karta byla vytvořena na jednom čipu, samozřejmě kromě nezbytných krystalů a čipů s obrazovou pamětí. Tato grafická karta byla poprvé použita v počítačích řady IBM PS/2. V těchto počítačích se dokonce provedlo zabudování grafické karty přímo na základní desku počítače, tuto konstrukci společnost IBM později použila i v dalších modelech, samozřejmě s využitím výkonnějších grafických karet či grafických čipů.

Veškeré dnes vyráběné grafické karty jsou zpětně kompatibilní s VGA. Grafické režimy VGA jsou i dnes používány v některých systémech, při inicializaci graficky orientovaných operačních systémů (a také při nefunkčnosti ovladačů grafických karet ve starších operačních systémech Microsoft Windows.

3.2.7 SVGA

Termínem SVGA nebo také Super VGA jsou obecně označovány všechny grafické karty vytvořené pro počítače řady IBM PC, které umožňovaly či umožňují používat rozšířené grafické režimy, tj. režimy s rozlišením vyšším, než je nejvyšší rozlišení karty VGA: 640x480 pixelů.

Grafické karty SVGA nabízejí grafické režimy s vyšším rozlišením než 640x480 pixelů, typicky jsou dostupná rozlišení 800x600 pixelů a 1024x768 pixelů. Tyto režimy je možné používat s různou barevnou hloubkou a s různou obnovovací frekvencí obrazu i řádkovou frekvencí.

3.2.8 Standard VESA

V předchozích dvou kapitolách se práce věnovala grafickým kartám typu SVGA a XGA. Tyto karty nabízely poměrně mnoho funkcí i grafických režimů, jejich největší nevýhodou však byla vzájemná nekompatibilita – neexistoval jednotný způsob nastavení rozšířených grafických režimů, stejně jako se lišily režimy přístupu do video paměti.

Byl tak zaveden standard navržený sdružením VESA (Video Electronics Standards Association). Do tohoto sdružení se dříve či později zapojili prakticky všichni výrobci grafických čipů a významné softwarové firmy. Výsledkem práce tohoto sdružení bylo několik verzí norem, které byly v počítačové literatuře zkráceně nazývány pouze termínem VESA, například VESA 1.1 či VESA 2.0. V těchto normách byly uvedeny zejména čísla rozšířených grafických režimů tak, aby je bylo možné nastavit pomocí služeb takzvaného VESA BIOSu. Kromě toho byly specifikovány služby pro zjištění parametrů grafických karet, čili výrobce grafického čipu, typu grafické karty, podporovaných rozlišení a podporovaných obnovovacích frekvencí obrazu.

Standard VESA byl v minulosti použit zejména v operačním systému DOS, v dnešní době má význam pouze jako nejmenší společný prvek, který by (alespoň teoreticky) měly podporovat všechny moderní grafické karty.

3.3 Grafické akcelerátory

Spolu s vývojem grafických programů, operačních systémů a her bylo zřejmé, že chudý grafický režim a dvojrozměrný obraz nebude patřit budoucnosti. S narůstajícími operacemi a výpočty, jež zdokonalovaly výslednou scénu, rostl úměrně i nárok na procesor. Během vykreslování grafických efektů se zároveň provádějí i další výpočty, například fyzikální simulace či výpočet trajektorie a případných kolizí pohybujících se těles. V tomto případě je vhodné ponechat co nejvíce výpočetního výkonu mikroprocesoru pro výpočet simulací a co nejvíce grafických algoritmů provádět na specializovaném zařízení – grafickém mikroprocesoru GPU (Graphics Processor Unit).

Tato kapitola čerpá především ze zdrojů [1, 2, 3, 16].

Zprvu je však nutné, aby byla nastíněna grafická rozhraní, ve kterých většina trojrozměrných aplikací pracuje.

3.3.1 Moderní grafická rozhraní:

Téměř všechny postupy, které jsou dnes implementovány na grafických akcelerátorech, které práce představí později, je možné vyjádřit či popsat pomocí funkcí dvou nejpoužívanějších rozhraní pro 3D grafiku: OpenGL[™] a DirectX. Těmto rozhraním, jenž velmi usnadňují práci vývojářům 3D aplikací se bude práce věnovat níže.

3.3.1.1 OpenGL[™]

Knihovna OpenGL[™] (Open Graphics Library) byla navržena firmou SGI (Silicon Graphics Inc.) jako aplikační programové rozhraní (Application Programming Interface – API) k akcelerovaným grafickým kartám respektive k celým grafickým subsystémům. Jedná se o průmyslový standard, který specifikuje API pro tvorbu grafického rozhraní (počítačové grafiky). Používá se při tvorbě počítačových her, CAD programů, aplikací virtuální reality.

3.3.1.2 DirectX

Tato programátorská knihovna, obsahující nástroje pro tvorbu počítačových her i multimediálních aplikací, byla vytvořena firmou Microsoft, pro použití pod operačním systémem Windows jako reakce na již zavedené profesionální rozhraní OpenGL.

Verze 1.0 byla vyvinuta v roce 1995, dnes je očekávána verze DirectX 10 pro nový operační systém Windows Vista.

3.3.2 Chronologický vývoj grafických akcelerátorů

Stav trhu na poli grafických karet nebyl vždy tak průhledný, jako tomu je nyní, kdy dominují dvě, respektive tři firmy. Objevilo se mnoho výrobců a společností, které vyrobily „nový nejlepší a nejrychlejší grafický čip v dějinách“, z nichž některé zmizely stejně rychle, jako se na trhu objevily. Práce se bude zabývat jednotlivými obdobími, jenž přinesly převratnou novinku či přímo kartám, který by v letmé historii akcelerátorů neměly chybět a zároveň při té příležitosti představí společnosti, jenž do vývoje zasahovaly.

3.3.3 1995 první grafický akcelerátor

3.3.3.1 3Dfx

První grafický akcelerátor představila do té doby neznámá společnost **3Dfx**. Byla založena v roce 1994 a již v následujícím roce uvedla první a svým způsobem revoluční čip Voodoo Graphics.

3.3.3.1.1 Voodoo Graphics

Karta byla koncipována jako přídatný akcelerátor do PCI slotu, což byl důsledek snazší technologie výroby. Voodoo tak získalo výhodu nižší ceny (nižší náklady) i lepšího chlazení, kdy karta nepotřebovala žádný chladič. Karta obsahovala 4 MB paměti a hlavně díky dobrému marketingu byla velmi úspěšná. Rozlišení, v němž akcelerátor pracoval, bylo fixní – 640x480 pixelů.

3.3.4 1997

3.3.4.1 nVidia vystrkuje růžky

V tomto roce se na trhu začala silně projevat konkurence pro 3Dfx – společnost nVidia.

Tato kalifornská společnost byla založena v roce 1993, přičemž první čip uvedla až v roce 1995.

3.3.4.1.1 RIVA 128

Prvním čipem, kompatibilním s rozhraním Direct3D byla RIVA 128 (NV3). Byla to zkratka plného názvu Realtime Interactive Video and animation Accelerator. Karta nabízela řešení kompletně integrované do jednoho čipu – VGA jádro, 2D, 3D akcelerace, video a nechyběl ani RAMDAC⁴.

3.3.5 1998

3.3.5.1.1 Voodoo 2

Tento rok přinesl nové Voodoo 2. Karta nabízela až trojnásobný výkon, obsahovala dvě texturovací jednotky, díky čemuž bylo možné aplikovat multitexturing (schopnost vykreslit více textur na jeden pixel).

Přestože nebyla Voodoo 2 tak revoluční, jako Voodoo Graphics, přinesla s sebou jednu velkou novinku. Zapojení 2 akcelerátorů do jednoho počítače – SLI (dnes touto technologií disponuje nVidia a po počátečních problémech již dokáže SLI režim poskytovat dvojnásobný výkon).

3.3.5.1.2 RIVA TNT

Pod taktovkou společnosti nVidia, byl v červnu 1998 vydán zcela nový čip RIVA TNT (NV4). TNT značilo TwiNTexel, tedy schopnost vykreslit dva (bilineární single-textured) texely v cyklu.

Klíčem k úspěchu NV4 byla rovnováha mezi výbornou kvalitou obrazu a velmi slušným výkonem.

3.3.6 1999

3.3.6.1.1 GeForce

Nedlouho poté co nVidia představila TNT2 se objevila další novinka v podání této firmy – grafická karta GeForce. Tato karta bylo převratná z mnohých důvodů. Obsahovala GPU, tedy Graphics Processing Unit, což je čip s podporou akcelerace T&L (Transform & Lighting) operací kompatibilních s Microsoft DirectX 7. Tato jednotka umožňovala urychlení práce při převodu informací a zpracovávání světla. Čím výkonnější systém však byl, tím méně byla jednotka T&L znát.

Dalším, neméně důležitým přínosem byla implementace DDR⁵) řadiče, který umožnil použití DDR pamětí, což byl velice úspěšný krok, který dočasně odboural limitaci výkonu grafického čipu propustností paměťové sběrnice.

3.3.7 2000

3.3.7.1.1 Radeon 256

V tomto roce vydal i jeden z mála konkurentů nVidie – společnost ATI novou grafickou kartu – Radeon 256. nVidia v tomto roce taktéž nezůstala pozadu a představila GeForce 2, což bylo de facto vylepšené GeForce 1.

3.3.8 2001 – DirectX 8.0 nastupuje

3.3.8.1.1 GeForce 3

V tomto roce si nVidia podržela svůj technologický náskok a představila kartu GeForce 3 (NV20), která svým hrubým výkonem nedala šanci nikomu. Navíc měla optimalizovanou paměťovou sběrnici a plně podporovala DirectX 8.0. Ten v sobě obsahoval podporu pro nové programy, které se zpracovávaly přímo na grafické kartě.

3.3.8.1.2 Shadery

Shadery jsou malé programy určené pro zpracování přímo na grafické kartě. Takovéto programy se píší nejčastěji v jazycích CG (nVidia, univerzální), HLSL (Microsoft, pro DirectX), GLSL (pro OpenGL) a později jsou překladačem přeloženy do assembleru (strojový kód) přímo pro danou grafickou kartu. V současné době jsou tři typy shaderů: vertex, pixel a geometry.

3.3.8.1.2.1 Vertex shader

Program, který se provede na každý vrchol (vertex) vstupní geometrie. Nejčastěji provádí transformaci (násobí vrchol World, View a Projection maticemi) a výsledkem je pozice na obrazovce (screen space). Může ale také provádět různé pohyby s vrcholy, např. pro hladinu vody. Vždy do

programu vstoupí jeden vrchol, je upraven a zase vystoupí, nelze tedy vrcholy přidávat či odebírat.

3.3.8.1.2.2 Pixel shader

Po transformaci vrcholů získáme místa na obrazovce, na které je nutno nanést texturu. Tento program se provede na každý pixel na obrazovce, proto jsou tyto operace náročné na hardware a pixel shadery jsou tak co nejjednodušší a maximálně optimalizované.

3.3.8.1.2.3 Geometry shader

Umožňuje přidávat a odebírat vrcholy a tím upravovat geometrii. Využití geometry shadery lze pro vytvoření ostnů na modelu dinosaura nebo zarůstání plotu trávou v reálném čase. Tyto shadery zatím běží pouze na DirectX 10 pod Windows Vista v REF. Nutno dodat, že v současnosti podporují DirectX 10 pouze grafické karty nVidia GeForce 8. Reálné užití je však zatím vzdáleno.

3.3.8.1.3 GeForce 4

Krátce po uvedení NV20, si nVidia pospíšila a uvedla na trh nový model – NV25 známý jako GeForce 4 Titanium. Podporoval nové pixel shadery verze 1.3 (oproti 1.1) pro DirectX 8.1, měl opět vylepšenou optimalizaci přístupu k pamětem a několik dalších menších změn a přinášel opět velmi vysoký hrubý výkon.

3.3.8.1.4 Radeon 8500

Tato karta (R200) byla vydána společností ATI jako soupeř pro GeForce 4. S podporou nových shaderů se v době vydání však výkonově rovnaly GF3, díky dobré práci softwarového týmu však byly vylepšeny ovladače, a tak se nový Radeon blížil k výkonnostní třídě nejrychlejší karty od nVidie.

3.3.9 2002 – DirectX 9.0

Nová revize rozhraní od Microsoftu s sebou přinesla výrazné změny na trhu grafických karet. Doposud suverénní nVidia se dá říci “usnula na vavřínech” a nedokázala zareagovat na příchod nového Radeonu.

3.3.9.1.1 Radeon 9700Pro

Tato revoluční karta (čip R300) doslova smetla všechna řešení ze stolu nVidie. Byl to zcela výkonově neohrožitelný čip, díky kterému ATI na rok převzala nadvládu nad trhem s grafickými kartami. R300 plně podporoval DirectX 9.0 (Pixel a Vertex Shaderů 2.0) a vysokým hrubým výkonem ničil soupeře. Nově se také začal používat konektor externího napájení a karty začaly vyžadovat kvalitní aktivní chladič.

3.3.9.1.2 Série GeForce FX

Reakcí na R300 od ATI měla být vlajková loď nVidie – GeForceFX 5800Ultra. Naprogramované jednotky, které měly zpracovávat nové instrukce v DX 9.0, byly značně problematické a karta tak měla velmi nízký výkon při nejnovějším DirectX 9.0.

Posledním krokem nVidie bylo rychlé přepracování stávající karty a tak vznikl čip NV35 – GeForceFX 5900. Díky rozšíření paměťové sběrnice, výměně některých jednotek se výkon zvýšil, na první místo to však nestačilo.

3.3.10 2004

Dva roky trvalo, než se nVidia dokázala vzpamatovat z nástupu R300, v dubnu roku 2004 se však vrátila ve velkém stylu. Obě konkurenční firmy začaly přecházet na nový typ sběrnice PCI-E, a tak se AGP verze vydávaly se zpožděním a některými úpravami.

3.3.10.1.1 GeForce 6

Architektura GeForce 6 (NV40) je velmi efektivní, konkurenční Radeony série X800 potřebují pro dosažení obdobného výkonu v průměru o 20% vyšší frekvenci jádra. GeForce 6 jako první generace karet společnosti nVidia nabídla možnost grafického multiprocessingu (SLI) pro domácí PC. Podpora SLI režimu byla také velmi dobrým marketingovým trikem, na který neměla ATI žádnou adekvátní odpověď. Podpora Shader Modelu 3.0 je spíše také teoretická, karty této generace výkonem sotva dostačovaly pro komplexní SM 2.0, ale opět to byl krůček navíc oproti konkurenci.

3.3.10.1.2 Radeon X

Čtrnáct dní poté co nVidia představila svou řadu 6, byl představen nový čip ATI, X800Pro. Ve skutečnosti to byl notně upravený čip R300, výkonem se však dokázal rovnat nejrychlejším kartám nVidie.

3.3.11 2005, 2006

V těchto letech byly vydány nové čipy od každé společnosti, nVidia představila G70 alias GeForce 7900, což je ve skutečnosti upravený čip NV40. ATI vydala nové čipy téměř souběžně s nVidií, Radeon X1800XT (R520) a později Radeon X1900XTX (R580). Z hlediska vývoje se nejedná o revoluční čipy, obě firmy se snaží hlavně optimalizovat svá řešení a minimalizovat výkonové nedostatky, čili výkon se zde posunuje v řádu procent. Podrobnějšímu náhledu se bude práce věnovat v následující kapitole.

4. Hodnocení současného stavu

4.1 Architektura grafických karet současnosti

S pohledem na výběr grafických karet, kdy jsou na výběr desítky různých typů karet je zřejmé, že správný výběr není v žádném případě jednoduchý. Jednotlivé typy se liší v mnoha parametrech, kterými se práce bude zabývat, zprvu však práce velmi zjednodušeně popíše práci grafické karty a její architekturu, z níž se taktéž odvíjí výkonová křivka.

4.1.1 Nejnovější čip – nová architektura

Nejnovějším grafickým čipem, který byl představen je první čip pro DirectX 10, GeForce 8 od společnosti nVidia. Microsoft, se kterým na vývoji s DX 10 pilně spolupracovala i nVidia, komplexně změnil klasický systém renderování, což by byl pro stávající architektury velmi tvrdý oříšek. Změna klasického pojednání architektury tedy byla nutná.

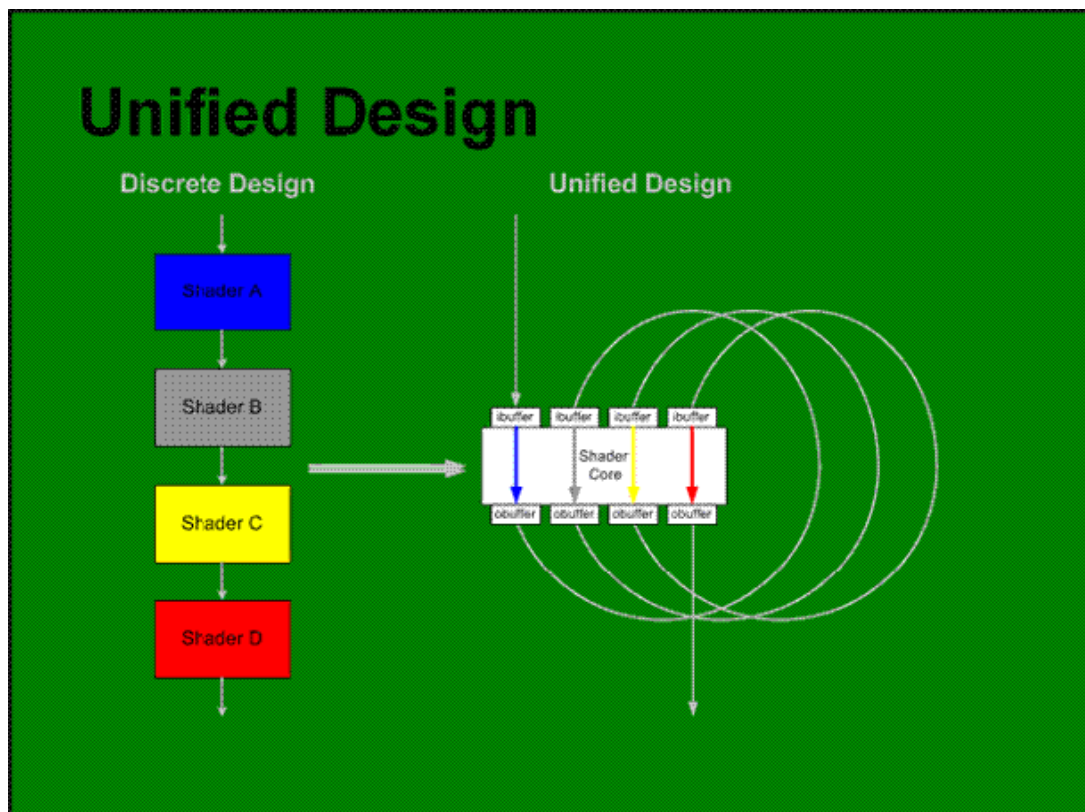
Až doposud zpracovávaly prakticky všechny GPU data víceméně stejným způsobem, tedy sekvenčně. I přestože programovatelné shadery daly programátorům velkou volnost, stále jsou specializované jednotky a pevně dané pořadí, v jakém data pipeline⁶⁾ procházejí. To se s novou architekturou mění.

Grafická karta nese označení GeForce 8800 (G80). Je dlouhá 27 centimetrů (jinak řečeno je obrovská) a obsahuje 681 milionů tranzistorů. Jen pro srovnání – předchozí top generace obsahovala 279 milionů tranzistorů, což je navýšení téměř o 250 procent.

Při zachování stejného výrobního procesu si to však vybírá daň v podobě obrovské spotřeby a nutného velikého chlazení, jenž zabere další slot.

4.1.1.1 Unifikované shadery

Rozdíl mezi dedikovanými (přiřazené k určitým typům operací) a unifikovanými shadery ilustruje následující obrázek.



Obr. 1 - Rozdíl při zpracování shaderů u G80

DirectX 10 přináší Shader Model 4.0 a s ním unifikované instrukce pro shadery. Fakt, že místo několika specializovaných typů jednotek je více univerzálních, má své výhody i nevýhody. Z hlediska vývoje čipu je jednodušší mít jeden typ shaderů (byť značně sofistikovaný), návrh se snáze ladí a lépe se dá škálovat výkon. DirectX 10 také specifikuje nový typ shaderů (v pořadí třetí) – Geometry shadery. Bylo prakticky nemožné vybalancovat dříve používaný poměr vertex : geometry : pixel shaderů a z toho důvodu musela přijít unifikace.

Unifikovaný shader “zaskočí” tam, kde je zrovna potřeba, čili celé GPU stále jede na prakticky plný výkon. Podmínkou skutečného dosažení maximálního výkonu je dostatečná rezerva výkonu ROP⁷⁾ (raster operations pipelines units) a velmi rychlý přístup k pamětem.

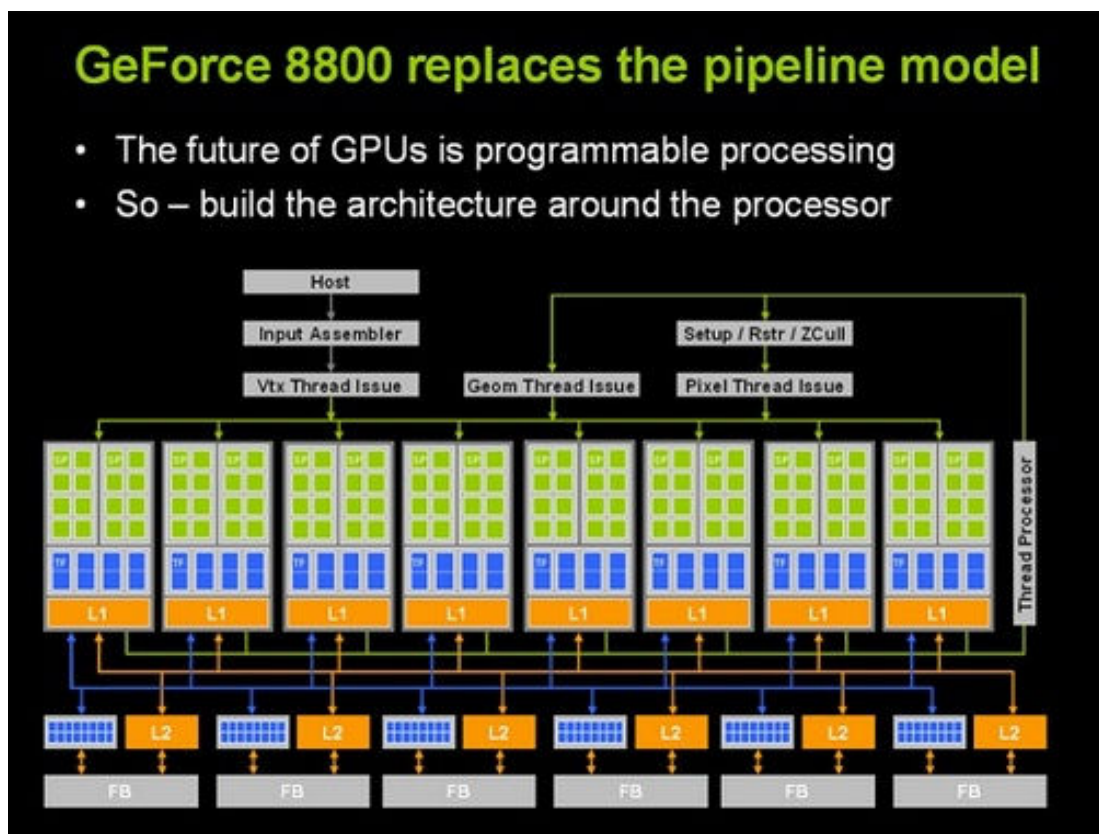
Nevýhodou unifikovaných shaderů je výrazně větší počet tranzistorů na shader (proto oněch 681 milionů). Uvážíme-li, že ve 3D teoreticky jedou všechny shadery naplno, je jasné, odkud se bere nenasytnost karty.

4.1.1.2 Struktura GPU

Složitou architekturu jádra G80 znázorňuje obrázek 1. Čerpáno z [1].

Data od procesoru do karty přenáší PCI-E sběrnice, na obrázku značená jako *Host. Input assembler* se stará o příjem geometrických dat, konvertuje je do FP32 (32bitová přesnost) a provádí též indexování.

Indexování se hodí při provádění opakovaných operací, například při instancingu, což je generování mnoha objektů z jedné vstupní dat, například listy na stromech. Zelené jednotky jsou označovány jako scalar stream procesory (procesory pro skalární proudové zpracování dat) – na obrázku SP. Právě ony jsou srdce karty, které se stará o všechny shaderové operace. Z obrázku je vidět, že tyto stavební kameny jsou sdruženy do bloků po šestnácti, přičemž každý blok sdílí jednotky pro adresování textur a filtrování textur a blok L1 cache. Frekvence těchto procesorů je 1,35GHz.



Obr. 2 - Struktura GPU u G80

Dle nVidie tak má nejvýkonnější karta současnosti hrubý výkon cca 520 gigaflops⁸⁾. Pro srovnání – jeden z nejvýkonnějších procesorů pro PC, Intel Core2 Extreme x6800 má výkon zhruba 47 gigaflops. Tato čísla je nutno brát s rezervou, spíše než čísla je zajímavý fakt, že dnešní GPU poskytují o mnoho vyšší výkon, než CPU.

V architektuře G80 má každá ROP jednotka 64bitový přístup k pamětem, celková šířka paměťové sběrnice je tedy 384 bitů u nejrychlejší verze čipu. Karty předchozí generace měly maximální šířku 256 bitů, propustnost dat se tedy opět navýšila k závratným číslům.

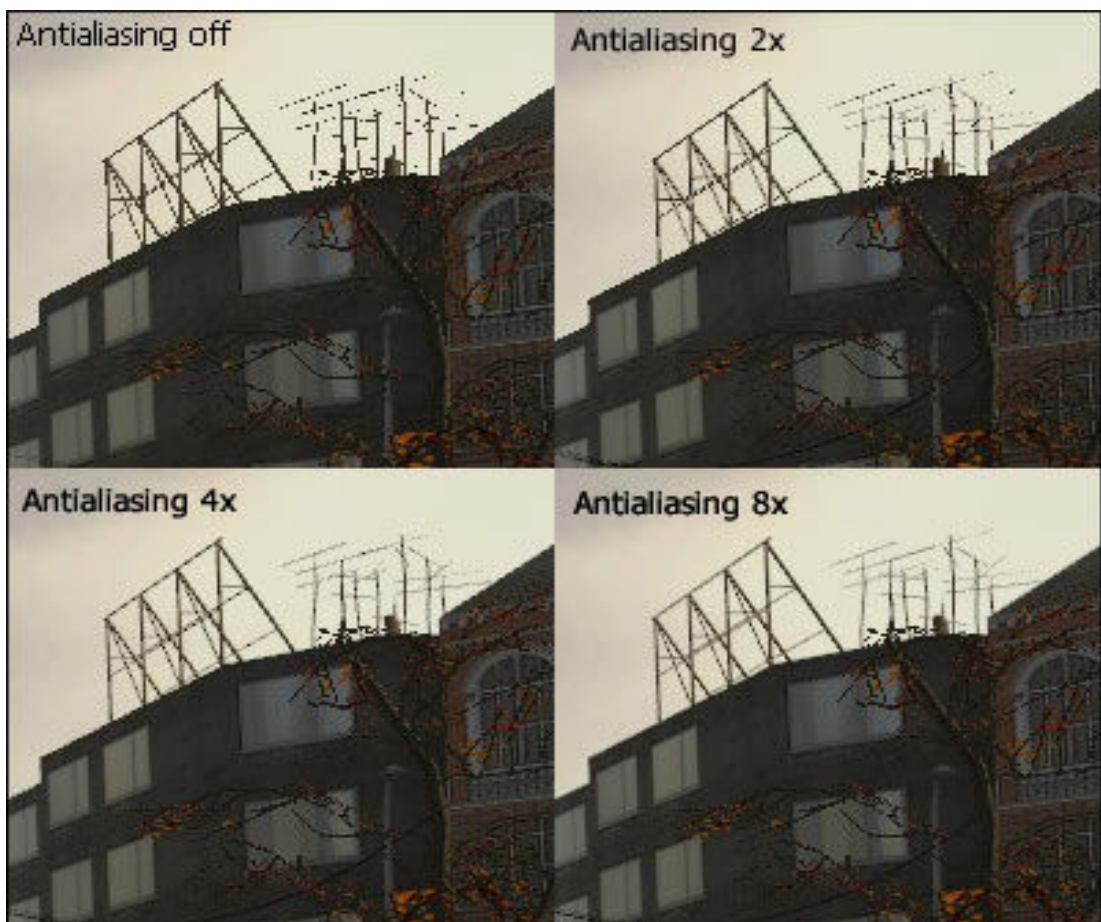
4.1.1.3 Vyhlazování a filtrování obrazu

4.1.1.3.1 Full Scene Antialiasing

V předchozích kapitolách se práce příliš nevěnovala vyhlazování hran (FSAA) obrazu, které do domácích počítačů přinesla společnost 3Dfx. Reálné užití antialiasingu bylo možné pouze u high-endových karet a to v omezené míře, jelikož aplikace vyhlazování je velmi náročná a málokterý čip byl pro to schopen poskytnout výkon.

FSAA je vyhlazení zubatých hran, jenž vznikají při filtrování scény, čímž dosáhneme mnohem lépe vypadající obraz bez ostrých zubů. Patrné jsou při pohledu z různých úhlů či při pohledu z větší dálky. Je logické, že čím vyšší rozlišení a vyšší hodnota antialiasingu, tím kvalitnější obraz.

Po prozkoumání detailů střechy a antén z obrázku 3, lze lehce usoudit, jak jsou v případě použití FSAA 4x zuby na hranách objektu téměř „vyhlazené“.



Obr. 3 - Rozdílné hrany při použití FSAA u hry Half-Life 2

4.1.1.3.1.1 Adaptivní transparentní antialiasing

Klasický antialiasing má tu nevýhodu, že za pomoci současných metod neumí vyhlazovat transparentní textury. Nejčastěji to je tráva, listí stromů, ploty a jiné objekty, které mají velmi jednoduchou geometrii. Proto se místo jejich modelování použije transparentní (průhledná) textura.

Tato nová metoda, představena u G70, bere v úvahu alpha složku dané textury, která určuje její průhlednost. Pokud takový objekt najde, použije na něj antialiasing, i když se jedná o vnitřní část textury.

4.1.1.3.1.2 Antialiasing u G80

Díky novému algoritmu, který byl nazván Coverage Sampling, je možné aplikovat 16x CSAA, přičemž nVidia tvrdí, že 16xQ mód má dosahovat podobného počtu snímků za vteřinu jako 4x multisampling, ovšem s vyšší kvalitou výsledného obrazu.

Vše asi není ideální, jelikož nVidia píše ve svých materiálech: „v některých částech scény, např. tam, kde jsou použity stencil shadow volume efekty, bude 16x mód nahrazen 4x multisamplingem“. Až v nezávislých recenzích se pozná, jak se bude chovat nový algoritmus v praxi.

4.1.1.3.2 Anisotropic Filtering

Anisotropní filtrování textur ovlivňuje zobrazení vzdálených ploch. Pro zvýšení výkonu jsou plochy, u nichž není příliš důležitý ostrý detail, velmi často zjednodušovány. Každý povrch má pak, v závislosti na vzdálenosti pozorování, různou kvalitu textury povrchu. Obraz se nám u velmi vzdálených jeví jako rozmazaný či jednolitý.

Moderní grafické karty jsou schopny aplikovat až šestnáctinásobné filtrování (16xAF) s praktickým neznatelným dopadem na výkon. Je nutno říci, že rozdíl mezi 4x a 8xAF je téměř neznatelný, natož extrémní 16x AF.

Na obrázku je názorně vidět, jak vypadá hra Half-Life 2 s vypnutým a se zapnutým AF. Detaily zebry a dlažebních kostek mluví za vše.



Obr. 4 - Rozdílná filtrace textur. Vlevo trilineární filtrování, vpravo použité AF 4x

4.1.1.3.2.1 Úhlově optimalizované anisotropní filtrování

Termín úhlově optimalizovaná AF je označení pro formu AF, která při nastavení například AF 16x realizuje plnohodnotnou šestnáctinásobnou filtraci pouze na texturách, které jsou zobrazeny pod určitým úhlem. Textury zobrazené pod jiným úhlem jsou obvykle filtrovány v nižší kvalitě (AF 2x či 4x). K tomu vedou výrobce dva důvody, první je ten, že pro úhlově optimalizovanou filtraci dostačuje jednodušší texturovací jednotka (tím pádem nižší počet tranzistorů v čipu) a druhý je výkon. Žádný grafický čip, který by neměl texturovací jednotku oddělenou od pixel pipeline, není schopen provozovat plnou anisotropní filtraci s akceptovatelně nízkým propadem ve výkonu.

4.1.1.3.3 *High Dynamic Range (HDR)*

High Dynamic Range znamená velký rozsah zobrazovaných barev, rozsah jasu.

Engine pro HDR zahrnuje několik různých faktorů. Pro jeho použití je třeba znát odrazivost povrchů, která může být uložena například ve formě textury, pro kterou se pak napíše nějaký konkrétní program do pixel shaderu. Dále se používají efekty jako bloom, motion blur či depth of field. Bloom efekt rozmazává hrany objektu umístěných před světelným zdrojem, motion blur způsobuje, že rychle se pohybující předměty vypadají věrohodněji, tedy rozmazaně s neostrými hranami. Funkce Depth of Field navozuje efekt zaostření do dálky. Z toho vyplývá, že předměty stojící blízko jsou rozmazané a předměty ve správné vzdálenosti jsou velmi ostré.



Obr. 5 - Rozdílné nasvětlení scény u hry Far Cry při zapnutém HDR

4.1.2 Parametry grafických karet – pojmy

V této kapitole práce objasní pojmy a parametry, které je možno nalézt u většiny popisků grafických karet. Ať už se jedná o elektronický obchod, časopis o hardwaru či recenzi na webových stránkách.

Rozdílné parametry tří grafických karet jsou znázorněny v Tabulce 2, jedná se o nejnovější model nVidie (G80) s předchozí generací (G71) a top modelem společnosti ATI v podání R580+.

Model	8800GTX	7900GTX	X1950XTX
Jádro	G80	G71	R580+
Počet tranzistorů	681 milionů	279 milionů	384 milionů
Počet shaderů	128	8 vertex, 24 pixel	8 vertex, 48 pixel
Počet ROP	6x8	16	16
Šířka paměťové sběrnice	384 bitů	256 bitů	256 bitů
Propustnost paměti	86,4 GB/s	51,2 GB/s	64 GB/s
Frekvence jádra	575 MHz	650 MHz	650 MHz
Frekvence shaderů	1350 MHz	650 MHz	
Frekvence paměti	900 MHz	800 MHz	1000 MHz
Paměť	768 MB GDD3	512 MB GDDR3	512 MB GDDR4
Fillrate	36 800 MTex/s	15 600 MTex/s	10 400 MTex/s
Verze shaderů	4.0	3.0	3.0
Technologie	90nm	90nm	80nm
Zdroj – minimum	450W	350W	350W

Tabulka 2 – Příklad uváděných parametrů grafických karet (zdroj:[8])

4.1.2.1 Počet tranzistorů

Počtu tranzistorů není nutné věnovat příliš pozornosti. S novějšími jádry stoupá jejich počet, jak už z tabulky vyplývá, nejnovější karta nVidia však v tomto ohledu vyčnívá. Je to dáno novou architekturou s unifikovanými shadery. Nová architektura má pozitivní dopad na výkon, avšak negativní dopad na počet tranzistorů. Na jednu jednotku pro výpočet shaderů, je jich potřeba mnohem více.

Nevýhoda většího počtu tranzistorů je zajisté velikost karty a samozřejmě spotřeba.

4.1.2.2 Počet shaderů

Shadery jsou programy, jenž zpracovávají instrukce přímo na grafické kartě a výrazně tak navyšují výkon. Do příchodu jádra G80 od nVidie byla jejich práce přímo určená, každý shader měl určenou jistou činnost. Jak je patrné z tabulky, nově již shadery nejsou specializovány, ale unifikovány. Dělají tedy to, co je právě potřeba. Nárůst jejich počtu spolu s rozdílným pojetím práce dává tušit, že G80 bude při práci se shadery velmi rychlá.

4.1.2.3 Počet ROP

GeForce 8800 GTX má 6 ROP, z nichž každá může zpracovávat najednou 4 pixely. Pokud se zpracovávají zároveň barvy a Z souřadnice, zvládá tedy ROP 24 pixelů za takt. Pokud se pracuje pouze se Z souřadnicemi, je výstup při 4x MultiSamplingu roven 48 pixelů za takt.

4.1.2.4 Šířka paměťové sběrnice

Již od roku 2002 mají nejrychlejší karty každé generace 256bitovou šířku paměťové sběrnice. Teoreticky se dá říci, že s rostoucí šířkou nám geometricky stoupá výkon, tedy při přechodu ze 128bitové šířky na 256bitovou bychom měli dostat dvojnásobný výkon. Praxe je však trochu odlišná, pouze při opravdu velkém rozlišení (tzv. 1600x1200 pixelů), zapnutém FSAA (4x), může být výkon dvojnásobný, jelikož je potřeba přenést obrovské množství dat. Paměťovou sběrnici je de facto „hrdlo“ grafické karty, kterým putují data.

Jak již práce zmiňovala, u karty GeForce 8800 GTX má každá ROP jednotka 64bitový přístup k pamětem. Při počtu šesti jednotek dostáváme neobvyklou šířku sběrnice 384 bitů. (například pomalejší model nové generace – GF 8800 GTS má „pouze“ 5 ROP jednotek a tedy 320bitovou sběrnici).

4.1.2.5 Propustnost paměti

Datová propustnost neboli propustnost paměti úzce souvisí se šířkou paměťové sběrnice a frekvencí paměti. Čím vyšší je průchod sběrnice a čím

rychlejší jsou paměti, tím vyšší je datový přenos. Názorný příklad je o odstavci výše, kde autor přirovnává paměťovou sběrnici hrdlu láhve.

Frekvence paměti je například u testovaného Radeonu X1950XTX 1000 MHz při šířce paměťové sběrnice 256 bitů. Rychlost paměti je nutno vynásobit dvěma, jedná se o efektivní frekvenci díky technologii DDR⁵⁾.

Výpočet:

Efektivní frekvence paměti (MHz) * šířka paměťové sběrnice (B) = datová propustnost (GB*s⁻¹)

Příklad (X1950 XTX): $(1000 \times 2) \times 256 / 8 = 64 \text{ MB} \cdot \text{s}^{-1}$

4.1.2.6 Frekvence jádra

Výše frekvence jádra je závislá pouze na použité technologii výroby a kvalitním chlazení. Stropem 90nm výroby se zdá být 700 MHz (čehož dosáhneme pouze přetaktováním základního nastavení).

Čím vyšší frekvenci jádro má, tím vyšší výkon poskytuje. Rozdílným výkonům při měnící se frekvenci se bude práce zabývat dále.

4.1.2.7 Frekvence shaderů

Tato hodnota nebyla u grafických karet před pár měsíci uváděna, proto se jí práce nebude věnovat, je však logické, že čím vyšší frekvence bude, tím efektivněji a rychleji budou shadery pracovat. Rozdílná hodnota o více než dvounásobek u nového jádra je vypovídající.

4.1.2.8 Frekvence paměti

S příchodem technologie DDR se výkon a rychlost paměti zdvojnásobila, dnešním standardem je typ DDR3.

Rychlé DDR3 paměti u grafických karet jsou díky rychlému časování (1,1ns) schopné pracovat na velmi vysokých frekvencích. Čím nižší časování, tím vyšší maximální frekvence. Nejnovější paměti DDR4 Radeon X1950XTX pracují při časování 0,9ns na 1000 MHz (2000 MHz efektivně), což je proti grafickým kartám před dvěma lety dvounásobný skok. Maximální možné frekvence se odvíjejí od časování, přičemž platí:

$1000 \text{ MHz} / \text{časování paměti (ns)} = \text{maximální frekvence (MHz)}$

Příklad (X1950XTX): $1000 \text{ MHz} / 0,9 \text{ ns} = 1100 \text{ MHz}$

Paměti u toho Radeonu jsou tedy schopny pracovat až na 1100 MHz (2200 MHz efektivně).

4.1.2.9 Velikost paměti

Před třemi lety kolovaly internetem diskuze, že 256 MB paměti u grafické karty je příliš a karta je nedokáže efektivně využít. Čas pokročil a

nejnovější počín nVidie má rovných 768 MB, což už se téměř blíží paměti operační u běžného osobního počítače.

Je pravdou, že byly vydány některé hry, které dokáží naplnit 512 MB paměti a využít tak potenciál, mají však minoritní podíl na trhu a jejich procento závratně nestoupá.

4.1.2.10 Fillrate

Hodnota Fillrate nám udává, kolik pixelů je schopna grafická karta generovat za jednu vteřinu. Tento počet ovlivňuje počet texturovacích jednotek, nové jádro tím ukazuje svůj velký výkonový potenciál.

4.1.2.11 Verze shaderů

Nejnovější DirectX 10 představily novou, čtvrtou verzi shaderů, nová grafická karta ji samozřejmě podporuje, je ovšem otázkou, zda-li do té doby, než se začne verze 4.0 aplikovat, nebude karta zastaralá. Ve skutečnosti je dostačující verze 3.0 z prostého důvodu. Stávající karty stále nevyužívají plného potenciálu třetí řady, čili je zbytečné trvat na grafické kartě, která podporuje Shadery 4.0. nVidia jen dokázala svůj technologický náskok, v praxi se však nové shadery objeví za dlouho dobu.

Verze 3.0 je ve dnešní situaci standardem, přičemž u nich platí, že dokáží zobrazovat efekty stejně jako verze 2.0 (pouze pár efektů navíc), ale umí je zobrazovat rychleji.

4.1.2.12 Výrobní technologie

Pokrok výrobních technologií je více méně pravidelný, každé dva roky se stávající proces zdokonalí a pokročilejší technologie umožní zvyšovat frekvence či počet tranzistorů při nezměněné velikosti karty.

4.1.2.13 Zdroj – minimum

Jak práce již mnohokrát naznačila, nové karty se nemohou chlubit obrovskou spotřebou, kterou mají. Při koupi nové karty je potřeba pozorně přečíst technické specifikace, zda-li bude kartě dostačovat současný zdroj. U zdrojů se rozhodně vyplatí nešetřit, značkové zdroje s výkonem 400W je mnohdy mnohem výkonnější a lepší, než neznačkový zdroj o výkonu 500W.

4.2 Porovnání výkonu

V této kapitole bude práce porovnávat výkony vybraných akceleratorů v několika kategoriích, z nichž učiní závěry a případná doporučení pro daný segment. S ohledem na požadavky běžného uživatele se práce bude zabývat kartami pro běžný *kancelářský provoz*, kartami pro běžný multimediální počítač (*mainstreamové karty*) a *high-end karty* pro velmi náročné hráče.

Jednotlivé testy budou probíhat ve dvou fázích, první je syntetický, pomocí benchmarku. Jedná se speciální program 3DMark 06 vyvinutý společností FutureMark pro testování grafických karet. Obsahuje v sobě několik testů, které simulují herní prostředí s maximálním důrazem na využití veškerých možných efektů. Výsledek je bodový, čili je možné porovnat jednotlivé hodnoty a vyvodit závěry. Výhodou je, že výsledek neovlivňuje použitý procesor. Výsledky však mnohdy nejsou příliš vypovídající, již v minulosti si vývojové společnosti pomáhali ulehčením pomocí softwarových ovladačů, čili hlavním znakem výkonového potenciálu jsou hry.

Druhá fáze je reálné testování výkonu pomocí několika vybraných titulů, které k tomuto účelu výborně slouží. Svojí náročností a použitím nejnovějších efektů, umožňují roztřídit velmi často vyrovnanou řadu sobě si konkurujících výrobků.

Výsledné hodnoty jsou udávány ve snímcích za vteřinu (frames per second) neboli FPS. Čím vyšší číslo, tím rychleji tedy hra poběží a tím lepší zážitek a požitek z ní bude uživatel (hráč) mít. Je také důležité se zmínit, že výsledky jsou udávány v průměrných snímcích za vteřinu, minimální hodnoty (tedy ty kritické momenty) bohužel nejsou v některých testech zařazeny. Minimální průměrnou hodnotou pro velmi dobrý zážitek je číslo 30 FPS (optimální hodnota je 50 FPS, kde je rezerva pro složitější výpočet scén). Pokud hodnota FPS klesá pod 25, obraz se stává trhaným a hra tedy velmi „špatně hratelnou“.

Seznam her a jejich nastavení uvede práce k příslušnému testu, jelikož každý segment má jiný požadavek a jiné možnosti. Výběr se snažil pokrýt celé spektrum rozdílných rozhraní, použitých efektů či odlišných technologií pro výrobu her (enginů).

Veškeré hodnoty jsou čerpány ze zdrojů [1-7].

4.2.1 Karty vhodné do kanceláře

Uživatel, jenž je spíše svátečním hráčem a pracuje pouze v kancelářských programech má jiné preference než „hardcore hráč“.

Důležitým aspektem je hlučnost karty, respektive jejího chladiče. Nejlepší možností je pasivní chlazení a tím pádem nulový hluk, který je nespornou výhodou při práci s programy.

Vybrány byly jedny z nejlevnějších karet od obou hlavních vývojových společností na trhu, tedy od ATI a nVidie, vybavené pasivním chlazením. Integrované řešení nebylo vybráno z jednoduchého důvodu. Pro hraní her jsou naprosto nedostačující a nabízejí sotva poloviční výkon karet, které jsou vybrány v tomto segmentu. Integrovaná řešení jsou proto použitelná pouze do pracovních stanic.

4.2.1.1 Konfigurace

Procesor	Athlon64 3200+ (2GHz)
Operační paměť	1 GB DDR 400MHz
Základní deska	VIA KT890
Ovladače	ATi Catalyst 6.11 nVidia ForceWare 93.71
Operační systém	Windows XP Professional SP2

Vybrána byla levná a „zastaralá“ sestava, která odpovídá cenové kategorii testovaných karet.

4.2.1.2 Parametry grafických karet

Model	Radeon X1300PRO	GeForce 7300GT
Jádro	RV15	G73
Počet tranzistorů	101 milionů	177 milionů
Počet shaderů	2 vertex, 4 pixel	4 vertex, 8 pixel
Počet ROP	4	4
Šířka paměťové sběrnice	128bitů	128 bitů
Propustnost paměti	12,8 GB/s	10,7 GB/s
Frekvence jádra	600	350
Frekvence paměti	400 MHz	333 MHz
Paměť	256 MB GDD3	256 MB GDDR3
Fillrate	2400 MTex/s	1400 MTex/S
Verze shaderů	3.0	3.0
Výrobní technologie	90nm	90nm
DirectX	9.0c	9.0c
Přibližná cena (duben '07)	2100 Kč	2300 Kč

4.2.1.3 Testy

Syntetické

3DMark06 (patch 1.0.2)

DirectX

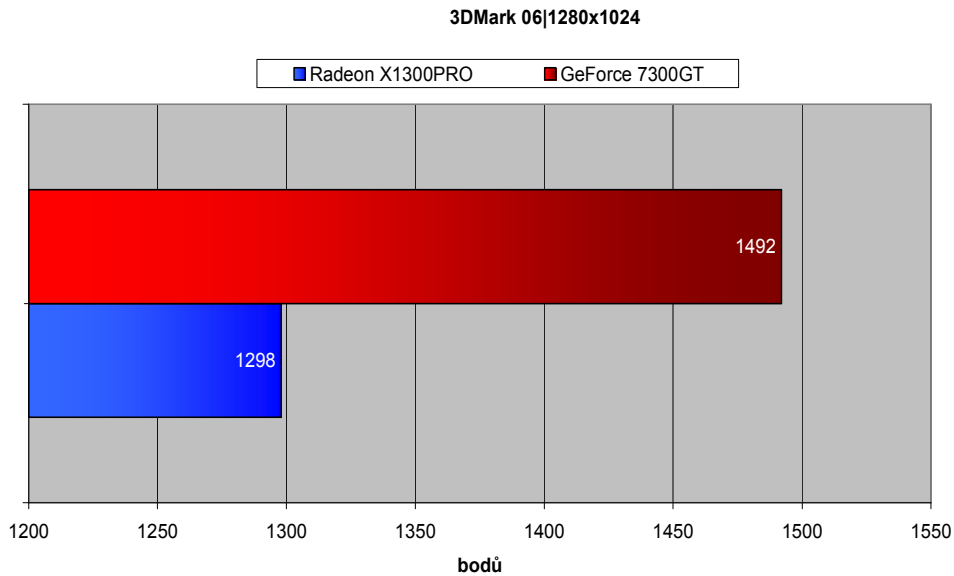
Far Cry (patch 1.32, Research)

F.E.A.R. (default)

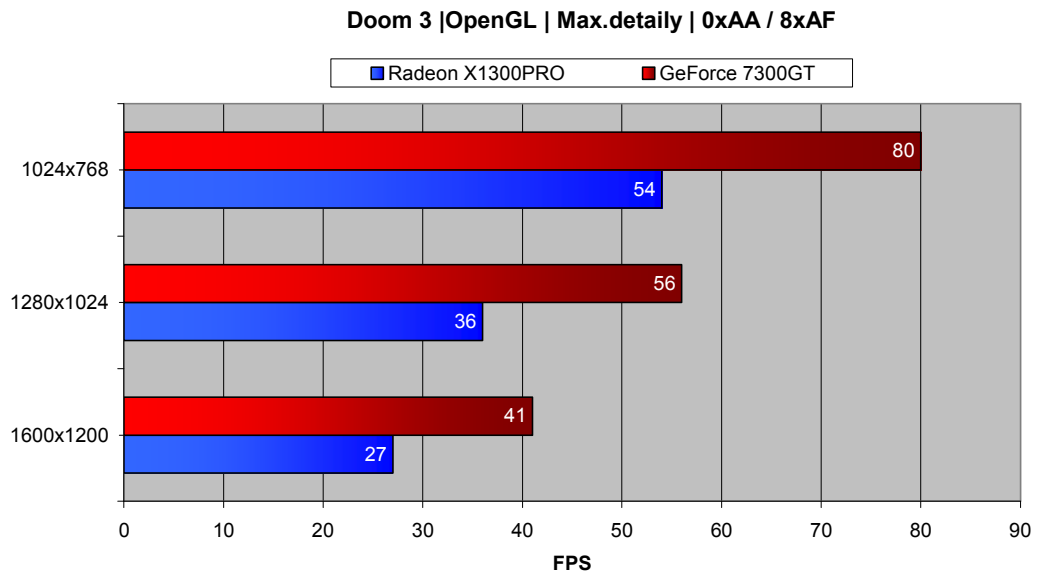
The Elder Scrolls IV: Oblivion

OpenGL

Doom 3 (Timedemo1)

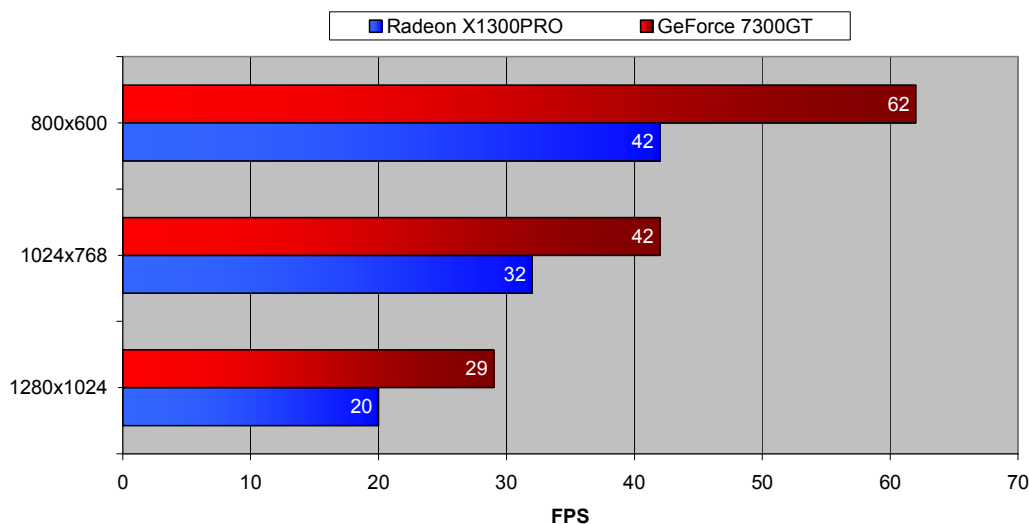


3DMark 06 nahrává do karet 7300GT, výhodou je zde vyšší počet pixel pipeline oproti Radeonu.



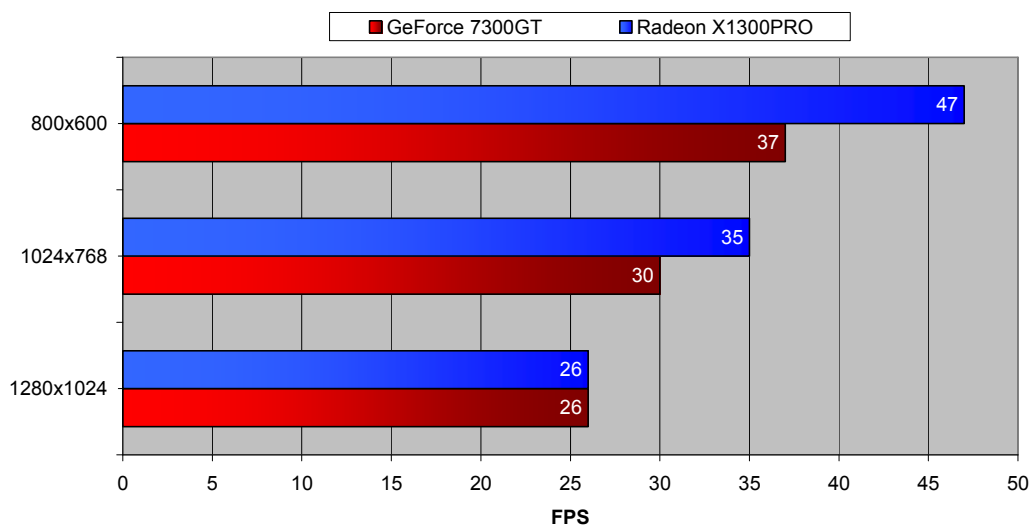
Na současné generaci low-endových grafických karet je Doom 3 velmi dobře hratelný a jak z grafu vyplývá, obzvláště na 7300GT hra běží velmi svižně. Je to dáno rozhraním OpenGL ve kterém je nVidia až na výjimky rychlejší.

F.E.A.R. |Direct3D | Střední detaily | 0xAA / 0xAF



Velmi náročná hra plná pokročilých a náročných efektů, které si zde vybírají svou daň. Na přijatelné rozlišení 1024x768 je hra při středních detailech hratelná na obou kartách, na vyšší detaily či rozlišení je však nutno zapomenout.

Oblivion |Direct3D | Nízké detaily | 0xAA / 0xAF



Nejnáročnější hra současnosti se dá na nízké rozlišení rozpohybovat na obou kartách, velmi dobře si zde vede Radeon X1300PRO. Bohužel je možné hrát pouze na nejnižší možné detaily, čímž hra ztrácí na své atraktivitě, najdou se však i tací, kteří nehledí pouze na dokonalou grafiku.

4.2.1.4 Závěr

Obě karty potvrdily to, co se od nich očekávalo. Pro občasné hraní jsou vyhovující, na nízké detaily a nízké rozlišení jsou schopny spustit téměř všechny hry. Závratně vysokých snímků za vteřinu však nejsou schopny a pro časté hráče tak mohou být zklamáním.

Tento druh karet by měl tedy koupit každý, komu nestačí integrovaná grafika pro občasné hraní nebo ten, kdo plánuje koupit nového počítače a má velmi omezený rozpočet.

4.2.2 Mainstream

Mainstream je označení pro střední výkonovou třídu neboli „zlatou střední cestu“. Svým výkonem tyto karty nijak neoslňují, nemají však problém s jakoukoliv hrou, pokud je uživatel ochoten snížit maximální detaily. Optimální rozlišení pro tyto karty bývá 1280x1024, ve kterém by měly být schopny na střední až vysoké detaily poskytnout velmi slušný požitek z hraní.

Některé z těchto karet již vyžadují přídavné napájení, spotřeba však není nikterak strašlivá a obyčejné zdroje ve většině případů postačí. Aktivní chladič se samozřejmostí, karty již mají vcelku slušný výkon, a tak velmi silně „topí“.

4.2.2.1 Konfigurace

Procesor	Intel Core 2 Duo E6600 2.4GHz
Operační paměť	Corsair PC2-6400 2GB DDR2
Základní deska	Asus P5N32-SLI
Ovladače	ATi Catalyst 6.11 nVidia ForceWare 93.71
Operační systém	Windows XP Professional SP2

Pro silnější grafické karty je potřeba silný procesor, jinak zůstane jejich výkonnostní potenciál nevyužit. Volba na Intel Core 2 Duo E6600 2.4GHz padla z důvodu, že se jedná o jeden z nejlepších procesorů vůbec.

4.2.2.2 Parametry grafických karet

Model	X1650XT	7600GS	7600GT	7900GS	X1950PRO
Jádro	RV560	G73	G73	G71	RV570
Počet tranzistorů	330 mil.	177 mil.	177 mil.	278 mil.	329 mil.
Počet shaderů	8 vertex, 24 pixel	5 vertex, 12 pixel	5 vertex, 12 pixel	7 vertex, 20 pixel	8 vertex, 36 pixel
Počet ROP	8	8	8	16	12
Šířka paměťové sběrnice	128bitů	128 bitů	128 bitů	256 bitů	256 bitů
Propustnost paměti	21,6 GB/s	12,8 GB/s	22,4 GB/s	42,2 GB/s	44,2 GB/s
Frekvence jádra	575 MHz	400 MHz	560 MHz	450 MHz	575 MHz
Frekvence paměti	675 MHz	400 MHz	700 MHz	660 MHz	690 MHz
Paměť	256 MB GDD4	256 MB GDDR3	256 MB GDDR3	256 MB GDDR3	256 MB GDDR4
Fillrate	4600 MTex/s	4800 MTex/s	6720 MTex/s	9000 MTex/s	6900 MTex/s
Verze shaderů	3.0	3.0	3.0	3.0	3.0
Výrobní technologie	80nm	90nm	90nm	90nm	90nm
DirectX	9.0c	9.0c	9.0c	9.0c	9.0c
Přibližná cena (duben '07)	3500 Kč	2000 Kč	3500 Kč	4200 Kč	4200–5000 Kč

4.2.2.3 Testy

Syntetické

3DMark06 (patch 1.0.2)

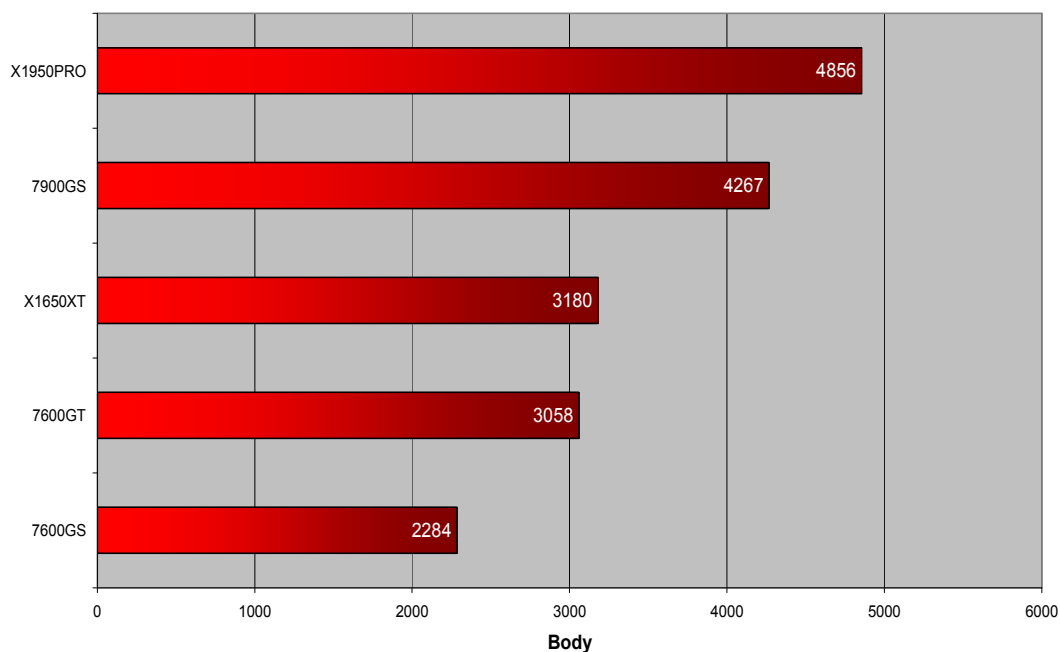
DirectX

Far Cry (patch 1.32, Research)

F.E.A.R. (default)

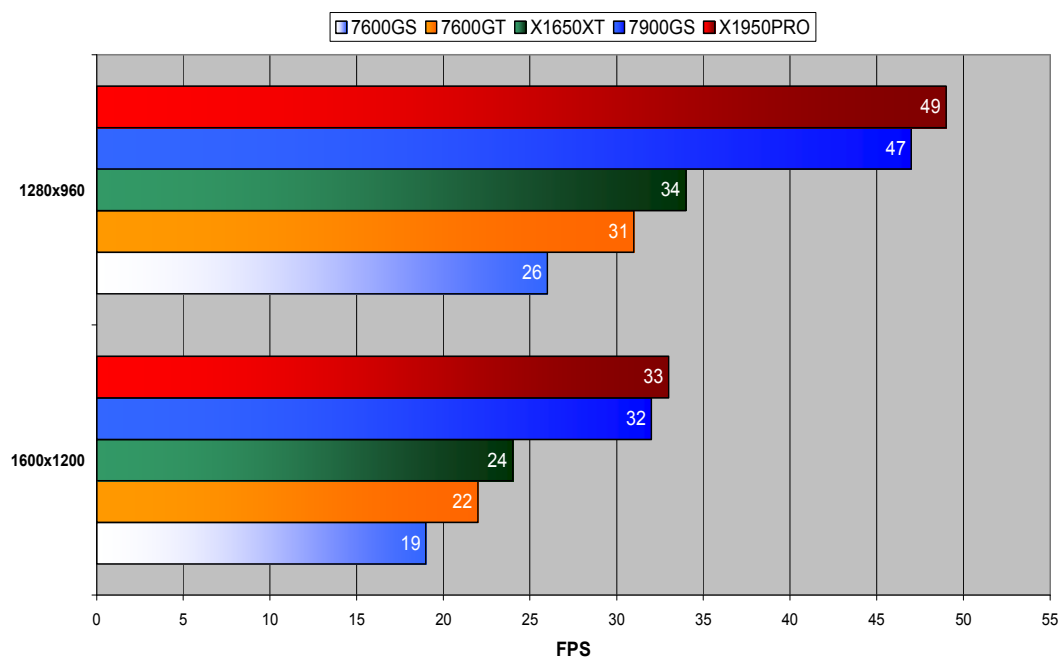
The Elder Scrolls IV: Oblivion

3DMark 06 | 1280x1024

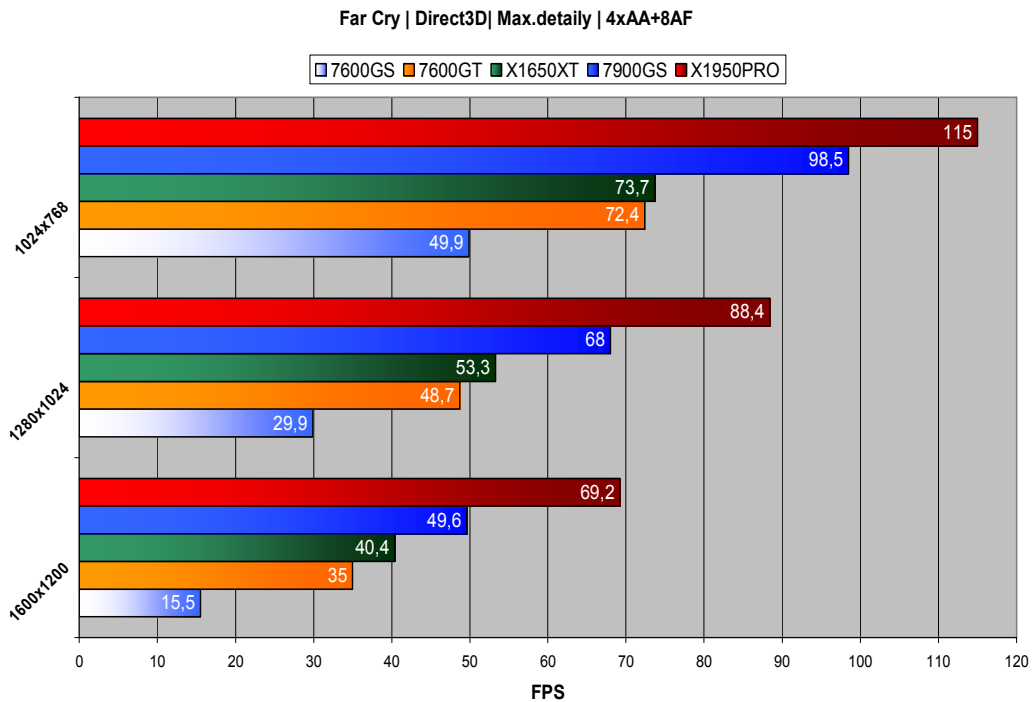


Výkony v 3Dmarku 06 již napovídají tomu, že výkonnost těchto karet je mnohem vyšší. Teoretická hrubá síla Radeonu X1950PRO se projevila i v praxi, působí na tomto poli víceméně suverénně. Oproti tomu je vidno, že je grafická karta 7600GS díky velmi pomalým taktům hluboko na chvostu.

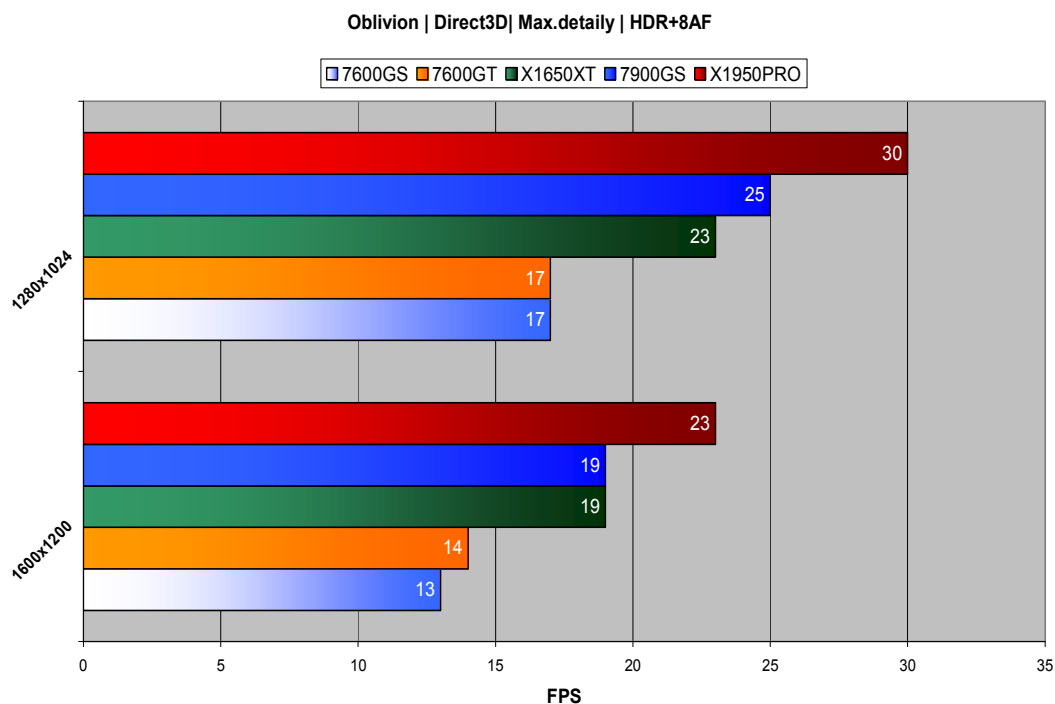
F.E.A.R. | Direct3D | Max.detaily | 4xAA+16AF



Vysoké rozlišení, maximální detaily a velmi náročné vyhlazování jasně ukázalo méně výkonné karty. X1950 i 7900GS poskytují vysoký výkon i při takto náročném nastavení, po zvýšení rozlišení na 1600x1200 však ztrácejí i ony. Na kartách typu 7600 jsou tato rozlišení problém, pro lepší hratelnost je nutné snížit detaily či vypnout (snížit) vyhlazování obrazu.



Střílečka z tropického ostrova, která okouzila celou herní scénu působivou grafikou, již nepředstavuje takový problém pro grafické karty, jako před třemi lety. Na maximální nastavení je to stále pastva pro oči, na základní rozlišení 1024x768 ji zvládají veškeré karty s přehledem. Zvyšování rozlišení odděluje karty s nižší frekvencí pamětí či nižší paměťovou sběrnicí, což se maximálně projevuje při rozlišení 1600x1200. V tomto rozlišení při nastavení maximálních detailů je potřeba přenést velké množství textur, což se stává osudné pro 7600GS.



Skvostné RPG (Role Playing Game), plným názvem The Elder Scrolls IV: Oblivion vyšlo před necelým rokem, ale rozumného počtu FPS při nejvyšších detailech, je u něj stále těžké dosáhnout s jakoukoliv kartou. Je otázkou, zda je engine tak náročný, či prostě neoptimalizovaný. Při pohledu na grafické podání Oblivionu je však tolerována i velká náročnost.

Na testovaná nastavení se uživatel příliš plynulé hry nedočká od žádné karty, opět zde vyčnívá Radeon X1950PRO, jenž si zachovává odstup při všech rozlišení.

4.2.2.4 Závěr

Ve všech výkonnostních testech s přehledem zvítězila grafická karta Radeon X1950PRO. Nejnáročnější hry pro ni nejsou velkou překážkou a při snížení detailů či vypnutí vyhlazování počet snímků stoupá velmi strmě. S ohledem na cenovou relaci je tato karta velmi výhodná, jelikož není o mnoho dražší než konkurenti. Pro mainstream je tato karta jistou koupí.

Naopak jistým zklamáním může působit výkon 7600GS. V testu byla uvedena verze s pamětí taktovanou na 400 MHz, v prodeji jsou však rychlejší verze s pamětí na frekvenci až 700 MHz, čímž se blíží výkonu 7600GT. Pro toho, kdo má hluboko do kapsy, je jasným favoritem Radeon X1650XT, který se za svůj výkon stydět nemusí a svou cenou mile překvapí. Středním proudem tedy vládne kanadská společnost ATI.

4.2.3 Nejrychlejší grafické karty současnosti – High-end

Maximální výkon, vysoká cena. Tyto vlastnosti charakterizují nejvyšší třídu grafických karet. Jsou exkluzivním zbožím, vývojové společnosti těmito kartami demonstrují sílu své současné výrobní a programové technologie a předhánějí se tak o místo na pomyslném trůnu na trhu grafických karet.

Pro tyto karty je potřebný velmi výkonný počítač s velmi rychlým procesorem, jenž dokáže dodávat potřebné množství dat grafickému jádru. V dnešní situaci neexistuje procesor pro osobní počítače, který by dokázal zásobovat kartu dle jejích požadavků a jak práce naznačila, výkony grafických čipů jsem o řád výše než výkony procesorů pro osobní počítače. Proto jsou testy prováděny na maximální rozlišení pro jasné rozčlenění pole.

4.2.3.1 Konfigurace počítače

Pro test toho segmentu byl zvolen prakticky nejrychlejší dvoujádrový procesor současnosti, Intel Core Duo2 E6600, přetaktovaný na frekvenci 3,2 GHz, velmi rychlé paměti i kvalitní základní deska.

Procesor	Intel Core 2 Duo E6600 3.2GHz
Operační paměť	2x1GB Corsair Dominator XMS2 1145MHz
Základní deska	Asus Striker (680i)
Ovladače	ATi Catalyst 6.11 nVidia ForceWare 93.71
Operační systém	Windows XP Professional SP2

4.2.3.2 Parametry grafických karet

Výběr karet byl jednoduchý. Čtyři nejrychlejší karty současnosti, podporující nejnovější technologie, mající velikou a rychlou paměť (7950GT2 obsahuje dvě jádra). Zároveň se práce bude zabývat výkony variant multislottedového řešení. SLI v podobě nVidie a CrossFire u ATI. Obě řešení se skládají ze dvou identických karet v počítači. Cena je tedy dvojnásobná při zachování stejných parametrů.

Zde jsou jejich parametry karet použitých v testu.

Model	Radeon X1950XTX	GeForce 7950GT2	GeForce 8800GTS	GeForce 8800GTX	GeForce 7900GXT
Jádro	R580+	2xG71	G80	G80	G71
Počet tranzistorů	384 mil.	278 mil.	681 mil.	681 mil.	279 mil.
Počet shaderů	8 vertex, 48 pixel	16 vertex, 24 pixel	96	96	8 vertex, 24 pixel
Počet ROP	16	16	20	20	16
Šířka paměťové sběrnice	256 bitů	256 bitů	320 bitů	384 bitů	256 bitů
Propustnost paměti	64 GB/s	38,4 GB/s	64 GB/s	84,7 GB/s	51,2 GB/s
Frekvence jádra	650 MHz	500 MHz	500 MHz	575 MHz	650 MHz
Frekvence paměti	1000 MHz	600 MHz	900 MHz	900 MHz	800 MHz
Paměť	512 MB GDD4	512 MB GDDR3	640 MB GDDR3	768 MB GDDR3	512 MB GDDR3
Fillrate	10 400 MTex/s	2x 12 000 MTex/s	24 000 MTex/s	36 800 MTex/s	15 600 MTex/s
Verze shaderů	3.0	3.0	4.0	4.0	3.0
Výrobní technologie	80nm	90nm	90nm	90nm	90nm
DirectX	9.0c	9.0c	10	10	9.0c
Přibližná cena (duben '07)	10 000 Kč	11 000 Kč	12 000 Kč	15 000 Kč	11 000 Kč

4.2.3.3 Testy

Syntetické

3DMark06 (patch 1.0.2)

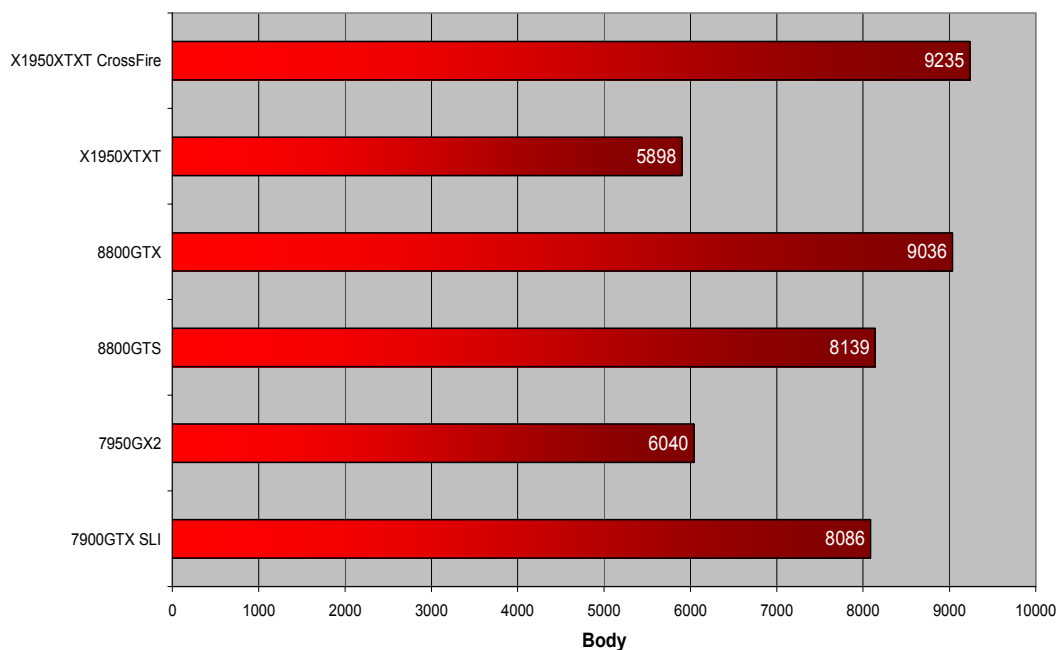
DirectX

Far Cry (patch 1.32, Research)

F.E.A.R. (default)

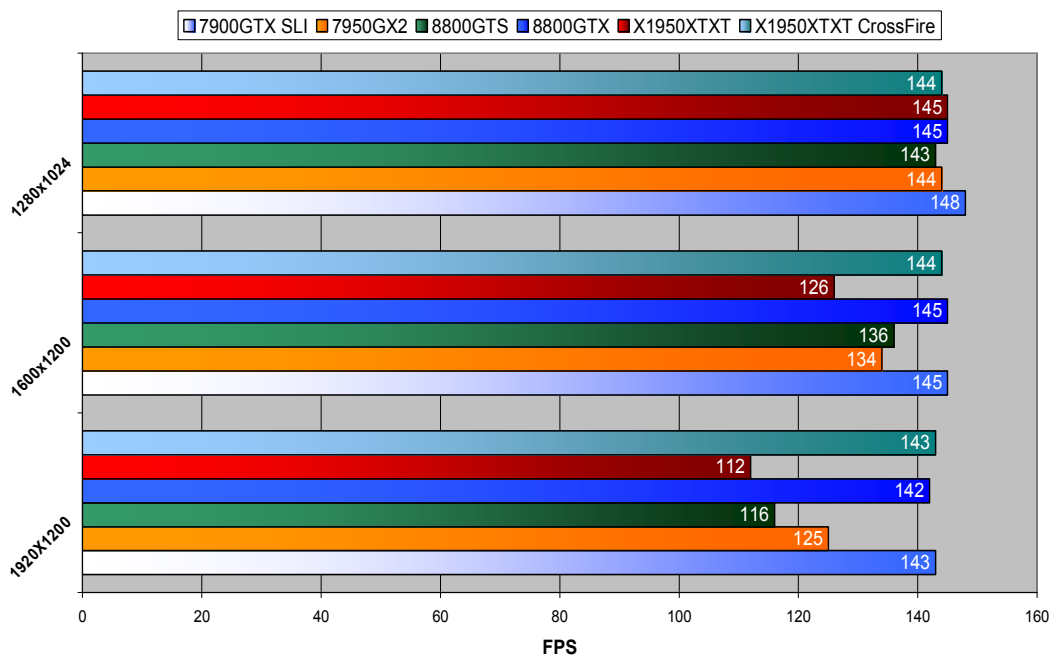
The Elder Scrolls IV: Oblivion

3DMark 06 | 1280x1024

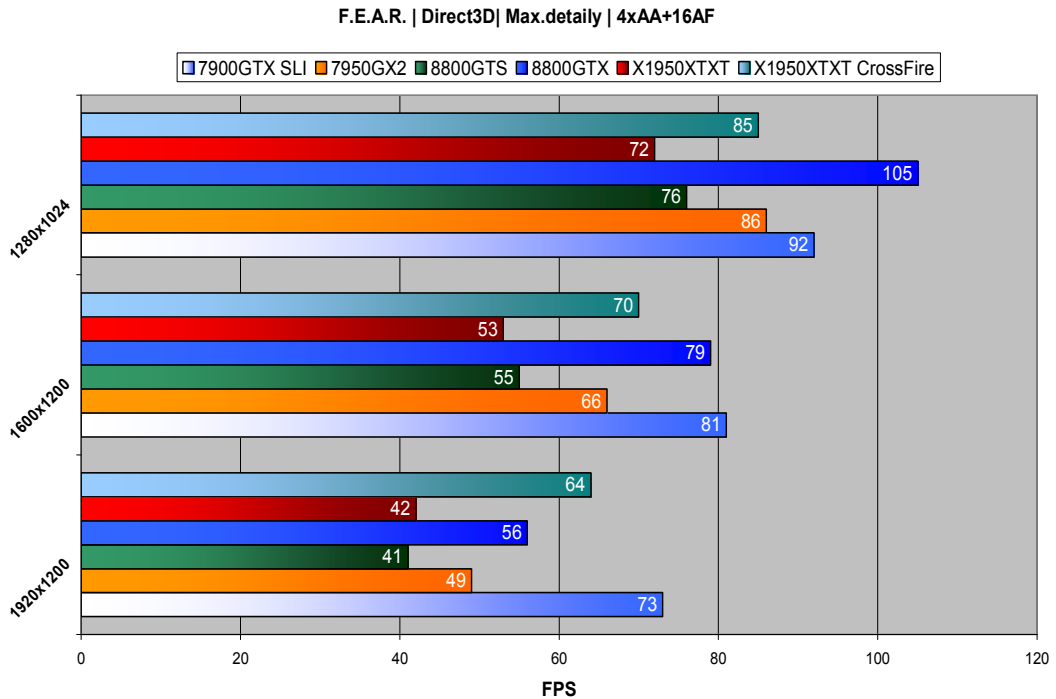


Hrubá síla těchto karet je skutečně impozantní. Rozdíl mezi generacemi je vcelku výrazný (3 000 bodů), nové grafické karty jsou soupeřem pouze ultimátní řešení, v podobě použití 2 nejrychlejších karet předchozí generace.

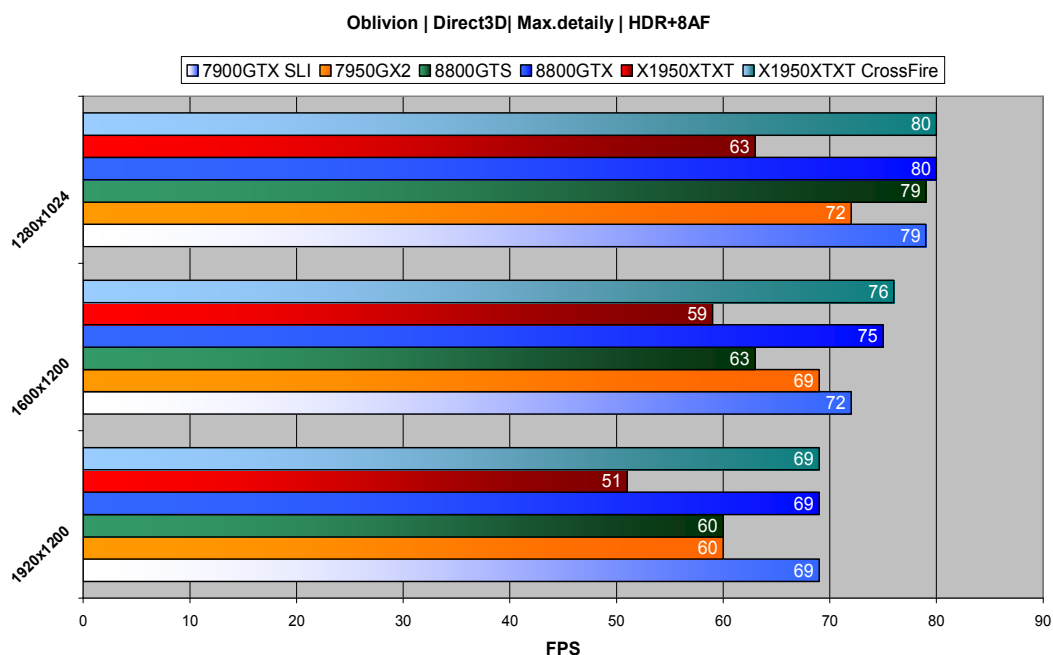
Far Cry | Direct3D | Max.detaily | 4xAA+8AF



Far Cry již delší dobu nepatří mezi nejnáročnější hry současnosti, což lze vyčíst z grafu. Na všech kartách maximálně plynulý na maximální možné details. Limit je v tomto případě pouze procesor, což je názorné u rozlišení 1280x1024, kdy karty vykreslují stejný počet FPS.



F.E.A.R. již dokáže karty „zahřát“. Se vzrůstajícím výkonem si velmi dobře vede multislottedé řešení od společnosti nVidia, následované CrossFire od ATI, které v tomto testu poráží i nový čip. Ten si však oproti minulé generaci drží slušný odstup, pomalejší derivace G80 – karta 8800GTS je s minulou generací výkonově totožná.



Oblivion, jeden z nejnáročnějších titulů řádně prověří všechny karty. S příchodem nové generace toto tvrzení už tak neplatí, i v extrémním rozlišení jako je 1920x1200 je hra nádherně plynulá a kartě nedělá žádný problém. Stejný výkon obou multislotových zařízení je pravděpodobně dán limitem procesoru (i přesto že byl použit nejrychlejší dvoujádrový procesor současnosti). Zajímavé by spíše byly hodnoty minimálních snímků za vteřinu, které jsou taktéž limitujícím a ukazujícím faktorem. Pro tento test bohužel nebyly k dispozici.

4.2.3.4 Závěr

Nová generace přinesla nové limity ve výkonu v DirectX 9.0c. Průměrně je 8800GTX až o 20% výkonnější než karty předchozí generace, je si však nutno uvědomit, že tento výkonnostní vzrůst není pouze v extrémním nastavení, jaké bylo použito v testu, ale při použití běžného rozlišení (1024x768) budou rozdíly propastnější. DirectX 10 k nám prakticky ještě nedorazil, čili výkon v DX10 nelze hodnotit a jedná se tak pouze o papírovou výhodu.

Pomalejší verze čipu G80 má téměř shodný výkon jako řada předchozí (G70), což je velké pozitivum, jelikož po zlevnění se tato velmi výkonná karta stane dostupnou i pro běžné uživatele.

Nejlepší volbou je GeForce 8800GTX. Má obrovský výkon v DirectX 9.0c, který je až 2x výkonnější než top-modely konkurence, má velikou paměť, nový typ architektury, jejíž lepší využití určitě přinese výkon navíc a má plnou podporu DirectX 10.

5. Očekávaný vývoj grafických karet

Vývojové společnosti v oblasti grafických karet, tedy nVidia, ATI a de facto i Intel se snaží každý rok uvádět výkonnější model.

5.1 Intel

U společnosti Intel je situace odlišná, ve skutečnosti uzme největší procentuální část z trhu z velmi prostého důvodu. Do většiny kancelářských počítačů se osazuje integrovaná grafická karta a právě těmto řešením se Intel věnuje. Zaměřuje se tak hlavně integrované čipy, tedy produkty, které budou velmi dobře zapadat do jeho filosofie (periferie – tedy komplexní soustava zařízení, nejlépe od Intelu). Neuvádí tedy každým rokem nový hit, ale neustále vyvíjí nové grafické čipy, které již nejsou pro smích jako před pěti či šesti lety.

5.2 nVidia

Nejmenovaným, ale zasloužilým králem je společnost nVidia, která je opět, po zaváhání ve vývoji nových karet v roce 2002, výrobcem číslo jedna. Již půl roku má na trhu nejvýkonnější čip, který drtí konkurenci. V prvním čtvrtletí roku 2007 byla ohlášena další nová verze jádra G80.

nVidia je první společností, která má grafickou kartu, jenž má plnou podporu DirectX 10. Jelikož pro novou verzi rozhraní nejsou vydány žádné tituly, veškeré optimalizace budou prováděny právě na hardwaru od nVidie, což má obrovské výhody. Vývojáři se tak mohou poučit z chyb ještě před vydáním herních titulů a mají tak čas a příležitost případné problémy odstranit. Zároveň bude software (hry, programy) optimalizován právě pro jádro G80 a vyšší, čímž se dosáhne vyššího výkonu.

Veškeré trumfy má tak ve svých rukou nVidia a pokud „neusne na vavřínech“ a bude se dál velmi pilně věnovat vývoji, ostatní firmy ji mohou velmi těžko ohrozit.

Poslední vydané karty, derivace čipu G80 spadající do mainstreamu, se dle prvních testů neukázaly v příliš dobrém světle, čili tomuto segmentu by se nVidia měla začít naplno věnovat, jelikož síla středního proudu jí může přinést dobré finanční výsledky a navýšení reputace.

5.3 ATI

Kanadská společnost ATI byla v červenci roku 2006 odkoupena firmou AMD, druhého největšího výrobce procesorů. Spolupráce těchto dvou vývojových společností by mohla přinést oživení trhu v podobě výkonného čipu, jenž by zpět přinesl tvrdý konkurenční boj, z něhož vždy bude profitovat potenciální kupec.

Poslední novinky ATI byly vždy o krok za nVidií. Poslední kartou, která byla o krok napřed, byl Radeon 9700Pro a to už je pět let zpátky. ATI (AMD) musí co nejdříve vyvinout čip, který bude plně podporovat DirectX 10. Jelikož nástup nVidie byl v tomto směru poměrně drtivý a tento čip již má půl roku na trhu, AMD se ocitá v těžké situaci.

Nejsilnější modely svých grafických karet prodává AMD za velmi nízkou cenu, čímž strhává zákazníky na svou stranu, je ovšem otázkou, zda tato strategie povede k finančním výnosům, nutným pro vývoj.

5.4 Vývoj architektury

Očekávaný vývoj již nastínil nový čip od nVidie, tedy maximální unifikace, která pomůže nelimitovanému výkonu v mnoha ohledech. Pevné rozdělení vertex a pixel shaderů, je daný poměr shaderů (např. 1:1) při výpočtech. Je zřejmé, že přiřazené jednotky pro určité operace, budou vždy ve výsledku pomalejší, jelikož při výpočtu náročné scény bude jejich počet omezen, ačkoli jiné jednotky budou čekat na přidělený „práce“. Rychlejší bude vždy maximální počet univerzálních jednotek, které budou volné pro daný výpočet.

Je pravděpodobné, že tímto směrem se budou ubírat i ostatní části jádra a architektura čipů tedy bude velmi flexibilní a výkonná.

5.4.1 Výpočet fyziky

Veškeré výpočty, které se týkají fyziky, v dnešní době vypočítává procesor. Z důvodu, který již práce nastínila (tedy, že CPU je o řády pomalejší než výkonné GPU) by tedy bylo logické, kdyby tyto výpočty mohla obstarat grafická karta a ulevit tak procesoru od složitých výpočtů. Přenos dat po sběrnících taktéž nepatří k nejrychlejším a brzdí generování scény, čili by se mnohé urychlilo. V minulosti se již objevily přídatné karty, které měly obstarat výpočet fyzikálního enginu. Z nedostatečné podpory ze strany výrobců her se však na trhu neuchytily a při pár pokusech (testech) vyšlo najevo, že s touto kartou výkon nestoupá, nýbrž klesá.

Vývojové společnosti by se tak měly zaměřit na převzetí některých úkonů od CPU a využít okamžiků, kdy karta není zatížena na sto procent. Hrubý početní výkon karet je obrovský, je tedy potřeba lépe využívat jejich potenciálu.

5.4.2 Optimalizace ovladačů

Důležitým krokem při vývoji nových čipů je maximální softwarová podpora. Ovladače pro grafickou kartu mnohdy dokáží zvýšit výkon až o deset či patnáct procent a není tajemstvím, že mnohdy karty papírově silně byly velkým zklamáním jen díky špatným ovladačům, jenž jejich výkon brzdily.

Není pravidlem, že s novým čipem jsou vydány nové stabilní a výborné ovladače, čili lepší softwarová implementace by z hlediska uživatele byla více než přínosná.

6. Závěr

Na počátku svého vývoje, byly grafické karty velmi pomalé a neschopné složitějších výpočtů. Bylo to dáno výkonem počítačů, technologií výroby i faktem, že byly brány jako pouhý doplněk.

V současnosti jsou grafické karty jednou z hlavních výpočetních jednotek, které dokáží ovlivnit výkon celého počítače. Díky konkurenčnímu boji se na trh dostávají velmi výkonné karty, jejichž cena se neustále snižuje. Nabídka je velmi široká a díky mnoha typům, které výrobci dodávají, je velká možnost výběru od dvou tisíc až po patnáct tisíc korun. Pro běžného uživatele postačují karty v cenové relaci čtyř až pěti tisíc korun, kdy je dosažen nejlepší poměr cena/výkon.

V nejbližší době se začnou vydávat hry s podporou DirectX 10, což pravděpodobně přinese realističtější ztvárnění grafické scény. Nyní je však brzy na závěry, bude trvat nejméně půl roku, než se objeví první tituly, jenž budou mít plnou podporu DirectX10. Není tedy důvod k obavám a rychlé koupi nového a drahého akceleratoru, jelikož ty současné pro běžného uživatele stačí.

Při výběru grafické karty hraje velikou roli obrovské množství parametrů, není tedy příliš jednoduché vybrat tu správnou. Výběr je veliký a strategický záměr výrobců s nepravdivým značením karet, pouze ztěžuje situaci. Volba by tedy měla být konzultována, neboť špatný výběr může negativně ovlivnit výkon celého počítače.

Grafické karty jsou dnes o řád výkonnější než současné procesory pro osobní počítače. V příštích letech bude jejich výkon stále stoupat a budou tak stále více ovlivňovat celkovou výpočetní sílu počítače v grafických aplikacích. Díky strmému vývoji jsou pouhý rok staré karty relativně „zastaralé“. Při požadavku na vysoký výkon ve stávajících aplikacích, je tedy nutná častá koupě nového a tedy drahého zařízení.

7. Seznam literatury

- 1) PC TUNING. All about PC hardware [online]. <<http://www.pctuning.cz>>.
- 2) SVĚT HARDWARE. Vše o hardwaru [online]. <<http://www.svethardware.cz>>.
- 3) CDR. Vše co se točí okolo vypalování CD a data storage [online]. <<http://www.cdr.cz>>.
- 4) DIGITALDOOM'S DIGIWORLD. Svět PC inteligentně [online]. <<http://www.ddworld.cz>>.
- 5) TOM'S HARDWARE. Computer Hardware News, Tests, Ratings, Reviews. [online]. <<http://www.tomshardware.com>>.
- 6) XBIT LABORATORIES. [online]. <<http://www.xbitlabs.com>>
- 7) GURU3D. Reviews, analysis, drivers and tweaks for 3D graphics cards. [online]. <<http://www.guru3d.com>>.
- 8) CZECH COMPUTER. [online]. <<http://www.czc.cz>>.
- 9) ALZASOFT. [ONLINE]. <<http://www.alza.cz>>.
- 10) ROOT. Linux, open source and freeware [online]. <<http://www.root.cz>>.
- 11) FAKULTA INFORMATIKY MU. Historie a vývoj programování grafických adaptérů pro PC. [online]. <http://www.fi.muni.cz/usr/jkucera/pv109/2003p/xfilipov_programovani_grafiky.html>.
- 12) TISCALI GAMES. [online]. <<http://games.tiscali.cz>>.

- 13) TWEAK TOWN. [online]. PC hardware, tweak guides, overclocking, cooling, CPU and others. <<http://www.tweaktown.com>>.
- 14) WIKI – Wikipedia, the free encyclopedia. [online]. <http://en.wikipedia.org/wiki/Main_Page>.
- 15) PCTUNING. STALKER Shadow Of Chernobyl - výkon grafických karet. [online]
<http://www.pctuning.cz/index.php?option=com_content&task=view&id=8491&Itemid=88&limit=1&limitstart=0>.
- 16) ROOT. Grafické karty a grafické akcelerátory. [online]. <<http://www.root.cz/clanky/graficke-karty-a-graficke-akceleratory-1/>>.
- 17) PCTUNING. Který procesor je lepší? 34 procesorů v 10 testech. [online].
<http://www.pctuning.cz/index.php?option=com_content&task=view&id=8617&Itemid=28&limit=1&limitstart=6>.

8. Slovníček

1) PIXEL

Představuje jeden svítící bod na obrazovce monitoru, resp. jeden bod obrázku, zadaný svou barvou. Body na obrazovce tvoří čtvercovou síť, díky níž se dá každý pixel přesně identifikovat dle souřadnic. Velikost pixelu záleží na zvoleném rozlišení v případě klasických analogových (CRT) monitorů, kde s vyšším rozlišením klesá velikost pixelu a tím se zvyšuje kvalita zobrazení. U LCD obrazovky je velikost pixelu dána fyzicky (tzv. nativní rozlišení) a jeho velikost je tak pevně vázána na dané rozlišení. Pokud se rozlišení změní, dojde k jisté deformaci a tím možnému zkreslení.

Při rozlišení 1024x768 u patnáctipalcového monitoru je velikost jednoho pixelu zhruba 0,3 mm.

2) PLATFORMA

Společnost Intel definuje platformu „jako integrovanou sadu komponent, která umožňuje zamýšlené způsoby použití, rozšiřuje stávající trhy, vytváří nové příležitosti a přináší koncovým uživatelům větší užitek než seskupení jednotlivých komponent“.

Platforma neznamena pouze procesor a příslušný software – zahrnuje hardware, software, služby a další (viz. níže) a je rovněž propojena do okolního digitálního prostředí. Systém platformy je klíčový předpoklad vzniku nové generace výpočetních systémů — přizpůsobených jednotlivci a využívajících všudypřítomné inteligentní elektronické prostředí. Platformy uživatelům zajišťují vyšší užitnou hodnotu.

3) PERIFERNÍ ZAŘÍZENÍ

Rozšiřující zařízení počítače, rozděluje se do více skupin, např. výstupní (monitor, tiskárna), a vstupní (myš, klávesnice) periferní zařízení.

4) RAMDAC

Random Access Memory Digital-to-Analog Converter konvertuje digitální obrazová data, které zpracovává grafický čip a následně je přeposílá jako analogové do monitoru. Rychlejší RAMDAC poskytuje kvalitnější a stabilnější obraz.

5) DDR

Zkratka pro double-data-rate synchronous dynamic random access memory, je typ paměti používaný v dnešních počítačích. Dosahuje vyššího výkonu než předchozí typ SDRAM tím, že přenáší data na

obou koncích hodinového signálu (vzestupného i sestupného). Tento přístup zvyšuje efektivní výkon téměř dvakrát bez nutnosti zvyšování frekvence sběrnice. DDR paměti na 100 MHz jsou tedy přibližně stejně rychlé jako SDR paměti na 200 MHz.

6) PIPELINE

Jedná se o výpočetní jednotku, která se stará o operace s pixely. Tato pixel pipeline ovšem dokáže paralelně pracovat s více pixely či texturami a tím efektivně urychlovat práci. Čím více pipeline pracuje, tím větší výkon poskytují, zpravidla však nedokáží být maximálně a hlavně efektivně využity.

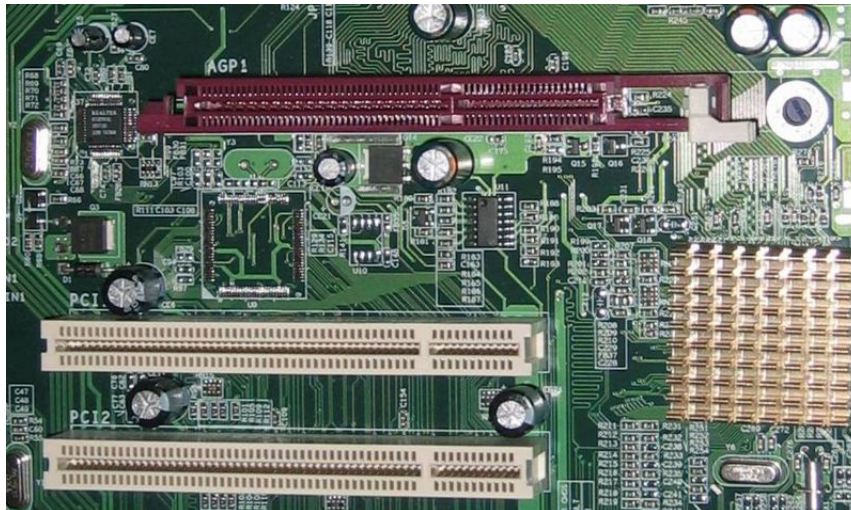
7) ROP

Jedná se o jednotku pro rastrové operace (Raster Outputs), která provádí antialiasing, stará se o komprese barev a Z-bufferu a zapisuje výslednou scénu do frame bufferu.

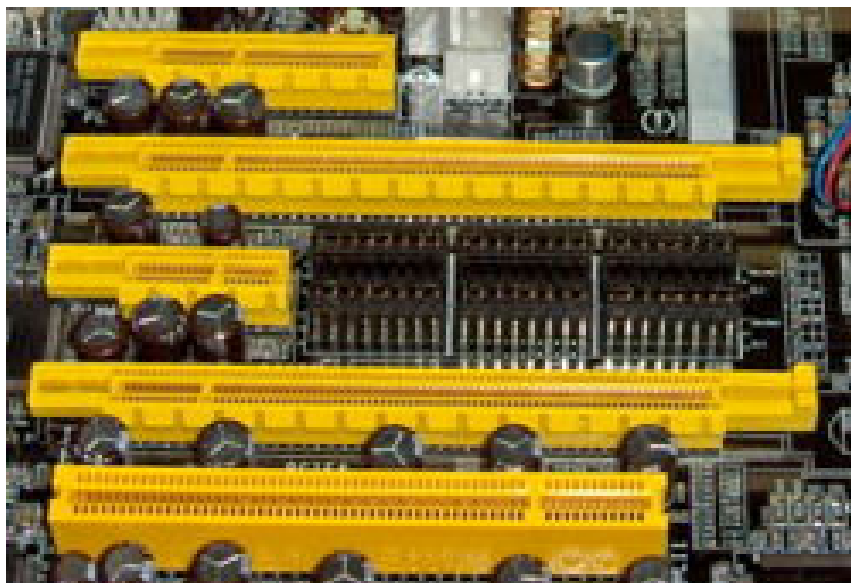
8) GIGAFLOPS

Je miliarda operací v pohyblivé řádové čárce za sekundu. Jedná se o měření rychlosti a čisté výpočetní síly matematické jednotky počítačového procesoru nebo GPU.

Příloha A Sběrnice PCI, AGP, PCI-E



Obr. 1 - 2 PCI sloty spolu se sběrnicí AGP
(vpravo chladič čipsetu základní desky)



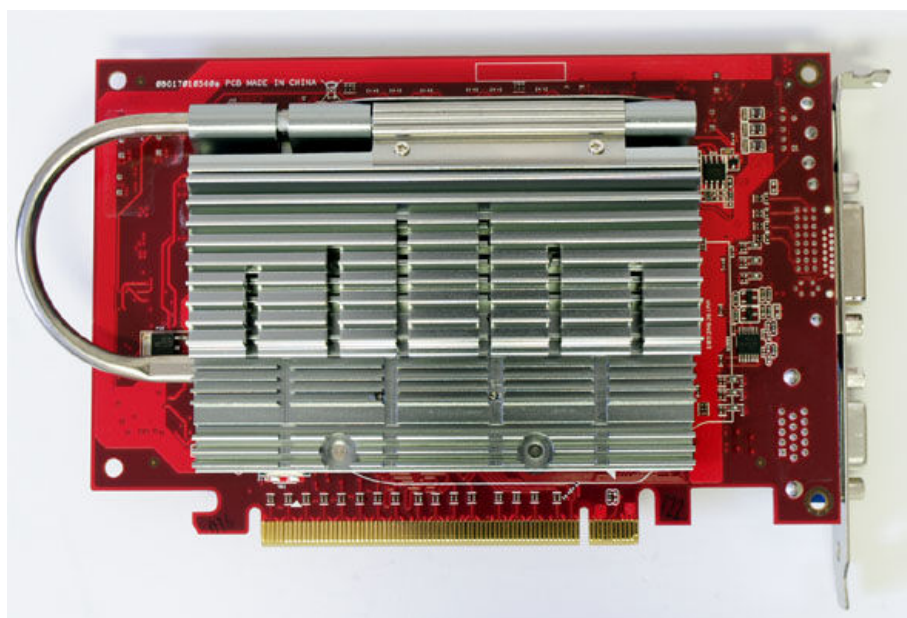
Obr. 2 - PCI-Express v možném SLI zapojení

Zdroj obrázků: 14)

Příloha B Pasivní a aktivní chlazení grafických karet



Obr. 1 - Přední strana chladiče pro X1300Pro



Obr. 2 - Zadní strana X1300Pro

Velmi dobře řešené pasivní chlazení v podání jedné z testovaných karet, Radeonu X1300Pro.



Obr. 3 - Aktivní dvouslotové chlazení u GeForce 8800GTX

Na tomto obrázku je chlazení nejrychlejší karty současnosti, GeForce 8800GTX. Dvouslotové chlazení není příliš uživateli oblíbeno, výkon si však žádá své v podobě obrovského chladiče.

Zdroj obrázků [1].

Příloha C Rozdíl Min a Max detailů



Obr. 1 - Stalker – minimální detaily

Rozdíl mezi nastavením grafických efektů zobrazují tyto dva obrázky. Na prvním je vidět neostrost textur, prakticky žádné detailní textury, hra tak vypadá jako kdyby vyšla před třemi lety. Zároveň je však možné si zahrát na prakticky jakékoliv grafické kartě.

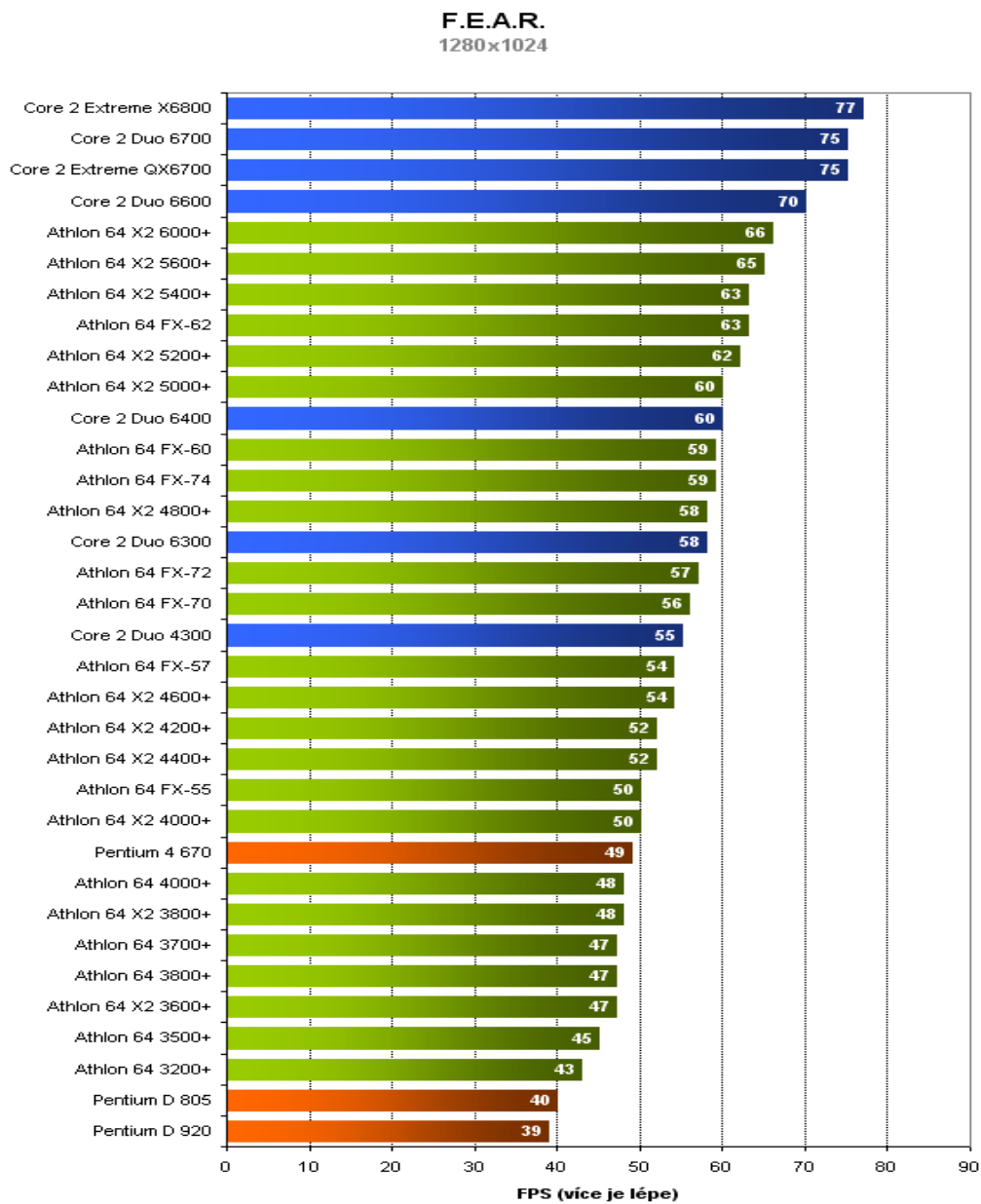
Oproti tomu na obrázku dvě jsou patrné detailní textury (auto) či bohatší vegetaci. Statický obraz však není příliš vypovídající. Na maximální detaily je rozpočítána tráva dle fyzikálních zákonů, reálné stíny, detail ve hledí pušky a jasné kontrastní nebe, způsobené technologií HDR.



Obr. 2 - Stalker – Maximální detaily (HDR)

Zdroj obrázků [15].

Příloha D Závislost grafické karty na CPU



Obr. 1 – Výkon 8800GTS s různými procesory (zdroj [17])

Dle obrázku je názorně vidět závislost velmi výkonné grafické karty užití v testu (8800GTS) na procesoru. I z tohoto důvodu je nutnost testovat grafické karty na maximálně výkonné sestavě. Důvod je zřejmý – v případě nižšího výkonu CPU dochází ke zkreslení.