

PCs buszrendszerek

A 'busz' nem más, mint egy összeköttetés a központi egység (CPU) és a perifériák között. Az első ISA buszok mai mércével mérve neveltségese, alig pár száz KB/s-os adatátvitelt bonyolítottak. Később kialakult a mai ISA buszok végleges, standard formája, a 16 bites ISA busz. Ez a maga 8 MHz-es frekvenciájával a gyakorlatban közel 5 MB/s-os csúcsebességre képes (az elméleti maximum 8 MB/s, ez azonban a fellépő késleltetések miatt nem érhető el). Ez a '80-as évek közepén még bőven megfelelt minden eszköznek, azonban az idő múlásával szükségessé vált egy jóval nagyobb sebességű buszrendszer kifejlesztése.

1991-ben kezdődött a PCI specifikáció gyakorlati megvalósításának kidolgozása az Intel laboratóriumaiban. Természetesen a vezető hardvergyártó cégek képviselőinek bevonásával folytak a fejlesztések, melynek gyakorlati eredményét (a PCI ipari szabványt) alig 4 év leforgása alatt közel 170 cég licenzelte... Így vált a legnépszerűbb PC-s buszrendszerre a PCI.

Az ISA 8 MHz-es órajelét 33 MHz-re emelve, a sávszélességet pedig 32 bitre növelve a PCI maximális átviteli sebessége közel 132 MB/s lett. Ez a 3D grafikus kártyákat leszámítva minden korszerű perifériát ki tud szolgálni. Ha azonban ez a közeli jövőben mégis szűkössé válna, bármikor bevezetésre kerülhet a 'dupla PCI', a PCI busz 66 MHz-es, 64 bites változata, mely elméletben 264 MB/s-os csúcsebességre képes.

A PCI másik előnye, hogy nem a jól bevált ISA buszrendszer leváltására tervezték, hanem annak kiegészítésére, 'patchelésére'. Így a PCI minden további nélkül alkalmazható bármely régebbi rendszerrel együtt. Külön chipkészlet, az ún. 'PCI bridge' felelős a CPU és a PCI perifériák közötti forgalom bonyolításáért. Speciális pufferrel rendelkezik, melynek segítségével a CPU a transzfer befejezése előtt más feladattal foglalkozhat: a puffer és a perifériák közötti tényleges adatátvitelt a CPU helyett a PCI bridge vezérli! Erről az ötletes megoldásról nevezik a PCI-t 'lokális busz'-nak, bár valójában csak egy átmenet a valódi lokális busz (CPU+cache+memória) és a periféria buszok (pl. ISA busz) között.

Továbbá a PCI buszon megvalósítható az ún. 'bus mastering' is, mely az intelligens perifériák működésének támogatását jelenti. Bus mastering tulajdonsággal rendelkező perifériák képesek a PCI buszon keresztül egymással kommunikálni: működést összehangolni, információt cserélni, akár nagy mennyiségű adatot is mozgatni. S ami a lényeg: mindezt a CPU teljes kikerülésével! Jogos lehet azonban a kérdés: ebbe az adatforgalomba hogyan tud a CPU 'beleszólni'? Miképpen tud az egymás közötti kommunikációba a 'külsős' CPU is bekapcsolódni? Nos, erre a problémára a 'PCI concurrency' technika jelent megoldást, melynek alkalmazásával a PCI bridge képes a CPU-periféria és a bus mastering periféria-periféria adatforgalmat egymástól függetlenül bonyolítani.

E sok misztikus fogalom után talán nem haszontalan egy példát felhozni :) Milyen perifériák használnak bus mastering-et? Az egyik legegyszerűbb, könnyen vázolható példa a korszerű PCI TV tuner kártyák kép megjelenítése. Az általam is használt miroVIDEO PCTV az antennáról fogott (vagy egyéb bemeneti portról kapott) analóg jelet a kártyára integrált BrookTree/848 chippel digitalizálja, és másodpercenként 25 képkockát küld át egyenesen a videokártyához (természetesen a CPU teljes kikerülésével, a 'helyi' PCI buszon át). A megjelenítés módjától függően az adatok mehetnek egyből a videomemóriába, vagy további feldolgozás céljából előbb a videochiphez is. Gyors fejszámolás: egy PAL formátumú jel (768 x 576 pixel, 32 bit színmélység, 25 frame/s) folyamatos bus mastering átvitele konstans 44 MB/s-ot használ a PCI 132 MB/s-os sávszélességéből... Tehát még bőven marad a többi perifériának !

A '90-es évek elején kifejlesztett PCI busz akkor még rendkívül nagyra vélt sávszélessége a real-time 3D alkalmazások térhódításával kezdett szűkössé válni. Az elméletben 132 MB/s-os csúcsebességre képes 32 bites, 33 MHz-es PCI valójában még burst módban sem tud 110 MB/s-os gyakorlati sebesség fölé kapaszkodni. Ez egy átlagos 128 bites, 100 MHz-es lokális busszal szerelt videokártya belső átviteli sebességének (max. 1.6 GB/s) csak töredéke. Könnyen belátható tehát, hogy a képadatok (főleg textúrák) rendszermemóriából a videokártyához való eljuttatása a PCI buszon csak alacsony felbontásban oldható meg.

Hogyan tovább?

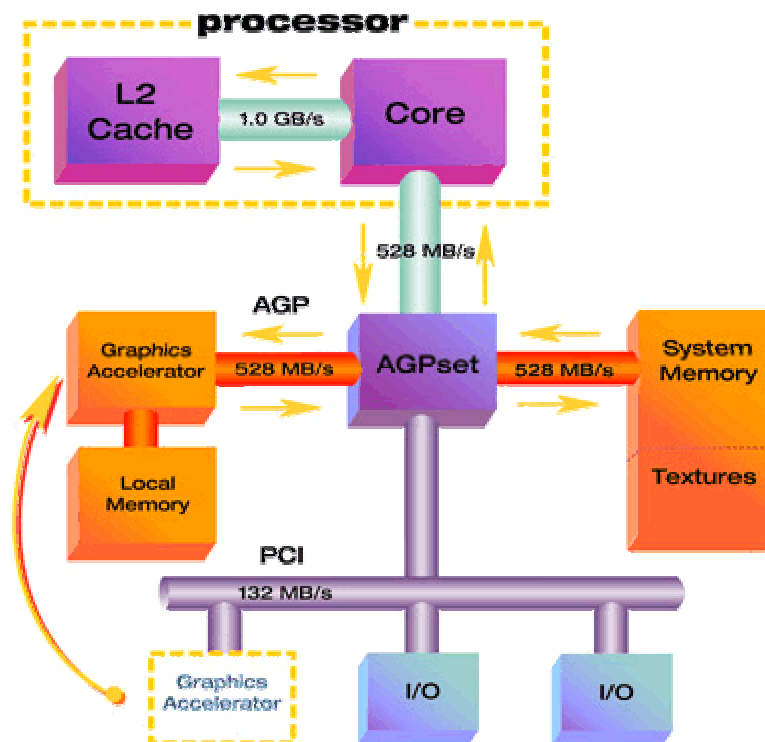
A kép bonyolultságának és a felbontásnak növelésével az adatokat a videokártya memóriájában kell tárolni, ha elfogadható sebességet akarunk elérni. Ez bizonyos határokon belül működik csak, hiszen pl. 1280x1024-es felbontásban, 32 bites színmélységű (32 bit Z-buffer, Double Frame Buffer) képfolyam folyamatos renderelése 15 MB videomemóriát igényel! Így egy manapság átlagosnak tekinthető 16 MB-os videokártya esetén mindössze 1 MB videomemória marad a textúráknak. Ezen aztán lehet segíteni az S3 Savage3D-ben debütált textúratömörítéssel, mely nem csak a textúramemória, hanem a sávszélesség

szükségére is gyógyír lehet. Ez a technológia azonban ront a textúrák minőségén, és sajnos egyelőre gyakorlatilag nulla a támogatottsága...

További gyorsító megoldás lehet az ún. 'bus mastering' megoldás használata. Bus mastering tulajdonsággal rendelkező perifériák képesek a PCI buszon keresztül egymással kommunikálni: működést összehangolni, információt cserélni, akár nagy mennyiségű adatot is mozgatni. S ami a lényeg: mindezt a CPU teljes kikerülésével. Ez a 3D gyorsító kártyáknál remekül alkalmazható a textúrák vagy egyéb képadatok mozgatására a rendszermemóriából a videokártya lokális memóriájába. Közben természetesen a CPU már az elkövetkező képek adataival foglalkozhat, hiszen a bus mastering csak a PCI buszt terheli, a CPU-t nem.

A feldolgozás tovább gyorsítható, ha a geometriát és a kép előállításához szükséges adatokat a processzor előre kiszámolja, és valamilyen feladatleíró nyelvvel szekvenciális utasítások sorozataként tárolja a memória meghatározott részén. Ez nagy segítség lehet bus mastering-et támogató rendszerben, hiszen így a kép-információkat teljes egészében a CPU kihagyásával lehet a videokártyához szállítani.

AGP



Picture Copyright by Intel Corporation (see developer.intel.com)

Az imént vázolt megoldások azonban egytől-egyig csak átmenetileg jelentenek gyógyírt a sávszélesség problémára. Végző megoldásként 1996-ban jelentették be az AGP-t (Accelerated Graphics Port). Az új szabvány a már létező 66 MHz-es PCI 2.1 specifikációra épül, azt egészíti ki néhány optimalizációt célzó újítással. Az AGP szoftverszinten megegyezik a PCI-jal, tehát lefelé kompatibilis vele, azonban minden más tekintetben eltér tőle.

A PCI eszközök egy lokális buszon helyezkednek el, s ez a busz az alaplapi chipkészletre csatlakozik. Az AGP teljesen független a PCI busztól, egyetlen portjával közvetlenül a chipkészlettel kommunikál (point-to-point kapcsolat). Sem logikailag, sem fizikailag nem egyeztethető össze a PCI specifikációval, ám mégis kompatibilis azzal.

Az első AGP-s videokártyák (pl. Matrox Millennium II AGP) minden tekintetben megfeleltek az AGP leírásnak, azonban csak PCI tranzakciókat bonyolítottak, semmilyen AGP specifikus műveletre nem voltak képesek. Jogos lehet a kérdés: minek akkor az AGP? Nos, elsősorban marketing fogás, hiszen egy hiper-szuper újdonságokat felvonultató termék jóval könnyebben eladható, mint egy bevált technológiával felvértezett - amiből kismillió van már a piacon. Nem szabad azonban azt hinni, hogy semmi értelme az ilyen 'AGP-s' videokártyáknak. Először is, az AGP független a PCI busztól, így a videokártya AGP-n keresztüli adatforgalma felszabadítja a PCI buszt. Kérdés, hogy ez mennyire számít, hiszen lehet, hogy a busz eleve nem volt terhelve :) Másodsorban, az AGP kétszer akkora órajelen működik, és a tranzakciók lebonyolítása is hatékonyabb, mint a PCI buszon.

Sebesség

A PCI busz burst módban 4 db 32 bites adatot képes átvinni 5 órajelciklus alatt (1 ciklus szükséges a címzéshez). Így a PCI effektív sebessége 33 MHz-en 110 MB/s. Az AGP eleve duplaakkora órajelen működik, ráadásul az ún 'sideband addressing'-nek köszönhetően egy 32 bites adat átvitele pontosan 1 órajelciklus alatt megy végbe. Ugyanis az AGP speciális címbusszal rendelkezik, ezen keresztül már jóval az adatok mozgatása előtt megtörténik a címek átvitele. Így végül az AGP effektív sebessége $32 \text{ bit} \times 66 \text{ MHz} = 264 \text{ MB/s}$.

Ha azonban utánaszámolunk, ezt az első ránézésre tekintélyes sávszélességet nem is olyan nehéz kihasználni. Vegyünk egy egyszerű példát: 3D lövöldözős játék, nagy teremben állunk, és forgunk körbe-körbe. Egy nagyon jó képminőségű játékhoz elengedhetetlen a legalább 512×512 pixeles textúrák és az 1280×1024 -es, 32 bites színmélységű felbontás használata, természetesen min. 30 FPS (30 képkocka/s) képfrissítéssel. Ha a geometria adatoknak lefoglalunk 4 MB-ot másodpercenként, akkor $260 / 30 = 8.66$ MB/s marad a textúráknak. Ez 512×512 -es textúrákból mindössze 8 db-ra elegendő! ($512 \times 512 \times 32 \text{ bit} = 1 \text{ MB}$) Más kérdés, hogy általában nem szokott minden képkockában cserélődni az összes textúra - de mi vegyük csak a legrosszabb eshetőséget :)

Szükség van tehát egy nagyobb sebességre, amit jelenleg az AGP 2x szabvány képvisel. Egy órajelciklus alatt 2 adatot képes mozgatni az új AGP port, mely minden más tekintetben megegyezik az eredeti AGP-vel. Az effektív sebesség ezzel a duplájára, 528 MB/s-ra nőtt. A jelenleg kapható második generációs 3D gyorsítók AGP változatai egytől-egyig támogatják az AGP 2x szabványt.

A sávszélesség éhség növekedése elvezetett a manapság teret hódító AGP 4x szabvány bevezetéséhez. Ez az AGP 2x-szel megegyező megoldás, azonban az eddigi adat- és címbuszt megduplázták, így effektíve 133 MHz-es sebesség érhető el. Elgondolkodtató, hogy az AGP 4x-et támogató alaplapokon a PC100-as memória és a chipkészlet közötti sávszélesség kisebb, mint a videokártyáé (AGP 4x: 1 GB/s, PC100: 800 MB/s).

Az eddigiekben felvázolt technikák lehetővé teszik a sávszélesség növelését, azonban a gyakorlatban nem jelentenek igazi megoldást a videokártyák textúra éhségének kielégítésére. Lássuk tehát, mit is tud valójában az AGP!

A PCI bus mastering lehetősége remekül alkalmazható kis mennyiségű adatok mozgatásánál, azonban nagyméretű textúrák átvitelénél nincs értelme minden egyes textúránál újra felprogramozni az alaplapi chipkészletet. Nem erre találták ki a bus mastering-et, ez tény. Emellett, a videokártya számára sem túl könnyű állandóan a chipkészlethez fordulni, ha újabb textúrára van szüksége, és ráadásul tudnia is kell, hogy a fizikai memóriában éppen hol helyezkedik el a kívánt adat. Ráadásul teljesen fölöslegesen töltögeti át a saját memóriájába a textúrákat. Ez úgy, ahogy van, alkalmatlan a feladatra.

Az igazi megoldást már jóval előbb kidolgozták, bár nem PC-re, hanem nagyteljesítményű grafikai munkaállomásokra (pl. Silicon Graphics). Minden címzési nehézséget el lehet kerülni, ha a videokártya memóriáját kiterjesztjük a rendszermemória meghatározott részére. Ezt természetesen az alkalmazások számára el kell rejtteni, mindent az AGP vezérlőre és a videokártyára kell bízni.

A rendszermemóriában lefoglalt címterületet hívja a szakma 'AGP aperture'-nek. Ez a címterület a rendszermemóriában a videokártya Linear Frame Buffer-ének (videomemória leképezése a virtuális memóriába) területe alatt helyezkedik el. A címterület nagyságát a BIOS setup-ban beállítható 'AGP Aperture Size' határozza meg, de a méretről meg lehet győződni Win98 alatt is, a Vezérlőpult / Rendszer / Eszközkezelő / Rendszereszközök-ben rákeresve az alaplapi AGP vezérlőre. Az AGP vezérlő által lefoglalt címtartomány nagysága pontosan az AGP aperture méretét adja.

Az AGP aperture-ben elhelyezkedő adatok kiterjesztett videomemóriába való elhelyezését a GART (Graphics Aperture Remapping Table) írja le. A Minden egyes AGP aperture hivatkozást a GART segítségével lehet átkonvertálni fizikai rendszermemória címmé. A GART az alaplapi chipkészletben helyezkedik el, és az operációs rendszer programozza fel (emiatt van szükség Windows alatt az AGP driverekre). A GART szerkezete a flexibilitás érdekében fizikailag nem meghatározott, szoftverszinten kell a megfelelő struktúrát felépíteni (AGP driver feladata).

Az AGP optimális működését DIME-nak (Direct Memory Execution) nevezik, mely egyszerűen csak az AGP aperture és a GART használatát jelenti. Az alapvető különbség a DIME és a klasszikus DMA (Direct Memory Access) megoldás között mindössze annyi, hogy DIME használatával nincs szükség minden memória műveletnél az alaplapi vezérlő felprogramozására, hiszen azt automatikusan végzi az AGP aperture és a GART segítségével. DMA módban a kevésbé effektív bus mastering technikával történik az adatátvitel - akárcsak a klasszikus PCI buszon -, mindössze a sávszélesség duzzad fel min. 264 MB/s-ra...

Buszrendszerek összehasonlítása

	ISA	ISA	EISA	VESA	PCI	AGP 1x	AGP 2x	AGP 4x
Sávszélesség	8 bit	16 bit	32 bit	32 bit	32 bit	32 bit	32 bit	32 bit
Buszsebesség	5.33 MHz	8.33 MHz	8.33 MHz	33 MHz	33 MHz	66 MHz	66 MHz	66 MHz
Max. átviteli sebesség	5.33 MB/s	8.33 MB/s	33 MB/s	132 MB/s	132 MB/s	264 MB/s	528 MB/s	1 GB/s
Bus mastering támogatás	Nincs	Nincs	Van	Van	Van	Van	Van	Van
Adat / cím paritás	Nincs	Nincs	Nincs	Nincs	Van	Van	Van	Van
Kártya ID (automatikus konfiguráció)	Nincs	Nincs	Van	Van	Van	Van	Van	Van

Forrás: Fiery, PROHARDVER!