

Mehr 3-D mit AGP

Vielleicht sollten Sie noch etwas warten, bevor Sie sich eine 3-D-Grafikkarte kaufen: Wenn Intels AGP-Initiative einschlägt, dann ist in einem Jahr alles veraltet. CHIP sagt Ihnen, was der neue AGP-Standard bringen soll.

Während der vergangenen fünf Jahren war die schnelle Grafikausgabe ein Leistungstreiber für die PC-Architektur. Der gute alte ISA-Bus war zusehends überlastet, Mikrokanal und EISA stritten um die Gunst der Entwickler und Käufer. Das VESA-Komitee (Video Electronic Standard Association) verpflanzte daraufhin den bekannten VL-Bus in die PC-Systeme. Bald darauf war der von Intel propagierte PCI-Bus (Peripheral Component Interconnect) geboren, und der VESA-Local-Bus verschwand in der Versenkung.

Beide Bussysteme verfügen über genügend Leistung, um die Grafikkarte fortwährend mit Daten für die Standard-2-D-Darstellung zu versorgen. Probleme, so zeigte sich jedoch bald, treten auf, wenn die Busse für 3-D-Datentransfers genutzt werden sollen. Daher kündigte Intel im Frühjahr einen neuen Vorstoß im Grafikbereich an: AGP (Accelerated Graphics Port). Auf der Entwicklerkonferenz Ende Mai trafen sich die wichtigsten Hersteller von Grafikchips und PC, um den neuen Standard zu verabschieden. Dabei wurde klar, daß AGP ein neuer Bus im PC ist – allerdings mit grundlegend anderen Eigenschaften als ein herkömmlicher PC-Bus.

○ Warum AGP?

Die heute gängigen 3-D-Grafikkarten stellen aufgrund des Verfahrens sehr hohe Anforderungen an den PCI-Bus. Verständlich wird dies, wenn man sich die verschiedenen Stufen ansieht, die Pro-

zessor und Grafikkarte beim 3-D-Rendering bewältigen müssen. Auf der ersten Stufe werden die Szenerie als mathematisches Modell aufgebaut, die Kameraposition und Lichtquellen festgelegt und daraus die resultierende Stützstellenansicht berechnet. In Stufe zwei werden die Polygone – also die Vielecke – errechnet, die für die definierte Szenerie notwendig sind. Stufe drei schließlich trägt eine Textur auf die jeweiligen Polygone auf, berechnet die sichtbaren Bereiche und stellt schließlich das Bild dar.

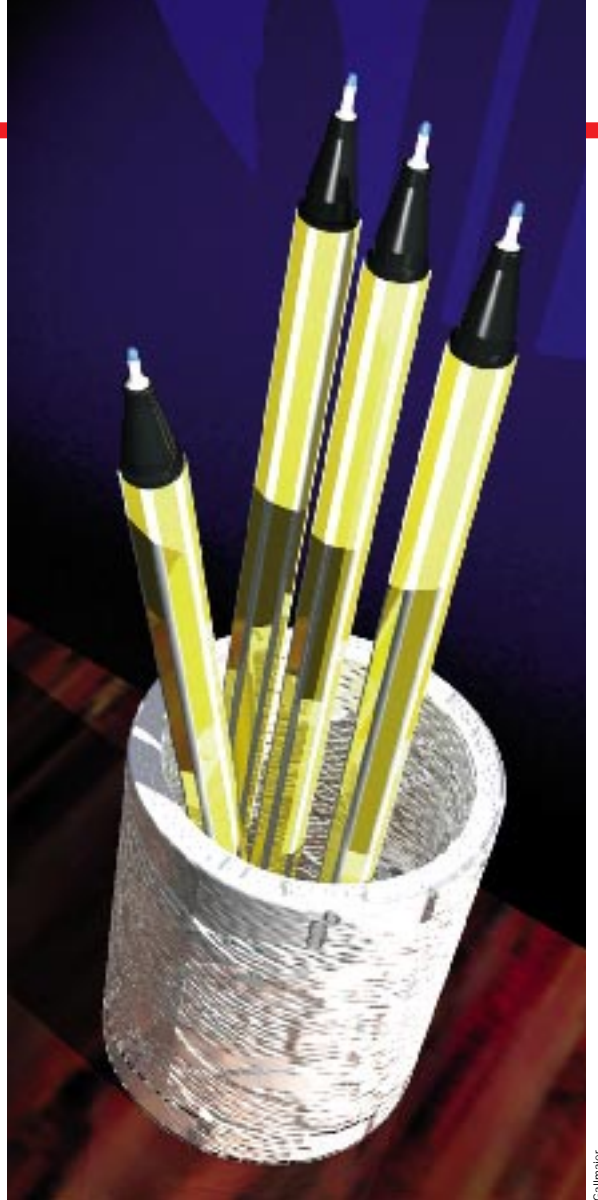
Gängige Prozessoren schaffen zwar die Berechnung aller Stufen; ein Beispiel sind die verschiedenen Spieletitel im Markt. Dennoch müssen Abstriche an Qualität oder Geschwindigkeit gemacht werden, wenn die 3-D-Kalkulation in einer mittelschnellen CPU wie einem Pentium 100 erfolgt. Aktuelle 3-D-Beschleunigerkarten übernehmen nur die Berechnungen der dritten Stufe. Dabei muß eine enorme Datenmenge zwischen Beschleunigerkarte und CPU ausgetauscht werden, unter anderem Polygonbeschreibungen (Koordinaten, Luminanz und Farbe) und deren Texturdaten.

Diese Daten überfluten damit den Bus zwischen Prozessor und Grafikkarte. Unter dem Diktat der maximalen Busbandbreite haben viele Software-Entwickler die Komplexität ihrer Szenen herabsetzen müssen, um eine gute 3-D-Darstellung zu bekommen.

○ Punkt, Punkt, Komma, Strich

Um das Problem der Busbandbreite in den Griff zu bekommen, wird AGP nicht nur schneller (siehe unten), sondern verlagert vor allem bestimmte Rechenoperationen: Die Berechnung der Polygone sowie die üblichen Schritte zur Aufrasterung und Ausgabe auf dem Bildschirm erfolgen dann auf der Beschleunigerkarte. Der Prozessor erstellt lediglich noch die Stützstellenansicht. Dadurch verringert sich die Datenmenge, die über den Bus geschaufelt werden muß. Die Polygondarstellung benötigt nach einer Schätzung von Microsoft im Vergleich zu den Stützstellen zwischen drei- und achtmal so viele Daten.

Doch damit nicht genug: Ein weiteres gewichtiges Problem ist die Speicherverteilung für 3-D-Anwendungen. Speicher ist nötig, um Texturen (Pixelbilder) auf die Polygone zu bringen. Diese Bilder, etwa die Imitation einer Holzmaserung, müssen irgendwo abgespeichert werden, um später perspektivisch richtig auf das jeweilige Polygon aufgebracht werden zu können. Während sehr teure 3-D-Karten diese Texturen on Board ablegen,



Gallmaier

greifen kostengünstigere Architekturen über den PCI-Bus auf den Arbeitsspeicher zu und legen einen Teil der Texturen für den schnellen Zugriff in einem nicht genutzten Bereich des Grafikkartenspeichers (Off-Screen) ab.

Da die Menge dieses Speichers naturgemäß beschränkt ist und zudem mit der Auflösung variiert, wäre es besser, die Texturen direkt aus dem Arbeitsspeicher zu laden. Das belastet aber den PCI-Bus, so daß zur Zeit ein Kompromiß geschlossen werden muß. Mit AGP kann hingegen der Hauptspeicher so schnell angesprochen werden, daß eine Einlagerung auf dem Off-Screen erst einmal unnötig wird.

Ähnliches wie für die Texturspeicherung gilt auch für den sogenannten Z-Buffer. Dieser Speicher enthält die Werte für die räumliche Tiefe der Objekte. Sobald sich ein Blickpunkt ändert, muß der gesamte Z-Buffer-Inhalt neu berechnet werden. Auch hier hängt es von der Anordnung des Speichers ab, wie viele Daten über den PCI-Bus transferiert werden müssen.

Mit dem AGP lassen sich diese Speicherprobleme elegant umgehen. Allerdings sind dafür einige Voraussetzungen nötig: Das PC-System braucht mindestens 16 Megabyte Arbeitsspeicher, wovon 2 bis 4 Megabyte nicht für Applika-

tionen oder das Betriebssystem genutzt werden. Außerdem ist der Arbeitsspeicher so organisiert, daß genügend schnell darauf zugegriffen werden kann. Eine 64-Bit-Speicherarchitektur mit SDRAM würde diesen Bedarf an Speicherbandbreite decken können. Damit entspricht der AGP einer UMA (Unified Memory Architektur). Allerdings unterscheidet sich der AGP von den ersten vorgeschlagenen UMAs durch das Konzept eines weiterhin auf der Grafikkarte

Der AGP elektrisch

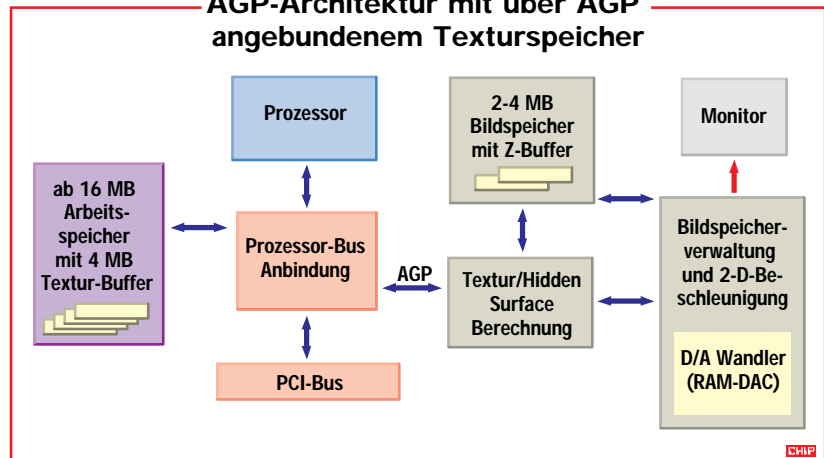
Zunächst wird der AGP eine neue Geschwindigkeitsklasse für den Datentransfer im PC einführen. Vom Konzept her basiert AGP auf dem verbreiteten PCI-Bus. Allerdings wird nur ein einziger Steckplatz vorhanden sein, um die elektrischen Eigenschaften zu vereinfachen und ein Busmanagement überflüssig zu machen. Falls der Grafikcontroller auf der Hauptplatine sitzt, wird gar kein Steckplatz sichtbar sein. Im Gegensatz zum heute üblichen 33-MHz-PCI-Bus wird der AGP bereits mit 66 Megahertz Taktfrequenz laufen; das soll mit der Einführung von PCI 2.1 wohl auch der normale PCI-Bus beherrschen.

Außerdem sind Datentransfers sowohl bei der steigenden als auch bei der fallenden Flanke des Taktsignals möglich, wodurch die Datentransferrate nochmals verdoppelt wird. Insgesamt schafft der AGP also den vierfachen Datendurchsatz eines heute üblichen PCI-Busses; das sind im besten Fall (Burst-Modus) rund 528 Megabyte pro Sekunde. Außerdem werden verbesserte Eigenschaften wie Seitenband-Kontrollsignale für eine Leistungssteigerung sorgen: Während beim herkömmlichen PCI-Bus die Daten stets zusammen mit den Adressen übertragen werden müssen, können per AGP mehrere Adressen und Daten ungeordnet übertragen werden. Dieses Multi-Pipeline-Konzept erlaubt es, auf verschiedene Bereiche des Arbeitsspeichers quasi simultan zuzugreifen. Allerdings werden dabei auch Änderungen am Design der Grafik-Chipsätze erforderlich.

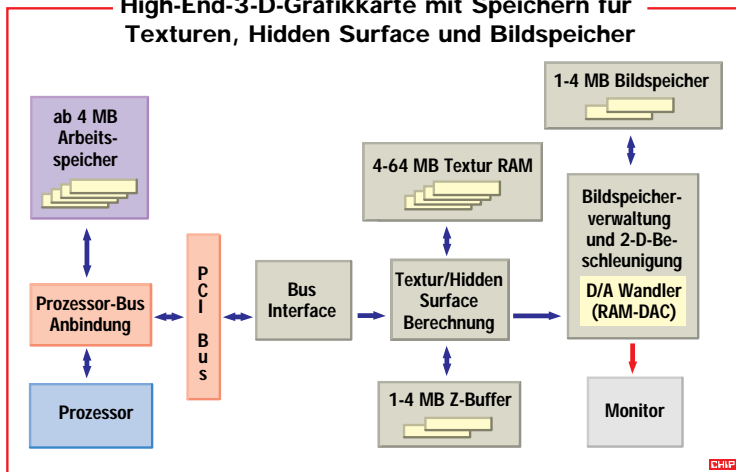
So schön die Theorie, so nüchtern ist die Praxis. Es ist zu erwarten, daß die ersten AGP-Karten stark vom höchstmöglichen Leistungswert abweichen, weil die Hersteller der Chipsätze nicht sofort auf die neuen Eigenschaften umsteigen können. Daß das Potential schneller Bussysteme anfangs kaum richtig ausgenutzt wird, hat bereits die Entwicklung der Grafikkarten mit VLB- und PCI-Interface gezeigt. Selbst heute gibt es nur wenige Karten, die den PCI-Bus vollkommen ausreizen; deshalb wird der erwartete Leistungsschub von AGP erst nach und nach einsetzen. Und nebenbei benötigt der staunende Kunde für AGP auch eine neue Hauptplatine.

Joe Curley, Jörg Lorenz

AGP-Architektur mit über AGP angebundenem Texturspeicher



High-End-3-D-Grafikkarte mit Speichern für Texturen, Hidden Surface und Bildspeicher



tionen oder das Betriebssystem genutzt werden. Außerdem ist der Arbeitsspeicher so organisiert, daß genügend schnell darauf zugegriffen werden kann. Eine 64-Bit-Speicherarchitektur mit SDRAM würde diesen Bedarf an Speicherbandbreite decken können. Damit entspricht der AGP einer UMA (Unified Memory Architektur). Allerdings unterscheidet sich der AGP von den ersten vorgeschlagenen UMAs durch das Konzept eines weiterhin auf der Grafikkarte

chen und ein Busmanagement überflüssig zu machen. Falls der Grafikcontroller auf der Hauptplatine sitzt, wird gar kein Steckplatz sichtbar sein. Im Gegensatz zum heute üblichen 33-MHz-PCI-Bus wird der AGP bereits mit 66 Megahertz Taktfrequenz laufen; das soll mit der Einführung von PCI 2.1 wohl auch der normale PCI-Bus beherrschen.

Außerdem sind Datentransfers sowohl bei der steigenden als auch bei der fallenden Flanke des Taktsignals mög-

lich, wodurch die Datentransferrate nochmals verdoppelt wird. Insgesamt schafft der AGP also den vierfachen Datendurchsatz eines heute üblichen PCI-Busses; das sind im besten Fall (Burst-Modus) rund 528 Megabyte pro Sekunde. Außerdem werden verbesserte Eigenschaften wie Seitenband-Kontrollsignale für eine Leistungssteigerung sorgen: Während beim herkömmlichen PCI-Bus die Daten stets zusammen mit den Adressen übertragen werden müssen, können per AGP mehrere Adressen und Daten ungeordnet übertragen werden. Dieses Multi-Pipeline-Konzept erlaubt es, auf verschiedene Bereiche des Arbeitsspeichers quasi simultan zuzugreifen. Allerdings werden dabei auch Änderungen am Design der Grafik-Chipsätze erforderlich.

Joe Curley, Jörg Lorenz



Internet: Die neuesten Spezifikationen zu AGP liegen auf folgender Site:
<http://www.teleport.com/~afgxpport>