



Technische Übersicht

AGP 8X

Die Weiterentwicklung der
Grafikschnittstelle



Mehr Grafikbandbreite

Niemand muss erst noch davon überzeugt werden, dass die Gesamtqualität eines PCs immer stärker von der effizienten Verarbeitung der Grafikdaten abhängt. Websites, Spiele und mittlerweile auch fast alle anderen Anwendungstypen stützen sich auf grafische Informationen, um dem Anwender intuitive, unmittelbar ansprechende Einblicke in die digitale Datenwelt zu vermitteln.

Neue Plattformen wurden entwickelt, um diese rasante Vermehrung visueller Informationen zu bewältigen. Zunächst definierte die AGP-Spezifikation (AGP = Accelerated Graphics Port) einen im Vergleich zu PCI optimierten Pfad für den Transfer von Daten zwischen dem Grafiks subsystem, der CPU und dem Hauptspeicher. Diese Schnittstelle spielt noch immer eine entscheidende Rolle. Inzwischen muss der AGP 4X-Bus jedoch mit immer höheren Anforderungen fertig werden:

- **Inhalt:** Grafische Szenen enthalten immer mehr komplexe Geometrie- und Texturdaten.
- **Präzision:** Die Entwickler fordern lautstark eine immer höhere Präzision der Daten. NVIDIAs GPUs (Graphics Processing Units) der nächsten Generation bieten 128 Bit Farbtiefe. Das ist die Präzisionsstufe der modernen Filmindustrie - Voraussetzung für spektakuläre Kinoeffekte in Echtzeit. Die größeren Datenformate erhöhen jedoch den Bandbreitenbedarf im gesamten System.
- **Interaktivität:** Szenenwechsel in Echtzeit belasten den AGP-Bus durch das dynamische Laden von Spieleumgebungen und mit dem Streaming von Spezialeffekten durch das Grafiks subsystem.
- **Weiterentwicklungen der Systeme:** Das Funktionsspektrum von Host-Plattformen und Grafiks systemen entwickelt sich ständig weiter, und zwar im Hinblick auf höhere Prozessorgeschwindigkeiten, mehr Hauptspeicherkapazität und -bandbreite sowie Multiprocessing-Funktionen. Damit muss sich auch die Nahtstelle zwischen diesen Systemen weiterentwickeln, denn nur so können Softwareprogramme die Fortschritte auf den anderen Gebieten auch tatsächlich nutzen.

Das neueste Release der AGP-Spezifikation - Version 3.0 - bildet durch die Einführung der AGP 8X-Schnittstelle einen Meilenstein der PC-Technologie, denn AGP 8X hält die Lösung für alle diese neuen Anforderungen bereit. AGP 8X verdoppelt die Bandbreite des Grafikbusses und ermöglicht damit eine drastische Verbesserung des Gesamtdurchsatzes bei den grafikintensiven Anwendungen von heute. Damit die Trends bei heutigen und künftigen Grafikanwendungen tatsächlich in vollem Umfang genutzt werden können, hat NVIDIA die erste und einzige Familie von AGP 8X-fähigen GPUs und Kernlogikprodukten in der

Branche auf den Markt gebracht, die das gesamte Spektrum von Einstiegs- bis zu Highend-Lösungen umfasst.

Diese Abhandlung beschreibt die Eigenschaften der neuesten, verbesserten AGP-Version und stellt im Detail dar, wie NVIDIA den neuen Standard umsetzen wird.

Die AGP-Spezifikation

Der AGP-Standard wurde zunächst mit dem Ziel entwickelt, hochleistungsfähige Schnittstellen zur Verbesserung der 3D-Grafikleistung zu definieren. Ein dedizierter Hochgeschwindigkeits-Port verbindet das Kernlogik-Chipset mit dem Grafik-Controller und schafft so einen direkten Übertragungsweg für Grafiktexturen in den und aus dem Hauptspeicher für den Fall, dass die Kapazität des lokalen Frame-Puffers ausgeschöpft ist oder eine neue Szene geladen werden muss. Dieses Design bietet mehrere Vorteile:

- AGP überträgt Texturdaten mit Geschwindigkeiten von mehreren Gigabyte pro Sekunde (GB/s), was die Geschwindigkeit des PCI-Busses bei weitem übertrifft (2,1 GB/s statt nur 132 MB/s bei PCI). Außerdem wird die Ausführung von Textur-Maps direkt aus dem Systemhauptspeicher unterstützt, so dass nicht mehr alle Texturdaten zuvor in den lokalen Grafikhauptspeicher geladen werden müssen.
- Die Spezifikation umfasst auch einen Seitenband-Adressmodus, mit dessen Hilfe die GPU neue Adressen und Anforderungen absetzen kann, noch bevor die vorhergehende Anforderung abgearbeitet ist.
- Die Belastung des PCI-Busses wird deutlich geringer, was die Performance der Geräte, die auf diesen Bus angewiesen sind, wesentlich erhöht (Festplatten-Controller, LAN-Karten, Video-Aufzeichnungssysteme usw.).

AGP Release 3.0

Seit ihrer Einführung im Jahr 1996 wurde die AGP-Schnittstelle Schritt für Schritt immer weiter aktualisiert. Ursprünglich sollte die Spezifikation Schwächen des weit verbreiteten PCI-Busses beheben und eine Schnittstelle definieren, die auf die Anforderungen von Grafikoperationen und Datentransporten zugeschnitten war. Bei Revisionen der Spezifikation ging es in erster Linie um eine Skalierung der Bandbreite:

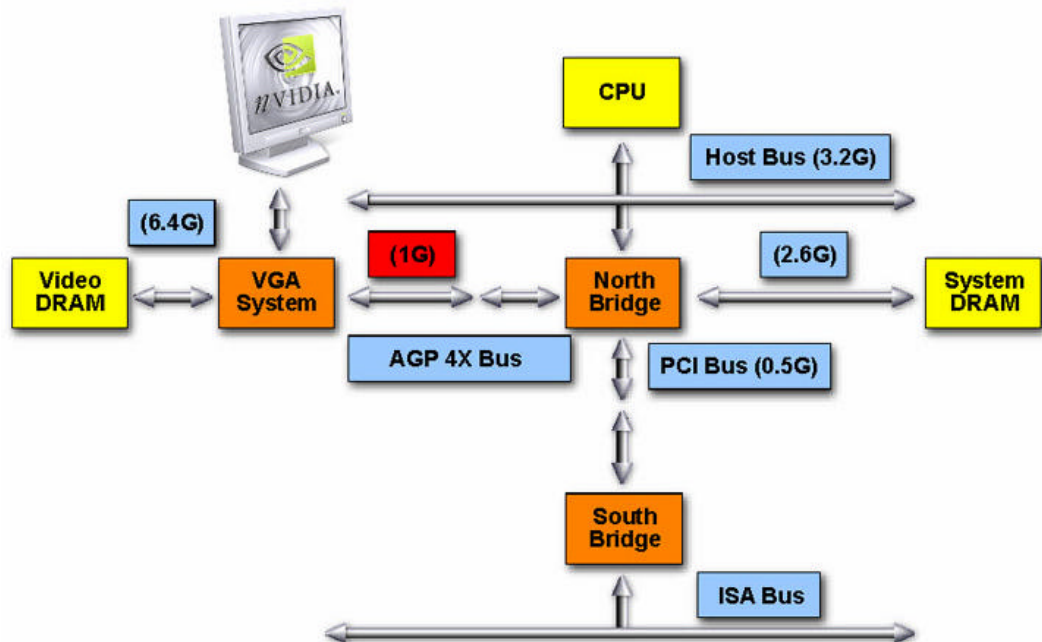
- In Version 1.0 der AGP-Schnittstelle 1.0 (AGP 1.0) wurden die Bandbreitenstufen AGP 1X und AGP 2X gleichzeitig eingeführt. AGP 1.0 enthielt also beide Schnittstellengeschwindigkeiten, wobei AGP 2X theoretisch doppelt so schnell war wie AGP 1X. Tabelle 1 unter "Die Performance von AGP 8X" stellt die verschiedenen Transferraten detailliert dar.

- AGP 4X, definiert in der AGP 2.0-Spezifikation, wurde zwei Jahre später eingeführt.
- AGP 3.0, die neueste Version, beinhaltet die AGP 8X-Bandbreite und führt den isochronen Betrieb sowie die AGP-Texturfunktionen (siehe ausführliche Beschreibung in den folgenden Abschnitten) ein.

Version 3.0 der AGP-Schnittstelle (AGP 3.0) verdoppelt die theoretische Performance bei Datentransporten über den Bus. Darüber hinaus umfasst diese neue Version der Spezifikation auch eine Reihe von neuen Funktionen, verzichtet dafür aber auf einige nicht benötigte Funktionen, um die Schnittstelle zu vereinfachen. NVIDIA unterstützt AGP 3.0 beginnend mit GPU- und Kernlogikprodukten, deren Markteinführung für den Herbst 2002 geplant ist.

Optimierung von Grafikfunktionen und Texturspeicher

Seit der Einführung von AGP 2.0 im Jahre 1998 haben sich Systemdesigns und insbesondere Grafiks subsysteme rasant weiterentwickelt. Mittlerweile erweist sich AGP 2.0 mit seiner AGP 4X-Bandbreite als Engpass beim Transport von Grafikdaten (siehe Abbildung 1).



Host bus = host-Bus
 Video DRAM = Video-DRAM
 VGA System = VGA-System
 System DRAM = System-DRAM
 AGP 4X Bus = AGP 4X-Bus
 PCI Bus = PCI-Bus
 ISA Bus = Bus ISA

Abbildung 1. Die AGP 4X-Bandbreite bildet einen Engpass im System. Bei AGP 8X dagegen ist die Grafikbandbreite dem Durchsatz im restlichen System (einschließlich Hauptspeicher) gewachsen.

Dank einer theoretischen Bandbreite von 2,1 GB/s gibt AGP 8X Entwicklern nun auch die Möglichkeit eines effizienteren Managements von Szenen mit komplexer

Geometrie und der dynamischen Umschaltung zu neuen Szenen in Echtzeit. Durch die Verringerung des Aufwands für das Speichern und Abrufen komplexer Texturen oder das Daten-Streaming aus dem Hauptspeicher erhöht AGP 8X den Gesamtdurchsatz des AGP-Busses. Daraus ergeben sich signifikante Performance-Steigerungen bei Visualisierungsfunktionen mit komplexen Texturen und Geometrien. Und dies wiederum ist die Voraussetzung für lebensechte Bildwelten am PC.

Isochronmodus für das Streaming

Die grundlegenden Bandbreitenverbesserungen bei AGP 3.0 wirken sich auf alle Grafikoperationen mit komplexen Geometrien, Texturen oder Streaming-Funktionen positiv aus. Eine neue Funktion von AGP 3.0 - der Isochronmodus - ist besonders vorteilhaft bei Grafikoperationen mit einem vorhersehbaren, ununterbrochenen Datenfluss. Die früheren Versionen der AGP-Schnittstelle können bei latenz-sensitiven Übertragungen die erforderliche Bandbreite nicht garantieren. Die "So-gut-wie-möglich"-Strategie dieser Schnittstellen erzielt zwar eine geringe Durchschnittslatenz bei einem hohen durchschnittlichen Durchsatz und funktioniert damit recht gut, birgt aber gelegentlich das Risiko außergewöhnlich langer Verzögerungen und kann daher zu Datenverlusten führen. Streaming-Anwendungen - Anwendungen also, die für Video-Broadcasts, für das Herunterladen aus dem Netz und ähnliche Aufgaben mit digitalen Datenströmen in Echtzeit arbeiten - tolerieren keine Datenverluste und erfordern vorhersehbare Transferraten. Darüber hinaus muss für kostengünstige Designs eine Möglichkeit gefunden werden, isochrone Datentransfers zu unterstützen, ohne dass dazu eine Aufstockung des teuren Datenpuffers erforderlich ist. Der Isochronmodus von AGP 3.0 erfüllt diese Anwendungsanforderungen in kostengünstiger Art und Weise.

Kompatibilität

Die AGP 3.0-Spezifikation bietet einen nahtlosen Aktualisierungspfad zu AGP 8X. Die mechanischen Busspezifikationen bleiben gleich. Die Geschwindigkeiten und die Funktionen von AGP 8X werden mithilfe zuvor nicht benötigter Stifte implementiert, und zwar so, dass AGP 8X-Karten in bestehenden AGP 2X- und 4X-Systemen genauso unterstützt werden wie in neuen Systemen, die die 8X-Schnittstelle in vollem Umfang nutzen. Die AGP 8X-Grafiklösungen von NVIDIA erkennen die AGP-Stufe des Host-Systems und konfigurieren die AGP-Schnittstelle automatisch so, dass sie im 3.0-Modus (mit der Geschwindigkeit 4X oder 8X) oder im 2.0-Modus (mit der Geschwindigkeit 2X oder 4X) läuft. Damit sind NVIDIAS neue Grafiklösungen uneingeschränkt 8X-fähig und mit 2X-, 4X- und 8X-Systemen kompatibel. Die NVIDIA-basierten Karten arbeiten automatisch mit der höchsten Geschwindigkeit, die das Host-System unterstützt.

Die Performance von AGP 8X

AGP 8X bietet die doppelte Bandbreite wie AGP 4X. Der Effekt der AGP 8X-Bandbreite auf die Anwendungs-Performance insgesamt hängt von der Art der Anwendung ab:

- ❑ **Statische Welten:** Bei Anwendungen, die in kleinen, virtuellen Umgebungen laufen (bei denen gewissermaßen "die ganze Welt" permanent im Grafikspeicher präsent ist), wird die AGP 8X-Transfergeschwindigkeit kaum, wenn überhaupt, zu Performance-Steigerungen führen.
- ❑ **Komplexe Welten:** Bei den schnellen "Fly-through"-Anwendungen und Spielen von heute bewirkt die verdoppelte AGP 8X-Geschwindigkeit signifikante Verbesserungen der Gesamt-Performance. Denn diese Anwendungen und Spiele müssen Geometrien und Texturen "vorhersehen" und in den Frame-Puffer laden und profitieren daher von dem besseren Durchsatz zwischen Hauptspeicher und Grafiksbus. Auch Anwendungen mit hochpräzisen Daten und umfangreichen Texturen profitieren, denn auch bei diesen werden größere Datenmengen aus dem und in den Hauptspeicher übertragen.

Tabelle 1. Vergleich zwischen AGP 4X und AGP 8X

	AGP 4X	AGP 8X
Byte pro Transfer	4 (32 Bit)	4 (32 Bit)
Taktrate	266,67 MHz	533,33 MHz
Busbandbreite	1,1 GB/s	2,1 GB/s

Zusammenfassung

Grafikanwendungen belasten sämtliche Systemkomponenten und erfordern ständige Weiterentwicklungen, damit die Umgebungen praxistauglich bleiben und keine Engpässe entstehen. AGP 3.0 mit seiner AGP 8X-Bandbreite stellt im Bandbreitenkontinuum einen revolutionären Schritt nach vorn dar. Die höhere Bandbreite und das bessere Bus-Design von AGP 8X ergänzen neue Entwicklungen auf dem Gebiet der Grafikkarte und ermöglichen eine kreativere und geschicktere Nutzung der Hauptspeicherressourcen, so dass komplexe Texturfunktionen und Visualisierungen weit wirksamer unterstützt werden können. Während die Kompatibilität mit vorhandenen AGP 2X- und 4X-Systemen gewahrt bleibt, erlauben die neuen AGP 8X-Lösungen einen nahtlosen Aktualisierungspfad zu:

- ❑ Filmischen Grafikeffekten in Echtzeit und einer besseren Hauptspeichernutzung bei komplexen Geometrien, Texturen und hochpräzisen Daten

- Anwendungen mit höherer Performance, für die dynamische "Welten" geladen bzw. Streaming-Funktionen ausgeführt werden müssen
- Einer Unterstützung des isochronen Transfers, wie er für zahlreiche Digitalmedienanwendungen charakteristisch ist, zum Beispiel für das Streaming von digitalen Geräten und Netzwerken
- Einer ausgewogenen System-Performance, bei der die CPU in intelligenter Weise von Grafikoperationen und Datentransfers entlastet wird

AGP 8X führt bei zahlreichen Spielen und Anwendungen, die mit komplexen Texturen und Szenen arbeiten, unmittelbar zu einer Steigerung der Performance. Die neueste Version der AGP-Schnittstelle schützt das Grafiksубsystem vor Bandbreitenproblemen und verfügt über ein Leistungspotenzial, das auf Jahre hinaus für neu entwickelte Anwendungen genügen wird.

In den neuen GPUs und PC-Plattformen von NVIDIA werden die AGP 8X-Funktionen implementiert. Alle Vorzüge der AGP 3.0-Spezifikation stehen damit in vollem Umfang zur Verfügung. Wie schon seit jeher nutzen die Lösungen von NVIDIA neue Technologien und Weiterentwicklungen, ohne dass deswegen Einbußen bei der Stabilität und Qualität des Gesamtsystems in Kauf genommen werden müssen. Und dank der UDA, NVIDIAS vereinheitlichter Treiberarchitektur, lassen sich die neuen Funktionen, wann immer sie auf den Markt kommen, ohne lästige Treiberneuinstallationen sofort nutzen.



Hinweis

ALLE DESIGN-SPEZIFIKATIONEN, REFERENZKARTEN, DATEIEN, DIAGRAMME, DIAGNOSEN, LISTEN UND SONSTIGEN DOKUMENTE VON NVIDIA (EINZELN ODER IN KOMBINATION MITEINANDER, IM FOLGENDEN "MATERIALIEN" GENANNT) WERDEN "WIE GESEHEN" ZUR VERFÜGUNG GESTELLT. NVIDIA ÜBERNIMMT KEINERLEI GEWÄHRLEISTUNG - WEDER AUSDRÜCKLICH NOCH IMPLIZIT NOCH GESETZLICH VORGESCHRIEBEN NOCH VON SONSTIGER ART - IM HINBLICK AUF DIESE MATERIALIEN UND SCHLIESST AUSDRÜCKLICH JEDE IMPLIZITE GEWÄHRLEISTUNG HINSICHTLICH HANDELSÜBLICHER QUALITÄT, EIGNUNG FÜR EINEN BESTIMMTEN ZWECK ODER WAHRUNG DER RECHTE DRITTER AUS.

Die vorgelegten Informationen sind nach Ansicht von NVIDIA korrekt und verlässlich. Die NVIDIA Corporation übernimmt jedoch keinerlei Haftung für die Folgen der Nutzung dieser Informationen oder für Patentverletzungen oder für die Verletzung der Rechte Dritter, die möglicherweise aus der Nutzung dieser Informationen entstehen. Weder implizit noch in anderer Weise wird eine Lizenz im Zusammenhang mit Patenten oder Patentrechten der NVIDIA Corporation gewährt. Die in dieser Veröffentlichung genannten Spezifikationen unterliegen unangekündigten Änderungen. Diese Veröffentlichung ersetzt alle früheren Informationen. Ohne ausdrückliche schriftliche Genehmigung der NVIDIA Corporation sind Produkte der NVIDIA Corporation nicht für den Einsatz als kritische Komponenten in lebenserhaltenden Geräten oder Systemen zugelassen.

NVIDIA und das NVIDIA-Logo sind eingetragene Warenzeichen und GeForce2 Go ist ein Warenzeichen der NVIDIA Corporation.

Andere Firmen- und Produktnamen können Warenzeichen der Unternehmen sein, zu denen sie gehören.

Copyright NVIDIA Corporation 2002



NVIDIA.

NVIDIA Corporation
2701 San Tomas Expressway
Santa Clara, CA 95050
www.nvidia.com