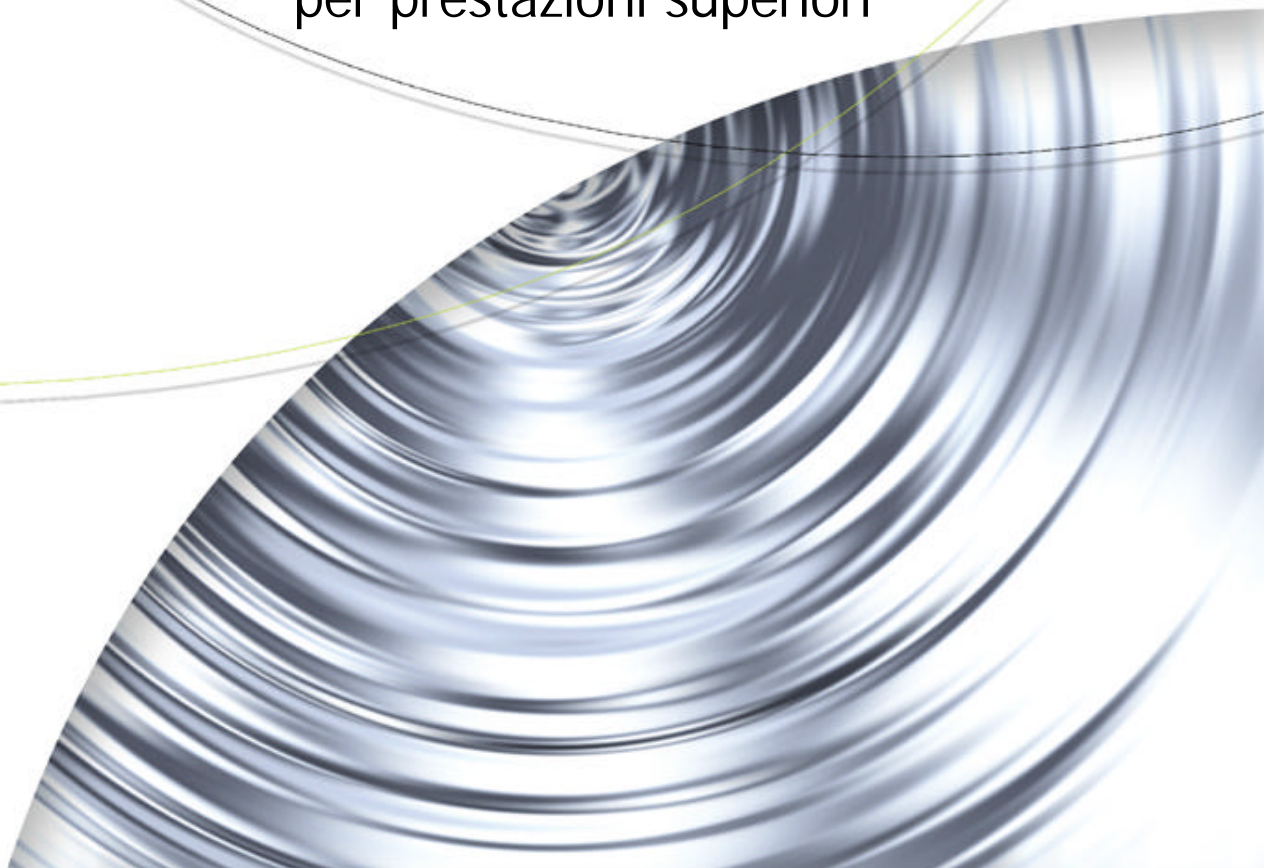




Note tecniche

Architettura della memoria alla velocità della luce II di NVIDIA

Svolte nella progettazione della memoria
per prestazioni superiori





GeForce4

Prestazioni di memoria LMA II

Creare degli ambienti 3D verosimili e in tempo reale su computer desktop è l'ambizione principale di migliaia di sviluppatori hardware e software dell'industria della computer graphics. Se gli sviluppi degli ultimi anni hanno compiuto notevoli passi in avanti nel migliorare la qualità della grafica 3D in tempo reale, una delle sfide fondamentali per fornire contenuti 3D interattivi è ancora presente: "Come incrementare le prestazioni grafiche dati i limitati aumenti della velocità della memoria?"

La larghezza di banda bus della memoria rimane un fattore decisivo per determinare la qualità e le prestazioni della grafica. L'architettura della memoria alla velocità della luce II di NVIDIA® (LMA II) comprende una quantità considerevole di successi rivoluzionari nell'architettura grafica. Questi progressi migliorano in modo considerevole l'efficacia della larghezza di banda della memoria delle unità di elaborazione grafica (GPU) della famiglia GeForce4, permettendo loro di fornire nuovi livelli di performance e una qualità d'immagine superiore.

Note: E' fondamentale capire le richieste del rendering di pixel sull'interfaccia della memoria, in modo da comprendere le varie sfide di resa dei mondi realistici 3D. Per l'elenco dei termini utili, consultare il Glossario.

Sfide della larghezza di banda della memoria

La larghezza di banda limitata della memoria rappresenta un grande ostacolo che limita il miglioramento delle prestazioni grafiche del computer. L'interfaccia della memoria viene sottoposta a richieste sempre maggiori e continue di larghezza di banda per eseguire, ad esempio, la scrittura dei pixel, la lettura dei pixel, l'aggiornamento video, le transazioni bus AGP e le letture della texture. Sfortunatamente, gli utenti finali notano un rallentamento delle prestazioni grafiche quando una di queste richieste multiple rimane nell'interfaccia della memoria.

Le sfide Fill Rate di pixel e larghezza di banda della memoria

Le architetture grafiche tradizionali rendono i pixel leggendo nei/dai vari buffer presenti nella memoria grafica. Questi buffer contengono i dati colore, vale a dire gli elementi alfa, rosso, verde e blu (ARGB) del colore del pixel, e contengono i valori Z, che rappresentano i parametri di profondità o di “visibilità” del pixel. Con la tecnica basilare di rendering si calcola un valore di colore per un dato pixel, si leggono il colore e i dati Z immagazzinati nel colore e nei buffer Z per quella posizione pixel, e si calcola il nuovo colore e il nuovo valore Z per quella posizione pixel, se necessario unendo e combinando eventuali valori di colore precedenti. Questo calcolo dipende dal valore Z e dai valori alfa. Una volta calcolati il nuovo colore e i valori Z, i nuovi dati devono essere immessi per iscritto nella memoria. Per quanto riguarda gli stencil buffer da 32 bit di colori e da 32 bit Z, queste operazioni comportano una quantità significativa di traffico bus di memoria per ogni pixel

In effetti, questo livello “di base” di rendering comporta 16 byte di larghezza di banda per ogni pixel:

Lettura colore a 32 bit + lettura 32 bit Z = 8 byte,
Scrittura colore a 32 bit + scrittura 32 bit Z = 8 byte, o 16 byte in totale

Il texturing e la complessità della profondità amplificano la larghezza di banda richiesta per ogni pixel di rendering di circa il 200 %. La maggior parte dei giochi attuali utilizza il multitexturing (applicando due o più texture ad un singolo pixel) in combinazione con il filtraggio texture bilineare o trilineare per ottenere un'immagine ricca di dettagli e altamente realistica. Questi dati texture devono essere letti dalla stessa memoria grafica utilizzata per immagazzinare il buffer Z e il buffer colore. Per spiegare a grandi linee come funziona la lettura della larghezza di banda della texture, considerare che ogni pixel comprende due texture (multitexturing) a filtraggio trilineare. Il filtraggio trilineare utilizza 8 campioni di texture per ogni texture di ogni pixel, e ogni campione di texture rappresenta 32 bit (4 byte) di dati. Ovviamente, ciò significa che occorre leggere 32 byte di dati a texture o 64 byte di dati a pixel per uno scenario di due texture. Siccome tutte le GPU forniscono lo stesso livello di cache texture, si supponga che metà dei dati della texture, in media, debba essere trasferita attraverso l'interfaccia esterna della memoria. (Abbiamo semplificato agli estremi questo problema complesso e dinamico, ma è una supposizione piuttosto ragionevole). In uno scenario di alta qualità con filtraggio trilineare, occorre leggere una media di 32 byte di informazioni dalla memoria grafica per ogni pixel reso.

La complessità della profondità moltiplica l'impatto di tutte le richieste di larghezza di banda descritte precedentemente perché, in un sistema di rendering normale, ogni pixel viene reso molte volte, in quanto alcuni oggetti si trovano di fronte ad altri. Sfortunatamente, la GPU non capisce se un pixel è nascosto finché non ha già eseguito il lavoro di rendering sul pixel nascosto, il che comporta una perdita preziosa di larghezza di banda.

Di conseguenza, la richiesta di nuova larghezza di banda per una modalità di visualizzazione 1280 x1024 x32bpp con un'applicazione ad una complessità di profondità media di 2,5 ed un'impostazione antialiasing di 2 AA campioni/pixel è:

Ris. orizz. x Ris. vert. x (Rendering di base + Letture texture) x Prof. x fattore AA = 1280 pixel/riga x 1024 righe/frame x (16 byte/pixel + 32 byte/pixel) x 2.5 x 2 =
315 milioni di byte a frame.

Effettuando il rendering a 60 f/sec., la richiesta della larghezza di banda della memoria sarà di 315 MB/frame x 60 f/sec., pari a **18,9 GB/sec. di larghezza di banda per il solo rendering del pixel!**

Tale richiesta diventa anche maggiore se consideriamo l'antialiasing 4x o le risoluzioni superiori, come 1600x1200. In particolare, una risoluzione di 1600x1200 con AA 4x richiede 921 milioni di byte per ogni frame o addirittura 55,3 GB/sec. di larghezza di banda della memoria.

Tuttavia, anche le attuali interfacce di memoria grafica più veloci forniscono solo circa 10,3 GB/sec. di larghezza di banda massima. Le GPU attuali consumano, virtualmente, tutta la larghezza di banda massima per effettuare unicamente il rendering dei pixel. Ciò ovviamente non comprende una risoluzione video più elevata, usando le impostazioni massime di antialiasing, né l'aggiornamento video o altre richieste, come le transazioni AGP. Questo livello di larghezza di banda è necessario su base sostenuta, non solo in linea di massima. Di conseguenza, la larghezza di banda della memoria rappresenta un limite alla performance e qualsiasi cosa riesca a fare la GPU per fare sì che la maggior parte del lavoro venga eseguito con una quantità prefissata di larghezza di banda dà luogo ad un'esperienza migliore per l'utente finale... nonché ad una grafica più ricca e frequenze di trama più veloci.

Architettura della memoria alla velocità della luce II di NVIDIA: una svolta nella larghezza di banda della memoria

L'architettura della memoria alla velocità della luce II di NVIDIA adotta molte tecnologie in attesa di brevetto in grado di migliorare l'efficacia di rendering dei pixel da parte delle GPU. Queste tecnologie comprendono un controller della memoria crossbar; un sottosistema cache di memoria Quad Cache; la compressione Lossless del buffer Z; un sottosistema di visibilità; l'azzeramento veloce dei buffer Z; il precaricamento automatico.

Controller della memoria crossbar

Siccome la grafica 3D dipende dalla larghezza di banda della memoria, il controller della memoria diviene fondamentale per migliorare le prestazioni. A prescindere dalla GPU, l'elemento più importante del sistema grafico è la memoria grafica locale. La memoria locale contiene i vari buffer: il frame buffer, il buffer Z/stencil, i buffer video, i buffer texture, i buffer per la geometria, e così via. Siccome prevede così tanti utilizzi e richieste, la memoria grafica locale rappresenta, solitamente, il sistema di memoria della larghezza di banda più importante di un computer. Di conseguenza, è fondamentale utilizzare le risorse disponibili nel modo più efficace.

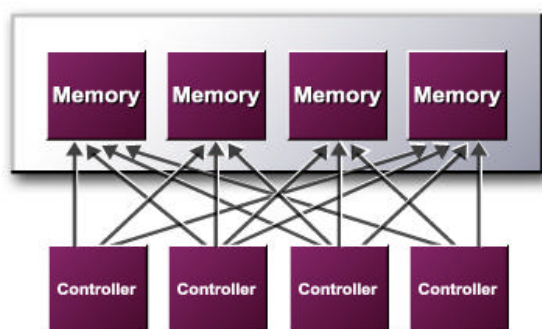
I controller di memoria tradizionali hanno raggiunto il punto in cui risultano sufficientemente efficaci a carichi di base, ottenendo più del 50 % della larghezza di banda della memoria massima dal frame buffer, a discapito di altre funzioni. Nei modelli attuali a doppia frequenza di dati (DDR), un normale controller di memoria a 128 bit accede, in realtà, a 256 bit di informazioni (perché i DDR trasferiscono il doppio delle informazioni in un singolo accesso).

Se il trasferimento di grosse quantità di dati in grandi blocchi rappresenta, solitamente, la soluzione ottimale – per scene complesse con centinaia di migliaia di poligoni a fotogramma – la realtà è piuttosto diversa. Per le applicazioni 3D all'avanguardia, la dimensione del triangolo medio (il blocco di realizzazione fondamentale di ogni grafica in tempo reale) può risultare molto piccola, talvolta solo di pochi pixel. Nel caso in cui un triangolo avesse la dimensione di due pixel e fosse composto di 32 bit di colore o di dati Z/stencil per ogni pixel, la quantità totale di dati per quel dato triangolo sarebbe di 32 bit x 2 pixel, o 64 bit.

Note: Solitamente, i pixel comprendono sia i dati relativi al colore che i dati Z/stencil, anche se immagazzinati in zone diverse della memoria. Di conseguenza, la lettura del colore avverrà a 64 bit o la lettura Z/stencil avverrà a 64 bit per il triangolo a due pixel di questo scenario, invece di avere una singola operazione di lettura a 128 bit.

Se i controller della memoria accedono alle informazioni solo a 256 bit, gran parte di questo accesso andrà persa, in quanto questo “precarico” o questa quantità di dati da trasferire provoca una perdita di gran parte della larghezza di banda potenziale del frame buffer. In questo esempio, un controller di memoria tradizionale a 128 bit risulta efficace solo al 25 %, “perdendo” il 75 % della larghezza di banda della memoria.

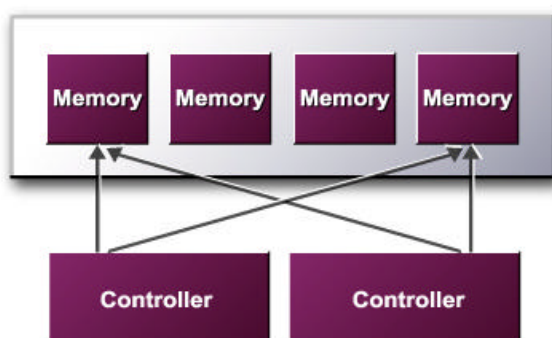
La GPU GeForce4 utilizza un controller della memoria crossbar radicale (vedi Figura 1), ottimizzato per accedere al frame buffer con uno schema di accesso a granularità fine. In questo modo, si garantisce un accesso individuale davvero efficace. La serie GeForce4 Ti comprende quattro controller indipendenti in grado di garantire prestazioni altamente qualificate. Il vantaggio dei quattro controller di memoria di GeForce4 Ti risiede nel miglioramento 4x della capacità della larghezza di banda.



L'interfaccia di memoria crossbar di GeForce4 Ti con quattro controller indipendenti fornisce una larghezza di banda della memoria quattro volte superiore alle architetture standard

Figura 1. **Interfaccia di memoria crossbar GeForce4 Ti**

L'interfaccia di memoria crossbar della GPU GeForce4 MX comprende una versione più economica che sfrutta due controller indipendenti (vedi Figura 2). Anche questa versione fornisce grossi vantaggi, essendo due volte superiore nel trasferire i dati rispetto ad un'applicazione non crossbar .



L'interfaccia di memoria crossbar GeForce4 MX con due controller indipendenti fornisce una larghezza di banda della memoria due volte superiore alle architetture standard

Figura 2. **Interfaccia di memoria crossbar GeForce4 MX**

I controller di entrambi i sistemi di memoria crossbar possono lavorare anche congiuntamente. In questo caso, la GPU riesce ad accedere a 256 bit di informazioni in un singolo ciclo.

Il vantaggio dell'interfaccia di memoria crossbar consiste nella flessibilità di accesso alle informazioni a 64 bit, a 128 bit o a 256 bit, in base al percorso di accesso più efficace in quel dato momento. In questo modo, l'efficacia di ognuno di questi accessi si avvicina alla perfezione, utilizzando a piena performance ogni aspetto del processore grafico e il frame buffer relativo.

Questo complesso sistema bilancia il carico continuamente, garantendo l'equilibrio di ogni aspetto del sistema di memoria e la gestione adeguata di ogni richiesta di memoria da parte del processore grafico. In situazioni di carico complesso, tipiche dei modelli di prossima generazione, l'architettura di memoria crossbar LMA II riesce veramente a fornire una larghezza di banda della memoria con una capacità da due a quattro volte superiore rispetto alle architetture standard.

Quad Cache

L'architettura LMA II comprende, inoltre, dei sottosistemi cache comunemente conosciuti come Quad Cache. Le cache non sono altro che buffer di accesso ad alta velocità in grado di immagazzinare piccole quantità di dati e di funzionare a larghezze di banda estremamente alte. Questi buffer speciali garantiscono che i dati vengano accodati e preparati per essere scritti in memoria fornendo, al contempo, una posizione adeguata affinché i dati del buffer possano essere letti dalla memoria.

La Quad Cache comprende quattro buffer indipendenti e ottimizzati, in grado di migliorare le prestazioni generali della pipeline grafica. Ogni cache immagazzina informazioni su primitive, vertici, texture e pixel. Queste cache vengono ottimizzate singolarmente per le informazioni

specifiche che devono gestire. I risultati o i dati immagazzinati precedentemente in queste cache possono essere prelevati quasi all'istante (non vengono recuperati dalla memoria o ricalcolati). Come risultato si ottiene un recupero veloce dei dati chiave che garantisce la massima performance della pipeline grafica.

Esempio: Nella sezione della larghezza di banda della memoria e del fill rate dei pixel, abbiamo mostrato la sequenza relativa al calcolo di una scena filtrata trilineare con due texture per ogni pixel, basato su una risoluzione 1280 x 1024 e a un ciclo di lettura /scrittura a 16 byte:

$$1280 \text{ pixel/riga} \times 1024 \text{ righe/frame} \times (16 \text{ byte/pixel} + 32 \text{ byte/pixel}) \times 2.5 \times 2 = 315 \text{ milioni di byte a frame.}$$

La cache texture Quad Cache riesce, da sola, a ridurre la quantità di byte di texture a pixel da 32 a 8! Ciò significa che l'esigenza di larghezza di banda a fotogramma passa da 315 milioni a 157 milioni di byte a fotogramma $(1280 \times 1024 \times (16+8) \times 2.5 \times 2 = 157 \text{ milioni})$.

La Quad Cache riduce la larghezza di banda della memoria totale utilizzata consentendo, al contempo, di rendere scene più belle e complesse.

Compressione Z Lossless

Il buffer Z tratta le informazioni di profondità o di visibilità per i pixel che verranno visualizzati in seguito al rendering. Le GPU tradizionali leggono e, se possibile, scrivono i dati Z per ogni pixel di rendering, trasformando il traffico del buffer Z, uno dei più grossi "consumatori" di larghezza di banda della memoria di un sistema grafico. Per mezzo di una metodologia proprietaria avanzata di compressione dati Lossless con un livello di compressione 4:1, la larghezza di banda della memoria consumata dal traffico del buffer Z può essere ridotta da un fattore di quattro. Questo tipo di compressione Z viene applicata negli elementi hardware e risulta completamente trasparente nelle applicazioni; la compressione e la decompressione avvengono in tempo reale tramite i motori di compressione /decompressione Z dell'architettura LMA-II. Siccome questo tipo di compressione è tipicamente Lossless, non si riduce la qualità dell'immagine e non si verifica una perdita di precisione. Il risultato di questa tecnologia si traduce in un sottosistema di memoria assolutamente efficace, che fornisce maggiori prestazioni senza compromettere la qualità dell'immagine.

Sottosistema di visibilità: Occlusione Z

Come indicato precedentemente, le architetture grafiche tradizionali effettuano il rendering di ogni pixel di ogni triangolo non appena vengono ricevuti, accedendo al frame buffer con ogni pixel, in modo da determinare i valori adeguati per il colore o per il parametro Z (o profondità) di ognuno di questi pixel. Questo metodo fornisce dei risultati corretti, ma è necessario procedere al rendering di ogni pixel, siano essi visibili oppure no. I normali modelli attuali presentano una complessità di profondità media di due, il che significa che è necessario rendere due pixel per ogni pixel da visualizzare. Ciò significa che, per ogni pixel visibile, il processore grafico deve compiere tale lavoro due volte, dato che la GPU dovrà accedere al frame buffer due volte, provocando un dispendio considerevole della larghezza di banda del frame buffer per effettuare il rendering dei pixel che non verranno visualizzati.

L'architettura LMA II adotta una tecnologia sofisticata di occlusione Z, in grado di determinare se un dato pixel verrà visualizzato. Nel caso in cui l'unità di occlusione Z determini che il pixel non verrà visualizzato, quel pixel non sarà sottoposto al rendering e non si verificherà l'accesso al frame buffer, con un conseguente risparmio della larghezza di banda del frame buffer. In base alla complessità di profondità della scena, ciò può comportare un'ottimizzazione incredibile dell'efficacia. Nei modelli attuali con una complessità di profondità media di due, questa tecnica potrebbe ridurre la necessità di larghezza di banda fino al 50 %. Nei modelli di prossima generazione con complessità di profondità quattro o più, i vantaggi possono essere enormi, fino ad un miglioramento 4x della capacità della larghezza di banda della memoria.

Una tecnica aggiuntiva che può essere impiegata come una "richiesta di occlusione". Fondamentalmente, l'applicazione fa sì che il processore grafico richieda di renderizzare una zona o una bounding box (parallelepipedo che comprende gli oggetti considerati) per testarne la visibilità. Se la GPU determina di occludere tale zona, ogni figura geometrica e rendering rappresentativo di quella zona può essere tralasciata, offrendo un potenziale ordine di aumento di ampiezza nella fill rate. I personaggi posti dietro a pareti o le scene esterne di una galleria possono essere semplicemente "mandati in occlusione" e tralasciati, senza sprecare larghezza di banda preziosa della memoria o il tempo di elaborazione della GPU per renderizzarli.

L'occlusione Z e la richiesta di occlusione rappresentano due tecnologie importanti che amplificano effettivamente la larghezza di banda di una GPU. Queste tecnologie aumentano considerevolmente la capacità della larghezza di banda della memoria offerta dal frame buffer e consentono di utilizzare il frame buffer in modo più efficace. In alcuni casi, i vantaggi sono quattro volte superiori rispetto le architetture precedenti anche se, in pratica, il vantaggio di queste tecnologie che amplificano la larghezza di banda della memoria si traduce in un miglioramento medio del 50 - 100 %.

Pre caricamento automatico

Un altro "onere" nascosto riguardo la larghezza di banda della memoria e spesso trascurato consiste nei vari ritardi dovuti alla gestione della pagina all'interno del chip della memoria. Le GPU di NVIDIA provviste di LMA II presentano una caratteristica speciale, detta "pre caricamento automatico", destinata a recuperare questa perdita di larghezza di banda. Per apprezzare i vantaggi del pre caricamento automatico, è importante capire che le memorie di accesso random dinamico (DRAM) sono disposte in file, colonne e "moduli", in modo da semplificare la struttura dell'unità di memoria e dell'interfaccia di memoria. Questa disposizione consente al computer di accedere ai dati contenuti in memorie molto grandi (una memoria 4Mx32 è pari a 16 MB di memoria su un singolo chip di memoria) con pochi cavi di controllo, riducendo gli sforzi totali sia per le memorie, sia per la GPU.

Lo svantaggio è che è possibile accedere immediatamente solo alla fila o alla colonna corrente di un modulo attivo. Se la GPU volesse leggere da (o scrivere su) una zona diversa del chip di memoria, dovrà dire alla memoria di chiudere il modulo corrente e di riattivare, in un secondo tempo, il modulo con le nuove impostazioni della fila e della colonna. Questo processo può impiegare fino a 10 DRAM, in quanto è necessario "pre caricare" un modulo in seguito alla chiusura dello stesso e prima della sua riattivazione. Ciò significa che, spesso, nessun dato si sposta attraverso il bus mentre la GPU attende che la memoria sia pronta. Questa penalità di tempo pari a 10 deve essere pagata ogni qualvolta si rende necessario modificare una fila o una colonna.

Le GPU della famiglia GeForce3 riescono a dire all'unità di memoria di "precaricare" attivamente le zone della memoria non in uso ma che si prospetta vengano utilizzate entro breve. (Questi tempi brevi, in termini di memoria, sono generalmente misurati in milionesimi di secondo o microsecondi). Il vantaggio consiste nel fatto che la GPU non dovrà attendere la fase di precaricamento quando dovrà accedere a queste zone. Dovrà attendere unicamente la fase di attivazione, che varia in base al fornitore di memoria e all'unità di memoria ma che, solitamente, è di 2 o 3 temporizzazioni. Il risultato consisterà in un minore tempo di attesa della GPU nei confronti della memoria e di un maggiore tempo di rendering dei pixel, che si traduce in prestazioni grafiche più veloci per l'utente.

Fast Z-Clear

L'azzeramento dei "vecchi" dati del buffer Z è un altro compito che consuma larghezza di banda preziosa della memoria. Siccome i dati del buffer Z sono usati per determinare se i dati del pixel dovranno essere "sovrascritti," tale procedimento dovrà risultare accurato, altrimenti l'utente visualizzerà oggetti non adeguati. Sfortunatamente, l'azzeramento del buffer Z (che consiste nello scrivere degli 0 in tutte le posizioni per cancellare i dati esistenti) spreca tempo e larghezza di banda della memoria. Le GPU della famiglia GeForce4 adottano la tecnologia Fast Z-Clear per ridurre il tempo necessario per azzerare i "vecchi" dati nel buffer Z. I risparmi potenziali di larghezza di banda per una visualizzazione 1600x1200, con un rendering del buffer Z di 32 bit per ogni pixel a 60 Hz, è di 460 MB/sec. ($1600 \times 1200 \times 60 \text{ Hz} \times 32 \text{ bit}$). La tecnologia Fast Z-Clears incrementa le frequenze di trama fino al 10 %, senza compromettere la qualità dell'immagine!

Conclusione

Per affrontare i problemi legati alla larghezza di banda del pixel, LMA II introduce una serie di tecnologie in grado di ottimizzare la capacità della larghezza di banda della memoria e di fornire prestazioni grafiche per computer davvero innovative. La combinazione delle architetture di memoria crossbar più efficaci e sofisticate con Quad Cache, l'avanzata compressione Z Lossless per ridurre il consumo di larghezza di banda, il precaricamento automatico, l'azzeramento veloce del buffer Z e una metodologia altamente avanzata per evitare il rendering e consumare larghezza di banda su pixel non visibili presenta vantaggi considerevoli. Le GPU provviste dell'architettura LMA II utilizzano la larghezza di banda della memoria in modo più efficace rispetto alle architetture grafiche precedenti.

I vantaggi offerti dall'architettura LMA II preparano il cammino ad esperienze grafiche 3D in tempo reale, ricche di dettagli e sempre più dinamiche. Migliorando la capacità di comunicazione fra l'host e la grafica, gli sviluppatori di contenuti possono continuare ad aumentare la ricchezza geometrica e la complessità visiva delle loro scene a nuovi livelli. In questo modo, gli utenti finali possono godere degli ambienti, delle immagini e degli effetti più realistici e verosimili mai visti prima. Date libero sfogo al divertimento!



Glossario

Profondità bit

La profondità bit indica il numero di bit di precisione per il colore e i valori Z associati ad ogni pixel sullo schermo. Maggiori sono i bit di precisione, migliori sono il realismo visivo e l'accuratezza del frame di resa. Le due profondità bit più comuni nella grafica hardware moderna sono i 16 bit e i 32 bit. Ognuno di questi valori può essere associato al colore o ai valori Z. Un colore a 32 bit (ad esempio) è solitamente usato per rappresentare i valori rosso, verde, blu e alfa (o trasparenza), fino a 8 bit ad elemento, o 256 "valori" per ognuno di questi elementi. Un valore Z a 32 bit è solitamente assegnato come 24 bit di precisione Z (o precisione di profondità) e 8 bit di precisione stencil o "maschera".

Complessità della profondità

La complessità della profondità misura la complessità di una scena. Indica il numero di rese necessarie di ogni pixel prima di ottenere il frame. Ad esempio, l'immagine di resa di una parete presenta una complessità di profondità pari a uno. L'immagine di una persona di fronte alla stessa parete presenta una complessità di profondità pari a due. L'immagine di un cane dietro la persona ma di fronte alla parete presenta una complessità di profondità pari a tre, e così via. Con l'aumentare della complessità della profondità, maggiori sono la capacità di resa e la larghezza di banda necessarie per rendere ogni pixel o scena. La complessità media di profondità delle applicazioni grafiche attuali è pari a 2 - 3, il che significa che per ogni pixel che si desidera visualizzare, il processore grafico dovrà effettuare il rendering due o tre volte.

Fill rate

La fill rate rappresenta la velocità con cui vengono tracciati i pixel nella memoria video. La fill rate è una misura comune utilizzata per indicare le capacità di elaborazione del pixel dei processori grafici 3D di oggi. La fill rate è solitamente misurata in milioni di pixel al secondo (Mpixel/sec.). Nel 1997, 50 - 70 Mpixel/sec. erano considerati il massimo raggiungibile. Nel 2002, i principali processori grafici 3D saranno in grado di trattare oltre 1200 Mpixels/sec. Se questo miglioramento rappresenta un passo in avanti incredibile, siamo ancora lontani dal poter affermare di poter creare

un ambiente 3D del tutto convincente. La resa di pixel ad una velocità così elevata comporta un eccessivo consumo di larghezza di banda della memoria.

Frame al secondo

I frame al secondo (f/sec.), o frequenza di trama, indicano quante volte al secondo la scena viene aggiornata dal processore grafico. Frequenze di trama elevate comportano un'animazione più realistica e regolare. Solitamente, 30 f/sec. forniscono un livello accettabile di animazione, ma aumentando le prestazioni a 60 f/sec, i risultati mostrano un'interazione ed un realismo decisamente migliori. Oltre i 75 f/sec. risulta difficile cogliere miglioramenti di performance. Se si visualizzano le immagini più velocemente rispetto la frequenza di aggiornamento del monitor, si ottiene una potenza ridotta della capacità di elaborazione grafica, in quanto il monitor non è in grado di aggiornare i propri fosfori (o display) oltre la frequenza di aggiornamento.

Larghezza di banda della memoria

La larghezza di banda della memoria indica la velocità di trasferimento dei dati fra il processore grafico e la memoria grafica. Le limitazioni della larghezza di banda della memoria sono uno dei problemi principali da superare se si vogliono creare ambienti 3D davvero realistici. Una resa 3D davvero sorprendente necessita di una risoluzione elevata, di una profondità di colore a 32 bit a frequenze di trama elevate, con una geometria ricca, un texture mapping sofisticato e un vertex e pixel shading complesso.

Risoluzione

La risoluzione indica il numero di pixel su uno schermo. Risoluzioni più elevate possono creare un ambiente 3D più realistico, perché si possono visualizzare maggiori dettagli della scena. I display più moderni sono in grado di visualizzare almeno 1280 pixel orizzontali x 1024 pixel verticali, mentre i display più grandi e più costosi sono in grado di visualizzare 2048 x 1536 pixel. La maggior parte delle applicazioni grafiche supporta varie risoluzioni, che consentono all'utente finale di lavorare a risoluzioni più elevate (ottenendo, quindi, un maggiore livello di dettagli), ma anche con un maggiore carico sul sistema di elaborazione grafica.

Texture Mapping

La texture mapping è una tecnica che permette di proiettare un'immagine 2D (solitamente un bitmap) su un oggetto 3D. La texture mapping consente di aumentare in modo considerevole i dettagli visualizzati, senza aumentare troppo la quantità di poligoni conseguente. Dato il maggiore realismo ottenibile con un piccolo aumento di costo computazionale, la texture mapping è una delle tecniche più comuni per visualizzare oggetti 3D in modo realistico. Per mappare la texture di un pixel, i dati della texture per tale pixel devono essere letti nel processore grafico, il che comporta un consumo della larghezza di banda della memoria.



Le informazioni fornite devono ritenersi accurate e affidabili. Tuttavia, NVIDIA Corporation non si ritiene responsabile delle conseguenze dovute all'utilizzo di tali informazioni o di eventuali infrazioni di brevetti o altri diritti da parte di terzi e derivanti dall'utilizzo di tali informazioni. Non sono ammesse concessioni, in modo implicito o esplicito, di brevetti o di diritti su brevetti appartenenti a NVIDIA Corporation. Le specifiche menzionate nella presente documentazione sono soggette a modifiche senza preavviso. La presente documentazione sostituisce qualsiasi informazione fornita precedentemente. I prodotti di NVIDIA Corporation non possono essere utilizzati in condizioni critiche sui dispositivi di supporto o sui sistemi senza il consenso specifico e per iscritto di NVIDIA Corporation.

Marchi di fabbrica

NVIDIA, GeForce4, GeForce4 Ti e il logo NVIDIA sono dei marchi di fabbrica esclusivi di NVIDIA Corporation.

Ulteriori nomi di prodotti o di aziende possono essere dei marchi di fabbrica delle aziende relative con cui sono associati.

Copyright

Copyright NVIDIA Corporation 2002.



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