



So oder so ähnlich
sie die **Fermi-
GeForce** aus
(Fotomontage).

Geforce GTX 380 mit DirectX 11

Anfang Januar stellte Nvidia seine neue GeForce-Generation vor. Wir beleuchten Technik, voraussichtliche Leistung und die Verspätung des potenziellen Radeon-HD-5000-Killers.

GameStar.de
DirectX 11 im Detail
► Quicklink: 6472
DirectX-11-Spiele
im Überblick
► Quicklink: 6632

Kurz nach dem Jahreswechsel in der Stadt des Spiels und der Sünde: Zwischen lebendigen Löwen, einarmigen Banditen und einer David-Copperfield-Inszenierung zeigt Nvidia im MGM Grand Hotel einer Handvoll gespannter Journalisten seine neue GeForce-Grafikkarte mit DirectX-11-Unterstützung. Eines vorweg: Der neue Nvidia-Chip ist von den Eckdaten her ein beeindruckendes technisches Monument. In der Praxis muss er sich allerdings erst beweisen – und das wird noch dauern. Im Vergleich zur Vorgängergeneration steigt die Zahl der für die Spieleleistung wichtigen Shader-Prozessoren gleich von 240 auf 512 um mehr als das Doppelte! Der alte GDDR3-Speicher geht endlich in Rente und wird durch den zweimal schnelleren GDDR5 ersetzt, den AMD für seine Radeons schon seit geraumer Zeit verwendet. Nvidia verspricht nicht weniger als die bei weitem schnellste Grafikkarte mit lediglich einem Prozessor – und damit die neue Leistungsreferenz.

Leider können wir das bisher nicht mit eigenen Benchmarks überprüfen, Nvidia wollte noch keine Testgeräte rausrücken. Das

sei zu früh, so die Firmenvertreter vor Ort. Und weiter: Weder stünden die Taktraten fest, noch sei der Kühler endgültig. Erst im März soll Nvidias DirectX-11-Chip den Weg in unsere Spiele-PCs finden – über sechs Monate später als ursprünglich angekündigt. Zunächst sind mit der voraussichtlich rund 500 Euro teuren GeForce GTX 380 und der GeForce GTX 360 (400 Euro) zwei High-End-Varianten geplant. Ab Sommer dürften günstige Mittelklasse-Karten (GeForce GTS 300) folgen.

Auf den folgenden vier Seiten beantworten wir die dringlichsten Fragen zur neuen GeForce im Detail. Was wird die Karte in Spielen leisten? Wie unterscheidet sie sich von AMDs starker DirectX-11-Konkurrenz Radeon HD 5000? Warum hat sich Nvidias DirectX-11-Debüt so verzögert?

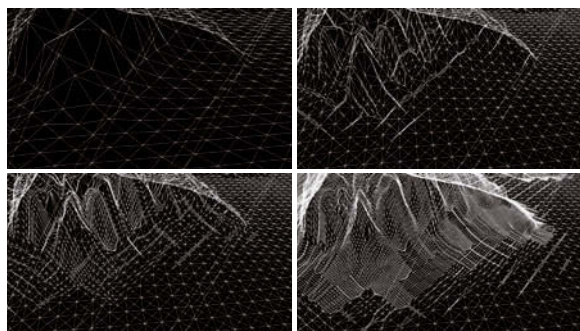
Unendliche Geschichte

Auf der Cebit 2009 werfen Windows 7 und DirectX 11 ihre Schatten voraus. Neugierige Journalisten, darunter die Redakteure von GameStar, recherchieren auf allen Kanälen (Grafikkarten-Hersteller, Gerüchteküche, allabendliche Messe-Partys, befreundete

Techniker, Entwicklungsleiter Spielstudios) nach der kommenden 3D-Generation. Nach und nach stellt sich heraus, dass AMD pünktlich zum Windows-7-Start am 23. Oktober 2009 seine DirectX-11-Radeons verkaufen können wird. Letztlich kommen die Karten sogar einen Monat früher, allerdings in anfangs winzigen Stückzahlen. Nvidia posaunt auf der Cebit: Bis zum Erscheinungstermin von Windows 7 ist auch die DirectX-11-GeForce fertig. Die besser Informierten haben Zweifel, besonders die Techniker der Grafikkarten-Hersteller signalisieren Skepsis. Stand der Akte DirectX 11 Ende März 2009: AMD ist zuerst dran, aber Nvidia kontert bis Weihnachten.

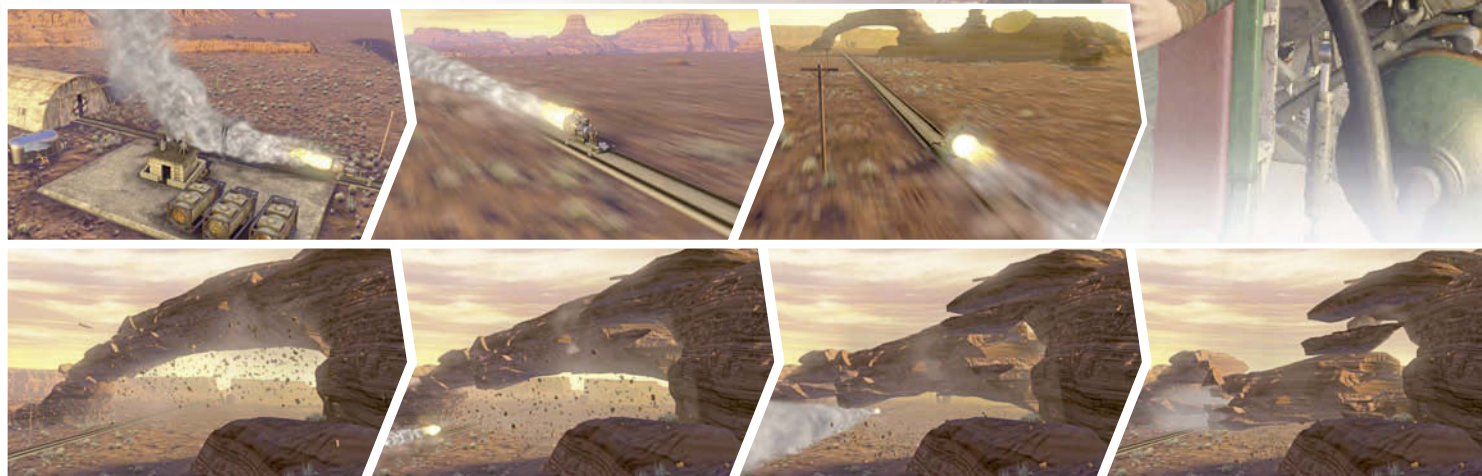
Anfang Juni 2009 zeigt AMD auf der Computex-Messe in Taiwan stolz einen Wafer – eine große Siliziumscheibe, aus der Computerchips gefertigt werden. In diesem Fall nicht irgendeinen, sondern einen mit den mit Spannung erwarteten DirectX-11-Chips. Die ersten Prozessoren sind also gefertigt, der sogenannte Tape-Out passiert. Auf Basis dieser Chips wird die Radeon HD 5000 zur Marktreife weiterentwi-

ckelt und schließlich am 23. September offiziell vorgestellt. Von Nvidia gibt es auf der Computex keine neue Wasserstandsmeldung. Fast zwei Monate später, am 29. Juli 2009, berichtet die oft richtig, aber auch manchmal falsch liegende Internet-Gerüchteküche SemiAccurate.com als Erste über einen Tape-Out von Nvidias kommendem Fermi-Grafikprozessor mit DirectX-11-Unterstützung. SemiAccurate-Redakteur Charlie Demerjian rechnet vor, wie lange es von der Fertigung der ersten Charge bis zum verkaufsfertigem Produkt normalerweise dauert und prophezeit: Frühestens Anfang 2010, eventuell erst Mitte 2010 wird Nvidia zurückschlagen können, mit dem auf der Cebit angekündigten Angriff zu Weihnachten sei es Essig. Am 15. September, kurz vor der Markteinführung der ersten DirectX-11-Karten durch AMD, legt SemiAccurate neue Informationen nach: Der Anteil der funktionierenden Fermi-Chips auf einem Wafer läge unterhalb von 2 Prozent. Üblich für so einen komplexen Baustein, der in einem neuen Fertigungsprozess entsteht, seien wenigstens 30 Prozent. Davon unbe-



Im Drahtgittermodus wird ersichtlich, wie die Supersled-Demo **Tessellation** von DirectX 11 einsetzt, um die Polygonzahl der Gebirge dynamisch anzupassen.

Traditionell fertigt Nvidia mindestens zu jeder neuen DirectX-Generation einen entsprechende Technik-Demo an. Die spektakuläre **Supersled**-Demo nutzt viele Funktionen von DirectX 11 und soll später kostenlos zum Download bereit stehen. Die Idee des Raketenschlittens orientiert sich an NASA-Versuchen aus den 1950er-Jahren.



Der Felsüberhang hält dem Druck des Fahrtwindes der Rakete nicht stand und fällt in sich zusammen. Die physikalischen Berechnungen laufen über **PhysX** auf der Fermi-Geforce.

eindrückt dementiert Nvidia Probleme mit der DirectX-11-Geforce.

Jetzt wartet die Technikwelt gespannt auf die Nvidia-Hausmesse, die GPU Technology Conference Ende September. In seiner Eröffnungsrede am 30. September 2009 präsentiert Nvidia-Boss Jen-Hsen Huang die neue Fermi-Architektur der Weltöffentlichkeit. Der Schwerpunkt seiner Präsentation widmet sich allerdings nicht den Spielefähigkeiten des neuen Chips, sondern dessen professionellem Einsatz in Großrechnern, zum Beispiel zur Klimaforschung. Nur vage führt Huang aus, dass Fermi auch als Geforce neue Leistungsrekorde aufstellen und besondere Funktionen für Spieler bieten werde. Zwei Tage später geht ein Lauffeuer um den Globus: Die von Jen-Hsen Huang vor der versammelten Presse als funktionstüchtig bezeichnete Karte ist ein Fake! Eine der beiden Strombuchsen hat gar keine Verbindung zur Platine, Teile der Luftauslässe sind verdeckt und ein Bauteil blockiert die SLI-Verbindungsstücke. Funktionstüchtige Karten scheint Nvidia aber zu haben, denn eine vorgeführte Technik-Demo läuft offensichtlich

auf Fermi-Hardware (wobei sich die Performance auch durch ein SLI-System der Vorgängergeneration hätte simulieren lassen).

In der Folge versuchen wir herauszufinden, wann denn nun die neue DirectX-11-Geforce endlich erscheint. Wir stoßen auf Bestätigungen von Demerjians These: nicht vor der Cebit 2010.

Was Nvidia jetzt sät, nennen Werbestrategen »FUD«, die englische Abkürzung für Furcht, Ungewissheit und Zweifel: Mit gezielt gestreuten Aussagen soll der (in diesem Fall) technologische Vorteil der Konkurrenz kleingeredet, potenzielle Kunden verunsichert werden. Am 8. Dezember spricht Michael W. Hara, Nvidias Vize-Präsident für Investor Relations, auf einer Konferenz für Finanzanalysten DirectX 11 die Bedeutung ab: »Nur wegen DirectX 11, ein paar zusätzlichen Effekten und mehr Spieleleistung kauft sich keiner eine neue Grafikkarte«. Auch Nvidias PR-Abteilung wird offenbar angewiesen, in den Redaktionen für Ruhe zu sorgen. Bei GameStar gehen erboste Anrufe bezüglich unserer Berichterstattung ein, in denen Aussagen fallen, wie PhysX sei viel wich-

tiger als DirectX 11. Wie ist der Stand heute, im Januar 2010? AMD kämpft weiter mit Liefer-schwierigkeiten, will nach eigener Aussage aber bereits 800.000 DirectX-11-Grafikchips verkauft und bis Februar seine komplette Grafikkarten-Serie auf DirectX 11 umgestellt haben. Irgendwann ab März gibt es dann auch die neue Geforce im Handel. Eins ist klar: Nvidia steht unter Druck, diese Karte muss ein Knaller werden.

Generalüberholte Chip-Architektur

Zu Beginn der Präsentation des GF100-Grafikprozessors in Las Vegas am 10.01.2010 entschuldigt sich Nvidia bei den Fans für die lange Wartezeit auf eine DirectX-11-Geforce. Vier Jahre habe man an der dem Chip zugrunde liegenden Fermi-Architektur gearbeitet, um die Seele eines Supercomputers im Körper eines Grafikchips unterzubringen, so blumig drückt sich Nvidias Drew Henry an diesem besonderen Tag aus. Herausgekommen ist der mit Abstand komplexeste PC-Baustein – der GF100 besteht aus über 3 Milliarden Schaltwerken! Zum Vergleich: AMDs DirectX-11-Chip RV870 hat

2,15 Milliarden Transistoren, der GT200b auf der Geforce GTX 285 rund 1,4 Milliarden und ein Core i7-Vierkerner lediglich 731 Millionen. Am meisten Platz belegen die von 240 auf 512 mehr als verdoppelten Shader-Prozessoren, von Nvidia in Anlehnung an die gleichnamige Programmierschnittstelle für Anwendungen abseits reiner Grafikberechnungen (Physik, Video-Konvertierung) nun CUDA-Cores genannt. Problem: Je nach Zählweise kommt eine Radeon HD 5870 auf 1.600 beziehungsweise 320 Shader-Einheiten. Endgültig vergleichen lassen sich AMDs und Nvidias Angaben also bisher nicht.

Die 512 Shader sind in 16 Blöcken mit je 32 Recheneinheiten organisiert, so dass Nvidia die Leistung der einzelnen Geforce-Modelle in diesen Schritten anpassen kann. Generell sind bei der Prozessorherstellung oft auch einzelne Teile des Chips defekt. Dann erfüllt der jeweilige Chip zwar nicht die Anforderungen des Spitzenmodells, lässt sich aber noch als preis- und leistungsreduzierte Variante verkaufen. Das künftige Flaggschiff, wahrscheinlich Geforce GTX 380 genannt,



Nur mit **Raytracing** lassen sich derart realitätsnahe Spiegelungen umsetzen – prinzipiell funktioniert diese Art der 3D-Darstellung physikalisch korrekt. Um die Luxusautos in Echtzeit zu berechnen, genügt aber auch die Performance der kommenden Geforce nicht.



wird vermutlich auf alle 512 Shader zurückgreifen, obwohl die für den Profi-Bereich gedachten Tesla-Fermis nur 448 aktive Shader haben (14 von 16 Blöcken). Wenn es bei 512 Shadern für die GTX 380 bleibt, dann erwarten wir bei der GTX 360 zwei deaktivierte Blöcke, also 448 Shader. Jeder Block hat vier Textureinheiten, der gesamte Chip also 64 und damit 16 weniger als der GT200-Vorgänger und AMDs RV870. Weil aber längst nicht mehr die Textureinheiten, sondern in erster Linie die Shader die Spieleleistung beeinflussen, dürfte dies in der Praxis keine Nachteile bedeuten.

Ein Problem von aktueller Spielegrafik allgemein identifiziert Nvidia in der Darstellung von geometrischen Formen, die uns im Alltag und in Computerspielen überall begegnen. Heutzutage wird jeder Pixel fast genauso penibel bearbeitet wie im Kino, die eigentliche Struktur von Objekten aber nur grob geschnitzt. Es

macht in der Nahansicht eben einen Unterschied, ob zum Beispiel eine Tastatur über eine noch so trickreiche Textur angedeutet oder mit Hunderten von Polygonen in 3D ausmodelliert wurde. Von der Geforce FX 5800 (2003) bis zur GTX 285 hat sich die Pixel-Shader-Performance laut Nvidia um das 150fache erhöht, die Geometrieleistung allerdings nur um das Dreifache. Allein durch die Anforderungen von DirectX 11 ergebe sich eine um das Achtfache gesteigerte Geometrieleistung. Die wichtigste Neuerung von DirectX 11 sieht Nvidia in der Tessellation-Funktion, mit der die Polygonzahl von Objekten dynamisch angepasst werden kann, wie die beeindruckende Raketenschlitten-Demo zeigt (in den Bildunterschriften gehen wir näher auf die Technik-Tricks der Demo ein). Ein Widerspruch zur FUD-Kampagne im Vorfeld – oder die tatsächliche Meinung von Nvidia?

Natürlich unterstützt der GF100-Chip auch alle anderen Bestandteile von DirectX 11, die wir in der Vergangenheit bereits ausführlich erläutert haben. Darunter: höhere Performance durch Optimierung auf Mehrkernprozessoren, saubere Berechnung von sich überlappenden, transparenten Objekten, bessere Texturkompression und Schattendarstellung (mehr dazu lesen Sie auf www.GameStar.de).

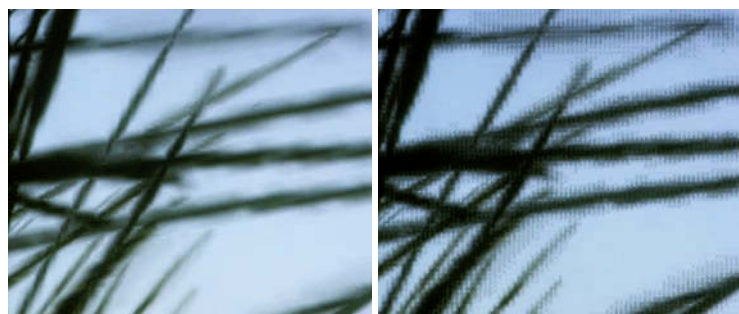
Wie bereits erwähnt, kommt die Fermi-Architektur nicht nur auf Geforce-Grafikkarten zum Einsatz. Gerade im umsatzträchtigen Profi-Segment will Nvidia seine Position mit dem neuen Chip stark ausbauen. Deshalb arbeiten auf dem GF100 einige Schaltkreise, die es in dieser Form vielleicht gar nicht gäbe, wenn er ausschließlich als Geforce, also rein für Spieler, konzipiert worden wäre. Spiele können dennoch davon profitieren. Dazu gehören vor allem die für einen Grafikprozessor einmalig groß ausgelegten L1- und L2-Caches – kleine, rasend schnelle Zwischenspeicher, in den die einzelnen Shader ihre Rechenergebnisse ablegen und Informationen miteinander austauschen können. Ferner kann der GF100 viel schneller zwischen seinen beiden Betriebsarten Grafik und CUDA hin- und herwechseln als der GT200. Wenn die Grafikkarte in Spielen außer der Grafik noch Physikberechnungen oder die Wegfindung der KI übernehmen soll, summieren

sich diese Wartezeiten schnell – je kürzer, desto besser also. Zusätzlich beherrscht Fermi wie Intels aktuelle Hauptprozessoren eine Art Hyperthreading, die jedoch nur in bestimmten Situationen anspringt.

Moderne Speicher-Konfiguration

Mit seiner DirectX-11-Generation wechselt Nvidia endlich auf den energieeffizienten und zugleich äußerst schnellen GDDR5-Videospeicher. Pro Taktzyklus überträgt er mit vier Datenpaketen doppelt so viele wie die auf GTX-200-Karten eingesetzten GDDR3-Module. Um Kosten zu sparen, schrumpft die Speicher-Anbindung von den monströsen 512 Bit bei der Geforce GTX 285 auf 384 Bit. Bei gleichem Speichertakt wie die GTX 285 hätte die neue Geforce trotzdem mehr Bandbreite zur Verfügung, was sich vor allem in hohen Auflösungen und mit aktivierter Kantenglättung positiv auf die Leistung auswirkt.

Das 384-Bit-Interface setzt sich aus sechs 64-Bit-Controllern zusammen und bestimmt so die Stufen des Speicherausbaus. Auf der Geforce GTX 380 erwarten wir 1.536 MByte, für die Geforce GTX 360 gehen wir von 1.280 MByte aus. Die Profi-Varianten Tesla unterstützen sogar bis zu 6,0 GByte und die ECC-Korrektur, mit der Speicherfehler im laufenden Betrieb erkannt und behoben werden können (auch in den chipinternen Cache-Speichern).



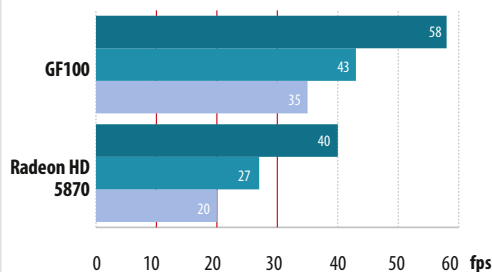
Nvidia hat neben Leistung und Technik auch die **Bildqualität** überarbeitet. Unter anderem soll die 32fache Kantenglättung (links) nur noch wenig langsamer arbeiten als der gröbere 8x-Modus.

Nvidias-Benchmarks

Unigine DirectX 11

Der Benchmark Unigine nutzt die erweiterten Geometrie-Fähigkeiten (Tessellation) von DirectX 11 besonders intensiv.

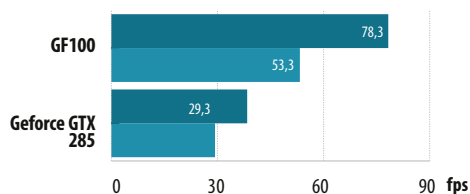
Maximum Durchschnitt Minimum



Dark Void

Auch in Dark Void verspricht Nvidia mehr als die doppelte Leistung einer GTX 285. Wir halten das weitgehend für realistisch.

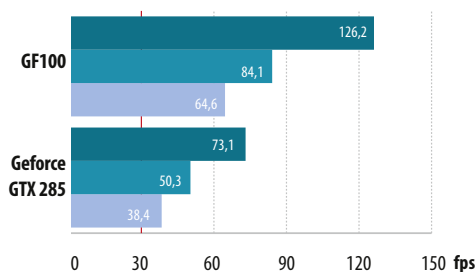
Durchschnitt Minimum



Far Cry 2 DirectX 10, Ranch Small

Den Benchmark konnten wir mit eigenen Augen verfolgen. Die gewaltige Leistungssteigerung muss der Test aber noch bestätigen.

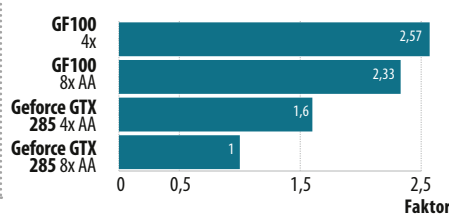
Maximum Durchschnitt Minimum



H.A.W.X. Kantenglättung

Im Vergleich zur GeForce GTX 285 soll die neue GeForce bei hohen Bildqualitätseinstellungen wesentlich schneller arbeiten.

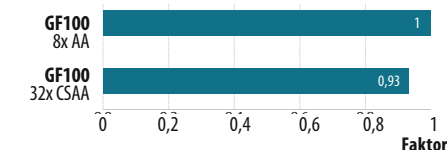
Vielfaches



CSAA-Kantenglättung Durchschnitt

Laut Nvidia kostet 32x Kantenglättung kaum mehr Leistung als 8x.

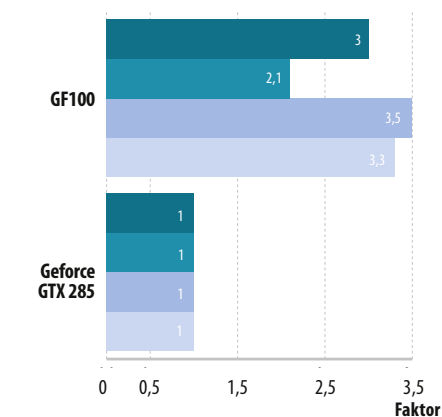
Vielfaches



GPU Compute

In diesen theoretischen Tests zur Performance abseits von Spiele-Grafik liegt die GF100 um ein Vielfaches vor der GeForce GTX 285.

PhysX Fluid Dark Void Raytracing KI



Potenzielle Leistungsreferenz

Von den Herstellern selbst angefertigte Benchmarks genießen wir naturgemäß mit Vorsicht. Allerdings haben sowohl AMD als auch Nvidia aus den Benchmark-Skandalen der letzten Jahre gelernt und legen meist einigermaßen realistische Performance-Vorhersagen vor. In Kombination mit den technischen Informationen trauen wir uns eine eigene Prognose zu: Nvidia wird sein Versprechen einlösen und mit dem kommenden Spitzenmodell GeForce GTX 380 die Radeon HD 5870 in Spielen deutlich hinter sich lassen – aber auch einen entsprechenden Aufpreis verlangen. Allerdings plant AMD bereits eine höher getaktete Radeon HD 5890 als Antwort auf Fermi. Im Vergleich zur GeForce GTX 285 erwarten wir abhängig von Auflösung und Bildqualität einen Vorsprung der GTX 380 von 50 bis 100 Prozent – also einen fetten Leistungssprung. Faustformel: Je höher die Qualitätseinstellung, desto größer fällt der Vorsprung aus. Ob

die neue GeForce auch AMDs Zwei-Chip-Grafikkarte Radeon HD 5970 schlagen kann, lässt sich noch nicht beurteilen. Die GeForce GTX 360 könnte ungefähr das Leistungsniveau einer Radeon HD 5870 erreichen.

Im Dunkeln liegen für uns neben den Taktfrequenzen der endgültigen Shader-Zahl und der Speicherkonfiguration auch Stromverbrauch und Lautstärke. Für die Tesla-Karte mit 448 Shadern gibt Nvidia eine Leistungsaufnahme von 225 Watt an. Mit 512 Shadern und den von uns erwarteten höheren Taktraten dürften es also wenigstens 250 Watt sein – fast so viel wie bei einer GeForce GTX 295 und rund ein Drittel mehr als bei der Radeon HD 5870. Wegen des außerordentlich hohen Stromverbrauchs können wir uns nicht vorstellen, dass die GeForce GTX 380 so leise arbeitet wie AMDs aktuelle Ein-Chip-Grafikkarten. Allerdings fiel die Karte in Nvidias Präsentationsrechner nicht durch eine besonders hohe Geräuschkulisse auf. Aufschluss kann aber nur ei-

ne Messung in unserer Schallschutzkammer liefern.

Gespannte Erwartung

Selten hat ein Grafikprozessor im Vorfeld seiner Markteinführung so hohe Welle geschlagen wie der GF100. In praktisch allen Punkten wird die neue GeForce mindestens auf Augenhöhe mit AMDs Radeon HD 5000 liegen. Abseits der bereits erwähnten Neuerungen unterstützt nun auch Nvidia das Spielen auf mehr als zwei Monitoren. Das Pendant zu AMDs Eyefinity tauft Nvidia GeForce 3D Vision Surround – ja, das Ganze

soll auch in 3D funktionieren. Ob die Leistung dafür reicht, wird sich zeigen. Jetzt sind wir aber gespannt auf den Test, auf den Vergleich mit der Radeon HD 5000. Sie werden den Bericht allerdings erst im übernächsten Heft lesen können, vorher gibt es keine Testsamples. Selbstverständlich halten wir Sie auf GameStar.de und in der nächsten GameStar-Ausgabe weiter auf dem Laufenden. Wenn Sie bestimmte Fragen zur neuen GeForce haben oder spezielle Benchmarks sehen wollen, dann schreiben Sie an daniel@gamestar.de. **DV**

GeForce GF100

► **Angeschaut** ► Hardware **Grafikkarte** ► Termin **1. Quartal 2010**
► Hersteller **Nvidia** ► Status **zu 90% fertig**

Daniel Visarius: Mit der GeForce GTX 380 wird Nvidia wieder an AMD vorbeiziehen. DirectX 11 und komfortablen Mehrschirmbetrieb beherrschen dann beide, aber Nvidia hat noch seine proprietären Extras GeForce 3D Vision und GPU-PhysX in der Hinterhand. Bei Lautstärke und Stromverbrauch wird sich zeigen, wer den Vorschlaghammer in der Hand hält und wer das Florett.



daniel@gamestar.de