



Arbeitsbeschleuniger

Der richtige Grafikprozessor für Bürocomputer und Workstations

Die einfache integrierte Prozessorgrafik ist besser als ihr Ruf: Sie bringt ein hoch aufgelöstes Bild auf einen oder mehrere Bildschirme und dekodiert moderne Videoformate. Bei Workstations gelten andere Anforderungen, die erst dedizierte Grafikkarten erfüllen.

Von Benjamin Kraft

Die Zeiten, in denen eine Grafikkarte am Arbeitsplatz nur das Bild auf den Monitor brachte, sind lange vorbei. Die Fülle völlig unterschiedlicher Funktionen erschwert die Auswahl des passenden Modells. Dieser Weg-

weiser zeigt, was bei Büro-PC, Workstation und Server wichtig ist.

Im Büro: Integrierte GPU oder Steckkarte?

In Bürocomputern dominieren Grafikprozessoren, die direkt in die CPU integriert sind, sogenannte IGP (Integrated Graphics Processor). Bei Intel stecken sie in allen Mainstream-Desktop-CPUs, nur die Vielkern-Modelle der HEDT-Serie (High-End Desktop), die eigentlich von der Xeon-Familie für Server abstammen, verkauft die Firma grafiklos. AMD gibt die Prozessorgrafik nur bestimmten, vor allem günstigen Produktreihen mit: den älteren A-Modellen sowie den aktuellen Ryzen-G- und Athlon-GE-Prozessoren. Dass in Office-PCs oft allein die Prozessorgrafik zum Einsatz kommt, hat gute

Gründe: Sie ist billig, leise und für Alltagsaufgaben völlig ausreichend.

Überhaupt leisten IGP erstaunlich viel: So übernehmen sie beispielsweise die Aufbereitung der grafischen Oberfläche des Betriebssystems, und moderne Browser fühlen sich nicht zuletzt deshalb so flott an, weil sie die Beschleunigungsfunktionen der GPU nutzen.

Seit der HD-Graphics-4000-Serie, die in den Prozessoren der Core-i-4000-Generation (Haswell) steckt, können Intels integrierte GPUs 4K (3840 × 2160 Pixel) mit augenschonenden 60 Hz ausgeben. Bei AMD gelingt das seit den APUs der Ax-7000-Familie (Kaveri). Voraussetzung bei beiden: Das Mainboard besitzt einen DisplayPort der Generation 1.2 oder einen HDMI-2.0-Ausgang. Letzteren findet man an Hauptplatinen für AMDs

Ryzen-GPUs, im Intel-Lager sind sie hingegen selten, weil auch aktuelle Intel-iGPUs noch nicht HDMI 2.0 beherrschen. Deshalb müssen die Hersteller einen zusätzlichen Chip auf die Platine löten. Mehr als 4K Auflösung schaffen integrierte Grafikeinheiten derzeit nicht.

GPUs entlasten außerdem die CPU-Kerne bei der Videowiedergabe, indem sie den Datenstrom dekodieren. Ältere Kompressionsverfahren wie DivX, H.264 und Microsofts VC-1 gehören schon lange zur Grundausstattung. Anspruchsvollere Formate wie H.265 oder VP9 erfordern neuere Hardware. Bei AMD können APUs seit Carrizo oder Bristol-Ridge für Sockel FM2+ und AM4 diese Formate entschlüsseln; Intel baut die Fähigkeiten seit der Skylake-Generation (Core-i-6000) ein. Selbst die billigen Gemini-Lake-Modelle der Atom-Familie (Celeron N4100), die man in extrem günstigen Komplett-PCs findet, bringen 4K- und Video-Fertigkeiten mit [1]. Aktuelle Preis/Leistungs-Lieblinge für unter 80 Euro sind auf AMD-Seite der günstige Athlon 200GE [2], bei Intel der Pentium Gold G5400 – beide Einstiegsprozessoren besitzen zwei echte und zwei logische CPU-Kerne und bringen auf GPU-Seite alles mit, was man im Office braucht.

Wer die Grafikfähigkeiten seines ansonsten ausreichenden Arbeitsplatz-PCs aufmotzen will oder nur mehr Anschlüsse braucht, steckt eine Grafikkarte nach. Ein guter und sparsamer Allrounder fürs Büro bleibt die Nvidia GeForce GT 1030 [3], die es als passiv gekühlte und damit lautlose Version ab 75 Euro gibt. Zudem beherrscht sie alle aktuellen Video-Codecs. Den neuen lizenzfreien Codec AV1, der als Nachfolger von VP9 und Konkurrent von H.265/HEVC positioniert wird, kann indes noch keine verfügbare Grafikkarte hardwarebeschleunigt wiedergeben. Das wird voraussichtlich erst GPUs gelingen, die ab 2020 erscheinen.

Die meisten Hersteller statten ihre GeForce GT 1030 mit DVI und HDMI 2.0 aus. Weil DisplayPort 1.4 aus unserer Sicht zukunftssicherer ist, raten wir zum Kauf einer Karte mit diesem Signalausgang. An ihm gibt sie mit nur einem Kabel ein 5K-Signal mit 60 Hz aus. Doch Vorsicht: Einige ältere 5K-Displays nehmen Bildsignale nur über zwei parallel genutzte DisplayPorts entgegen, die keine Budget-Karte hat. Ob DDR4- oder der schnellere GDDR5-Speicher zum Einsatz kommt, ist für den Office-Einsatz hingegen unerheblich.



Schon der billige AMD Athlon 200 GE bringt alles mit, was der Büro-PC braucht.

Wer drei Signalausgänge braucht, kann zu AMDs Radeon RX 550 greifen. Sie liefert bei ansonsten gleichen Fähigkeiten mehr 3D-Performance, nimmt aber auch mehr Leistung auf. Eine lautlose, passiv gekühlte Version gibt es nicht.

Workstation: Rendern, Rechnen, Retuschieren

Für den professionellen Einsatz bieten AMD und Nvidia spezielle Karten an, die hier Radeon Pro WX, da Quadro heißen. Die derzeit verfügbaren Modelle kosten bei AMD zwischen 110 und etwa 1800 Euro, Nvidia verlangt 120 bis über 10.000 Euro. Ihre Domäne: professionelle Fotobearbeitung in 10 Bit, Industriedesign per CAD und komplexe Berechnungen.

Das Besondere an den Workstation-Karten sind ihre Treiber: Sie sind für bestimmte Anwendungen zertifiziert, etwa AutoCAD, Solidworks, Siemens NX und PTC Creo. Durch sie leisten die Profi-

Modelle in diesen Anwendungen mehr als ihre Consumer-Varianten. Das zeigt sich beispielsweise bei der Darstellung von Drahtgittermodellen. Anwendungen wie der Pathtracer Luxmark oder professionelle Video-Encoder, die rein von der Shader-Performance abhängen, benötigen übrigens keine teure Karte – hier tut es auch ein günstigeres Modell aus dem Heimanwenderportfolio. Es kommt folglich auf die Software an, welche Karte man kaufen sollte.

Die Profi-Karten setzen grundsätzlich auf die gleiche Vega-, Polaris- oder Pascal-Architektur wie ihre Gamer-Gegenstücke. Sie bringen also auch deren Video-Encoder- und Decoder mit, die alle aktuellen Formate beherrschen und Inhalte in 10 Bit Farbtiefe und mit hohem Kontrastverhältnis (HDR) ausgeben können. Außerdem haben Sie moderne DisplayPort-1.4-Anschlüsse. Allerdings nutzen die Quados oft eine andere Anzahl an Shader-Einheiten. Anders bei AMD, wo jede Radeon Pro WX auch ein Radeon-RX-Gegenstück besitzt.

Beim Preis-Leistungsverhältnis hat AMD auf den ersten Blick entlang der gesamten Produktlinie die Nase vorn; besonders extrem ist das bei den Spitzen-Quados, die schon mal das Zweifache einer von den technischen Daten äquivalenten Radeon Pro kosten. Dabei liefern die AMD-Karten eine höhere theoretische Rechenleistung – erst recht bei doppelter Genauigkeit (FP64), denn Nvidia limitiert die FP64-Performance, um den Abstand zu den Spitzenmodellen der Quadro-, Titan- und Tesla-Familien zu wahren.

In OpenCL-Anwendungen spielen die AMD-Karten diesen Vorteil oft aus. Au-

Für Photoshop und Video

In den allermeisten Fällen reicht die Prozessorgrafik für die Fotobearbeitung und den Videoschnitt, weil die CPU die meiste Arbeit stemmt. Zwar werben die Hersteller gern damit, ihre GPUs könnten bestimmte Funktionen um ein Vielfaches gegenüber der CPU beschleunigen, aber in Wahrheit handelt es sich um eher selten genutzte Filter.

Ähnlich sieht es beim Videoschnitt aus, denn die eingebauten Video-Encoder-Einheiten VCE (AMD), QuickSync (Intel) und NVENC (Nvidia) werden nur

von sehr wenigen Anwendungen unterstützt.

Wer Bilder in 10 Bit Farbtiefe bearbeiten will, muss hingegen eine teurere Profi-Karte kaufen: Zwar geben Radeons und GeForces via DisplayPort tatsächlich 10 Bit pro Farbkanal aus, doch muss das auch die Anwendung unterstützen. Photoshop tut das ausschließlich über den OpenGL-Pfad – und den passenden Treiber liefern Nvidia und AMD nur für die Quados respektive die Radeon Pros mit.

ßerdem können sie anders als Nvidias Pascal-Karten parallel Grafikbeschleunigung und Berechnungen ausführen. In der CAD-Praxis bringt aber Nvidia mehr Leistung auf die Straße, wie der Workstation-Benchmark SPEC Viewperf [4] zeigt. Diesen Performance-Vorsprung will man mit den neuen Quadro-RTX-Modellen der Turing-Generation vergrößern, unter anderem durch den Einsatz von Tensor-Cores und Einheiten zur Raytracing-Beschleunigung. Allerdings sollen die neuen Karten bis zu 10.000 Euro kosten, etwa die Quadro RTX 8000 mit 48 GByte GDDR6-RAM. Dafür leistet sie über ihre Shader-Einheiten 16 TFLOPS und soll beim Raytracing 10 Gigarays pro Sekunde erzielen.

Auch an der Softwarefront kann Nvidia mit einem Trumpf aufwarten: der proprietären Programmierschnittstelle (API) CUDA für komplexe Berechnungen auf der GPU. Wer eigene Anwendungen entwickelt, findet bei Nvidia das bessere Rüstzeug und viel mehr Dokumentation als bei AMD. Zudem hat die Firma viel in Entwicklerbeziehungen investiert. Zwar setzt AMD mit OpenCL auf ein offenes API, unterstützt die Entwickler aber bei Weitem nicht so gut wie Nvidia bei CUDA. Beispielsweise liefert Nvidia viele hochoptimierte Bibliotheken, sodass man gleich loslegen kann, während es für OpenCL in einigen Fällen gar kein Gegenstück gibt – man müsste also bei null anfangen.

Weil schon die kleinsten Modelle der jeweiligen Profi-Produktfamilien Bilder mit 10 Bit pro Farbkanal ausgeben und damit besonders feine Farbverläufe zeichnen können, eignen sie sich für farbechte Fotobearbeitung an kalibrierbaren 10-Bit-Monitoren oder medizinische Bildgebungsverfahren.

Radeon Pro WX 4100 (270 Euro), WX 5100 (400 Euro) und Quadro P2000 (430 Euro) spielen etwa in derselben Leistungsklasse, mit leichtem Vorteil für AMD. In CAD-Benchmarks liegt die Nvidia-Karte allerdings mit der teureren AMD Radeon Pro WX 7100 (620 Euro) gleichauf, und beide halten einen gewissen Abstand zur Quadro P4000.

Bei den teureren Modellen steigt nicht nur der Preis, sondern auch der Speicherausbau. Sind es bei der Radeon Pro WX 8200 (1000 Euro) noch 8 GByte, haben die Quadro P5000 (1600 Euro) und Radeon Pro WX 9100 (1800 Euro) 16 GByte RAM an Bord, die P6000 (4700 Euro) 24 GByte. Damit sind gerade die



Nvidias Quados P6000 bietet zu einem hohen Preis viel Rechenleistung, die sie vor allem in CUDA-Anwendungen voll ausschöpft.

Spitzenmodelle in Kombination mit ihrer hohen Rechenleistung für große Datenmengen prädestiniert. Weitere Besonderheit: Alle vier können Speicherfehler per ECC entdecken und korrigieren.

Die knapp 9000 Euro teure Quadro GV100 hat sogar 32 GByte HBM2-Speicher an Bord, nutzt bereits die jüngere Volta-Architektur und enthält Tensor-Cores. Diese leisten bis zu 118,5 TFLOPS und beschleunigen KI-Anwendungen und maschinelles Lernen stark. Zudem halbiert sich ihre Performance bei Berechnungen in doppelter Genauigkeit (FP64) gegenüber einfacher Genauigkeit (FP32) nur von 14,8 auf 7,4 GFLOPS; die GeForce- und die kleineren Quadro-Karten schaffen bei FP64 nur ein Zweiunddrittel der FP32-Rechenleistung.

Im Rechenzentrum: Rechnen und Lernen

Mehr noch als bei den Workstations kommt es bei Grafik- und Beschleunigerkarten für den Server-Einsatz auf den konkreten Einsatzzweck an. Die Anforderungen sind äußerst unterschiedlich. Eine allgemeine Empfehlung lässt sich damit kaum formulieren. So benötigen komplexe Berechnungen wie Wettersysteme eine hohe FP64-Leistung, die etwa Nvidias 9000 Euro teure Tesla V100 liefert. Auf dem Papier ist sie mit 14 GFLOPS via Shader und einer Tensor-Leistung von 112 TFLOPS etwas langsamer als die ähnlich teure Quadro GV100 mit gleicher GPU.

Dass sie mit einer Grafikkarte verwandt ist, erkennt man von außen nicht – als reine Beschleunigerkarte verfügt die Tesla V100 über keinerlei Signalausgänge.

Machine Learning (ML) hingegen setzt nicht unbedingt eine teure Karte voraus. Für viele Anwendungen genügt eine geringere Genauigkeit wie FP16 – und da sieht es für AMD gut aus, denn die Rechenleistung verdoppelt sich gegenüber FP32. Bleibt das Problem, dass die meisten ML-Anwendungen CUDA nutzen, sodass AMDs Performance-Vorteil theoretisch bleibt.

Bei Nvidias Quados der Pascal-Generation ergibt sich hingegen kein solcher Geschwindigkeitsvorteil, wohl aber bei den Volta-Karten Quadro GV100 und Titan V [5] sowie den Modellen der neuen Serien Quadro und GeForce RTX 2000. Letztere sind immer noch günstiger als Titans oder Teslas. Allerdings verbieten Nvidias Nutzungsbedingungen den Einsatz von GeForce- und Titan-Grafikkarten im Rechenzentrum – das gilt übrigens auch für Anbieter von Spiele-Streaming-Diensten.

Fazit

Bevor man eine neue Grafikkarte sucht, sollten die Anforderungen feststehen. Die Software diktiert, welche Hardware überhaupt infrage kommt.

Wer nur Korrespondenz erledigt, im Web surft und Videos schaut, kommt gut mit einer halbwegs aktuellen IGP hin. Eine Workstation-Karte, die viel rechnen und rendern soll, braucht viele Shader-Einheiten – aber auch Anwendungen, die mit dem API laufen, das der GPU-Hersteller favorisiert. In der Hinsicht hat Nvidia mit CUDA einen massiven Vorteil. Im Server-Bereich gilt das umso mehr. AMD bietet das bessere Preis-Leistungsverhältnis, aber nur, wenn man mit OpenCL arbeitet oder die Shader-Einheiten sinnvoll einspannen kann. (bkr@ct.de) **ct**

Literatur

- [1] Streaming-Schreibmaschinen, Zwei Mini-PCs mit Celeron N4100 ab 200 Euro, c't 19/2018, S. 40
- [2] Carsten Spille, Wichtiger Lückenfüller, AMDs günstiger Zweikern-Prozessor Athlon 200GE, c't 22/2018, S. 48
- [3] Rudolf Opitz, Multimedia-Nachhilfe, Günstige Grafikkarten für 4K-Displays und Video, c't 15/17, S. 146
- [4] Martin Fischer, Profis für alle Fälle, Workstation-Grafikkarten von AMD und Nvidia, c't 6/2017, S. 142
- [5] Martin Fischer, Pina Merkert, Volles Rohr, Alleskönner-Grafikkarte Nvidia Titan V, c't 9/2018, S. 92