

Andreas Stiller, Martin Fischer

Prozessorgeflüster

Vom wahren Norden und Polarstern

Leistungsfähige GPUs haben sie dank weitaus besserer Trainingsmöglichkeiten wieder hoffähig gemacht, die künstliche Intelligenz, nun in Gestalt von „Deep Learning“. Derzeit gehört das Spielfeld fast allein Nvidia, doch neue Start-ups und auch Alt-GPU-Konkurrent AMD wollen mitspielen.

Seit man mit den GPUs neuronale Netze effizient trainieren kann, kommen neuronale und neuromorphe Techniken immer mehr in Mode. Machine oder Deep Learning sowie Cognitive Computing sind die großen Buzz Words, und alle Großen sind dabei: IBM, AT&T, Microsoft, Google, Amazon, Facebook, eBay ... Zahlreiche Cloud-Angebote und APIs sind im Angebot: IBM Watson, AT&T Speech, Google Prediction, Microsoft Azure und Amazon Machine Learning ...

IBM hat dabei nicht nur Watson im Spiel, sondern kaufte im März 2015 AlchemyAPI hinzu, einen Deep-Learning-Spezialisten. Daneben entwickelt IBM im Rahmen des SyNAPSE-Programms der DARPA auch sogenannte neuromorphe Chips. Das sind Chips, die sich an der Arbeitsweise von Hirn-Neuronen orientieren. Der von IBM im Sommer 2014 vorgestellte TrueNorth-Chip erkennt mit seiner Million Neuronen inzwischen Bilder sehr gut, sogar weit besser als gedacht, so IBM-Forscher Jun Sawada auf der SC15.

Diverse Wettbewerbe haben sich auf dem Gebiet von Machine Learning inzwischen etabliert, etwa die „Large Scale Visual Recognition Challenge“ (ILSVRC), die von führenden amerikanischen Universitäten (Stanford, Princeton, Columbia ...) seit fünf Jahren ausgeschrieben wird. 2015 haben sich daran immerhin rund 50 Teams weltweit beteiligt, zu meist aus chinesischen Unis und Forschungseinrichtungen. Baidu wurde zwischenzeitlich allerdings wegen Schummelei disqualifiziert. Mitte Dezember 2015 gab es dann die Ergebnisse anlässlich eines Treffens zur 29. Jahreskonferenz der Neural Information Processing Systems (NIPS) in Montreal. In der Hauptklas-

se (Object detection with provided training data) gewann ein Team von Microsoft Research aus China in 194 von 200 Kategorien.

Auf diesem Szene-Treffen rund um „Brains, Mind and Machines“ gab Stanford-Professor und Nvidia-Cheftechnologie Bill Dally einen Überblick über High Performance Hardware für Machine Learning – klar, welche Hardware bei ihm im Vordergrund stand. Aber auf diesem Gebiet erwächst allmählich neue Konkurrenz, etwa das Start-up Nervana Systems.

Diese Firma wurde vor einem Jahr von drei Qualcomm-Abtrünnigen rund um CEO Naveen Rao in San Diego gegründet. Sie hat nicht nur ein Open Source Deep Learning Framework namens Neon entwickelt, das nach eigenen Angaben auf Nvidia-GPUs dreimal so schnell sein soll wie Nvidias eigene Algorithmen, sondern auch einen eigenen Chip in Arbeit, der diesen Job noch weit effizienter ausführen soll.

Viel weiß man über den Chip noch nicht, nur das, was Rick Merritt von EETimes über ihn herausgefunden hat: Er ist für TSMCs 28-nm-Prozess ausgelegt und soll so groß sein wie handelsübliche GPUs. Er besteht aus einer Vielzahl von „Tensoren“, das sind kleine Rechen-Arrays mit direkt per 3D-Stacking angekoppeltem Speicher. Verknüpft über zahlreiche 25G-SerDes soll er einen Durchsatz von 2 Terabit/s erreichen.

Auf diesem Gebiet sind aber noch etliche weitere Start-ups aktiv, so wie SSLabs Inc. rund um den rührigen deutschen Physiker Axel Kloth. Seine in Pacifica ansässige Firma bastelt an mehreren Coprozessoren für Big Data und High Performance Computing mit

neuartiger Prozessor- und Speicherarchitektur. Darunter ist auch ein Neural Net Coprocessor, optimiert für Bild- und Videoauswertung und Deep Learning.

Im Berater-Board dieses Start-ups findet man viele altbekannte Namen, insbesondere auch John L. Gustafson – ein „bunter Hund“ der Szene. Er hat in letzter Zeit durch sein Buch zur Universal Number „End of Error“ auf sich aufmerksam gemacht. Zuvor war er in seiner bewegten Vita auch mal Chef von AMDs Grafikkabteilung. Diesen Job hat jetzt Jim Anderson inne, der wie Gustafson auch davor Intel zu Diensten war.

Kanadisches Nordlicht

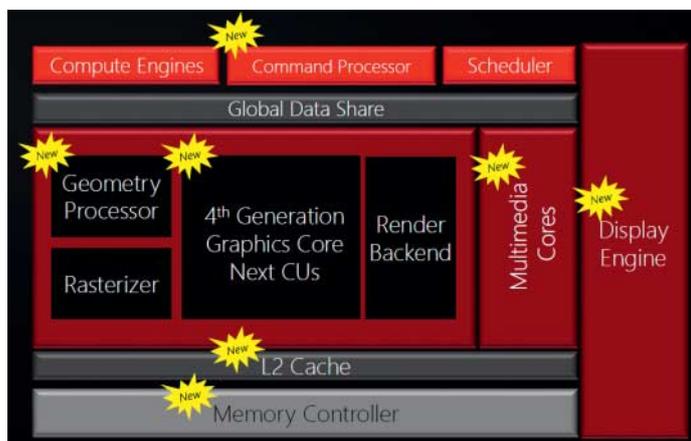
Ihm zur Seite steht der renommierte Chefarchitekt Raja Koduri. Aufmerksame AMD-Fans wunderten sich gegen Jahresende, was Koduri da Merkwürdiges mit dem Marketing-Drahtzieher Chris Hook über den Polarstern twitterte. Jetzt weiß man es: „Polaris“, so heißt die neue GPU-Architektur, die Nvidias lange dominierende Maxwell-2.0-GPUs – vielleicht sogar den kommenden Pascal – Mitte des Jahres in die Schranken weisen soll.

Damit das klappt, hat AMD die bestehende GCN-Architektur an entscheidenden Stellen umgekrempelt. Polaris-GPUs haben offenbar einen besseren Hardware-Scheduler und können Shader-Programme und Texturkompression effizienter ausführen. Dazu kommt noch die neue Display-Engine mit HDMI 2.0a und DisplayPort 1.3 (für HDR-Bilder und 5K-Displays) und eine aufgebohrte Video-Engine, die 4K-Filme im H.265-Format mit 60 fps kodiert oder dekodiert. Das ist vor allem für kommende Video-On-Demand-Streams (Netflix!) und 4K-Blu-rays wichtig.

Ach so, den großen Trommelwirbel haben wir glatt vergessen: Polaris-GPUs sind die ersten, die AMD im 14-nm-FinFET-Prozess herstellen lässt und damit offenbar vom Fertiger TSMC auf Globalfoundries umschwenkt. Der Schritt vom aktuellen 28-nm-Prozess ist gewaltig, und entsprechend sollen die Effizienzvorteile sein. In einem ersten Beispiel vergleicht AMD eine Polaris-GPU mit einer gleich schnellen Maxwell-v2-GPU, die für dieselbe Bildrate 140 statt 86 Watt schluckt. Plötzlich wirkt die bis dato hocheffiziente Maxwell-Architektur wie ein Heizkraftwerk.

Mit Polaris lassen sich also nicht nur fette Gamer-Grafikkarten bauen, sondern auch hocheffiziente Mobil-Grafikchips oder Steam-Machine-Grafikbeschleuniger. Das könnte für AMD der Schlüssel sein, dem großen Konkurrenten Nvidia wertvolle Marktanteile abzutrotzen. Denn Nvidia wird die Anfang April erwartete Pascal-Generation wohl vorrangig für den Supercomputer-Markt optimieren – und bei TSMC mit 16-nm-FinFET-Strukturen fertigen. Sicher ist: Sowohl AMD als auch Nvidia werden Chips auf den Markt werfen, die endlich genug Power zum 4K- und VR-Spielen haben werden. 2016 wird also das spannendste Jahr im Grafikkartenmarkt seit 2011 – und für PC-Gamer sowieso.

(as@ct.de/mfi@ct.de)



Alles „New“:
Die Polaris-Architektur ist AMDs wichtigstes Grafikkchip-Design seit 2011.