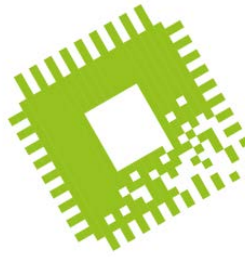


Bit-Rauschen



Nvidia-Beschleuniger A100, TSMC-Fab in den USA, ARM-Supercomputer

Der Nvidia-Chef holt neue GPU-Chips aus dem Backofen, Donald Trump holt sich taiwanische Spitzentechnik ins Heimatland und ein ARM-Chip könnte sich die Top500-Führung holen.

Von Christof Windeck

Wer zu spät kommt, den bestraft das Leben: Das bekommt nun die Firma Intel zu spüren, deren Xeon-Prozessoren noch immer kein PCI Express 4.0 (PCIe 4.0) beherrschen. Deshalb baut Nvidia in sein 200.000 US-Dollar teures Rechenmonster DGX A100 (siehe S. 132) nun AMD-Epyc-Prozessoren ein. Mit den paar Tausend Chips für die teuren Nvidia-Kraftpakete wird AMD zwar nicht viel Geld einnehmen, gewinnt aber eine Menge Prestige: Ausgerechnet der Grafikchip-Erzkonkurrent Nvidia kauft bei AMD ein und sortiert Intel aus.

Als virales Appetithäppchen wenige Tage vor dem A100-Start ließ der selbstbewusste Nvidia-Chef Jen-Hsun „Jensen“ Huang ein YouTube-Video veröffentlichen, das ihn selbst in seiner Küche zeigt. Unter dem Titel „Was hat Jensen denn da gebacken?“ ist zu sehen, wie er ein A100-System aus seinem teuren Dacor-Edelbackofen holt. Der lustig gemeinte Einfall wird manchem alten Nvidia-Kunden sauer aufstoßen: Vor zwölf Jahren fielen reihenweise Notebooks und Grafikkarten mit den GPUs G84 und G86 (GeForce GT 8000/M) aus. Nvidia zeigte sich nicht sehr kulant. Seinerzeit lautete ein Tipp, solche Grafikkarten im Backofen zu reparieren, was tatsächlich manchmal half.

Der A100 – nun nicht mehr „Tesla“ genannt wegen der deutlich bekannteren Automarke – verspricht jedoch ein großer Wurf zu werden. Obwohl über 100 Start-up-Firmen KI-Chips entwickeln, steht Nvidia bei den Beschleunigern für Rechenzentren unangefochten weiter auf

dem Siegertreppchen. Und der Hunger nach KI-Rechenleistung ebbt keineswegs ab, ganz im Gegenteil. Mit dem CUDA-(X-)API, für das es schon zahllose Anwendungen gibt, hat Nvidia über Jahre hinweg eine etablierte Infrastruktur für Programmierer aufgebaut. Diese Hürde muss die Konkurrenz erst einmal nehmen.

KI-Datenformat BFloat16

Die A100-„Ampere“-Beschleuniger sind auch die ersten GPU-Chips, die das effiziente KI-Datenformat BFloat16 verarbeiten. Das ist allerdings nicht auf dem Mist von Nvidia gewachsen, weshalb man dort gleich noch ein paar weitere Tensor-Datenformate wie TF32 hinzuerfand. Vor allem Google hat BFloat16 bekannt gemacht und verwendet es in den hauseigenen TPUs, IBM will es im Power10 nutzen und Intel ab der Xeon-Generation Cooper Lake. Letztere soll bloß noch in wenigen Varianten auf den Markt kommen, wichtigster Käufer könnte Facebook werden. Auf dem Open Compute Summit kamen Hinweise zu Cooper Lake an die Öffentlichkeit, ein konkreter Starttermin allerdings nicht.

Ganz ohne zusätzliche Rechenbeschleuniger, aber mit neuartigen Scalable Vector Extensions (SVE) legen die 7,6 Mil-

lionen ARM-Prozessorkerne im japanischen Supercomputer Fugaku los. Die fast 160.000 Fujitsu-A64FX-Chips mit HBM-Speicher schaffen theoretisch 537 PFlops, im Linpack also wohl gut ein halbes Exaflop. Damit dürfte Fugaku am 22. Juni die Spitze der Top500-Liste der schnellsten Supercomputer erklimmen und auch die Green500-Liste der effizientesten anführen – falls kein chinesischer Kontrahent überholt.

US-Chips aus Taiwan

Nvidia bezieht seine Chips bekanntlich nicht aus Jensen Huangs Luxusbackofen, sondern lässt sie in dessen Geburtsland Taiwan bei TSMC backen. Für Donald Trumps Geschmack beliefert TSMC aber auch die Falschen, nämlich beispielsweise Huawei in China. Um das zu verhindern, setzte die US-Regierung schon verschiedene Daumenschrauben an, gibt nun aber Zuckerbrot statt Peitsche: TSMC erhält Subventionen, um eine insgesamt rund 12 Milliarden US-Dollar teure Chip-Fab in Arizona aus dem Boden zu stampfen. 2021 soll es losgehen, 2024 sollen die ersten Wafer mit 5-Nanometer-Chips durch die Öfen laufen. Laut TSMC hilft auch der Bundesstaat Arizona mit Fördermitteln, denn es stehen rund 1600 Jobs in Aussicht. Die neue Fab wird aber nicht allzu groß, rund 20.000 Wafer-Starts pro Monat sind geplant. TSMC hat bereits eine US-Fab, das WaferTech-Werk in Camas, Washington – nicht weit vom Flughafen Portland, Oregon. Es ist aber eine ältere Fab für 200-Millimeter-Wafer und 160-Nanometer-Strukturen.

Trump, selbsternannter Meister aller Dealmaker, will zwei Fliegen mit einer Klappe schlagen: TSMC an die Kandare nehmen und fortschrittliche Fertigungstechnik ins Homeland holen. Intel als ehemaligem Primus bei der Chip-Fertigungstechnik traut man diesbezüglich selbst in der Heimat nicht mehr allzu viel zu, Globalfoundries aber auch nicht. Politiker der Demokraten haken nun nach, ob beim TSMC-Deal alles korrekt läuft.

Als Trostpflaster für Intel erweist sich der Core i9-10900K tatsächlich als Gaming-Champion, siehe S. 86. Allerdings sind die aktuellen Ryzens in fast allen anderen Disziplinen schneller und deutlich effizienter; und der zum Jahresende erwartete Ryzen 4000 mit Zen 3 könnte Intel auch bei Spielen überholen.

(ciw@ct.de)



Nvidia-Chef Jensen Huang holt ein HGX-A100-System aus seinem Backofen.

Bild: YouTube/Nvidia