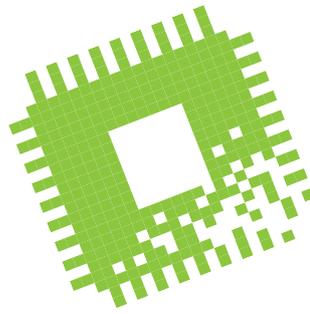


Bit-Rauschen

Server-CPU-Angriffe von Amazon, Lieferschwierigkeiten bei Intel



Nun wird es ernst bei den ARM-Servern: Amazon stellt den Graviton2 in Dienst. Intel will angeblich Prozessoren bei Samsung fertigen lassen und noch eine weitere KI-Firma schlucken.

Von Christof Windeck

Jetzt bricht der Kampf um die Rechenzentren wirklich los: Amazons Cloud-Sparte AWS nimmt die ersten Server mit dem selbst entwickelten ARM-Prozessor Graviton2 in Betrieb. Dieser Bolide hat mit dem letztjährigen Testballon Graviton kaum etwas gemein, sondern wurde konsequent als Intel-Xeon-Gegner gebaut. Der Graviton2 hat mehr Kerne als jeder Xeon, nämlich 64 wie der AMD Epyc 7002 „Rome“. AWS plant, Cloud-Instanzen mit bis zu 512 GByte RAM auf Graviton2 zu hosten. Die sollen in manchen Anwendungen um bis zu 40 Prozent schneller sein als „M5“-Instanzen auf Intel-Xeons – und gleichzeitig um 20 Prozent billiger.

Laut dem Analyst Patrick Moorhead sind die 64 Kerne des Graviton2 über einen Mesh-Interconnect verbunden, haben je 1 MByte L2-Cache sowie 32 MByte gemeinsamen L3-Cache. Acht Speicherkanäle (DDR4-3200) sowie 64 PCIe-4.0-Lanes stehen bereit. Die CPU-Kerne verarbeiten auch Int8- und FP16-Datenformate für schnelle KI-Algorithmen. Der Speicher-Controller verschlüsselt das RAM standardmäßig per AES-256.

Den 7-Nanometer-Chip mit rund 30 Milliarden Transistoren hat die 2015 in Israel zugekaufte Amazon-Sparte Annapurna Labs mit entwickelt, allerdings auf Basis des ARM-Referenzdesigns Neoverse N1. Es handelt sich also gerade nicht um einen aufwendig optimierten Spezialkern, sondern sozusagen um ARM-Kerne von der Stange. Der Graviton2 dürfte nicht nur Intels Xeon-Profiten drücken, sondern er gräbt auch den wenigen verbliebenen Firmen, die ARM-Serverprozessoren entwickeln, das Wasser ab. Denn welche CPU-

Stückzahlen bleiben am Ende noch für einen Cavium ThunderX3 oder Ampere eMag übrig, wenn Cloud-Giganten wie Amazon, Google, Facebook oder Microsoft auf Basis von ARM-Blaupausen gleich selbst Prozessoren nach Wunsch bauen? Quasi nebenbei hat Amazon übrigens gleich noch die ersten Elastic-Cloud-(EC2-)Instanzen In1 angekündigt, die der ebenfalls hausgebackene „Inferentia“-Prozessor für KI-Inferencing befeuert.

Kunde droht mit Bestellung

Für Intel könnte der Graviton2 keine Zeiten einläuten, weil Intel mehr als die Hälfte aller Xeons an große Cloud-Rechenzentren wie eben Amazon AWS verkauft. Bisher hat Intel jedoch ganz andere Schwierigkeiten, nämlich zu wenige Prozessoren. Nachdem Intel-Vorzeigekunde Dell in sehr deutlichen Worten seinen Zulieferer kritisierte, weil Lieferschwierigkeiten das mögliche Wachstum bremsen, schrieb Intel abermals einen Entschuldigungsbrief. Intel-Chef Bob Swan erklärte später wortreich vor Analysten der Großbank Credit Suisse, dass man die auch 2019 hohe Nachfrage unterschätzt habe. Schon 2018 und auch 2019 habe man jeweils mehr als 1 Milliarde US-Dollar zusätzlich für mehr Lieferkapazitäten investiert, sie sollen nun fast 50 Prozent höher sein als Anfang 2018 geplant. Es reicht demnach aber trotzdem nicht, auch weil man viel mehr unterschiedliche Chips fertige als früher – außer CPUs, Chipsätzen, Netzwerkchips und Flash-Speicher etwa auch FPGAs, LTE-Modems und KI-Beschleuniger. Zum wiederholten Mal hieß es, die 10-Nanometer-Fertigung laufe nun aber besser als erhofft.

Rund 20 Prozent aller Intel-Chips stammen eigentlich von Auftragsfertigern, das ist laut Bob Swan schon seit Jahren so. Das galt bisher aber nicht für Xeon- und Core-CPU. Nun gibt es Gerüchte, laut denen sich das 2020 ändern könnte: Demnach will Intel bei der Core-i-Generation „Rocket Lake“ eine von Samsung zugekaufte 14-Nanometer-CPU mit einer

10-Nanometer-GPU kombinieren. Intel wolle dazu die eigentlich für die 10-nm-Generation Tiger Lake entwickelten CPU-Kerne vom Typ Willow Cove auf 14 nm „zurückportieren“ samt ihrer AVX-512-Erweiterungen. Vielleicht lässt man bei Samsung aber auch nur billige Tremont-Celerons und -Atoms fertigen, um eigene Kapazitäten freizuschaukeln.

Intel hat in den vergangenen Jahren bereits mehrere KI-Spezialfirmen zugekauft, darunter das Silicon-Valley-Startup Movidius sowie die israelischen Firmen Mobileye und Nervana. Nun will Intel sich angeblich noch Habana Labs einverleiben, ebenfalls aus Israel. Deren Gaudi-Chip soll Nvidias Tesla V100 beim KI-Training schlagen. Wenn aber Amazon und Google eigene KI-Chips entwickeln, fragt man sich wiederum, wer die von Intel überhaupt kaufen soll.

Panasonic wirft unterdessen bei der Chip-Fertigung das Handtuch. Die japanischen Halbleiterproduzenten kämpfen ja schon seit Jahren mit der Wirtschaftlichkeit und versuchen, sich mit Fusionen zu retten – Beispiel Renesas, die vereinigten Chip-Sparten von NEC, Hitachi und Mitsubishi. Panasonic hatte schon 2014 mehrere Chip-Fabs in eine gemeinsame Gesellschaft mit dem US-amerikanisch-israelischen Auftragsfertiger TowerJazz abgespalten. Nun verkauft Panasonic die restliche Fertigung an die taiwanische Firma Nuvoton, die ihrerseits aus der bekannteren Firma Winbond entstanden ist. (ciw@ct.de) **ct**

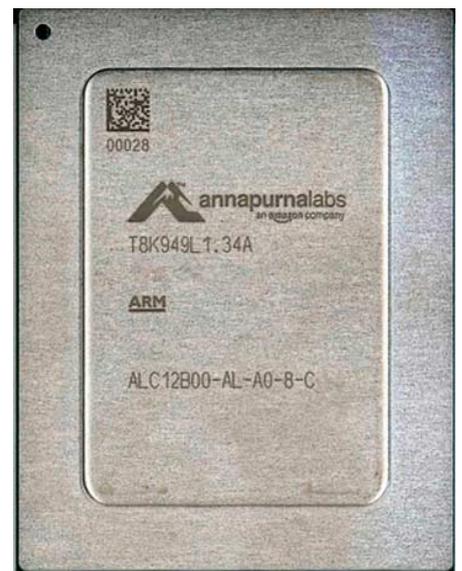


Bild: Amazon AWS

Amazons hausgemachter Graviton2-Prozessor mit 64 ARM-Kernen kratzt an Intels Xeon-Übermacht bei Cloud-Servern.