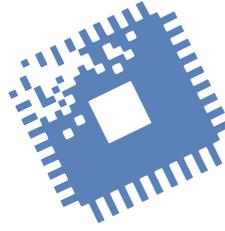


# Bit-Rauschen



## Ex-Hacker will AMD-Chips für KI zum Fliegen bringen

**AMD baut prima Grafikchips, meint George Hotz, aber katastrophale Software. Qualcomms schnellster Snapdragon für Windows-Notebooks zeigt sich. Intel lagert zu viel an TSMC aus.**

Von Christof Windeck

Der vor 15 Jahren als iPhone-Hacker „geohot“ bekannt gewordene George Hotz macht heutzutage in KI. 2016 gründete er die Firma Comma.ai für autonomes Fahren. Das hat bisher noch nicht so recht funktioniert, weil die Comma.ai-Software openpilot an gesetzlichen Vorgaben in den USA scheiterte. Für die weitere Entwicklung braucht Comma.ai möglichst viel billige Rechenleistung fürs Training von KI-Modellen.

Dazu setzt Hotz auf AMD-Hardware, lässt an den dafür von AMD bereitgestellten Tools wie ROCm aber kein gutes Haar. Stattdessen entwickelt er in seiner anderen Firma tiny corp das schlanke KI-Framework tinygrad und passt es an AMD-Rechenbeschleuniger an. Dafür sammelte tiny corp über 5 Millionen US-Dollar Risikokapital ein.

Ein erstes Ziel ist es, AMD-Hardware noch 2024 im KI-Benchmark „MLPerf“ des Industriegremiums ML Commons unterzubringen. Das hat AMD bisher aus eigener Kraft nicht geschafft. Die MLPerf-Veröffentlichungen laufen wie Wettbewerbe in zwei Runden pro Jahr. Dabei können die konkurrierenden Hersteller aus unterschiedlichen Benchmarks für das Training von KI-Modellen oder deren Anwendung (Inferencing) wählen oder auch mehrere einreichen. Die bisherigen Runden hatten oft den Charakter „Alle gegen einen“, nämlich gegen die starken

Nvidia-Rechenbeschleuniger A100 und H100.

George Hotz hat einige originelle Ideen. Er verteilt Arbeitspakete via GitHub an interessierte Entwickler und nennt dazu jeweils Honorare (Bounties) zwischen 200 und 10.000 US-Dollar. Wer bei tiny corp arbeiten möchte, kann sich nicht einfach bewerben, sondern muss eine der ausgelobten Aufgaben erfolgreich erledigen.

Tiny corp baut in seine kompakte KI-Trainingsmaschine „tinybox“ keine teuren AMD-Beschleuniger der Baureihe Instinct MI ein, sondern je sechs Gaming-Grafikkarten vom Typ Radeon RX 7900 XTX. Sie stellen insgesamt 144 GByte Grafikspeicher bereit und kooperieren in der tinybox mit einem AMD Epyc mit 32 CPU-Kernen. Hotz verspricht 768 TFlops FP16-Rechenleistung für 15.000 US-Dollar, also ähnliche Performance wie eine deutlich teurere Nvidia H100 mit 80 GByte HBM3-RAM. Die tinybox mit sechs 355-Watt-Beschleunigern dürfte allerdings deutlich



Bild: tiny corp/George Hotz

**George Hotz packt in jede „tinybox“ sechs Radeon-RX-Grafikkarten. Das Bild zeigt vier tinyboxes, jede soll 15.000 US-Dollar kosten und dicht an die Performance einer Nvidia H100 herankommen.**

mehr Strom schlucken als eine Workstation mit einer H100.

### Starker Notebook-ARM

Mit dem ARM-Zwölfkerner Snapdragon X Elite will Qualcomm endlich den Durchbruch bei Windows-11-Notebooks schaffen – und die Performance sieht gut aus, wenn auch gemischt. Denn mittlerweile sind Geekbench-6-Werte von unterschiedlichen Snapdragon-X-Elite-Notebooks aufgetaucht, die aber noch nicht verkauft werden. Genau wie bei den Konkurrenten AMD Ryzen 8040 und Intel Core Ultra 100 hängt die tatsächlich nutzbare Rechenleistung demnach von der Konfiguration des Notebooks ab, also der vom Hersteller eingestellten Thermal Design Power (TDP) und der Kühlung.

Qualcomm führte im vergangenen Dezember Referenzgeräte mit sehr hohen Rechenleistungen vor, wollte dabei aber die TDP-Einstellungen nicht verraten. Ein Lenovo-Notebook mit Snapdragon X Elite zeigte im Geekbench 6 nun deutlich niedrigere, aber immer noch ordentliche Werte. Etwas später tauchten Ergebnisse eines Samsung Galaxy Book 4 mit Snapdragon Elite X1E80100 und bis zu 4 GHz auf: Der könnte gegen einen Core Ultra 9 185H bestehen und würde ihn bei KI-Apps deutlich schlagen. Man munkelt aber, das Galaxy Book 4 solle über 1800 US-Dollar kosten.

### Intels TSMC-Chips

Intels Finanzchef David Zinsner sprach Anfang März auf einer Anlegerkonferenz der Bank Morgan Stanley. Nebenbei merkte er an, dass Intel zurzeit mehr vom Auftragsfertiger TSMC zukaufe als eigentlich geplant. Bisher war bekannt, dass TSMC etwa das GPU-Chiplet des Core Ultra 100 zuliefert und beim kommenden Lunar Lake auch das CPU-Chiplet. Nun gibt es Spekulationen, laut denen TSMC auch die CPU-Chiplets der Anfang 2025 erwarteten Mobilversionen von Arrow Lake produzieren soll. Dabei wirbt Intel damit, für Arrow Lake erstmals die haus-eigene Fertigungstechnik 20A einzusetzen – aber möglicherweise nur für die CPU-Chiplets der Desktop-PC-Varianten Arrow Lake-S.

Auf lange Sicht will Intel mehr Chiplets in eigenen Werken produzieren – aber genaue Pläne der Chiplet-Mischungen für die Arrow-Lake-Nachfolger Panther Lake, Nova Lake und Beast Lake verrät Intel wie üblich nicht. (ciw@ct.de) **ct**