Forschung

Neuromorphe Systeme als kommende AI-Plattform?

Zwei Forschungsprojekte beim U.S. Air Force Research Lab und an der Uni Heidelberg

von Hartmut Rehmsen

Die US-Luftwaffe tüftelt mit IBM an einem vom Gehirn inspirierten AI-Supercomputer

An das menschliche Gehirn angelehnte Supercomputing-Systeme könnten in Zukunft die Plattform für „Artificial Intelligence“ bilden. Deshalb hat das U.S. Air Force Research Lab jetzt IBM beauftragt, ein neues neurosynaptisches Supercomputing-System mit bis zu 64 Millionen Neuronen zu entwickeln, das komplexe AI-Aufgaben mit bisher nicht erreichter Geschwindigkeit und Energieeffizienz betreiben soll. Auch Deutschland mischt bei der Neuro-Forschung mit: Im April erfolgte der Spatenstich für das „European Institute for Neuromorphic Computing“ an der Universität Heidelberg.

Ende Juni haben IBM und das US Air Force Research Laboratory (AFRL) bekannt gegeben, dass sie an einem von Gehirn inspirierten Supercomputer auf Basis des vom Truenorth-Neurosynaptic-System bekannten 64-Chip-Arrays arbeiten; Truenorth ist ein neuromorpher Prozessor der zweiten Generation von IBM, der im Rahmen des Synapse-Programms der Darpa entwickelt worden ist.

Die skalierbare Plattform, die IBM für das AFRL bauen soll, wird über ein End-to-End-Software-Ökosystem verfügen, das dann ein tiefes Neuronales-Netzwerk-Lernen („Deep Learning“) und eine Informationsfindung ermöglicht. „AFRL war Early Adopter von Truenorth für die Umwandlung von Daten in Entscheidungsvorlagen“, wird Daniel S. Goddard, Informationsdirektor am U.S. Air Force Research Lab, in der Presseinformation zitiert.

Fortgeschrittene Mustererkennung

Die fortgeschrittene Mustererkennung des Systems und die sensorische Rechenleistung werden in der geplanten Endausbaustufe äquivalent zu 64 Millionen Neuronen und 16 Milliarden Synapsen sein, während die Prozessorkomponente das Energieäquivalent einer schwachen Glühbirne verbraucht - nur 10 Watt.

Forscher bei IBM glauben, dass das ein solches vom Gehirn inspiriertes, neuronales Netzwerk-Design viel effizienter für Mustererkennung und integrierte sensorische Verarbeitung sein kann als Systeme auf Basis herkömmlicher Prozessor-Chips. Das AFRL will Anwendungen des Systems in embedded, mobilen, autonomen Anwendungen untersuchen, bei denen heute Größe, Gewicht und Leistung („Size, Weight and Power“, kurz SWaP) entscheidende Faktoren sind.

Die „Linkslastigkeit“ des Computers beseitigen

IBM möchte mit seinem neuen Supercomputer de facto die Funktionen der rechten und linken Gehirnhälfte quasi miteinander kombinieren. Während konventionelle Computer derzeit die Verarbeitung und Analyse von Daten und Sprache beherrschen, also eher der linken Seite des menschlichen Gehirns entsprechen, kann das neurosynaptische System Truenorth Bilder, Video, Audio und Text aus mehreren, verteilten Sensoren effizient und in Echtzeit in „Daten“ umwandeln und für ganzheitliches Erfassen von Situationen und Intuition nutzen – wie die rechte Gehirnhälfte.

Dharmendra S. Modha, IBM Fellow und Chef-Wissenschaftler für Gehirn-inspiriertes Computing

AFRL will also die „Wahrnehmungsfähigkeit“ des Systems (quasi die „rechte Hälfte des Gehirns“) mit der klassischen Datenverarbeitung („linke Hälfte des Gehirns“) herkömmlicher Computersysteme kombinieren. Die enorme Skalierbarkeit des Systems soll dann „Datenparallelität“ ebenso ermöglichen (wobei mehrere Datenquellen parallel gegen das gleiche neuronale Netzwerk laufen können) wie auch „Modellparallelität“ (wobei unabhängige neuronale Netze ein Ensemble bilden, das parallel zu denselben Daten ausgeführt werden kann).

„In den vergangenen sechs Jahren hat IBM die Anzahl der Neuronen pro System von 256 auf mehr als 64 Millionen erweitert - eine jährliche Steigerung von 800 Prozent über sechs Jahre“, sagte Dharmendra S. Modha, IBM Fellow und Chef-Wissenschaftler für Gehirn-inspiriertes Computing im Forschungszentrum Almaden des Computerkonzerns.

Das neue System passt mit vier Höheneinheiten (also 7 Zoll) in ein Standard-Server-Rack; acht solcher Systeme ermöglichen die noch nie dagewesene Größenordnung von 512 Millionen Neuronen pro Rack. Ein einzelner Prozessor im System besteht aus 5,4 Milliarden Transistoren, die in 4.096 Neuralkerne organisiert sind. So wird ein Array von 1 Million digitalen Neuronen geschaffen, die über 256 Millionen elektrische Synapsen miteinander kommunizieren.

Enorm schnelle Mustererkennung

Damit lassen sich enorme Geschwindigkeiten bei der Mustererkennung erreichen, zum Beispiel für die Klassifizierung von Bildern.

Gemessen wurde für den vom „Canadian Institute for Advanced Research“ definierten Muster-Datensatz CIFAR-100 nahezu eine „State-of-the-Art-Genauigkeit“ – und zwar bei laufender Erkennungsrate von mehr 1.500 Frames pro Sekunde mit nur 200 mW (effektiv >7.000 Frames/s pro Watt). Das ist um Größenordnungen besser bei Geschwindigkeit und Energieverbrauch, als es mit konventionellen Computern möglich wäre. Übrigens: Die CIFAR-10- bzw. CIFAR-100-Datensätze bestehen aus Farbbildern mit 32×32 Pixel, die in 10 bzw. 100 Klassen unterteilt werden sollen.

Truenorth wurde ursprünglich unter der Schirmherrschaft von DARPA im Projekt „Systems Neuromorphic Adaptive Plastic Scalable Electronics“ (Synapse) in Zusammenarbeit von IBM Research mit der Cornell University entwickelt. Im Jahr 2016 erhielt das Truenorth-Team den ersten Misha-Mahowald-Preis für Neuromorphic Engineering – und Truenorth wurde in das Computer History Museum aufgenommen. Die Forschung mit Truenorth wird derzeit von mehr als 40 Universitäten, Regierungslabors und Industriepartnern auf fünf Kontinenten durchgeführt.

Der neuromorphe Computer „Brain Scales“

Das neuartige Computersystem „Brain Scales“ am Kirchhoff-Institut für Physik.

Einen anderen Weg gehen die Forscher an der Universität Heidelberg mit ihrem System „Brain Scales“; dieser neuromorphe Computer ging im März 2016 am Kirchhoff-Institut für Physik online. Jetzt erhalten die Heidelberger Wissenschaftler im „Human Brain“-Projekt sogar ein neues Forschungsgebäude; der Spatenstich für das „European Institute for Neuromorphic Computing“ (EINC) erfolgte im Frühjahr 2017.

Der Neubau für das EINC der Universität Heidelberg soll in rund zweijähriger Bauzeit auf dem Campus „Im Neuenheimer Feld“ errichtet werden. Mit einer Nutzfläche von rund 2.200 m² wird er auch eine große Maschinenhalle beherbergen; er entsteht in unmittelbarer Nachbarschaft zum Neubau des „Centre for Advanced Materials“, dem materialwissenschaftlichen Forschungszentrum der Uni. Bislang werden diese Forschungsarbeiten in einem Container-Bau durchgeführt.

Die Macher (von links) griffen an der Uni Heidelberg zum Spaten: Baudirektor Bernd Müller, Landtagsabgeordneter Albrecht Schütte, Staatssekretärin Gisela Splett, Ministerin Theresia Bauer, Bürgermeister Jürgen Odszuck, Katrin Tönshoff (Hopp-Stiftung), Hans-Peter Wild, Beate Spiegel (Tschira-Stiftung), Uni-Rektor Bernhard Eitel, Karlheinz Meier (Kirchhoff-Institut) und Architekt Michael Ohnemus.

Der Institutsneubau wird dann Platz bieten für die Wissenschaftler um Prof. Dr. Karlheinz Meier vom Kirchhoff-Institut für Physik. Unter seiner Leitung wollen sie gemeinsam mit Kollegen aus verschiedenen Forschungseinrichtungen in Europa eine technologische Plattform für neuromorphes Rechnen realisieren, deren Prototyp im vergangenen Jahr vorgestellt und zur Nutzung freigegeben wurde.

Was sind Neuromorphe Computer?

Neuromorphe Computer übertragen Architekturen und Prinzipien der neuronalen Informationsverarbeitung im Gehirn auf Siliziumchips.Als „neuromorph“ werden solche IT-Systeme bezeichnet, die auf elektronischen Modellen neuronaler Schaltkreise basieren. „In ihrem Aufbau orientieren sie sich an neurobiologischen Strukturen des Nervensystems und funktionieren daher fundamental anders als numerische Simulationen auf konventionellen Hochleistungscomputern“, so Prof. Meier.

„Das Heidelberger Brain-Scales-System verlässt dabei die von Alan Turing und John von Neumann in der ersten Hälfte des 20. Jahrhunderts formulierten Paradigmen einer programmgesteuerten Maschine und deren Umsetzung durch räumlich getrennte Speicher und Rechenwerke“, erläutert Prof. Meier. „Es ist vielmehr ein direktes physikalisches Abbild des biologischen Vorbildes, dessen Zellen, Verbindungen und Kommunikation durch analoge und digitale Schaltungen unter Verwendung moderner Mikroelektronik realisiert werden.“

18 Mio. Euro für ein neues Forschungsgebäude

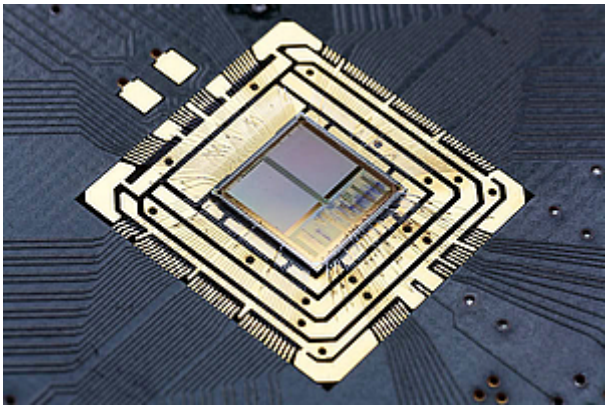

Bild eines neuromorphen Chips, konstruiert von der Heidelberger Arbeitsgruppe unter der Leitung von Prof. Karlheinz Meier. Der Chip umfasst 384 Neuronen sowie 100.000 Synapsen.

Die Hälfte der Kosten für den EINC-Neubau in Höhe von rund 18 Mio. Euro stammt aus Mitteln des „Europäischen Fonds für regionale Entwicklung“ (EFRE) im Zuständigkeitsbereich des baden-württembergischen Wissenschaftsministeriums, das weitere 2 Mio. Euro beisteuern wird. Mit einer Summe von sechs Millionen Euro sichern drei private Förderer die Finanzierung des Neubaus: Der Heidelberger Unternehmer Dr. Hans-Peter Wild, Ehrensenator der Heidelberger Uni Ruperto Carola (der ältesten Universität Deutschlands und einer der forschungsstärksten in Europa), unterstützt die geplante Infrastruktur mit 3 Mio. Euro; mit jeweils 1,5 Mio. Euro tragen die Stiftungen der beiden SAP-Gründer Klaus Tschira und Dietmar Hopp zum Bauvorhaben bei. Die Uni selbst bringt über 1 Mio. Euro aus Eigenmitteln auf.

„Hier entsteht Neues“, betont der Rektor der Universität Heidelberg, Prof. Dr. Bernhard Eitel. „Mithilfe dieser großzügigen finanziellen Unterstützung sind wir in der Lage, ein modernes Forschungsgebäude mit großer Maschinenhalle für die kognitiven Computer der Zukunft zu realisieren.“

Das „Human Brain“-Projekt

Das im Jahr 2013 gestartete „Human Brain“-Projekt ist eines von zwei „FET Flagship“-Initiativen der EU-Kommission für zukunftsweisende Technologien. Das Großprojekt zielt darauf, ein integriertes Verständnis der Gehirnstrukturen und Gehirnfunktionen mit Hilfe neuartiger Informations- und Kommunikationstechnologien zu ermöglichen. Sechs Technologieplattformen bilden die Grundlage für großangelegte Kollaborationen von Wissenschaftlern, Experten aus der klinischen Praxis und Ingenieuren.

Wichtig dabei ist das neuromorphe Rechnen. Es spielt eine zentrale Rolle bei der Emulation von Lern- und Entwicklungsprozessen, die wiederum als Basis für die Entwicklung kognitiver Computer für Anwendungen im maschinellen Lernen dienen. Gebraucht wird dafür „Neuromorphic Engineering“, eine neue interdisziplinäre Richtung biologisch inspirierter Physik, Mathematik, Informatik und Ingenieurwissenschaften zur Verbesserung der AI. Dies dient der Entwicklung von Anwendungen wie dem maschinellen Sehen, „Head-Eye Systems“, maschinellem Hörverstehen sowie autonomen Robotern.

„Neuromorphe Chips“ sind Silikonchips, die nach dem Beispiel von Nervennetzen gebaut sind. Solche Chips sind Forschungsgebiet der Bionik und der Neuroinformatik. Erste Prototypen bilden künstliche Innenohre (Cochlea), Netzhäute (Retina) oder weiter entwickelte Geruchsdetektoren und Sehsysteme ab, aber auch Taktgeber für naturnahe Bewegungsabläufe für Roboter. In Mensch oder Tier implantiert, können diese Chips künftig vielleicht sogar als Ersatz für durch Krankheit oder Unfall beeinträchtige Körperteile genutzt werden; andere Einsatzgebiete sind sensorische Systeme für Roboter oder andere autonome Systeme.

Während bei klassischen Von-Neumann-Computern die Hardware sehr allgemein konzipiert ist (und die Software dementsprechend sehr speziell), sind neuromorphe Chips hochspezialisiert auf eine Aufgabe – mit entsprechend spezialisierter Hardware, die durch hohe Vernetzung (Interkonnektivität) und diverse Rückkopplungen charakterisiert ist. Letztlich wird sogar angestrebt, die selbstorganisierende Entwicklung des Gehirns in Hardware zu gießen.

Was passiert in Heidelberg?

Als neuromorph bezeichnet man Computer nach dem Vorbild des Gehirns. Sie weisen dann auch besonders attraktive

Eigenschaften auf: Sie sind energieeffizient, robust gegenüber Fehlfunktionen und können lernen. Sie eignen sich besonders gut zur Realisierung kognitiver Computer, die komplexe Daten analysieren und erlerntes Wissen für Vorhersagen nutzen können – etwa bei der Mustererkennung und Mustervorhersage. Allerdings sind sie bei der Datenverarbeitung nicht deterministisch, denn Daten werden nicht gespeichert, sondern gelernt und ggf. wieder vergessen. Daher sind neuromorphe Computer ungeeignet, um zuverlässig und konsistent Berechnungen durchzuführen.

Die Heidelberger Arbeitsgruppe hat als Ergebnis einer mehr als 15-jährigen Förderung durch die Europäische Kommission ein neuartiges Konzept für solche neuromorphen Computer entworfen und umgesetzt, das auf radikale Weise alle etablierten Prinzipien konventioneller Computer hinter sich lässt und durch neuartige physikalische Modelle neuronaler Schaltkreise ersetzt. Im Neuenheimer Feld wurde ein System mit vier Millionen elektronischen Neuronen und einer Milliarde Synapsen aufgebaut und in Betrieb genommen. Gemeinsam mit Arbeitsgruppen in Manchester (UK) und im eingangs erwähnten IBM Forschungslaboratorium Almaden (USA) gehört die Heidelberger Gruppe damit zu den führenden weltweit; der vom IBM entwickelte Truenorth-Chip ist komplementär zu den beiden europäischen Systemen.

Die europäischen Projekte Facets und Brain-Scales

Die europäischen Entwicklungen basieren auf den Projekten Facets und Brain-Scales, die von 2005 bis 2015 im Rahmen des Programms „Future Emerging Technologies“ (FET) von der Europäischen Kommission gefördert wurden, sowie auf dem britischen Projekt Spinnaker. Mit den beiden Maschinen Brain Scales und Spinnaker positioniert sich Europa mit eigenen Hardwareentwicklungen auf dem Gebiet des alternativen Computing.

Parallel zum Start des Heidelberger Brain-Scales-Systems wurde 2016 an der Universität Manchester (Großbritannien) ein System vergleichbarer Größe mit dem Namen Spinnaker in Betrieb genommen. Unter der Leitung des Informatikers Prof. Dr. Steve Furber, der in den 1980er-Jahren die ARM-Chiparchitektur mitentwickelt hat, wurde dort ein System aus 500.000 ARM-Kernen aufgebaut; zusammen bilden die Systeme in Heidelberg und Manchester die „Neuromorphic Computing Platform“ des Human-Brain-Projektes.

Das in Heidelberg genutzte System besteht aus 20 Siliziumwafern mit insgesamt vier Millionen Neuronen und einer Milliarde synaptischer Verbindungen. Lern- und Entwicklungsprozesse können dort mit einem tausendfachen Beschleunigungsfaktor nachgeahmt werden, so dass die Simulation eines Tages auf 100 Sekunden komprimiert wird. Neben der Grundlagenforschung im Bereich der Selbstorganisation neuronaler Netzwerke liegen potentielle Anwendungen auf dem derzeit schnell wachsenden Gebiet des kognitiven Computing und speziell einer energie- und zeiteffizienten Umsetzung des sogenannten „Deep Learning“, wie es von Unternehmen wie Google und Facebook auf konventionellen Computern zur Analyse großer Datenmengen entwickelt wird.

Hartmut Rehmsen

hat sich als Fachjournalist auf den IT-Einsatz in Unternehmen spezialisiert.

Bildnachweise:

IBM / Universität Heidelberg / Kirchhoff-Institut für Physik