

Energy Models for Accelerated Workloads

- Improving Power Management in the Future SOC Lab

Arbeitslasten wie Maschinelles Lernen, In-Memory Computing sowie ähnlich komplexe Anwendungen stellen vermehrt eine Herausforderung für den Betrieb von Rechenzentren dar. Ein Aspekt ist die intensive Nutzung von Ressourcen, welche das Energiemanagement in den Fokus rückt. Einerseits übersteigen die laufenden Kosten wie z.B. Energiekosten eines Rechenzentrums bei weitem die Anschaffungskosten der Hardware. Andererseits ist der Energieverbrauch von IT-Systemen auf vielfältige Weise ein limitierender Faktor geworden. Dieser bestimmt wie viele Server betrieben und welche Ausstattung diese haben können. Die erforderliche Kühlleistung ist ebenfalls eng an dem Energieverbrauch der zu kühlenden Systeme gekoppelt. Die Rechenzentrumslandschaft wird immer heterogener und der Trend zeigt, dass Server-Systeme immer mehr Energie benötigen. Das wird bereits im Future SOC Lab deutlich.

Das HPI Future SOC (Service-Oriented Computing) Lab ist eine Kooperation des Hasso-Plattner-Instituts mit den Industriepartnern Hewlett Packard Enterprise (HPE), Dell EMC, Fujitsu, SAP. Seine Mission ist die Förderung des Austauschs zwischen Forschungsgemeinschaft und Industrie. Am Lab wird interessierten Wissenschaftlern eine Infrastruktur von neuester Hard- und Software kostenfrei für Forschungszwecke zur Verfügung gestellt. Einige der Schwerpunkte sind Cloud Computing, Maschinelles Lernen, Parallelisierung und In-Memory Technologien.

Studierende sollen sich in diesem Masterprojekt mit Energiemanagement von Beschleunigern auseinandersetzen. Systeme wie die Nvidia DG-X1¹ mit 8 vernetzten V100-GPUs haben eine Reihe interessanter Eigenschaften. Neben eines *distributed shared memory* Speichermodells ist auch der Energieverbrauch in Abhängigkeit der Arbeitslastcharakteristiken von großem Interesse. Es existieren bereits Werkzeuge zum Aufzeichnen von Energieverbrauchswerten. Diese wurden bisher größtenteils auf CPUs und Hauptspeicher angewendet. Ziel ist es nun Methoden, Techniken und Werkzeuge zu evaluieren, mit denen es möglich ist den Energieverbrauch von Beschleunigern zu erfassen und auszuwerten. Ein oder mehrere dieser Verfahren sollen ausgewählt und in Betrieb genommen werden. Arbeitslasten wie maschinelles Lernen sollen herangezogen und Energieprofile vermessen werden. Ziel ist es die Ergebnisse in einem wissenschaftlichen Papier oder einem technischen Bericht zu veröffentlichen.

Kontakt

Bei weiteren Fragen stehen wir gerne zur Verfügung. Projektverantwortlich sind Bernhard Rabe, Felix Eberhardt, Jossekin Beilharz, Max Plauth sowie Andreas Polze.

¹ <https://www.nvidia.com/de-de/data-center/dgx-1/>