



Beiträge zur Steigerung der Energieeffizienz FPGA-basierter Anwendungen

Der Fakultät für Informatik und Elektrotechnik der Universität Rostock
zur Erlangung des akademischen Grades eines

Doktor-Ingenieur (Dr.-Ing.)

vorgelegte Dissertation von
Christoph Niemann
geb. am 05.08.1988 in Potsdam
aus Rostock

Rostock, 12. September 2022

Zusammenfassung

Über mehrere Jahrzehnte gelang es der Halbleiterindustrie, durch die Skalierung der Prozesstechnologie enorme Leistungssteigerungen zu erreichen. Auch der Energiebedarf für Rechenoperationen wurde dabei extrem gesenkt. Aktuell verlangsamt sich dieser Trend jedoch deutlich. Wichtige Zukunftstechnologien wie künstliche Intelligenz und moderne Kommunikationssysteme wie 5G und 6G verlangen wiederum nach einer wachsenden Rechenleistung. Gleichzeitig darf der Energiebedarf jedoch nicht beliebig steigen. Weltweit betrachtet lag der Energiebedarf der Informationsverarbeitungs- und Kommunikationssysteme 2012 bei 920 TWh. Dies entspricht etwa 5 % des gesamten Verbrauchs an elektrischer Energie. 2018 lag der Verbrauch bereits bei 2000 TWh. Die Herausforderung liegt also darin, trotz einer weniger dynamischen Entwicklung der Prozesstechnologie mehr Rechenleistung pro Energieeinheit bereitzustellen.

Eine Möglichkeit dazu ist die Verwendung spezialisierter Hardware. Dabei spielen Field Programmable Gate Arrays (FPGAs) aufgrund ihrer Flexibilität eine wichtige Rolle. Mit ihnen kann, verglichen mit Central Processing Units (CPUs), eine um mehrere Größenordnungen höhere Energieeffizienz erreicht werden. Im Vergleich zu Application-specific Integrated Circuits (ASICs) wiederum sind FPGAs bezüglich der Energieeffizienz unterlegen. ASICs bieten jedoch nicht die häufig benötigte Flexibilität eines FPGAs. Es werden also Verfahren benötigt, um die Energieeffizienz von FPGAs zu steigern. Wie beschrieben genügt es nicht, nur auf die weitere Skalierung der Complementary Metal–Oxide–Semiconductor (CMOS)-Technologie zu vertrauen. Vielmehr werden einerseits Methoden benötigt, um das Potential der gegebenen Prozesstechnologie eines FPGAs besser auszuschöpfen. Andererseits müssen Anstrengungen unternommen werden, um Anwendungen energieeffizienter auf die Hardwareressourcen von FPGAs abzubilden.

Die vorliegende Arbeit verschreibt sich dieser Aufgabe und leistet dazu zwei Beiträge. Zum einen wird ein Adaptive Voltage Scaling (AVS)-System vorgestellt, das gezielt auf die Besonderheiten von FPGAs abgestimmt ist. Dabei wird insbesondere die Sensorkalibrierung und -platzierung in den Mittelpunkt der Betrachtung gestellt. Eine Implementierung und experimentelle Evaluation zeigen das erhebliche Potential des erarbeiteten Konzepts zur Einsparung von Energie auf.

Zum anderen wird Approximate Computing als eine Möglichkeit, Anwendungen energieeffizienter umzusetzen, untersucht. Dazu werden zunächst verschiedene Gütemaße und Optimierungsziele für approximierbare arithmetische Strukturen diskutiert. Um die Vergleichbarkeit der vielfältigen vorgeschlagenen Lösungen in diesem Bereich zu verbessern, wird zudem eine Evaluationsmethodik erarbeitet. Für das gewählte Optimierungsziel wird ein gezielt für moderne FPGA-Architekturen optimierter approximierter Multiplizierer entworfen. Dieser ist bezüglich der Leistungsaufnahme in Relation zu mehreren Gütemaßen für die Qualität der Approximation paretooptimal.

Abstract

Over several decades, the semiconductor industry has succeeded in achieving enormous performance increases by process technology scaling. The energy required for computing operations has also been extremely reduced. However, this trend is currently slowing down considerably. Important future technologies like artificial intelligence and modern communication systems such as 5G and 6G require increasing computing capabilities. Meanwhile, the energy consumption needs to be limited. Viewed globally, the energy demand of information processing and communications systems was 920 TWh in 2012. This corresponds to around 5% of the total consumption of electrical energy. In 2018, the power consumption has increased to 2000 TWh. The challenge therefore lies in processing more computational workload per unit of energy despite a less dynamic evolution of process technology.

One way to do this is to use specialized hardware. FPGAs are an important technology for this strategy. They enable improvements in energy efficiency of multiple orders of magnitude when compared to CPUs. But, compared to ASICs, FPGAs are inferior in terms of energy efficiency. However, ASICs do not offer the required flexibility of an FPGA. Hence, there is an urgent need for approaches to increase the energy efficiency of FPGAs. As described, it is not sufficient to rely on further scaling of the CMOS technology. Therefore, methods are needed to better exploit the potential of the given process technology of an FPGA. Moreover, efforts must be made to map applications more energy-efficiently to the hardware resources of FPGAs. The present work is dedicated to this task and makes two contributions.

On the one hand, an AVS system is presented, which is specifically tailored towards the properties of FPGAs. In particular, sensor calibration and placement are the focus of attention. An implementation and experimental evaluation show the considerable potential of the developed concept for saving energy.

On the other hand, approximate computing is investigated as a possibility to implement applications more energy-efficiently. For this purpose, different measures of quality and optimization goals for approximate arithmetic circuits are discussed. In order to improve the comparability of approaches in this area, an evaluation methodology is developed. For the chosen optimization objective, an approximate multiplier specifically optimized for modern FPGA architectures is proposed. The evaluation shows, that it is pareto-optimal regarding its power consumption and the quality of the approximation.