

Virtualisierung rekonfigurierbarer Hardware zur Steigerung der Rechenleistung und Sicherheit in einer flexiblen Cloud-Architektur

Oliver Knodel¹

Abstract: Rechenleistung und Skalierbarkeit sind essentielle Bestandteile moderner Rechenzentren. Spezielle rekonfigurierbare Hardwarebeschleuniger wie FPGAs (Field Programmable Gate Arrays) stellen eine Möglichkeit dar, um sowohl Rechenleistung, als auch die Energieeffizienz zu steigern, da die Architektur direkt an die Problemstellung angepasst werden kann. Insbesondere bei der Hintergrundbeschleunigung von Cloud-Anwendungen können durch den Einsatz von FPGAs die Betriebskosten eines Rechenzentrums deutlich verringert werden. Eine weitere Einsatzmöglichkeit ist die Erhöhung der Sicherheit von Anwendungen in der Cloud. Dieser Beitrag erläutert, wie rekonfigurierbare Hardware flexibel in eine Cloud-Umgebung eingebettet werden kann. Die Zielstellung besteht darin, die *Ressource FPGA* einer breiten Nutzergruppe für unterschiedlichste Services bereitzustellen. Neben der Verwaltung von Ressourcen wird insbesondere die Möglichkeit einer Virtualisierung der Hardware diskutiert.

Keywords: Rekonfigurierbare Hardware, Cloud Computing, Virtualisierung, Ressourcenverwaltung

1 Motivation und Problemstellung

Das Konzept des Cloud Computing basiert darauf, dass Nutzer Zugang zu gemeinsamen Ressourcen oder Diensten erhalten, die nach Belieben allokiert und wieder freigegeben werden können. Eine möglichst geringe Interaktion mit dem Anbieter der Cloud oder dem Betreiber des Rechenzentrums stellt dabei ein wesentliches Kriterium dar [MG11]. Dieses Konzept der Elastizität in der Cloud ist ein essentieller Unterschied zu traditionellen Rechenzentren, bei denen der Nutzer eine feste Anzahl von Ressourcen mietet.

Ein Hauptproblem bei großen Rechenzentren, welche Clouds bereitstellen, besteht in ihrem enormen Energieverbrauch. Eine deutliche Steigerung der Rechenleistung bei zusätzlicher Einsparung von Energie kann durch heterogene Systeme erreicht werden, in denen rechenintensive Aufgaben auf spezielle Koprozessoren oder Hardwarebeschleuniger wie FPGAs ausgelagert werden [Pu14]. Eine tiefgreifende und flexible Integration von FPGAs in ein skalierbares Rechenzentrum, welches die Cloud Charakteristiken *on demand self service, broad network access, resource pooling, rapid elasticity* und *measured service* erfüllt [MG11], bildet daher ein Forschungsfeld, welches an Bedeutung gewinnt. Um eine solche Integration zu erreichen ist neben einer Ressourcenverwaltung für spezielle Hardwarekomponenten eine Virtualisierung des FPGAs selbst unumgänglich.

¹ Technische Universität Dresden, Fakultät Informatik, 01069 Dresden, oliver.knodel@tu-dresden.de

Im Folgenden wird in Abschnitt 2 ein Überblick zur existierenden Literatur und verwandten Arbeiten gegeben. Danach werden in Abschnitt 3 zunächst Sichtweisen sowie mögliche Nutzer einer Cloud mit integrierten FPGAs diskutiert, bevor ein Entwurf einer flexiblen Cloud-Architektur auf Hardware- und Softwarebasis aufgezeigt wird. Hierbei spielen sowohl die Möglichkeit eines sicheren Zuganges zur Cloud über FPGAs, als auch eine Virtualisierung der Hardware selbst zur Erhöhung ihrer Auslastung eine wichtige Rolle. Nach diesem kurzen Abriss des Forschungsgegenstandes werden in Abschnitt 4 der aktuelle Stand der Arbeit, sowie erste Ergebnisse und ein Ausblick präsentiert.

2 Literaturüberblick und verwandte Arbeiten

In Cloud-Systemen werden FPGAs derzeit hauptsächlich zur Hintergrundbeschleunigung von Kernanwendungen statisch eingesetzt. Beispiele sind die Anonymisierung von eingehenden Nutzeranfragen, bevor diese an die eigentlichen Cloud-Dienste weitergereicht werden [EV12]. Ebenso ist die Erhöhung der Sicherheit bei Datenbanken und sicherheitskritischen Teilanwendungen Gegenstand der Forschung [Mo11]. Beiträge mit einem stärkeren Fokus auf austauschbaren rekonfigurierbaren Rechenkernen und einer flexiblen Bereitstellung der Ressource für eine breite Nutzergruppe gewinnen aber an Bedeutung [Fa15]. Neben einer Steigerung der Rechenleistung kann insbesondere auch die Sicherheit durch den Einsatz rekonfigurierbarer Hardware erhöht werden [EV12].

Entscheidend dabei ist der Begriff der *Virtualisierung*, welcher im Bereich der rekonfigurierbaren Hardware in unterschiedlichsten Zusammenhängen genutzt wird. Eine tiefgreifende Virtualisierung stellt pvFPGA [Wa13] mit einer Integration eines FPGAs in eine paravirtualisierte Xen-Umgebung dar. Ein Framework, welches rekonfigurierbare Hardware direkt in eine Cloud-Architektur integriert, wird in der Arbeit von Byma et. al [By14] vorgestellt. Das System ermöglicht die Beschleunigung von Anwendungen mit nutzerspezifischen Rechenkernen. Ein Ansatz mit mehreren unterschiedlichen Rechenkernen auf demselben physischen FPGA wird von Fahmy et. al [Fa15] und Weerasinghe et. al [We15] vorgestellt.

In der Literatur sind ebenso Beiträge zu finden, welche sich mit dem Speichern und dem Wiederherstellen des Kontextes eines Hardwaredesigns auf einem FPGA beschäftigen. In [Jo10] stellen Jozwick et al. ein System vor, welches die Konfiguration der internen Registerzustände aus einem FPGA ausliest und diesen Kontext zu einem späteren Zeitpunkt wiederherstellt, wobei jedoch zusätzliche Logik erforderlich ist. Happe et al. [Ha15] zeigen in ihrer Arbeit einen Kontextwechsel ohne zusätzliche Logik über das Auslesen und Modifizieren der Konfigurationsdatei.

3 Einsatzmöglichkeiten rekonfigurierbarer Hardware in einer Cloud

Das Ziel der Arbeit besteht darin, aufbauend auf der aktuellen Forschung (Abschnitt 2) die Einsatzmöglichkeiten rekonfigurierbare Hardwarekomponenten in Cloud-Architekturen zu untersuchen. Die grundlegenden Forschungsfragen dabei sind:

- Einsatzmöglichkeiten für rekonfigurierbare Hardware und deren Verwaltung in einer universellen Cloud-Architektur.
- Hintergrundbeschleunigung von Anwendungen und optimale Auslastung durch Virtualisierung der rekonfigurierbaren Hardware.
- Steigerung der Sicherheit von Daten und dem Zugang zur Cloud durch homomorphe Verschlüsselung [Ge09].

Die Bereitstellung von rekonfigurierbarer Hardware in einer Cloud oder einem Mehrbenutzersystem erfordert zunächst die in Abschnitt 3.1 beschriebene Analyse der unterschiedlichen Nutzer und die Festlegung auf entsprechende Servicemodelle, bevor Architektur und Komponenten in Abschnitt 3.2 aufgezeigt werden. Schließlich wird in Abschnitt 3.3 eine mögliche Virtualisierung erläutert, um den FPGA in einen *virtualisierten* FPGA (vFPGA) zu überführen.

3.1 Nutzer und Interaktionsmöglichkeiten

Der Einsatz von FPGAs zur Beschleunigung von Anwendungen und Services in Rechenzentren ist aufgrund der komplexen Integration und Bereitstellung die Ausnahme. In den meisten Fällen werden mittels FPGAs Dienste in einer Cloud beschleunigt, welche direkt vom Eigentümer des Rechenzentrums bereitgestellt werden, wie beispielsweise die Suchmaschine *Bing* von Microsoft [Pu14]. Hier ist ein vollständiger Zugriff auf die Ressource FPGA aufgrund des Zugangs zum Gesamtsystem möglich. Sollen jedoch Dienste von Anbietern beschleunigt werden, die selbst nur Nutzer einer Cloud-Infrastruktur sind, wie in Abb. 1(a) gezeigt, ist eine Virtualisierung der FPGAs erforderlich.

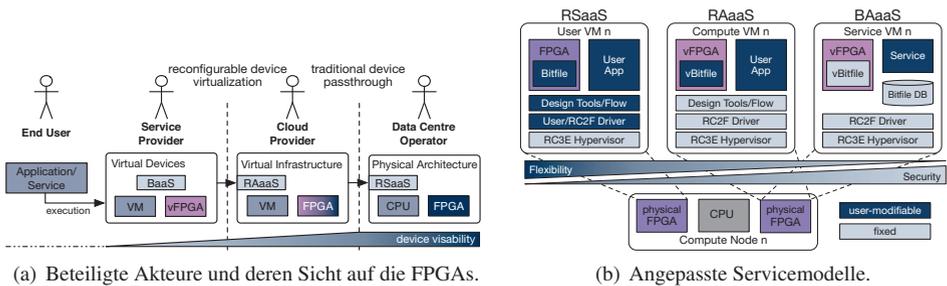


Abb. 1: Akteure und Interessengruppen sowie angepasste Servicemodelle in einer FPGA-Cloud.

Neben der Virtualisierung und der Bereitstellung eines Zugangs sind aber auch entsprechende Servicemodelle und eine flexible Einbettung von FPGAs in eine Cloud-Architektur erforderlich. Eine Ressourcenverwaltung, welche die Rechenlast auf die unterschiedlichen Knoten verteilt und den Zugang zu virtuellen FPGA-Ressourcen bereitstellt, ist das **Reconfigurable Common Cloud Computing Environment – RC3E**, welches im Rahmen der Promotion entstanden ist.

Die Ressourcenverwaltung basiert auf drei Servicemodellen, welche vom vollständigen Zugriff auf einen physischen FPGA bis hin zur Hintergrundbeschleunigung auf vFPGAs unterschiedlichste Stufen von Sicherheit und Flexibilität bieten (siehe Abb. 1(b)). Die Modelle sind im Einzelnen:

Reconfigurable Silicon as a Service – RSaaS: Voller Zugriff auf den gesamten physischen FPGA mit selbst definiertem Entwurfsablauf und sämtlichen Schnittstellen.

Reconfigurable Accelerators as a Service – RAaaS: Generieren einer eigenen Konfiguration auf einem vFPGA. Die Schnittstellen nach außen und der Entwurfsablauf sind vordefiniert.

Background Acceleration as a Service – BAaaS: Nur vordefinierte Konfigurationen auf einem vFPGA sind möglich, sodass eine sichere Hintergrundbeschleunigung von Anwendungen und Services umgesetzt werden kann.

Die FPGA-spezifischen Komponenten, wie das Hardwaredesign, welches die vFPGAs (siehe Abb. 2(c)) und die Softwareschnittstellen bereitstellt, bilden das ebenfalls während der Arbeit entstandene **Reconfigurable Cloud Computing Framework – RC2F**. Das Framework unterstützt die Möglichkeit der Virtualisierung von FPGAs, welche in Abschnitt 3.3 näher erläutert wird. Ergänzt wird das Virtualisierungskonzept durch die Nutzung des FPGAs als sicheren Zugang zur Cloud mittels des **Reconfigurable Secured Cloud Access – RSCA** in Abschnitt 3.4.

3.2 Entwurf einer flexiblen Cloud-Architektur mit sicherem Zugang

Eine Architektur, welche den unterschiedlichen Modellen gerecht wird, und somit unterschiedlichste Möglichkeiten der Interaktion mit der rekonfigurierbaren Hardware ermöglicht, bildet die Grundlage des RC3E. Des Weiteren ist der direkte Zugang zur Hardware ein wesentliches Kriterium, um einerseits die Rechenlast verteilen zu können und andererseits die Möglichkeit eines direkten Zugriffes in Form einer verschlüsselten Verbindung herzustellen.

Der Zugriff auf die Cloud wird über einen Managementknoten ermöglicht, der über Virtuelle Maschinen (VM) auf einem Rechenknoten den Zugang zu den vFPGAs bereitstellt, wie Abb. 2(a) zeigt. Jeder Rechenknoten besteht aus einem Host-System mit zwei mittels PCIe verbundenen FPGAs (siehe Abb. 2(b)). Die Allokation eines einzelnen FPGAs oder eines vFPGAs mit vordefiniertem RC2F Design (siehe Abb.2(c)) wird durch das Servicemodell RAaaS ermöglicht, welches für den Entwurf neuer Kerne zur Beschleunigung von Anwendungen vorgesehen ist.

3.3 Virtualisierung der FPGAs

Um eine dynamische Interaktion mit rekonfigurierbarer Hardware in eine derartige Cloud-Architektur zu ermöglichen, ist eine tiefgreifende Virtualisierung der FPGAs erforderlich.

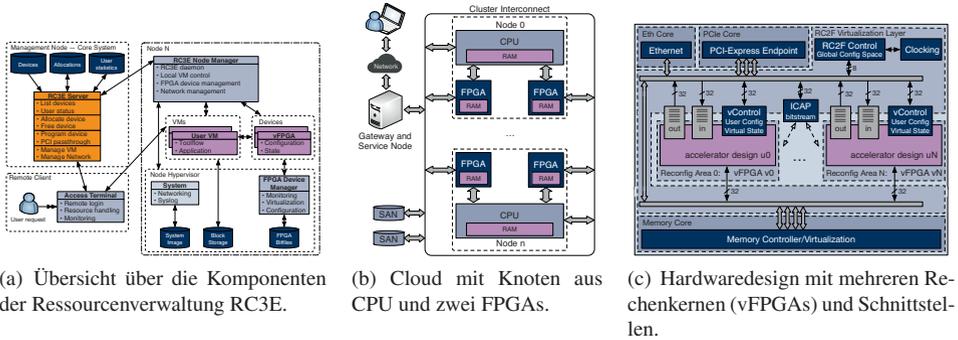


Abb. 2: Aufbau der Ressourcenverwaltung RC3E, der Knoten und des RC2F Hardwaredesigns.

Ziel ist es dabei, mit unterschiedlichsten Nutzern auf demselben physischen FPGA die Auslastung der Ressourcen zu maximieren. Wie bereits in Abschnitt 2 gezeigt, wird der Begriff Virtualisierung im Kontext von FPGAs für unterschiedliche Konzepte genutzt. Ein Grundgedanke des gewählten Ansatzes besteht darin, das Hardwaredesign innerhalb eines vFPGAs auf einem physischen FPGA auszuführen (analog Bare-Metal Virtualisierung).

3.4 Sicherheitskonzept unter Verwendung einer homomorphen Verschlüsselung

Die Architektur aus 3.2 bietet des Weiteren die Möglichkeit, mittels des internen Cloud-Verbindungsnetzwerks direkt auf die vFPGAs zuzugreifen. Auf diese Weise kann über den FPGA ein unmittelbarer Zugang zum Cloud-System ermöglicht werden. Durch Berechnungen mit Klartextdaten auf dem FPGA kann die Sicherheit durch eine Anonymisierung der Nutzerdaten ähnlich wie in [EV12] erreicht werden. Das Prinzip der homomorphen Verschlüsselung [Ge09] kann des Weiteren genutzt werden um Anwendungsteile auf den Host auszulagern.

4 Stand der Arbeit, erste Ergebnisse und Ausblick

Der Stand der Forschung auf dem Gebiet rekonfigurierbarer Hardware und mögliche Einsatzfelder von FPGAs in einer Cloud-Umgebung wurden bereits analysiert, wobei energieeffiziente Hintergrundbeschleunigung und Erhöhung der Sicherheit im Vordergrund standen. Essentiell bei der Entwicklung der flexiblen Cloud-Architektur ist eine Virtualisierung der rekonfigurierbaren Hardware, um Auslastung und Effizienz des Systems zu maximieren, eine Ressourcenverwaltung sowie ein Sicherheitskonzept. Die bisherigen Forschungsergebnisse können wie folgt zusammengefasst werden:

- Exemplarische Implementierung einer Cloud-Verwaltung (RC3E), in welche rekonfigurierbare Hardware mit unterschiedlichsten Servicemodellen integriert werden kann [KLS16].

- Erstellen eines Konzepts zur Virtualisierung von FPGAs sowie dessen Einbettung in ein Framework (RC2F) zur Beschleunigung von Hintergrundanwendungen [Kn15, KGS16] mit der Möglichkeit zur Migration eines Hardwaredesigns.
- Entwurf einer Architektur mit flexiblen Zugangsmöglichkeiten zur Cloud, welche die Sicherheit innerhalb des Systems erhöht [KLS16], indem der Zugang zum Prozessor eines Knotens über den FPGA erfolgen kann.

Literaturverzeichnis

- [By14] Byma, Stuart et al.: FPGAs in the Cloud: Booting Virtualized Hardware Accelerators with OpenStack. In: Field-Programmable Custom Computing Machines (FCCM), 22nd Int'l Symp. on. IEEE, 2014.
- [EV12] Eguro, Ken; Venkatesan, Ramarathnam: FPGAs for trusted cloud computing. In: Field Programmable Logic and Applications (FPL), 22nd Int'l Conf. on. IEEE, S. 63–70, 2012.
- [Fa15] Fahmy, Suhaib A et al.: Virtualized FPGA accelerators for efficient cloud computing. In: Cloud Computing Technology and Science (CloudCom), Int'l Conf. on. IEEE, 2015.
- [Ge09] Gentry, Craig et al.: Fully homomorphic encryption using ideal lattices. In: STOC. Jgg. 9, S. 169–178, 2009.
- [Ha15] Happe, Markus et al.: Preemptive Hardware Multitasking in ReconOS. In: Applied Reconfigurable Computing. Springer, 2015.
- [Jo10] Jozwik, Krzysztof et al.: A novel mechanism for effective hardware task preemption in dynamically reconfigurable systems. In: Field Programmable Logic and Applications (FPL). Int'l Conf. on. IEEE, 2010.
- [KGS16] Knodel, Oliver; Genßler, Paul; Spallek, Rainer: Migration of long-running Tasks between Reconfigurable Resources using Virtualization. In: ACM SIGARCH Computer Architecture News Volume 44 - HEART '16. ACM, 2016.
- [KLS16] Knodel, Oliver; Lehmann, Patrick; Spallek, Rainer: RC3E: Reconfigurable Accelerators in Data Centres and their Provision by Adapted Service Models. In: Cloud Computing, 9th Int'l Conf. on. IEEE, 2016.
- [Kn15] Knodel, Oliver et al.: Computing Framework for Dynamic Integration of Reconfigurable Resources in a Cloud. In: Digital System Design, Euromicro Conf. on. IEEE, 2015.
- [MG11] Mell, Peter; Grance, Timothy: The NIST definition of cloud computing. National Institute of Standards and Technology Gaithersburg, S. 23–27, 2011.
- [Mo11] Mondol, J-AM: Cloud security solutions using FPGA. In: Communications, Computers and Signal Processing (PacRim), Pacific Rim Conf. on. IEEE, S. 747–752, 2011.
- [Pu14] Putnam, Andrew et al.: A reconfigurable fabric for accelerating large-scale datacenter services. In: Computer Architecture (ISCA), 41st Int'l Symp. on. IEEE, 2014.
- [Wa13] Wang, Wei et al.: pvFPGA: Accessing an FPGA-based hardware accelerator in a paravirtualized environment. Hardware/Software Codesign and System Synthesis (CODES+ISSS), Int'l Conf. on, 2013.
- [We15] Weerasinghe, Jagath et al.: Enabling FPGAs in Hyperscale Data Centers. In: Cloud and Big Data Computing (CBDC), Int'l Conf. on. IEEE, 2015.