

Veröffentlichungen

2024/12/20

Paul Palomero Bernardo, Patrick Schmid, Christoph Gerum, and Oliver Bringmann. 2025. Compiler-aware AI Hardware Design for Edge Devices. In The 8th International Workshop on Edge Systems, Analytics and Networking (EdgeSys '25), March 30-April 3, 2025, Rotterdam, Netherlands. ACM, New York, NY, USA, 6 pages.

Scale4Edge at Digital-Gipfel 2024

2024/09/24



Scale4Edge Projektkoordinator Prof. Dr. Wolfgang Ecker, Distinguished Engineer, Infineon Technologies AG ist zum **Digital-Gipfel 2024 der Bundesregierung** eingeladen um unter anderem über Scale4Edge und den Projektbeitrag zu "Sicher und resilient in die Zukunft: Innovationen der Mikroelektronik für KI-Anwendungen von morgen" vorzutragen.

Um die immer größer werdenden Datenmengen zu bewältigen, werden Prozessoren benötigt, die nicht nur in puncto Rechenleistung, sondern auch hinsichtlich Energieeffizienz, Zuverlässigkeit, Robustheit und Sicherheit höchsten Anforderungen standhalten. Bei einem Speed Geeking können sich die Teilnehmenden zu innovativen Ideen im Bereich KI-Prozessoren und Open Source informieren und aktuelle Forschungsaktivitäten als Beiträge zu einer technologisch souveränen Wertschöpfung und Digitalisierung in Deutschland und Europa kennenlernen.

Der Beitrag von Wolfgang Ecker findet am Montag, den 21.10.2024 um 14:30 Uhr statt.

14:30
-
15:15

Sicher und resilient in die Zukunft: Innovationen der Mikroelektronik für KI-Anwendungen von morgen

Um die immer größer werdenden Datenmengen zu bewältigen, werden Prozessoren benötigt, die nicht nur in puncto Rechenleistung, sondern auch hinsichtlich Energieeffizienz, Zuverlässigkeit, Robustheit und Sicherheit höchsten Anforderungen standhalten. Bei einem Speed Geeking können sich die Teilnehmenden zu innovativen Ideen im Bereich KI-Prozessoren und Open Source informieren und aktuelle Forschungsaktivitäten als Beiträge zu einer technologisch souveränen Wertschöpfung und Digitalisierung in Deutschland und Europa kennenlernen.

Prof. Dr. Wolfgang Ecker, Senior Principal, Infineon Technologies AG

Prof. Dr. Armin Dietz, Professor und Leiter H2Ohm Institut, Technische Hochschule Nürnberg Georg Simon Ohm

Dr. Hans-Jörg Vögel, Head of Hardware/Software Co-Design, Bayerische Motoren Werke AG

Prof. Dr. Holger Blume, Professor für Architekturen und Systeme und geschäftsführender Leiter des Instituts für Mikroelektronische Systeme, Leibniz Universität Hannover

Prof. Dr. Mladen Berekovic, Leiter Institut für Mikroelektronische Systeme, Universität zu Lübeck

Prof. Dr. Dietmar Fey, Leiter Post Silicon Technologies, Friedrich-Alexander-Universität Erlangen-Nürnberg

Dr. Jens Krüger, Abteilungs- und Teamleiter Fraunhofer-Institut für Techno- und Wirtschaftsmathematik

Prof. Dr. Christian Mayr, Professor für Hochparallele VLSI-Systeme und Neuromikroelektronik, Technische Universität Dresden

Jacob Göppert, wiss. Mitarbeiter der Hahn-Schickard-Gesellschaft für angewandte Forschung e.V.

Speed Geeking
und Pitches

Moderation:
Dr. Uwe Vogel,
VDI/VDE Inno-
vation + Technik
GmbH



Raum Sirius

Unique Program Execution Checking: Formal Security Guarantees for RISC-V Systems - Alex Wezel, RPTU @ RISC-V Summit Europe 2024

2024/07/31

The seemingly endless stream of previously unknown microarchitectural attacks and security flaws in hardware systems is driving the need for more efficient and comprehensive verification techniques. Unique Program Execution Checking (UPEC) is a formal verification methodology that can be used to verify a wide variety of security requirements for hardware at the Register Transfer Level (RTL). With its white-box nature and reusable verification IP, UPEC complements the open-source ecosystem provided by RISC-V. We demonstrate the efficacy of UPEC through a set of selected case studies, covering different threat models and ranging from small RISC-V processors to entire Systems-on-Chips (SoCs). [Details and slides.](#)

The video is played on YouTube if you click on the image below.

Unique Program Execution Checking: Formal Security Guarantees for RISC-V Systems

Alex Wezel

RPTU Kaiserslautern-Landau

RISC-V

SUMMIT EU
JUNE 24 - 28 | MUNICH 2024

EDGE AI: Das verfügbare Potenzial ist grenzenlos ein Interview mit Wolfgang Ecker

von Infineon

2024/07/08

Aus Aus Industry.zero & Transformation for Industry Leaders am 9.7.2024 am 9.7.2024

Die jüngsten Erfolge in der generativen KI basieren auf einem Anstieg zentral verarbeiteter Daten, größeren neuronalen Netzen und mehr Rechenkapazität. Dies wirft Fragen zu Datenschutz, Kosten und Ressourcenverbrauch auf. Daher wird parallel ein anderer Ansatz verfolgt: die Dezentralisierung von KI-Architekturen nach dem Vorbild des Edge Computing – genannt Edge AI. Das Ziel: Daten nahe am Nutzenden und nicht in der Cloud verarbeiten. Wolfgang Ecker, Distinguished Engineer bei Infineon Technologies und Honorarprofessor der TU München, erklärt Vorteile, Einsatzmöglichkeiten und aktuellen Hürden von Edge AI.

Scale4Edge-Roadshow @ RISC-V Summit 2024

2024/06/18

Im Rahmen des RISC-V Summit 2024 wird die erste Scale4Edge-Roadshow durchgeführt.

Scale4Edge Partner präsentieren Poster, Vorträge und Demonstratoren.

Die Scale4Edge-Partner stehen für persönliche Gespräche bereit. Bitte vereinbaren Sie Termine mit Andreas Vörg <voerg@edacentrum [dot] de>

TETRISC SoC – Quad RISC-V Core Self-Adaptive High-Reliability ASIC from IHP GmbH

2024/05/30

The TETRISC SoC is a demonstrator of results that have been researched and developed in the Scale4Edge project. The video is played on YouTube if you click on the image below.



Success Story: Energy-efficient Keyword-Spotting

2024/03/27



Power-efficient edge AI applications have an enormous market potential. Key factors for a competitive hardware solution in this domain are energy efficiency and low silicon area in relation to computing performance. SpiNNcloud Systems GmbH is closely collaborating with TU Dresden in this domain for transferring cutting-edge research and development on AI hardware accelerators into commercially viable applications.

Within the Scale4Edge project, TU Dresden has developed the SpiNNedge AI accelerator IP for audio keyword spotting applications and beyond. Key focus of SpiNNedge is the reduction of on-chip memory and processing effort by hardware-optimized preprocessing and exploitation of sparsity in neural network based classification. With on-chip memory being one of the core drivers of silicon area and static power, SpiNNedge helps to significantly improve in both factors.

But an AI accelerator alone is worth almost nothing in isolation. Integration in the Scale4Edge ecosystem made SpiNNedge useable and created a commercial potential. The Scale4Edge RISC-V ecosystem core by MINRES runs the application with an offloading of key processing tasks for

keyword spotting to the SpiNNnEdge accelerator. The customizable RISC-V core solution could be perfectly adapted to the needs of the keyword spotting application. Moreover, Scale4Edge made a joint effort to provide software support for hardware extensions of microcontrollers in the widely adopted machine learning compiler Apache TVM. This is highly beneficial for the useability of an accelerator like SpiNNnEdge, as it bridges the gap between hardware IP and high-level software frameworks for machine learning that users employ to develop their edge AI applications.

TU Dresden has successfully implemented a test chip for audio keyword spotting, integrating the SpiNNnEdge accelerator with a customized MINRES RISC-V core into an overall processing chain from microphone input to detected keywords. The chip realizes a hierarchical processing approach, first detecting speech in the microphone input, and then starting keyword classification. Speech detection is performed by the ZEN accelerator module, which won a 1st prize in the BMBF German innovation competition "energy efficient AI systems". The chip has been implemented in GlobalFoundries 22 FDX technology, employing adaptive body biasing (ABB) IP for leakage power reduction, developed and provided by Racyics, a spin-off of TU Dresden.

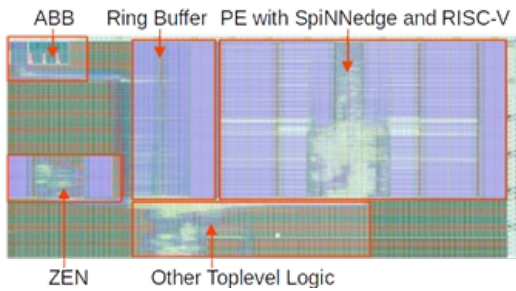


Figure 1: Core Layout of the keyword spotting test chip by TU Dresden. Core part is the processing element (PE) with TU Dresden's SpiNNnEdge accelerator and the MINRES RISC-V core.



Figure 2: Demo board for keyword spotting with the test chip by TU Dresden (white, center)

The processing element with the MINRES RISC-V core and the SpiNNnEdge accelerator performs real-time keyword spotting in 34 uW, achieving a classification accuracy on the widely-used Google speech command dataset of over 95%, which outperforms most of the existing low-power hardware solutions for this use case. A demo board has been implemented to showcase the solution.

The mission of SpiNNcloud Systems is to design energy-efficient AI systems by leveraging practical inspiration from the brain. Its core product is a holistic computing solution ranging from chip design, software development to deployment of full data center servers. Despite our large-scale supercomputers being built leveraging Arm IP, our constant commitment to pursue innovation in an Edge2cloud continuum has led us to evaluate different solutions at different scales. SpiNNcloud is highly interested in commercializing Edge AI solutions, in which the Edge IP built in Scale4Edge stands as a clear and attractive building block. The flexibility of these custom-extended RISC-V cores allow a path to achieve ultra efficient operations at the edge. Furthermore, the Scale4Edge Ecosystem provides a quick start for us into an efficient Edge exploration including know-how from compilers, functional safety, ISA-extensions and chip design methodology among others. Additionally, it introduces us to a wide variety of best practices that can boost development speed and reduce technical risks, all crucial to ensure commercial success. Strategically the consortium enables a fully European ecosystem, especially a German one, with technological independence that increasingly becomes more important to protect and foster companies in the field of computing.

Cont@ct:

TU Dresden | Dr.-Ing. Johannes Partzsch | johannes [dot] partzsch@tu-dresden [dot] de

MINRES | Eyck Jentsch | eyck@minres [dot] com

SpiNNcloud Systems GmbH | Matthias Lohrmann | matthias [dot] lohrmann@spinncloud [dot] com


<https://www.edacentrum.de/scale4edge/>

Further Scale4Edge partners and sub-contractors





Herunterladen:

 Success Story: Energy-efficient Keyword-Spotting

Auf dem Weg zu zukunftsfähigen Spezialprozessoren für die Technologiesouveränität in Deutschland

2024/02/07

Gemeinsame Auftaktveranstaltung der BMBF-ZuSE-Projekte KI-Mobil, KI-Power und Scale4Edge

Donnerstag, 15. Februar, 16:00 Uhr, Fritz-Walter-Stadion Kaiserslautern

Die Partner der BMBF-ZuSE-Projekte KI-Mobil, KI-Power und Scale4Edge freuen sich darauf, sie in Kaiserslautern zu treffen!

Teilnahmeregistrierung: <https://eveeno.com/130451080>

Vorläufige Agenda

Eröffnung

Prof. Wolfgang Kunz, RPTU

16:00

Grußwort

Prof. Arnd Poetzsch-Heffter, Präsident der RPTU

Einordnung in ZuSE

Dr. Eike-Christian Spitzner, VDI/VDE Innovation + Technik GmbH

„Die Rolle von Prozessoren bei den europäischen/deutschen Bemühungen um Halbleitersouveränität - Herausforderungen und Chancen“

Impulsvortrag

Prof. Norbert Wehn, RPTU

16:20

Podiumsdiskussion

Moderator:

Prof. Wolfgang Nebel, edacentrum

Teilnehmer:

Mario Brandenburg, parlamentarischer Staatssekretär BMBF

Prof. Armin Dietz, Technische Hochschule Nürnberg, Projektkoordinator KI-Power

Prof. Wolfgang Ecker, Infineon, Projektkoordinator Scale4Edge

Eyck Jentzsch, MINRES Technologies GmbH

Dr. Hans-Jörg Vögel, BMW Group, Projektkoordinator KI-Mobil

17:30

Pause

Projektvorstellungen

17:45

Prof. Wolfgang Ecker - **Scale4Edge**

Dr. Hans-Jörg Vögel - **KI-Mobil**

Prof. Armin Dietz - **KI-Power**

18:45

Poster und Demos aus den ZuSE Projekten mit Stehempfang

19:45

Gemeinsames Abendessen im Fritz Walter Stadion

22:00

Ende der Veranstaltung

Success Story: Multi-Compiler Support

Introduction

The Scale4Edge project aims to create an efficient RISC-V ecosystem for edge application optimization. It focuses on developing a versatile platform to provide cost-effective, application-specific edge devices and services for various market segments. This is achieved through fine-grained adaptation of generic components, covering CPU instructions, user/application software, and operating system/firmware levels. The ecosystem is highly scalable and customizable to individual applications, supporting various hardware architectures and non-functional properties like energy efficiency, fault tolerance, reliability, safety, and security.

The Scale4Edge ecosystem currently contains three different compilers, addressing different use-cases of the ecosystem:

CompCert is a formally verified compiler for safety relevant application of high assurance levels

X-LLVM is an extendible compiler for custom instructions based on Clang/LLVM

A custom configurable GCC compiler

CompCert - a Verified Compiler

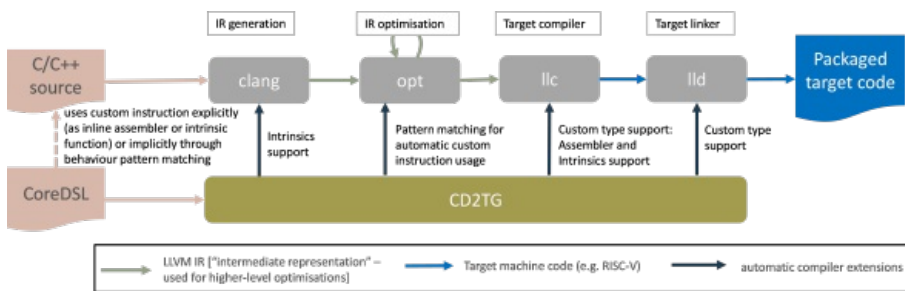
CompCert is a compiler for the C programming language (<https://www.absint.com/compcert/>). It accepts most of the ISO-C 99 language, with some minor exceptions and a few useful extensions. Its intended use is the compilation of life-critical and mission-critical software written in C and meeting high levels of assurance.

What sets CompCert apart from any other production compiler is that it is formally verified, using machine-assisted mathematical proofs, to be exempt from miscompilation issues. In other words, the executable code it produces is proved to behave exactly as specified by the semantics of the source C program. This level of confidence in the correctness of the compilation process contributes to meeting the highest levels of software assurance. Using the CompCert C compiler is a natural complement to applying formal verification techniques (static analysis, program proof, model checking) at the source code level: the correctness proof of CompCert guarantees that all safety properties verified on the source code automatically hold as well for the generated executable.

In 2022, the Association for Computing Machinery, ACM, presented the CompCert development team with the prestigious ACM Software System Award (<https://www.absint.com/releases/220511.htm>) and the ACM SIGPLAN Programming Languages Software Award (<https://www.sigplan.org/Awards/Software/>).

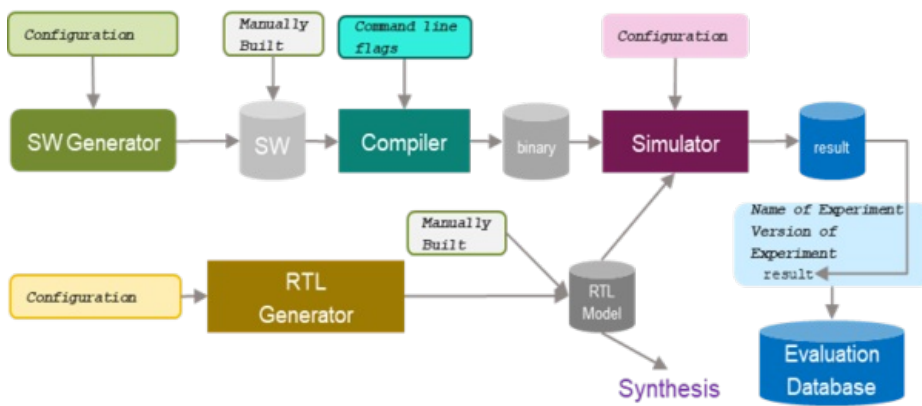
During the Scale4Edge project, CompCert was given a backend for RISC-V. CompCert for RISC-V generates code for the base instruction sets RV32I and RV64I with standard extensions M (Integer Multiplication and Division), F (Single-Precision Floating-Point), and D (Double-Precision Floating-Point). CompCert for RISC-V has been combined with picolibc (<https://github.com/picolibc/picolibc>), a C library designed for embedded 32- and 64-bit microcontrollers with small memory. After some changes proposed by the CompCert developers (mostly to avoid GCC/Clang specific language extensions), the picolibc source code can now be compiled with CompCert.

X-LLVM - an extendible Compiler for custom instructions



Existing toolchains with RISC-V support cannot yet be flexibly extended to quickly support ISA extensions. In addition, RISC-V extensions in the Scale4Edge project are described using a separate language called CoreDSL. An automated translation of these CoreDSL definitions into corresponding toolchain extensions, as far as possible, is therefore not only desirable for reasons of flexibility, but also almost inevitable for reasons of consistency. For this reason, the described hardware (processor core) and a virtual platform and a compiler with support of custom defined instructions can be generated from CoreDSL. We have successfully automated the modification of Clang/LLVM to support custom instructions throughout the whole software toolchain (compiler, linker debugger), as depicted in the figure. Based on a CoreDSL description of the custom instructions, the extendible compiler toolchain (called X-LLVM) implements a compiler generator (called CD2TG) to provide the defined custom instructions as assembler code or intrinsic function for explicit usage. If possible, custom instructions are also implicitly used by the compiler (e.g., MAC (Multiply Accumulate) or Zero Overhead Loop). If this were done based on a manual translation, there would be a potential for an incorrect or at least inconsistent translation at each individual step. By automating this process, these errors can be avoided. Another benefit of automation is the speed with which the different artifacts can be regenerated. For example, if a new CoreDSL description is created or an existing description is revised, the corresponding tool chains can be triggered directly to generate the different artifacts and thus be able to test the newly defined instructions directly and revise them further if necessary. The current open source release of X-LLVM is available at <https://github.com/DLR-SE/extendible-compiler>.

GCC - the power of optimization



gcc supports plenty of optimization opportunities, most can be controlled by over 120 command line flags.

The Infineon RISC-V RTL generator supports also plenty of configurations (superset of CoreDSL) to generate a huge number of RISC-V variants differing in function, performance in clock cycles and performance in clock period. From the perspective of compiler performance analysis, the number of clock cycles is relevant. The maximum clock period is provided by the synthesis – or better R2G – flow. Performance evaluation is done by RTL simulation but can be moved to FPGAs and Emulation as RISC-V or SoC features (Timers) are used to measure time. For evaluation, classical benchmarks but also generated software has been applied. Of course, the framework is also used to run regression over the complete generation chain and validate the compiler result for a specific code and application.

Results worth mentioning are a substantial impact of compiler optimization, more performance when comparing to other non-RISC-V cores but also a non-neglectable memory footprint overhead.

Cont@ct:

CompCert | Reinhold Heckmann | AbsInt Angewandte Informatik GmbH | heckmann@absint [dot] com

X-LLVM | Kim Grüttner | German Aerospace Center (DLR) | kim [dot] gruettnr@dlr [dot] de

GCC | Wolfgang Ecker | Infineon Technologies AG | wolfgang [dot] ecker@infineon [dot] com (wolfgang.ecker@infineon.com)

Further Scale4Edge partners and sub-contractors



Herunterladen:

Success Story: Multi-Compiler Support

Success Story: The Scale4Edge Hardware Verification and Validation Ecosystem for RISC-V Platforms

2023/05/05



For a commercially successful development of edge devices for a wide range of applications, a whole ecosystem of tools and hardware components is mandatory. Ensuring the safe, secure, and reliable design and operation of these devices is an indispensable, but challenging requirement for many of these applications. The Scale4Edge project aims to address this challenge by enabling a comprehensive RISC-V-based ecosystem that helps ensure the correct and secure functioning throughout the whole design process and lifetime of an edge system: The ecosystem encompasses tools for functional verification—both using virtual prototypes in an early stage of development and using formal methods—, security verification, software-based self-test (SBST), and system visualization and debugging. With the development of this hardware verification and validation ecosystem, the Scale4Edge project aims to create optimized and reliable edge devices that meet the high demands of modern industries.

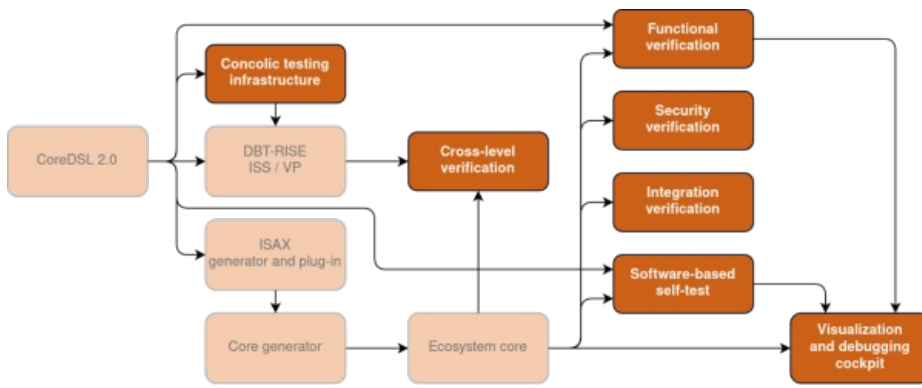


Figure 1: The Scale4Edge Hardware Verification Ecosystem

The Scale4Edge project has made significant strides in hardware verification and validation for RISC-V platforms, with multiple partners contributing to its success. The components of the hardware verification ecosystem are shown in Figure 1. The **University of Bremen** developed a comprehensive cross-level verification approach and a concolic testing infrastructure. **Siemens EDA**—an associated partner—created new functional verification tools for processor verification while the **University of Kaiserslautern-Landau (RPTU)** focused on formal security analysis with Unique Program Execution Checking (UPEC). The **FZI Research Center for Information Technology (FZI)** devised methods for specifying, generating, and verifying hardware properties in System-on-Chip (SoC) platforms, with the **University of Freiburg** and **Concept Engineering** contributing to in-field monitoring and testing (SBST) as well as debugging and visualization tooling. The collaborative efforts of the project partners around the Scale4Edge ecosystem core, developed by **MINRES** according to the ISO 26262 safety standard, resulted in a comprehensive hardware verification and validation flow with award winning methods and industry-proven EDA tools.

Herunterladen:

 [Success Story: The Scale4Edge Hardware Verification and Validation Ecosystem for RISC-V Platforms](#)

Das Projekt Scale4Edge wird unter den Förderkennzeichen 16ME0122K-140, 16ME0465, 16ME0900, 16ME0901 im Förderprogramm ZuSE durch das deutsche Bundesministerium für Bildung und Forschung (BMBF) gefördert.

Quell-URL: <https://www.edacentrum.de/scale4edge/node>