Application BASTRI
Fiches Equipes

TARAN (SR0905HR)

Architectures matérielles spécialisées pour l'ère post loi-de-Moore
CAIRN (SR0211OR) TARAN

Statut: Décision signée

Responsable : Olivier Sentieys

Mots-clés de "A - Thèmes de recherche en Sciences du numérique - 2024" : *Aucun mot-clé.*

Mots-clés de "B - Autres sciences et domaines d'application - 2024" : *Aucun mot-clé.*

Domaine : Algorithmique, programmation, logiciels et architectures
Thème : Architecture, langages et compilation

Période : 01/05/2021 -> 31/12/2025
Dates d'évaluation :

Etablissement(s) de rattachement : ENS RENNES, U. RENNES
Laboratoire(s) partenaire(s) : IRISA (UMR6074)

CRI : Centre Inria de l'Université de Rennes
Localisation : Centre Inria de l'Université de Rennes
Code structure Inria : 031133-0
CRI : Centre Inria de l'Université de Rennes
Localisation : École Nationale Supérieure des Sciences Appliquées et de Technologie
Code structure Inria : 031133-0

Numéro RNSR : 202124094C
N° de structure Inria: SR0905HR

Présentation

Energy efficiency has now become one of the main requirements for virtually all computing platforms. Computer architects are however facing new challenges for the next couple of decades, with the most prominent one being the end of CMOS scaling (Moore's law). Our belief is that the key to sustaining improvements in performance (both speed and energy) is ***domain-specific computing*** where all layers of computing, from languages and compilers to runtime and circuit design, must be carefully tailored to specific contexts.

In this new age, *the processor will be augmented by a bunch of hardware accelerators meant to perform specific tasks in a more efficient way.* ***Our research team focuses on designing accelerators that can prove energy-efficient and fault-tolerant****.*

Our main objective is to promote Domain-Specific Computing that requires the participation of the algorithm designer, the compiler writer, the microarchitect, and the chip designer. This cannot happen through individually working on the different layers discussed above. The unique composition of TARAN allows us to benefit from our expertise spanning multiple layers in the design stack.

Axes de recherche

Our research directions may be categorized into the following four contexts:

- **Accelerators**: Hardware accelerators will become

Contact

- **Responsable :** Olivier Sentieys
- **Tél :** 02.99.84.72.16
- **Secrétariat Tél :** 02.99.84.22.09

En savoir plus

- Site de l'équipe
- Site sur inria.fr
- Site du responsable
- Derniers Rapports d'Activité : 2021 , 2022 , 2023

Documents sur la structure

- Intranet
- Privés

Décisions

- 14815 (07/05/2021) : création
- 15209 (16/12/2021) : prolongation
- 16256 (03/07/2023) : modification

Localisation

- **Adresse postale :** Centre Inria de l'Université de Rennes 263, avenue du Général Leclerc Campus universitaire de Beaulieu 35042 Rennes Cedex France
- **Coordonnées GPS :** 48.116, -1.64

more and more common, and we must develop techniques to make accelerator design more accessible. The important challenge is raising the level of abstrac- tion without sacrificing performance. However, higher level of abstraction coupled with domain-specific knowledge is also a great opportunity to widen the scope of accelerators.

- **Accurate Computing**: Most computing today is performed with significant over-provisioning of output quality or precision. Carefully selecting the various parameters, ranging from algorithms to arithmetic, to compute with just the right quality is necessary for further efficiency. Such fine tuning of elements affecting application quality is extremely time consuming and requires domain knowledge to be fully utilized.
- **Resilient Computing**: As we approach the limit of CMOS scaling, it becomes increasingly unlikely for a computing device to be fully functional due to various sources of faults. Thus, techniques to maintain efficiency in the presence of faults will be important. Generally applicable techniques, such as replication, come with significant overheads. Developing techniques tailored to each application will be necessary for computing contexts where reliability is critical.
- **Embracing Emerging Technologies**: Certain computing platforms, such as ultra-low power devices and embedded many-cores, have specific design constraints that make traditional components unfit. However, emerging technologies such as Non-Volatile Memory and Silicon Photonics cannot simply be used as a substitute. Effectively integrating more recent technologies is an important challenge for these specialized computing platforms.

The common keyword across all directions is domain-specific. Specialization is necessary for addressing various challenges including productivity, efficiency, reliability, and scalability in the next generation of computing plat- forms. Our main objective is defined by the need to jointly work on multiple layers of the design stack to be truly domain-specific. Another common challenge for the entire team is design space exploration, which has been and will continue to be an essential process for HW design. We can only expect the design space to keep expanding, and we must persist on developing techniques to efficiently navigate through the design space.

Relations industrielles et internationales

Industrial Collaborations: The team collaborates (direct funding or collaborative projects) with large companies (Safran, Nokia Bell Labs, Orange, Kalray, Huawei, Thales, STMicroelectronics), various SMEs, or Institute like DGA (Rennes), CEA (Saclay, Grenoble), Onera (Toulouse) and CNES (Toulouse). At the international level, we currently collaborate with companies like Raytheon Technologies, Xilinx R&D, or Mentor Graphics.

International Collaborations: The team has formal or informal active international collaborations including University of British Columbia at Vancouver (low-precision AI accelerators), IIT Goa (AI accelerators), Colorado State University (HLS, polyhedral model), University College Cork (fault-tolerance, accelerators), University of Oxford (arithmetic), University of Massachusetts at Ahmerst (accelerators, FPGA), UCLA (arithmetic), EPFL (approximate computing), IMEC (fault tolerance), Université du Québec à Trois-Rivières (accelerators), Concordia University (optical NoC), University of Auckland (fault tolerance), University of Patras (heterogeneity), Southeast University China (heterogeneity). Formal projects include:

- IntelliVIS: Design Automation for Intelligent Vision Hardware in Cyber Physical Systems, INRIA Associated Team (2019-2021) with IIT Goa, India, Prof. Sharad Sinha on the design and development of artificial intelligence based embedded vision architectures for cyber physical systems. Key Investigators: S. Filip, O. Sentieys.
- DeLeES, Energy-efficient Deep Learning Systems for Low-cost Embedded Systems, University of British Columbia, Vancouver, Canada, Prof. Guy Lemieux, France-Canada Research Fund (FCRF). Key Investigators: S. Filip, O. Sentieys.
- LRS, Loop unRolling Stones: compiling in the polyhedral model, Colorado State University, US, Prof. Sanjay Rajopadhye. This collaboration led to two international jointly supervised PhDs (or 'cotutelles' in French) that started in Oct. 2019, one in France (C. Ferry) and one in US (L. Narmour). Key Investigators: S. Derrien, T. Yuki.
- Taran is involved in the Cooperation program PRoSFER between the

French Ecoles normales supérieures and East China Normal University in Shanghai (the coordinator of this program for Computer Science is P. Quinton). This program aims at preparing ECNU students to research, and to allow them to do a PhD in one of the ENS. Key Investigators: E. Casseau, S. Derrien, A. Kritikakou, P. Quinton.