Awards
ABOUT NSF
Search Awards
Recent Awards
Presidential and Honorary Awards
About Awards
How to Manage Your Award
Grant General Conditions
Cooperative Agreement Conditions
Special Conditions
Federal Demonstration Partnership
Policy Office Website

Award Abstract # 1551044

InTrans: A virtualized SoC platform architecture for mini autonomous drones

NSF Org:	IIS Div Of Information & Intelligent Systems
Awardee:	PRESIDENT AND FELLOWS OF HARVARD COLLEGE
Initial Amendment Date:	September 1, 2015
Latest Amendment Date:	September 1, 2015
Award Number:	1551044
Award Instrument:	Standard Grant
Program Manager:	Kenneth Whang kwhang@nsf.gov (703)292-5149 IIS Div Of Information & Intelligent Systems CSE Direct For Computer & Info Scie & Enginr
Start Date:	September 1, 2015
End Date:	August 31, 2018 (Estimated)
Total Intended Award Amount:	$100,000.00
Total Awarded Amount to Date:	$100,000.00
Funds Obligated to Date:	FY 2015 = $100,000.00
History of Investigator:	Gu-Yeon Wei (Principal Investigator) guyeon@eecs.harvard.edu Robert Wood (Co-Principal Investigator) David Brooks (Co-Principal Investigator)
Awardee Sponsored Research Office:	Harvard University 1033 MASSACHUSETTS AVE Cambridge MA US 02138-5369 (617)495-5501
Sponsor Congressional District:	05
Primary Place of Performance:	Harvard University 33 Oxford Street Cambridge MA US 02138-3846
Primary Place of Performance Congressional District:	05
Unique Entity Identifier (UEI):	
Parent UEI:	
NSF Program(s):	Information Technology Researc
Primary Program Source:	040100 NSF RESEARCH & RELATED ACTIVIT

Program Reference Code(s):	1640
Program Element Code(s):	1640
Award Agency Code:	4900
Fund Agency Code:	4900
Assistance Listing Number(s):	47.070

ABSTRACT

Building on recent results from the RoboBees project, this InTrans project explores the design of new computing modalities that enable scalable performance and energy efficiency across a wide range of energy- and weight-constrained applications, from miniature autonomous drones (e.g., RoboBees) to larger small-scale flying robots. The pursuit of energy-efficient compute performance remains an important objective, but computing modalities continue to evolve over time. Demand for mobility and portability led to laptops and then to smaller smart phones that pack the functionality and performance of yesterday's computer into one's hand. Technology is now at the cusp of another paradigm shift with the proliferation of "internet of things" (IoT) devices and small-scale drones. Despite the perception that CMOS technology scaling is reaching a plateau, these applications require ever higher levels of energy-efficient computing with ever shrinking form factors, demanding a new era of circuits, architectures, and systems.

One extreme example is the RoboBees project at Harvard, which set out to create a colony of autonomous robotic bees: <500mg, insect-sized, flapping-wing robots with sufficient sensing, computing, and actuation to achieve autonomous flight. This NSF-funded project designed, implemented, and tested a "brain chip" that met severe weight and real-time computation requirements. It comprises a general-purpose core with a collection of dedicated hardware accelerators for the specific sensing, image processing, and control algorithms needed for autonomous flight. While this chip significantly improves power and performance compared to those only using general-purpose cores, a platform that combines multiple one-off accelerators is inflexible to subsequent changes in computing needs. Future computing systems for autonomous drones must readily accommodate and adapt to changing workloads and application needs. A promising research direction is to develop a virtualized platform architecture constructed out of composable accelerators that can provide scalability via virtualization. In other words, next-generation computing systems will comprise computing elements flexible enough to operate across a range of applications while retaining the benefits of specialization. Via close collaboration with researchers at Intel, this InTrans project will (1) identify interesting applications to target within the broad range of embedded systems that combine sensing, computing, and actuation and (2) develop and demonstrate new computer architectures that readily scale performance and energy consumption across a broad range of platforms and system constraints.

PUBLICATIONS PRODUCED AS A RESULT OF THIS RESEARCH

Note: *When clicking on a Digital Object Identifier (DOI) number, you will be taken to an external site maintained by the publisher. Some full text articles may not yet be available without a charge during the embargo (administrative interval).*

Some links on this page may take you to non-federal websites. Their policies may differ from this site.

Brandon Reagen, Paul Whatmough, Robert Adolf, Saketh Rama, Hyunkwang Lee, Sae Kyu Lee, José Miguel Hernández-Lobato, Gu-Yeon Wei, David Brooks "Minerva: Enabling Low-Power, Highly-Accurate Deep Neural Network Accelerators" *International Symposium on Computer Architecture (ISCA)* , 2016

Yakun Sophia Shao, Sam Likun Xi, Vijayalakshmi Srinivasan, Gu-Yeon Wei and David Brooks "Co-Designing Accelerators and SoC Interfaces using gem5-Aladdin" *International Symposium on Microarchitecture (MICRO)* , 2016

PROJECT OUTCOMES REPORT

Disclaimer
This Project Outcomes Report for the General Public is displayed verbatim as submitted by the Principal Investigator (PI) for this award. Any opinions, findings, and conclusions or recommendations expressed in this Report are those of the PI and do not necessarily reflect the views of the National Science Foundation; NSF has not approved or endorsed its content.

The pursuit of energy-efficient compute performance remains an important objective, but computing modalities continues to evolve over time. Demand for mobility and portability led to laptops and then to smaller smart phones that pack the functionality and performance of yesterday’s computer into one’s

hand. Technology is now in the midst of another paradigm shift with the proliferation of "internet of things" (IoT) devices and practical implementation and application of deep learning algorithms with corresponding hardware. Despite a perceived plateauing of CMOS technology scaling, these applications require ever higher levels of energy-efficient computing with ever shrinking form factors; demanding a new era of circuits, architectures, and systems. One extreme example is the RoboBees project at Harvard, which had set out to create a colony of autonomous robotic bees – <500mg, insect-sized, flapping-wing robots with sufficient sensing, computing, and actuation to achieve autonomous flight. The NSF-funded RoboBees project designed, implemented, and tested a "BrainSoC" integrated circuit (IC) that met severe weight and real-time computation requirements. It comprised a general-purpose core with a collection of dedicated hardware accelerators for the specific sensing, image processing, and control algorithms needed for autonomous flight. While the BrainSoC significantly improved power and performance compared to using general-purpose cores, a platform that simply combines multiple one-off accelerators together is inflexible to subsequent changes in computing needs.

Intellectual merit This proposal sought to leverage the insights learned and design tools developed while building the BrainSoC for the Harvard RoboBees (previously funded by the CISE Expeditions in Computing program). Efforts focused on developing flexible yet efficient computing hardware for a broader class of energy- and weight-constrained computing technologies like cognitive IoT. Research led to development of design tools and architectural techniques in order to demonstrate systems comprising scalable composable accelerators for a broad range of applications – from extremely small-scale micro air vehicles (MAVs) to other static and dynamic monitoring applications to various embedded systems that combine sensing, computing, and actuation (e.g., smart medical devices). There are multiple intersecting layers to this work; from (i) understanding application-dependent constraints and requirements to (ii) exploring algorithmic options and power/performance tradeoffs to (iii) evaluating and accommodating system-level effects of combining and sharing specialized compute modules to (iv) finding the appropriate level(s) of granularity for acceleration.

Over the course of the project, there were two main activities and accomplishments: Development of IoT applications that can run on extreme energy-constrained hardware accelerators that implement fully-connected deep neural networks (FC-DNN). Porting, fabrication, and testing of a 16nm FC-DNN accelerator that fits within a thumb-drive formfactor to support a variety of demonstrations and capable of always-on operation owing to its high energy efficiency.

Broader impacts The original RoboBees project was funded by the National Science Foundation (NSF) Expeditions in Computing program. As a mechanism to extend the impact of Expeditions projects to industry, the NSF launched the "Innovation Transition" (InTrans) program. Via close collaboration with researchers at Intel, this InTrans project sought to (1) identify interesting applications to target within the broad range of embedded systems that combine sensing, computing, and actuation and (2) maximize opportunities to transfer research outcomes to the industry partner. Research results were presented at international conferences and directly shared with industry partners via lab visits and periodic phone and in-person meetings.

Last Modified: 11/29/2018
Modified by: Gu-Yeon Wei

RESEARCH AREAS

National Science Foundation, 2415 Eisenhower Avenue, Alexandria, Virginia 22314, USA

Tel: (703) 292-5111, FIRS: (800) 877-8339 | TDD: (800) 281-8749