可重組式系統設計與應用
Reconfigurable System Design and Applications

Chun-Hsian Huang (黃駿賢)

huangch@nttu.edu.tw
Dept. of Computer Science and Information Engineering
National Taitung University

2019/11/13@NPTU

About Me

Chun-Hsian Huang (黃駿賢)

- https://sites.google.com/site/chunhsianhuang/home
- 現職: 國立臺東大學資訊工程系副教授(2016/08~至今)
- 學歷: 國立中正大學資訊工程系博士(逕讀博班)
- 經歷:
  - 助理教授, 國立臺東大學資訊工程系
  - 助理研究員, 國家中山科學研究院第五研究所
  - 博士後研究員, Intel-臺大創新研究中心, 國立臺灣大學

XILINX Applications Products Developers Support About (0)

Xilinx - Adaptable. Intelligent. > News > Press > 2017 Press Releases > Xilinx and IBM First to Double Interconnect Performance for Accelerated Cloud Computing with New PCI Express Standard

Xilinx and IBM First to Double Interconnect Performance for Accelerated Cloud Computing with New PCI Express Standard

Technology milestone between FPGA-based accelerator and POWER CPUs to boost acceleration of data center applications

May 16, 2017

SAN JOSE, Calif., May 16, 2017 /PRNewswire/ -- Xilinx, Inc. (XLNX) today announced an achievement in PCI Express® Gen4 capability. Together with IBM, the two companies are first to double interconnect performance between an accelerator and CPU through the use of PCI Express Gen4 compared to the existing widely-deployed PCI Express Gen3 standard. Gen4 doubles the bandwidth between CPUs and accelerators to 16 Gbps per lane, thereby accelerating performance in demanding data center applications such as artificial intelligence and data analytics.

Since the introduction of PCI Express in 2003, Xilinx has been a leader in PCI™ interconnect-based solutions, offering PCI Express compliance across its All Programmable FPGA families. Today IBM and Xilinx have achieved Gen4 interoperability between Xilinx® 16nm UltraScale+™ devices and IBM POWER9 processors, demonstrating the first-ever PCIe Gen4 capability in a programmable device.

"This leadership in PCI Express is another reason that POWER architecture is being deployed in modern data centers," said Bradley McCredie, vice president and fellow at IBM. "IBM is excited to leverage the underlying performance of PCIe Express Gen4 for CAPI 2.0 which eases the programming experience for application developers."

"We believe in open standards," said Ivo Bolsens, CTO at Xilinx. "It's gratifying to see this milestone between our companies that will alleviate significant performance bottlenecks in accelerated computing, particularly for data center computing."

Source: https://www.xilinx.com/news/press/2017/xilinx-and-ibm-first-to-double-interconnect-performance-for-accelerated-cloud-computing-with-new-pci-express-standard.html

物聯網 / 系統開發

Intel FPGA助攻Microsoft Azure的人工智慧發展

鄭斐文 2018-05-24 讚0 分享 轉寄 列印

Microsoft於Microsoft Build大會上，發表Azure機器學習硬體加速模型以供預覽，搭載Project Brainwave，與Microsoft Azure機器學習SDK整合。可利用Azure的大規模佈建Intel FPGA(現場可程式邏輯閘陣列)技術，使其模型獲得人工智慧(AI)推論效能。

英特爾可編程解決方案事業群副總裁兼總經理Daniel McNamara表示，英特爾是整體技術供應商，透過與Microsoft的密切合作實現AI。AI可應用於從訓練到推論、從語言識別到影像分析等各種使用情境，而Intel擁有豐富的硬體、軟體和工具，可實現這種全面的工作負載。

資料科學家和開發人員可輕鬆將深度神經網路(Deep Neural Networks；DNN)用於各種即時工作負載，包括製造、零售和醫療領域的工作負載，橫跨全球最大的加速雲端。並可訓練模型，然後在Project Brainwave上佈建，無論是在雲端或網路邊界皆可運用FPGA。

Project Brainwave釋出採用Intel FPGA的可編程硬體以提供即時AI，開創AI的未來。以FPGA為基礎的架構具經濟性和高效率，且傳輸量極高，能夠執行ResNet 50而不需要批次處理。AI客戶不必在高效能或低成本之間選擇。

利用適用於Python的Azure機器學習SDK，客戶能夠用資料來重新訓練以ResNet 50為基礎的模型，將影像識別任務專門化。就即時AI工作負載而言，運算強度需要專用的硬體加速器。Intel FPGA可讓Azure針對任務準確配置硬體，發揮最高效能。

Source:
https://www.digitimes.com.tw/iot/article.asp?cat=130&cat1=40&id=0000532124_V8O0QH3E3VCTKN3KW6AAR

intel Newsroom Top News Sections News By Category All News Search Newsroc

News Byte

July 3, 2018

Contact Intel PR

INTEL AI AT BAIDU CREATE: AI CAMERA, FPGA-BASED ACCELERATION AND XEON SCALABLE OPTIMIZATIONS FOR DEEP LEARNING

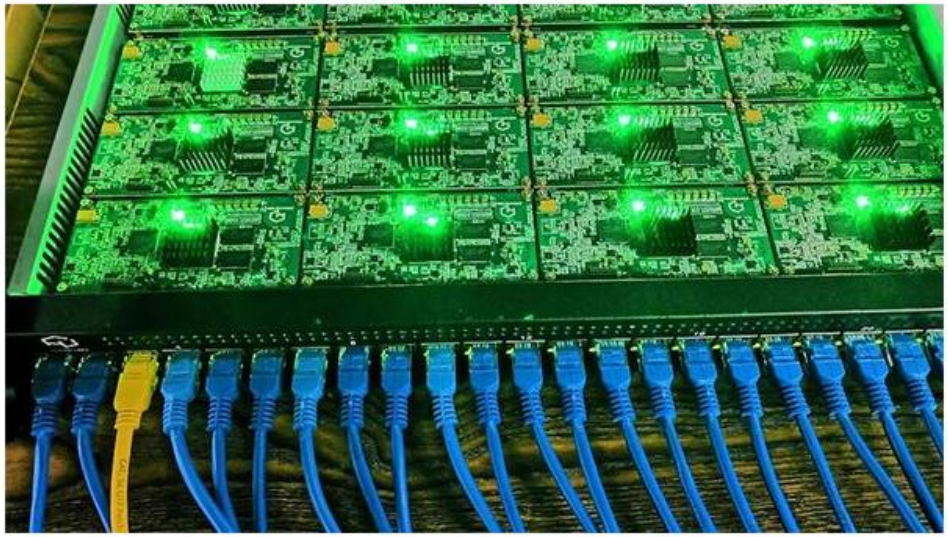
Intel Vice President Gadi Singer announces a series of collaborations with Baidu on artificial intelligence at Baidu Create on Wednesday, July 4, 2018, in Beijing. (Credit: Intel Corporation)
» Download full-size image

What's New: Today at Baidu* Create in Beijing, Intel Vice President Gadi Singer shared a series of collaborations with Baidu on artificial intelligence (AI), including powering Baidu's *Xeye** a new AI retail camera with Intel® Movidius™ vision processing units (VPUs); highlighting Baidu's plans to offer workload acceleration as a service using Intel® FPGAs; and optimizing PaddlePaddle*, Baidu's deep learning framework for Intel® Xeon® Scalable processors.

Source: https://newsroom.intel.com/news/intel-ai-baidu-create-ai-camera-fpga-based-acceleration-xeon-scalable-optimizations-deep-learning/#gs.LRDdaTJg

高出GPU服务器30倍，前NASA项目的超算大拿要做FPGA性能的颠覆者

由 judyzhong 于 星期四, 11/22/2018 - 11:06 发表



作者：四月，机器之能

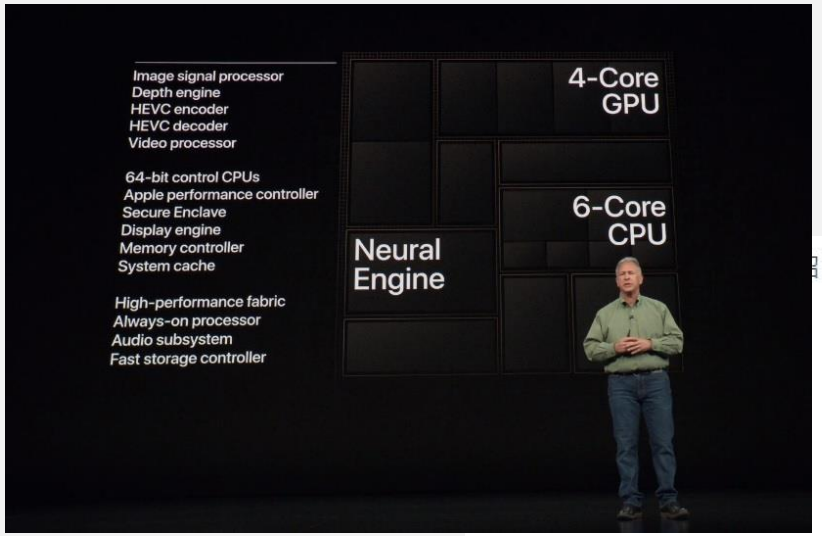
他们自诩为"搭积木的人"——"FPGA是乐高积木，用最少的积木搭建出整个高楼大厦——这就是我们的能力。"

在大多数芯片从业者看来，因为批量开发难度大和成本过高，FPGA一直作为"技术验证者"的配角存在。但雪湖团队试图打破这一观念，他们希望凭借多年的开发经验积累和自研开发工具将FPGA芯片推向人工智能舞台的中央。

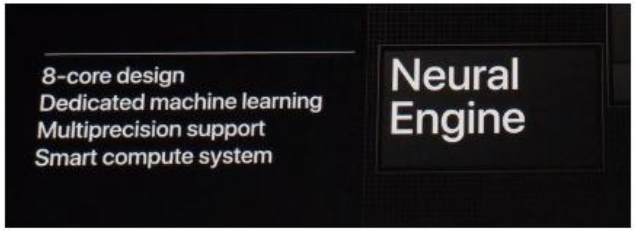
"我们不是一家卖Know-how、卖算法的公司，甚至可以说我们不是一家AI芯片公司。我们把自己定义成一家异构计算公司"——这是张强为雪湖写下的注脚。

6月底的一个夜晚，北京颐和园被灯光装点得美轮美奂。这座千年园林里正在上演一场充满科技感的发布会，美

Source: http://xilinx.eetrend.com/d6-xilinx/news/2018-11/13915.html



，在於那塊 Neural Engine。



這是一塊 FPGA（Field-Programmable Gate Array 現場可程式化邏輯閘陣列），一種自定義能力強、適合客製化計算需求，在機器學習領域廣泛採用的積體電路。

和 A11 雙核相比，A12 的 Neural Engine 核數暴增到 8 顆，是為了：

1. 在本地處理神經網路任務，比如圖像辨識、物體追蹤、空間位置感應。

現場展示的籃球即時數據分析技術 Homecourt Shotscience，可即時監測球員的位置、姿態，球出手時的高度、角度和時機等諸多訊息。

2. A12 Neural Engine 第二個重要任務是為 CPU 和 GPU，以及兩個單元的不同核心之間分配任務，更高效處理任務，提高性能、降低功耗。

第二代 Neural Engine 的存在，不僅體現蘋果晶片設計的強大，更證明一件事：為了在裝置實現甚至創造新的用戶需求，讓 iPhone 比前一代更盡善盡美，蘋果願意深入晶片的層面。

Source: https://technews.tw/2018/09/14/apple-chips-and-ai-are-really-amazing/

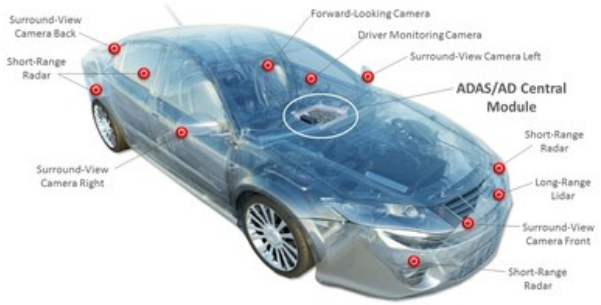
Daimler AG Selects Xilinx to Drive Artificial Intelligence-Based Automotive Applications

Xilinx and Daimler to Develop Ultra-Efficient AI Solutions for Future Mercedes-Benz Models

Jun 26, 2018

SAN JOSE, Calif., June 26, 2018 /PRNewswire/ -- Xilinx, Inc. (NASDAQ: XLNX) and Daimler AG today announced that the two companies are collaborating on an in-car system using Xilinx technology for artificial intelligence (AI) processing in automotive applications. Powered by a Xilinx automotive platform consisting of system-on-a-chip (SoC) devices and AI acceleration software, the scalable solution will deliver high performance, low latency and best power efficiency for embedded AI in automotive applications today.

Media Kit



Xilinx has a strong pedigree in automotive. For more than 12 years the company has shipped over 40 million cumulative automotive units to automakers and Tier 1 suppliers.

"We are accelerating our product development using AI technology by engaging our global development centers with Xilinx experts," said Georges Massing, director user interaction & software, Daimler AG. "Through this strategic collaboration, Xilinx is providing technology that will

Source: https://www.xilinx.com/news/press/2018/daimler-ag-selects-xilinx-to-drive-artificial-intelligence-based-automotive-applications.html

Navy orders diminishing manufacturing sources (DMS) FPGAs to keep F-35s and other military aircraft flying

PATUXENT RIVER NAS, Md. – The U.S. Navy is buying more than $40 million worth of obsolescent field-programmable gate arrays (FPGAs) to ensure the long-term operation and maintenance of the F-35 joint strike fighter and other military aircraft for the U.S. Air Force, Navy, Marine Corps., and allied military forces.

Author — John Keller

Nov 20th, 2018

Navy orders diminishing manufacturing sources (DMS) FPGAs to keep F-35s and other military aircraft flying

PATUXENT RIVER NAS, Md. – The U.S. Navy is buying more than $40 million worth of obsolescent field-programmable gate arrays (FPGAs) to ensure the long-term operation and maintenance of the F-35 joint strike fighter and other military aircraft for the U.S. Air Force, Navy, Marine Corps., and allied military forces. Officials of the Naval Air Systems Command at Patuxent River Naval Air Station,

Source: https://www.militaryaerospace.com/computers/article/16726578/navy-orders-diminishing-manufacturing-sources-dms-fpgas-to-keep-f35s-and-other-military-aircraft-flying

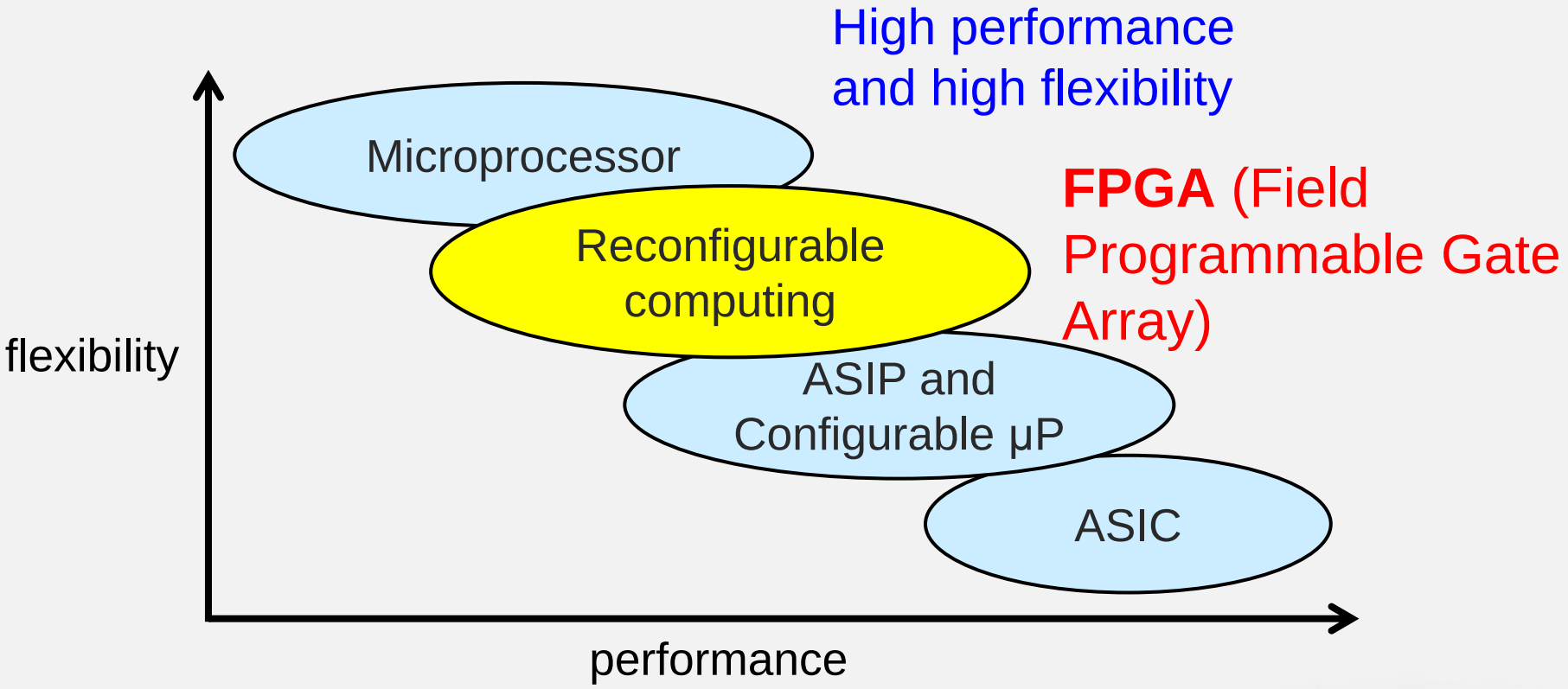
Why Reconfigurable Computing?

- Memory access and I/O bottleneck in microprocessor architecture performance
- Increasingly higher NRE (non-recurring engineering) costs and longer design time for nanometer ASIC development
- Applications are getting complex and dynamic, ASICs, ASIPs, and configurable processors are inflexible in coping with dynamic changes

Source: 教育部ATP課程-PAL聯盟「可重組式晶片系統雛形設計與應用」

Why reconfigurable computing?

- Applications are getting complex and dynamic



Source: 教育部ATP課程-PAL聯盟「可重組式晶片系統雛形設計與應用」

Benefits of Reconfigurable Computing

- Flexible enough to cope with dynamic changes
- Application specific design increases resource utilization and performance
- Spatial parallelization is superior to temporal parallelization (μp) in accelerating application performance
  - 10X ~ 100X speedup of data-intensive applications

Source: 教育部ATP課程-PAL聯盟「可重組式晶片系統雛形設計與應用」

What is Reconfigurable Computing?

- A system or application functions can be changed by configuring a fixed set of logic resources through memory settings
  - "General purpose custom hardware"

    - Seth Copen Goldstein
  - "On the fly ASIC"

    - Fadi Kurdahi
  - "Logic-on-demand"

    - Virtual Computer Corporation (VCC)

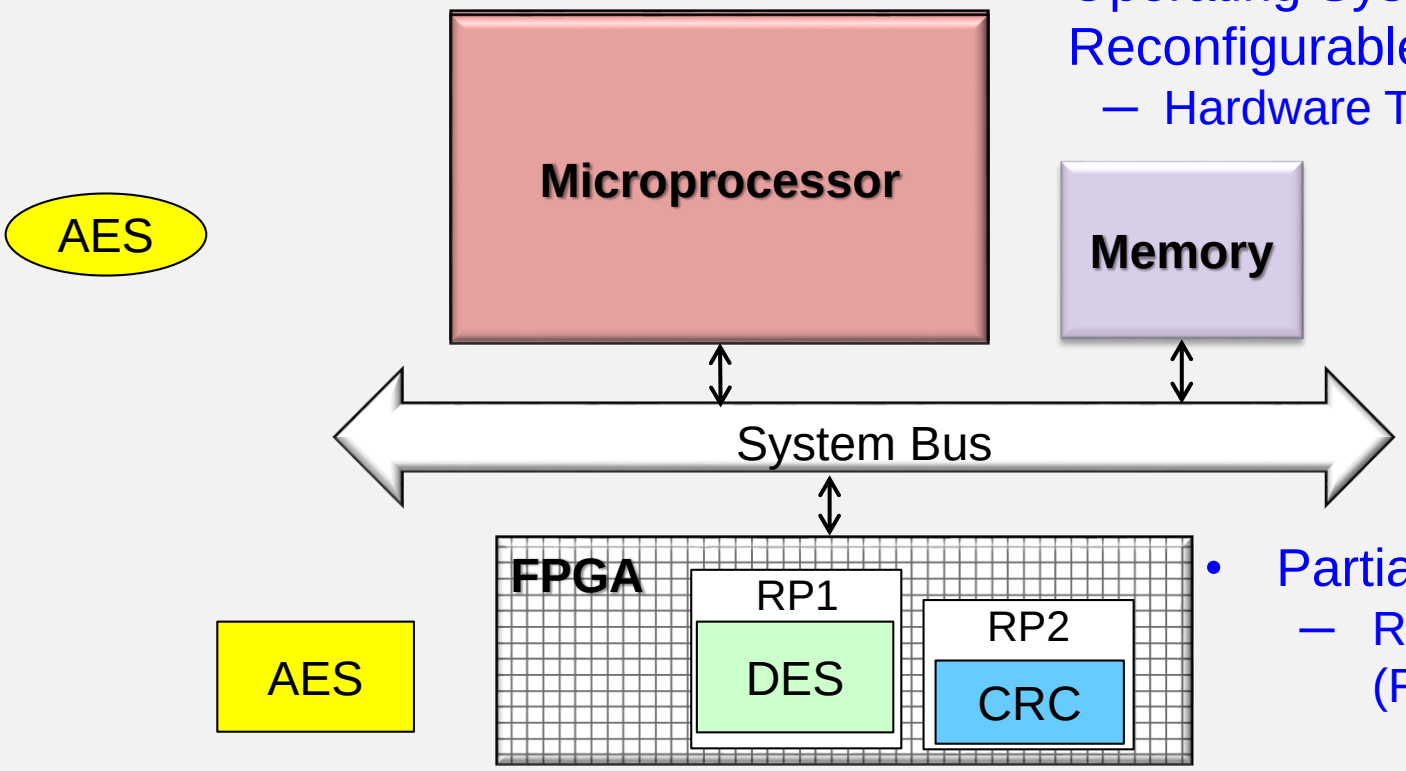
Source: 教育部ATP課程-PAL聯盟「可重組式晶片系統雛形設計與應用」

Reconfigurable System = RC + MP

- Most reconfigurable systems or platforms either have a hard microprocessor core embedded or soft microprocessor cores can be configured
  - Altera: ARM 922T (hard), Nios II (soft)
  - Xilinx: ARM, PowerPC 405 (hard), Microblaze (soft)

Reconfigurable System

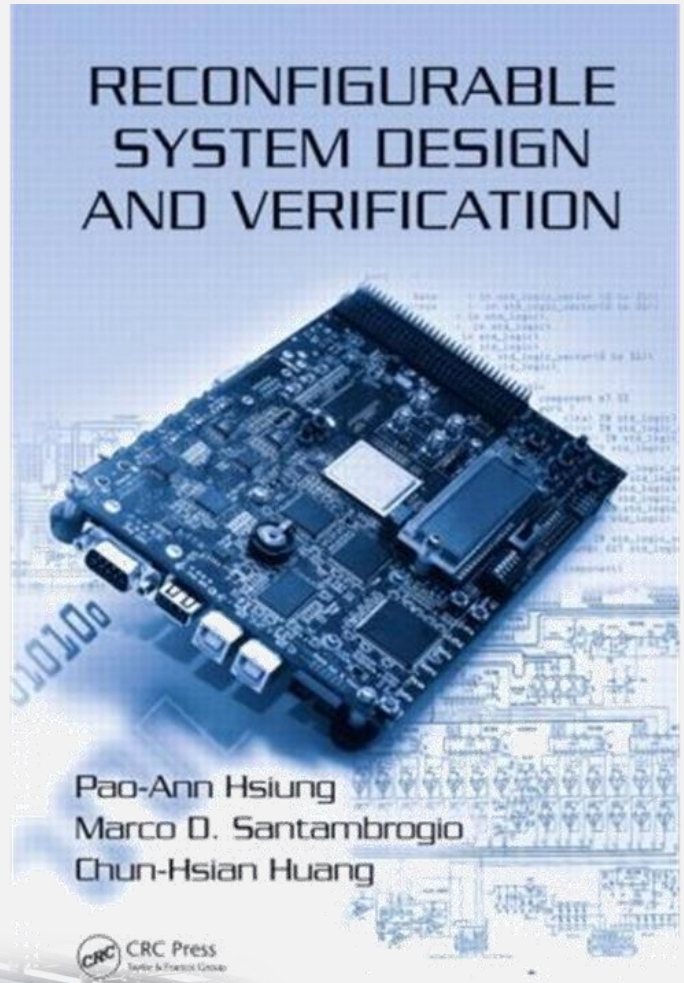
- Operating System for Reconfigurable Systems (OS4RS)
  - Hardware Task



- Partial Reconfiguration
  - Reconfigurable Partition (RP)

- Increase system flexibility and computation power
- Reduce the overall system execution time

Reconfigurable System Design and Verification



P.-A. Hsiung, M. D. Santambrogio and **C.-H. Huang**, Reconfigurable System Design and Verification, CRC Press, USA, ISBN: 978-1420062663, 2009.

https://www.amazon.com/Reconfigurable-System-Design-Verification-Pao-Ann/dp/1420062662

Adaptive and Service-oriented Embedded System for Information Security Applications

- Computers & Electrical Engineering, Vol. 73, pp. 145-154, January 2019. (SCI)
- IEEE/IFIP International Conference on Embedded and Ubiquitous Computing (EUC), pp. 188-192, October 2015.
- IEEE International Conference on Consumer Electronics–Taiwan (ICCE-TW), pp. 276-277, June 2015.

Introduction

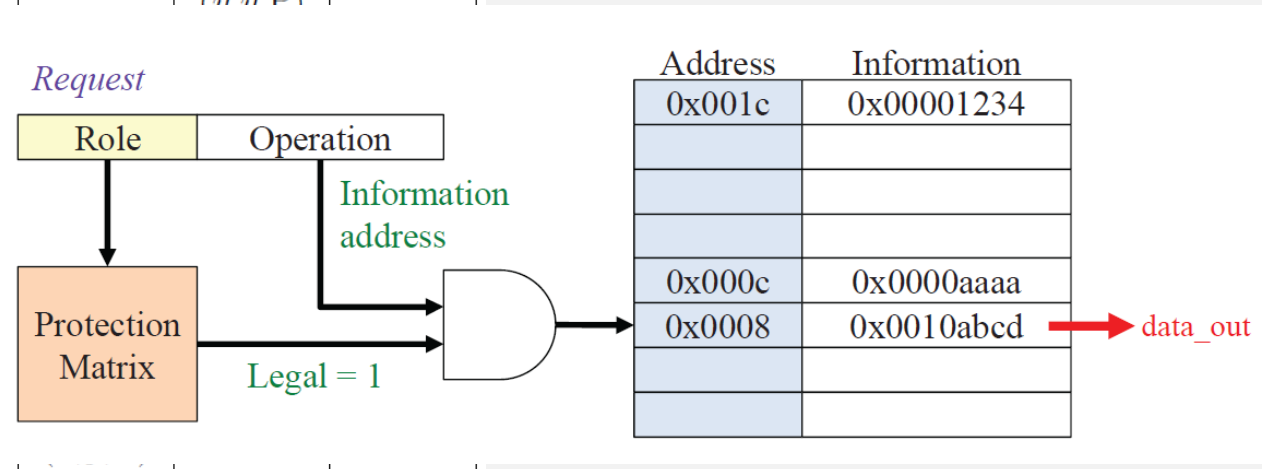
- Software is easily modified than hardware
  - A pure software-based architecture is highly prone to illegal access.
- The pure software-based architecture also incurs the performance/security trade-off problem.
- The inflexible feature of hardware implementation, the system adaptivity to changing environments is restricted.

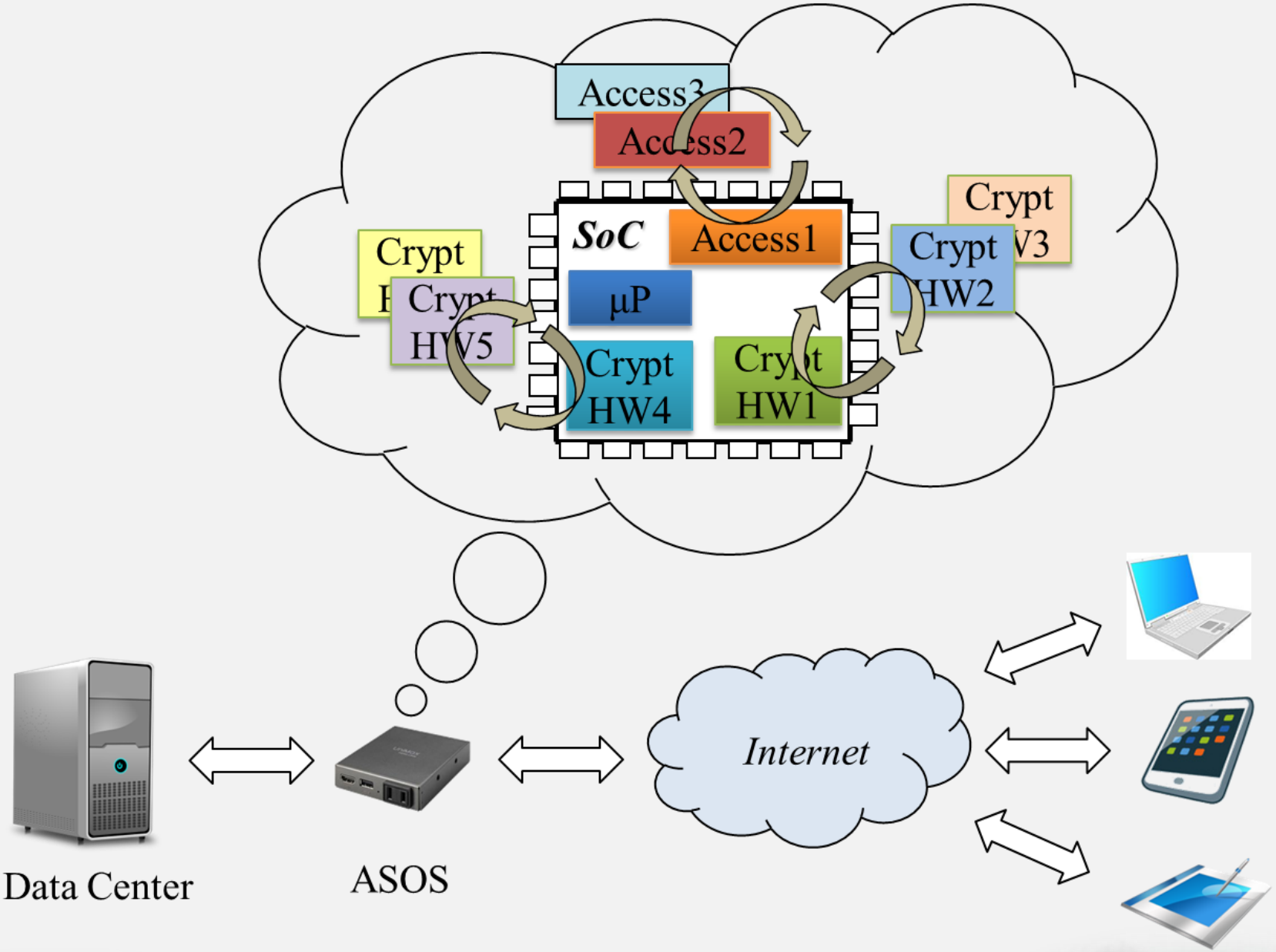
Protect Matrix

Table 1: Protection matrix of ASOS

	on	off	create	grant	take	destroy
$c_1(x,y)$	(x,x,A)		y			
$c_2(x,y)$	(x,x,A)	(y,y,P) (y,y,M)				y
$c_3(x,y)$	(x,x,A)	(y,y,P)		(y,y,P)		
$c_4(x,y)$	(x,x,A) (y,y,P)				(y,y,P)	
$c_5(x,y)$	(x,x,P)	(y,y,M)				
$c_6(x,y)$	(x,x,P) (y,y,M)					
$c_7(x,y)$	(x,x,A) (y,y,P)					
$c_8(x,y)$	(x,x,A) (y,y,P)					
$c_9(x,y)$	(x,x,P) (y,y,M)					
$c_{10}(x,y)$	(x,x,P) (y,y,M)				(x,y,W)	

A: Administrator, P: Personnel office, M: Manager, W: The permission of writing the personal information, R: The permission of reading the personal information.

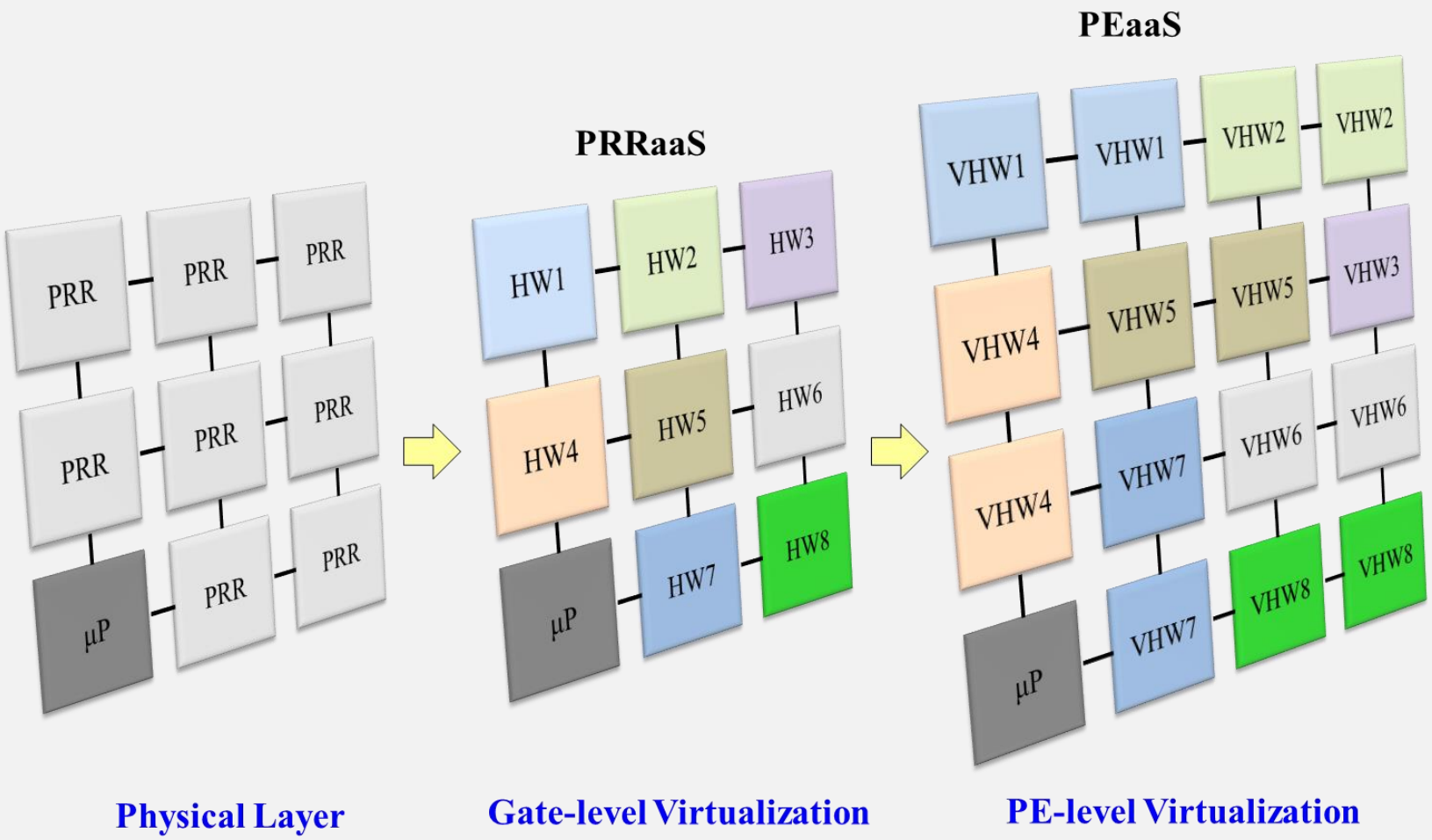




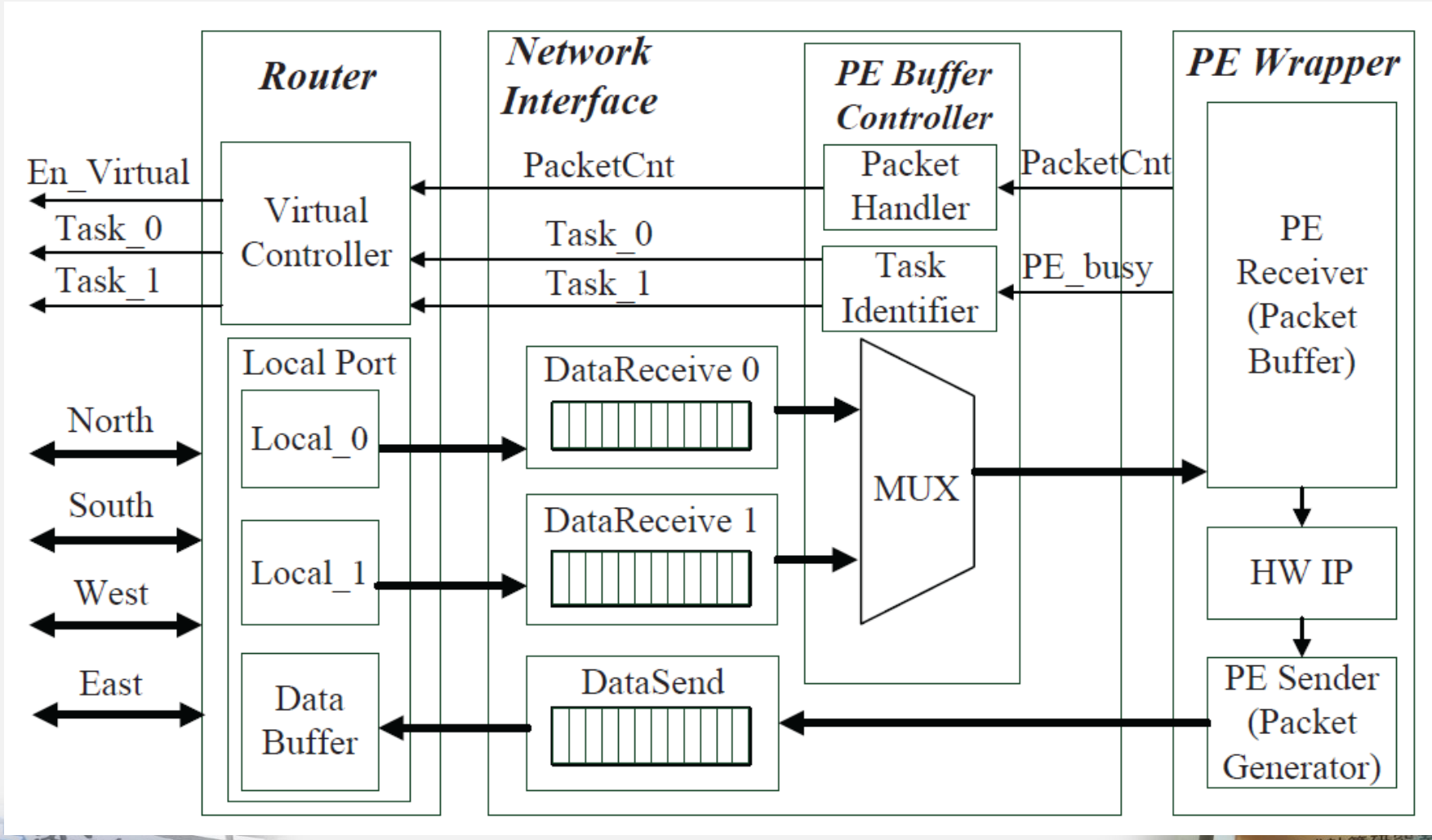
ASOS: Adaptive and Service-Oriented hardware/software System

Service-oriented architecture design for virtualization in Network-on-Chip

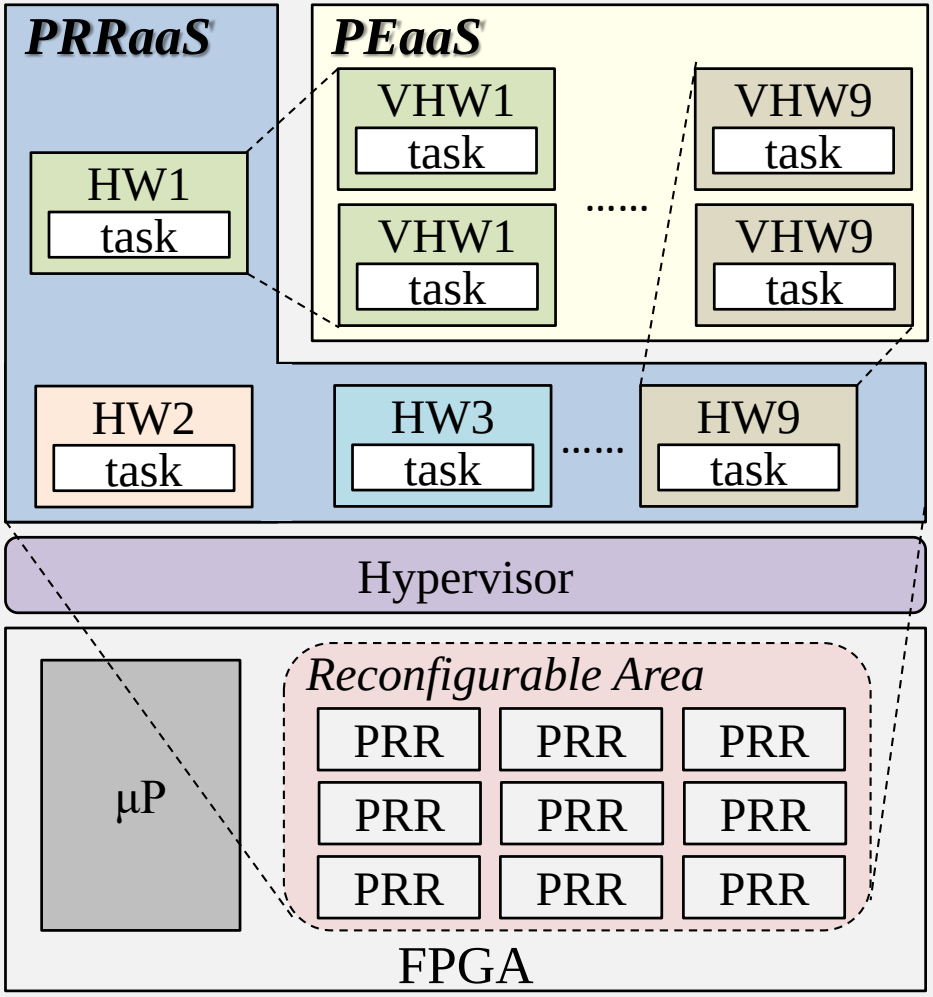
- International Computer Symposium (ICS), pp. 108-113, December 2016.
- International Workshop on FPGAs for Software Programmers (FSP), pp. 73-74, September 2015.


PEaaS
PRRaaS
PRR
PRR
PRR
PRR
PRR
PRR
PRR
PRR
μP
HW1
HW2
HW3
HW4
HW5
HW6
HW7
HW8
μP
VHW1
VHW1
VHW2
VHW2
VHW4
VHW5
VHW5
VHW3
VHW4
VHW7
VHW6
VHW6
VHW7
VHW8
VHW8
μP
Physical Layer
Gate-level Virtualization
PE-level Virtualization

Router, network interface, and PE wrapper



Service-oriented system framework



Hierarchical and Dependency-Aware Task Mapping for NoC-based Systems

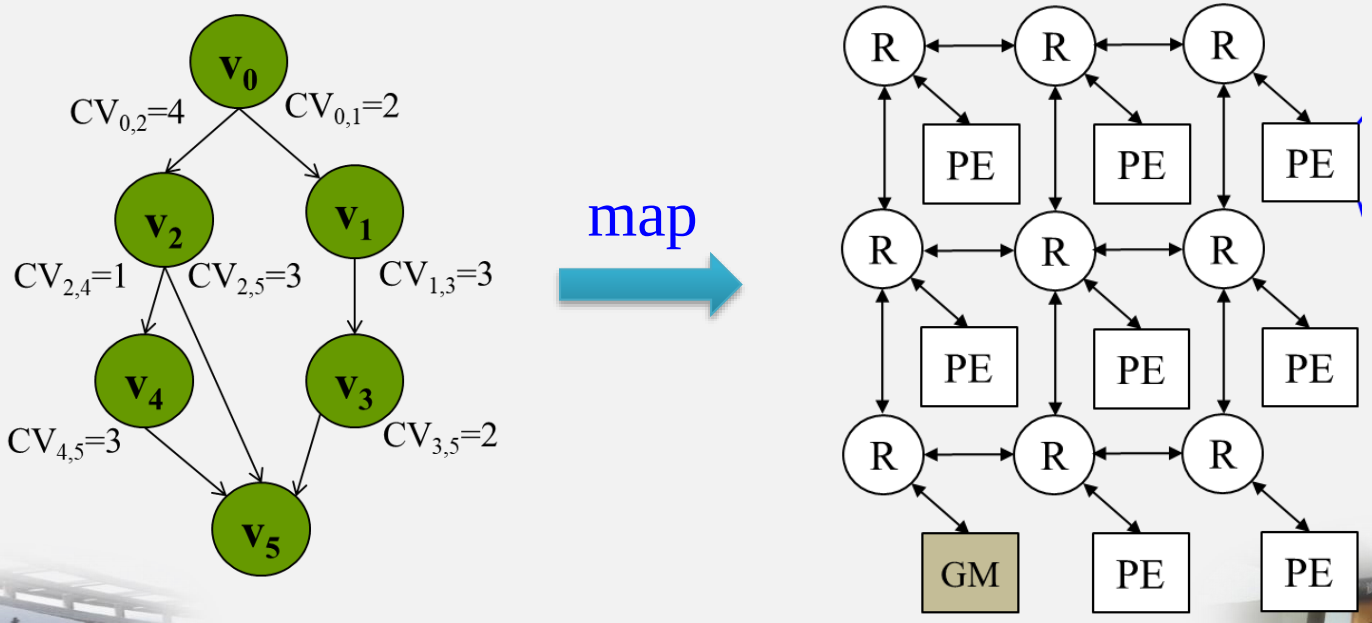
- Journal of Systems Architecture (SCI). (Major revision)
- International Workshop on Network on Chip Architectures, pp. 1-6, October 2018.
- Microprocessors and Microsystems, Vol. 51, pp. 297-312, June 2017. (SCI)

Motivation

- Try to integrate the concepts of spatial mapping and inter-task dependency (temporal) to enhance the elasticity of task mapping
- Propose a Hierarchical and Dependency-Aware (HDA) task mapping for NoC-based Systems
  - Region selection phase
  - Hierarchical task mapping phase

Problem Formulation

- Given an NoC modeled by AG G' = (V' ,E') and an application modeled by ACG G = (V, E)
  - Find a mapping function map() to minimize the total processing time



Proposed Method - HDA

- Region selection phase
  - Ensure that every mapped task can receive the data from its parent tasks and transmit the results to its child tasks
  - The reduction of the number of required PEs in a region enhances the elasticity of future task mapping
- Hierarchical task mapping phase
  - PE reuse scheme: when previously mapped task finishes, its corresponding PE is released and can be mapped by another unmapped task
  - Lower communication latencies and higher utilization of PEs

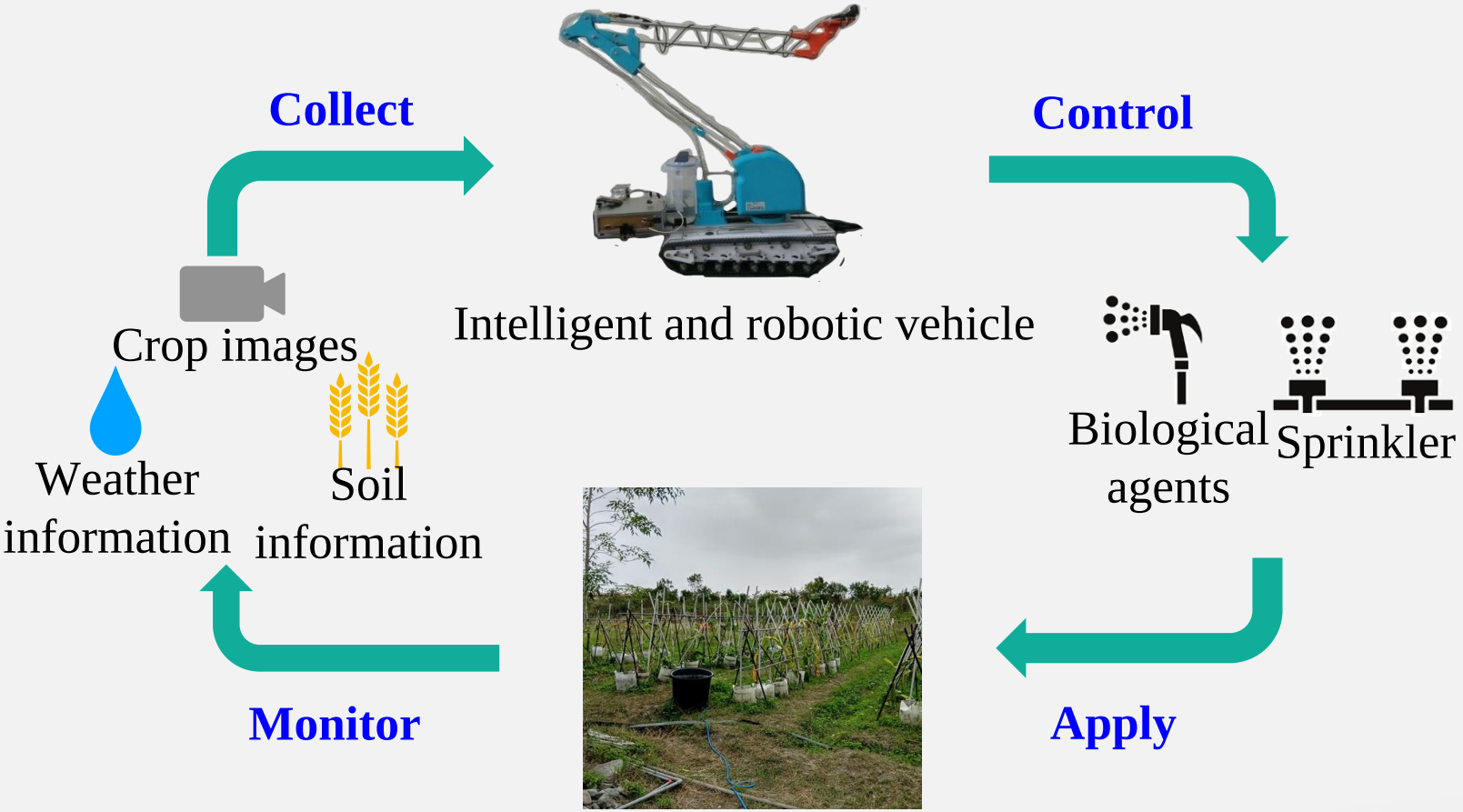
An FPGA-based Intelligent Robotic Vehicle System for Agricultural Cyber Physical Systems

- IEEE International Conference on Consumer Electronics (ICCE), January 2020.
- 第24屆大專校院資訊應用服務創新競賽「AIoT兩岸交流應用組」- 第一名, 2019.
- VLSI Design / CAD Symposium, August 2019.
- Taiwan Academic Network Conference, 2018. (最佳口頭報告獎)

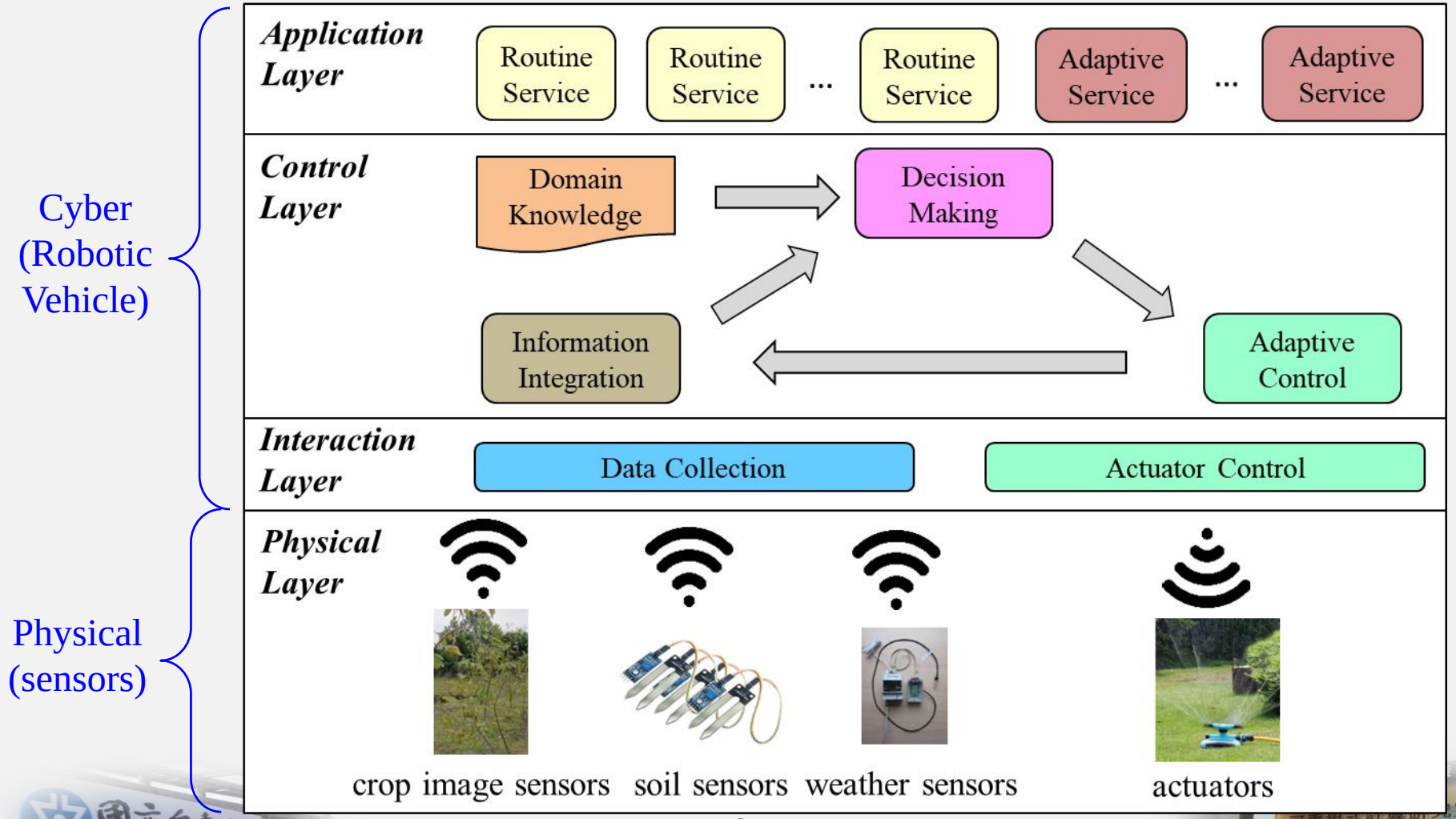
Research Issues

1. What is the system design that we should realize, so that the physical world can be integrated with the cyber system efficiently?
2. How can the data collection and the control of actuators be performed in real-time, while the system can be adapted to varying environmental conditions?
3. What is the decision model that can make this system autonomous?

Agricultural Cyber Physical Systems



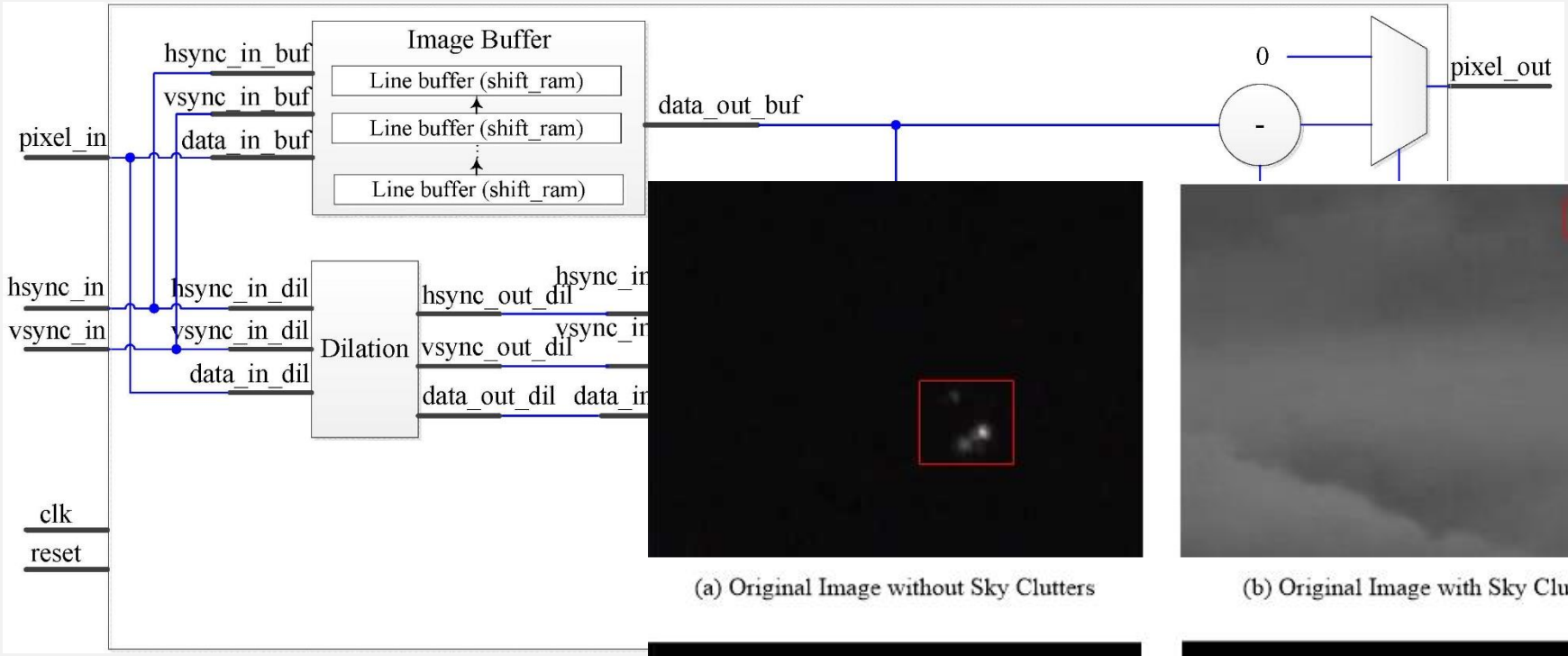
Layered Agricultural CPS



題目：
應用於農業網宇實體系統之智能機器人車系統設計

https://youtu.be/7hceoc3ISss

C.-H. Huang, "A Reconfigurable Point Target Detection System Based on Morphological Clutter Elimination," *Journal of Systems Architecture* (JSA), Vol. 60, Issue 8, pp. 644-654, September 2014. (SCI)



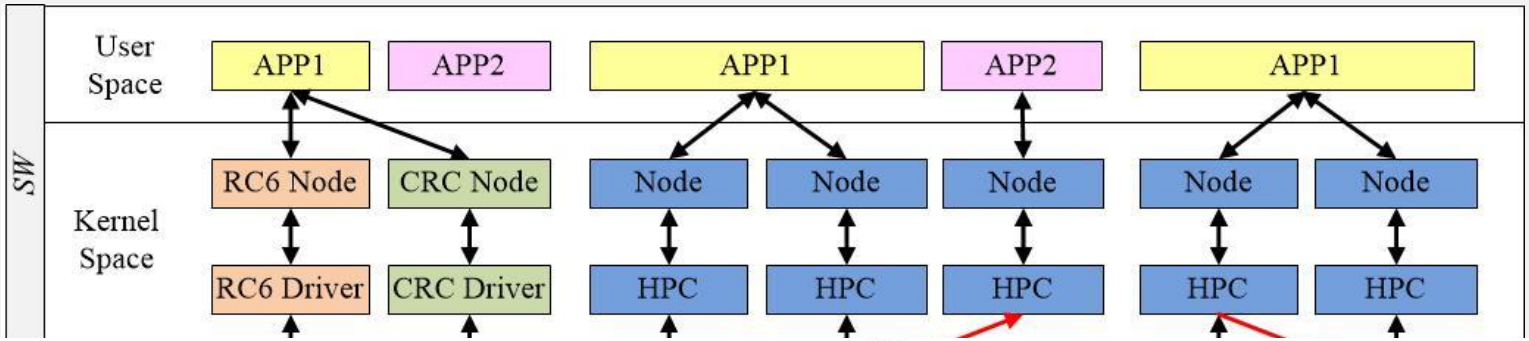
(a) Original Image without Sky Clutters

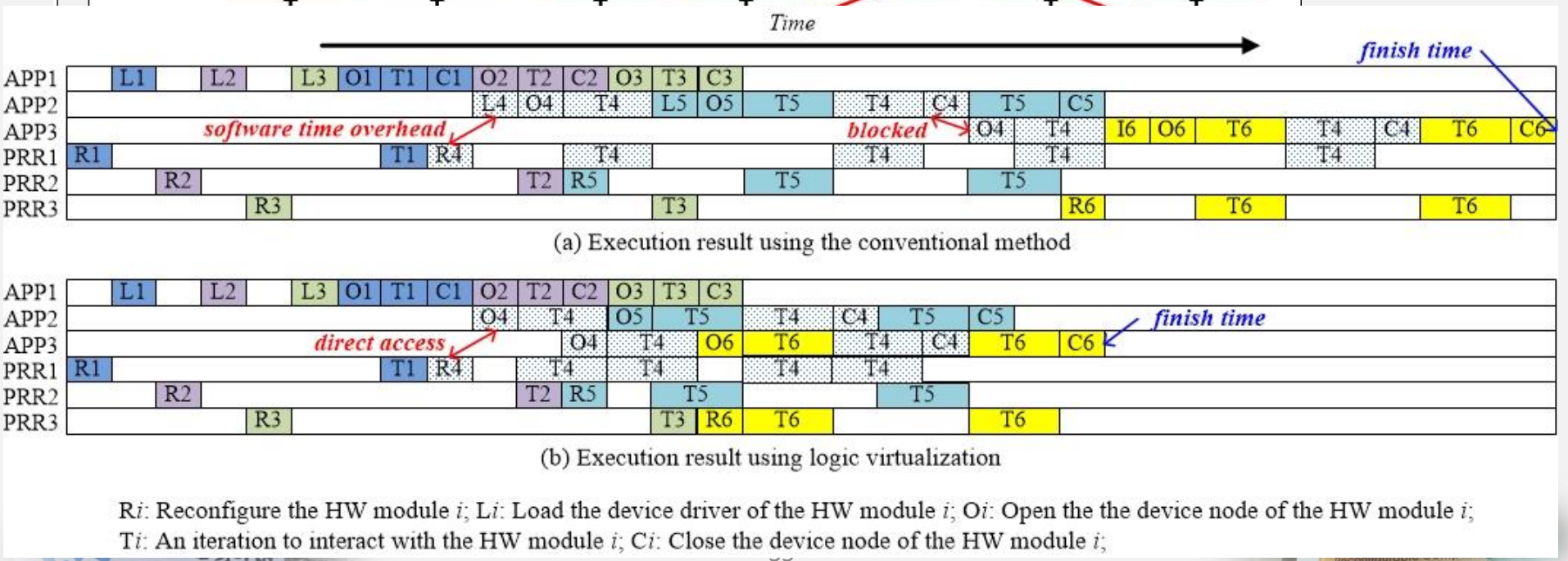
(b) Original Image with Sky Clutters

(c) Point Detection Result of Figure 9(a)

(d) Point Detection Result of Figure 9(b)

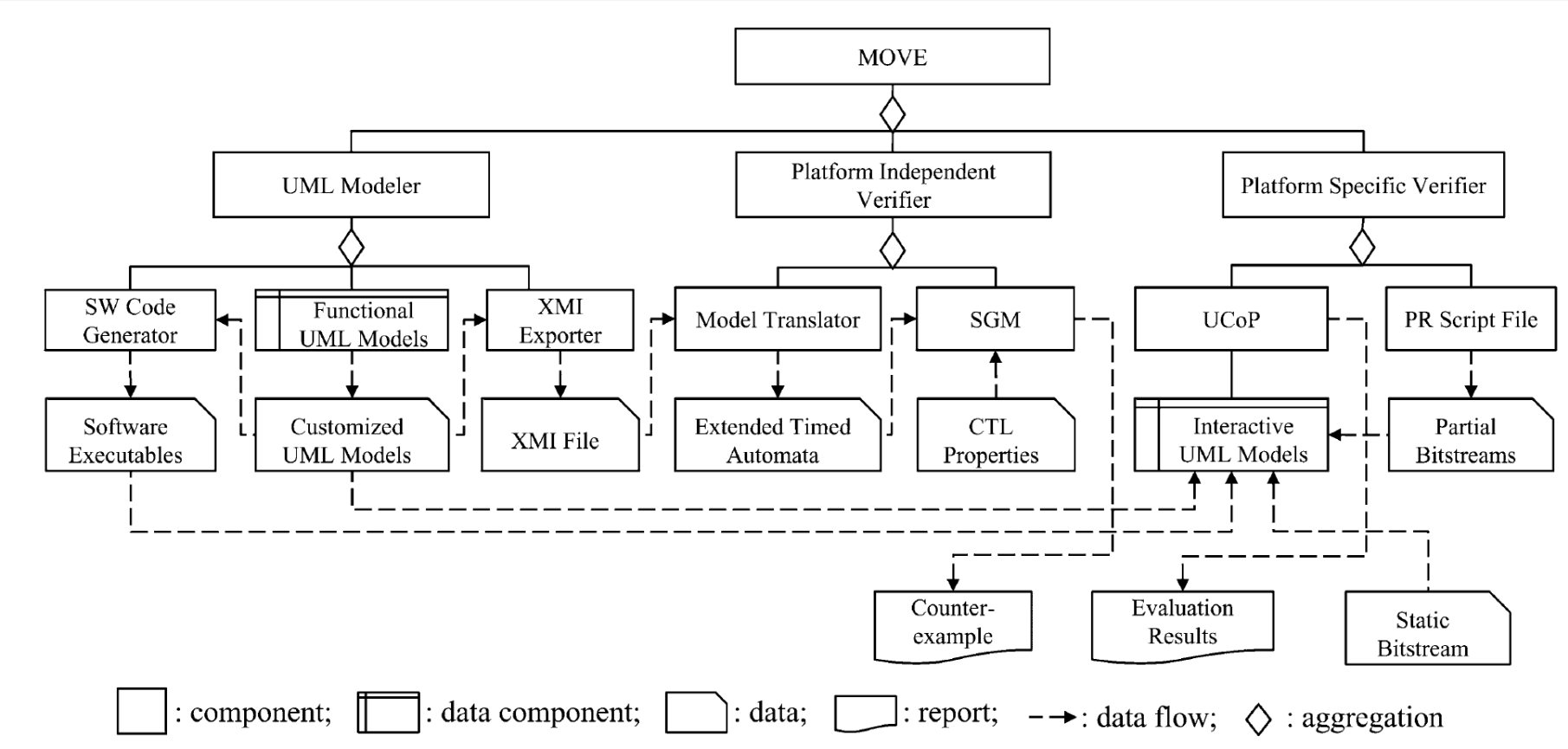
C.-H. Huang and P.-A. Hsiung, "Virtualizable Hardware/Software Design Infrastructure for Dynamically Partially Reconfigurable Systems," *ACM Transactions on Reconfigurable Technology and Systems* (TRETS), Vol. 6, No. 2, Article 11, July 2013.(SCI)





R*i*: Reconfigure the HW module *i*; L*i*: Load the device driver of the HW module *i*; O*i*: Open the the device node of the HW module *i*;
T*i*: An iteration to interact with the HW module *i*; C*i*: Close the device node of the HW module *i*;

C.-H. Huang and P.-A. Hsiung, "Model-Based Verification and Estimation Framework for Dynamically Partially Reconfigurable Systems," *IEEE Transactions on Industrial Informatics* (TII), Vol. 7, No. 2, pp. 287-301, May 2011. (SCI)



References

1. **C.-H. Huang**, H.-Y. Chen, Y.-Y. Tzeng, and P.-Y. Li, "Adaptive and Service-oriented Embedded System for Information Security Applications," *Computers & Electrical Engineering*, Vol. 73, pp. 145-154, January 2019. (SCI)
2. **C.-H. Huang**, C.-Y. Wang, and P.-A. Hsiung, "Elastic Superposition Task Mapping for NoC-based Reconfigurable Systems," *Microprocessors and Microsystems*, Vol. 51, pp. 297-312, June 2017. (SCI)
3. **C.-H. Huang**, "A Reconfigurable Point Target Detection System Based on Morphological Clutter Elimination," *Journal of Systems Architecture* (JSA), Vol. 60, Issue 8, pp. 644-654, September 2014. (SCI)
4. **C.-H. Huang** and P.-A. Hsiung, "Virtualizable Hardware/Software Design Infrastructure for Dynamically Partially Reconfigurable Systems," *ACM Transactions on Reconfigurable Technology and Systems* (TRETS), Vol. 6, No. 2, Article 11, July 2013.(SCI)
5. **C.-H. Huang** and P.-A. Hsiung, "Model-Based Verification and Estimation Framework for Dynamically Partially Reconfigurable Systems," *IEEE Transactions on Industrial Informatics* (TII), Vol. 7, No. 2, pp. 287-301, May 2011. (SCI)

References

6. **C.-H. Huang**, P.-J. Chen, H.-T. Tseng, Z.-M. Zhang, H.-L. Hong, and Y.-Z. Tu, "Design of an Intelligent Robotic Vehicle for Agricultural Cyber Physical Systems, " *Proc. IEEE International Conference on Consumer Electronics* (ICCE), January 2020. (accepted for publication)
7. **C.-H. Huang**, C.-Y. Chen, and H.-Y. Huang, "Hierarchical and Dependency-Aware Task Mapping for NoC-based Systems," *Proc. 11th International Workshop on Network on Chip Architectures*, pp. 1-6, October 2018.
8. **C.-H. Huang**, K.-W. Tseng, and P.-A. Hsiung, "Service-oriented Architecture Design for Virtualization in Network-on-Chip," Proc. International Computer Symposium (ICS), pp. 108-113, December 2016.
9. **C.-H. Huang**, H.-Y. Chen, T.-F. Huang, Y.-Y. Tzeng, P.-Y Li, and P.-S. Wu, "A Self-Adaptive System for Vehicle Information Security Applications*," Proc. 13th IEEE/IFIP International Conference on Embedded and Ubiquitous Computing* (EUC), pp. 188-192, October 2015.
10. **C.-H. Huang**, K.-W. Tseng, C.-C. Lin, F.-Y. Lin and P.-A. Hsiung, "Virtualization Architecture for NoC-based Reconfigurable Systems," *Proc. Second International Workshop on FPGAs for Software Programmers* (FSP), pp. 73-74, September 2015.
11. **C.-H. Huang**, H.-Y. Chen, T.-F. Huang, P.-W. Chen, and Y.-Y. Tzeng, "A Vehicle Information Security System with Self-Adaptation," *Proc. IEEE International Conference on Consumer Electronics–Taiwan* (ICCE-TW), pp. 276-277, June 2015.
12. **C.-H. Huang**, "An FPGA-Based Point Target Detection System Using Morphological Clutter Elimination," *Proc. IEEE International Symposium on Circuits and Systems* (ISCAS), pp. 2436-2439, May 2013.

Thanks for your attention!

huangch@nttu.edu.tw