IARIA

Keynote at ICONS 2021

Human Skill Transfer for AI Accelerator Programming and Its Application to Humanoid Two-Handed Robots

Prof. Chung-Ta King
king@cs.nthu.edu.tw
Department of Computer Science
National Tsing Hua University, Taiwan

About Me

- 1988: Ph.D., Michigan State University, USA
- 1988-1990: Assistant Professor, NJIT, USA
- 1990-1997: Associate Professor, NTHU, Taiwan
- 1997-: Professor, NTHU
- 2009-2012: Chair, Dept of CS, NTHU
- 2014-2015: Associate Dean, College of EECS, NTHU
- 2019-: Vice President and Chief of Staff, NTHU
- Research interests: embedded & mobile computing, computer architecture, parallel & distributed systems
- Conference organizer: Cluster 2016, ICPADS 2014, CloudCom 2012, ESWeek 2011, …

Acknowledgements

- This talk is based the outcomes obtained so far from a 4-year project (2019-2022) funded by the Minister of Science and Technology, Taiwan, under the Moon-Shoot Project

- Team members:
  - Professors Cheng-Wen Wu, Jen-Yuan Chang, Jing-Jia Liou, Chih-Tsun Huang, and Hung-Kuo Chu of National Tsing Hua University, Taiwan
  - Over 20 research staffs and graduate students

Growing Importance of Service Robots

https://www.pinterest.com/pin/127719339416902130/

Medical care, logistics, farming, household, companion, education, entertainment, …

Growing market size to > $20 Billion in 2024

Humanoid, 5-fingered hands desirable for adapting to varying tasks and dynamic environments

Challenge: Complexity in robot controls due to very high *degree-of-freedom* (DOF) and complex working environments ➔ application developments

Fast Development of Robot Applications

Common practices:

- Programming by programmers
- Moving arms to set actions
- Training and imitation

To handle dynamics:

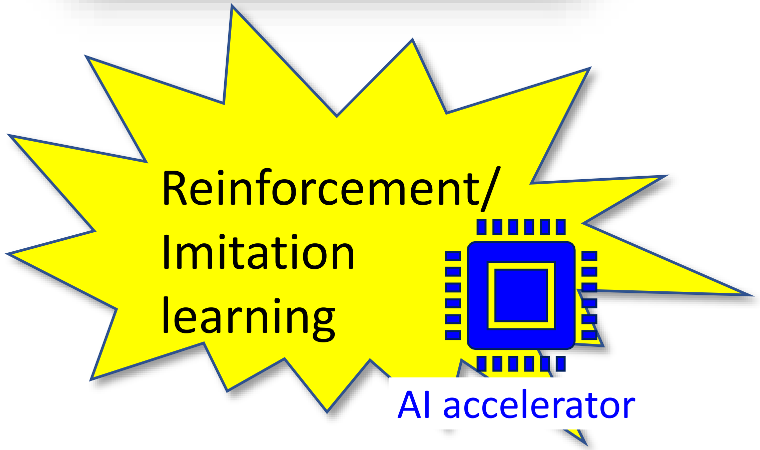
- Sensing: vision, touching
- Planning and decision
- Control

Imitation for Fast Application Development

Human demonstrates:
- What are sensed (vision, touch, …)?
- What actions are performed in response?

→ Without expertise in robotics

Robot imitates:
- Not just clone trajectories and actions (*behavior cloning*)
- Must be able to adapt and generalize



Problems with Imitation

Scattered tools relying on manual interfacing	Training with physical robots too slow and unsafe	Process from NN models to AI accelerators too long	Design space exploitation difficult and inefficient

Need a systematic flow for fast development of robot applications
-- from human demonstration onto accelerator chips

Human Skill Transfer for AI Accelerators

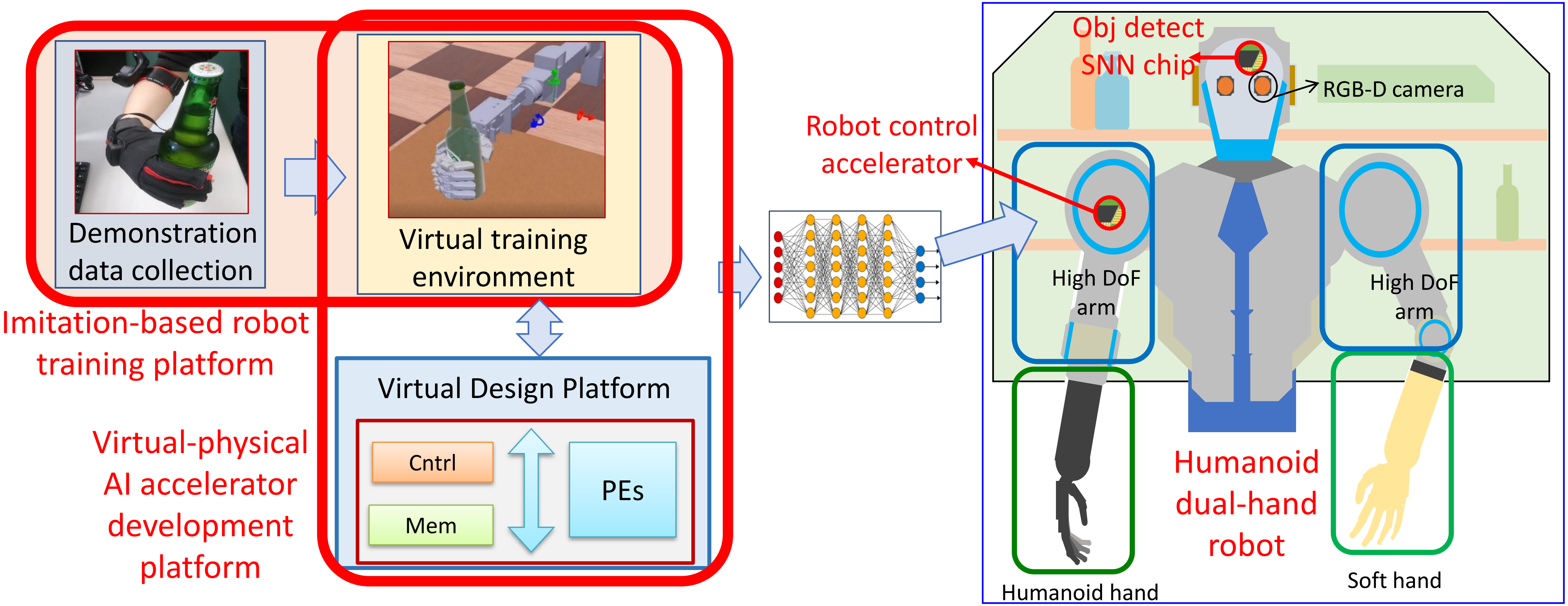
Two design flows:

- From human skills to AI model for robot control
  - Collect data from human demonstration to train robot control models in a virtual environment
- From AI model to AI accelerator chips
  - Compile and optimize the AI control model to run on AI accelerators

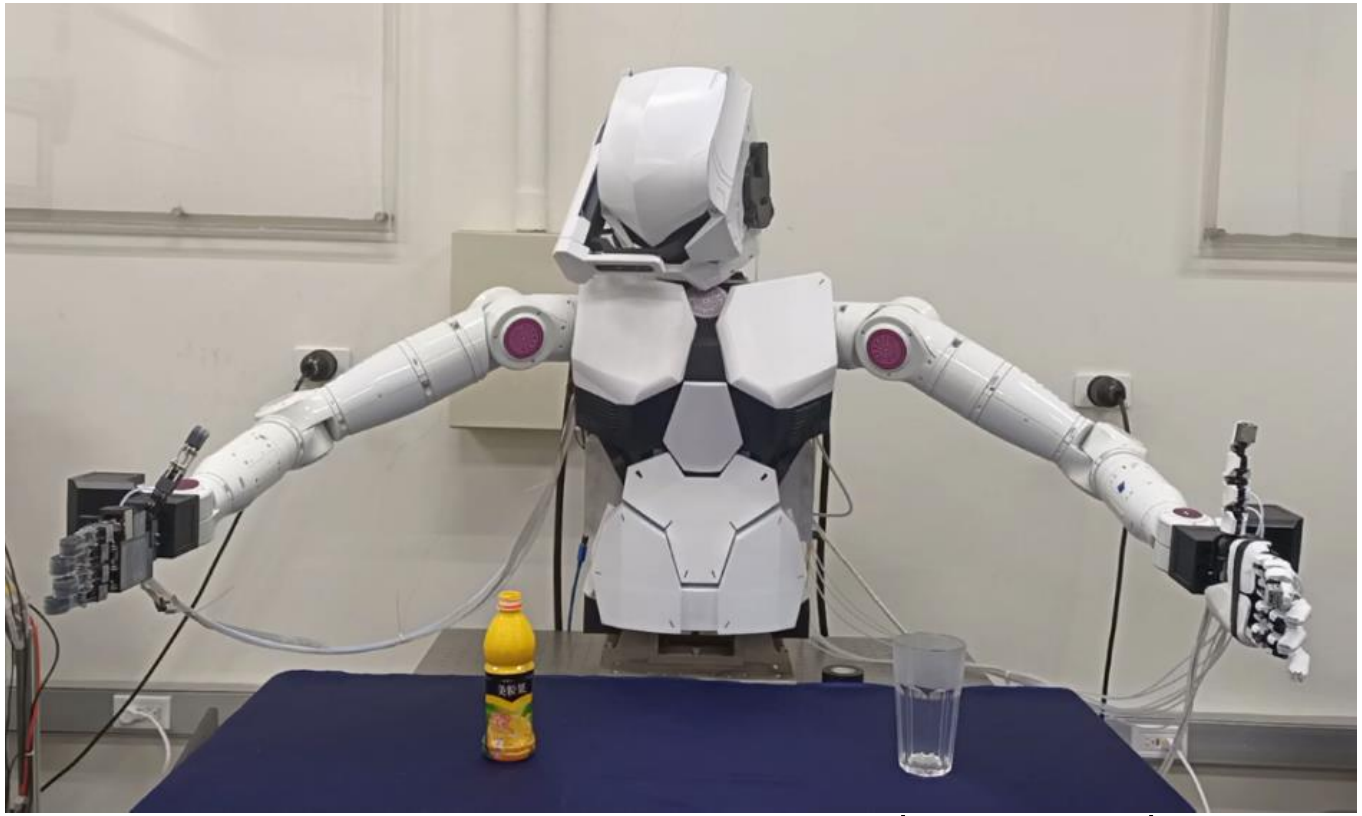
A humanoid, dual-hand robot:

- Controlled by AI accelerators that run neural networks

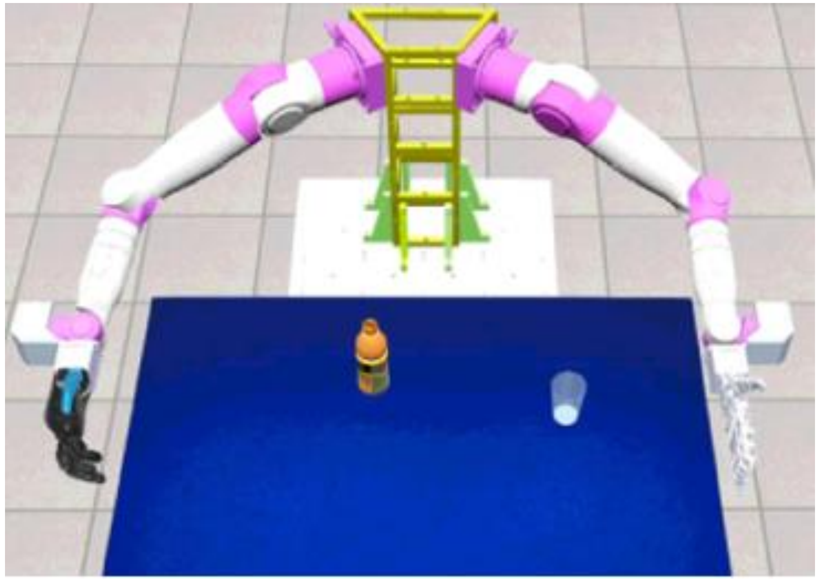
Project Overview



Demonstration of Trained Robot Operations

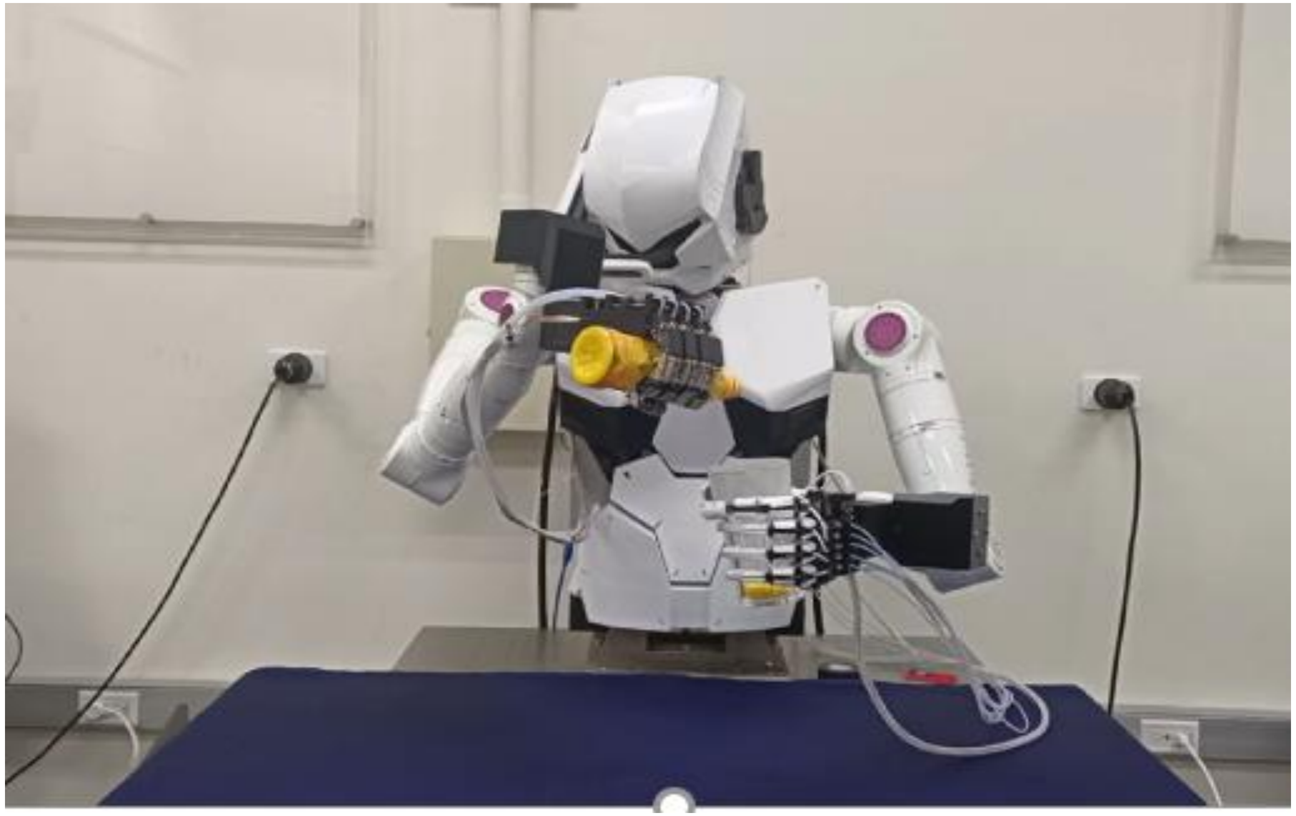


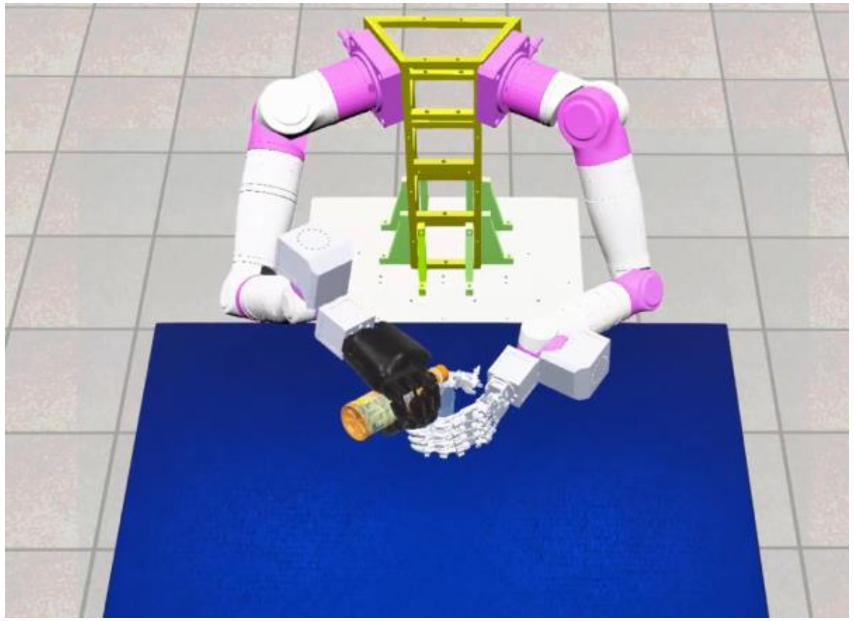
Operations by physical robot (3X speed)



Operations in simulator

Demonstration of Trained Robot Operations


Operations by physical robot (3X speed)

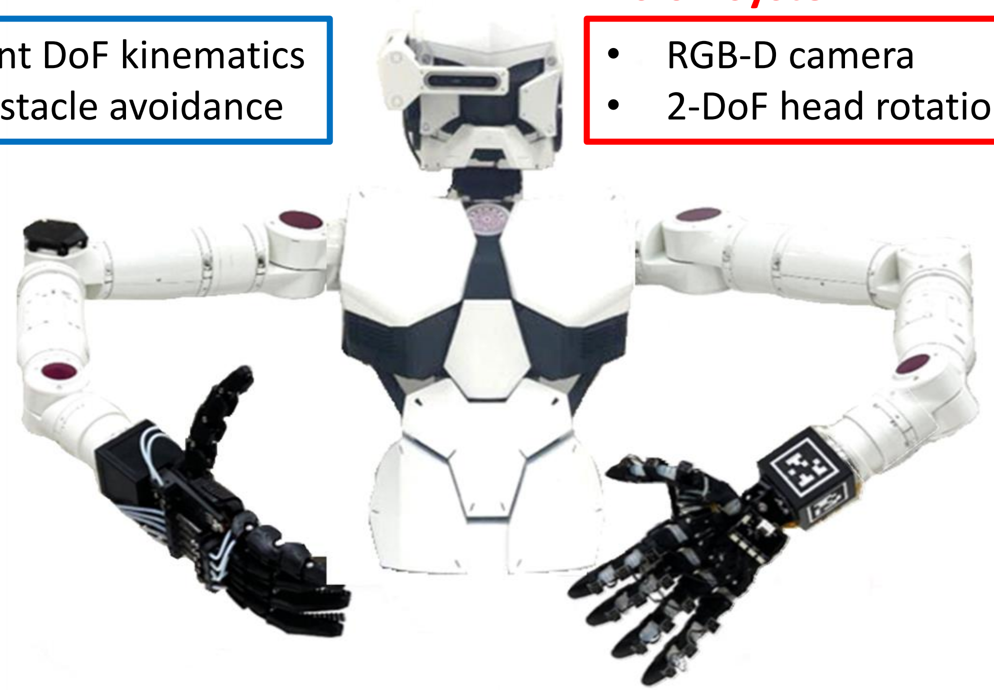

Operations in simulator

Humanoid Dual-Hand Robot

(Prof. Jen-Yuan Chang)

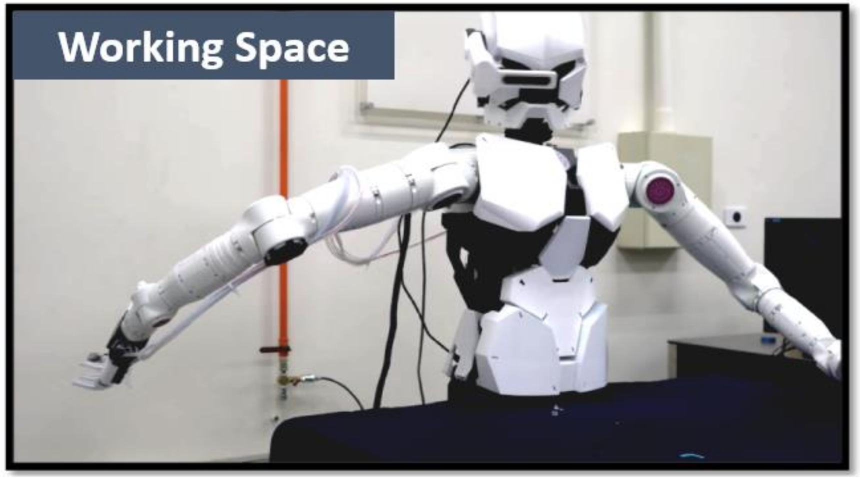
7-DoF arms
- Redundant DoF kinematics
- 7-DoF obstacle avoidance

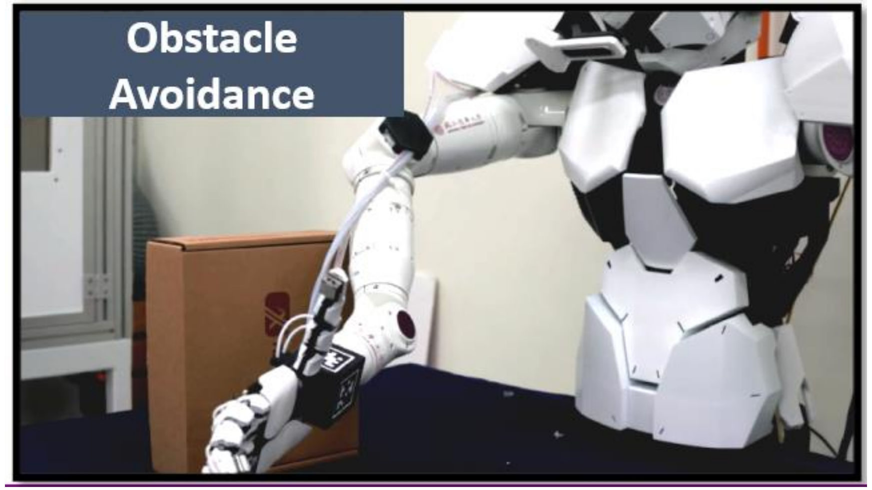
Vision system
- RGB-D camera
- 2-DoF head rotation



Wire-drive 5-fingered hand
- Underactuated wire-driven 5-fingered gripper
- Touch sensor feedback

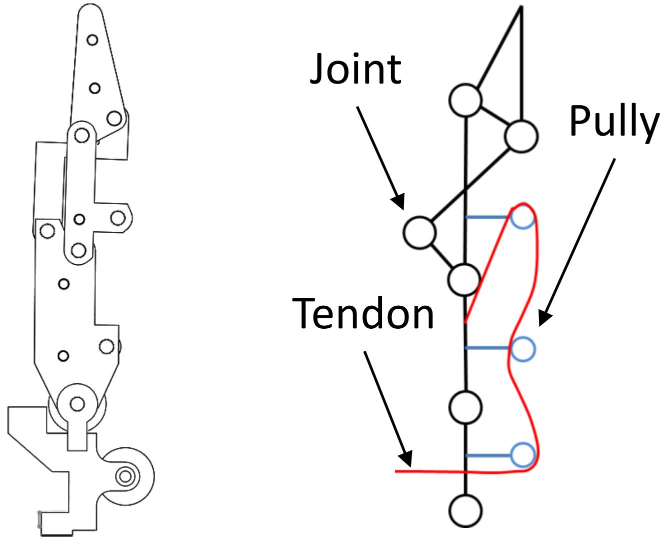
Soft hand + IMU
- Flexible 5-fingered gripper
- Pneumatic actuation
- IMU for angle feedback





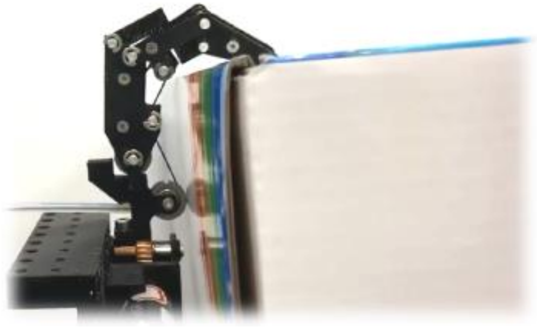
Self-Adaptive Finger Joints

(Prof. Jen-Yuan Chang)



Underactuated design

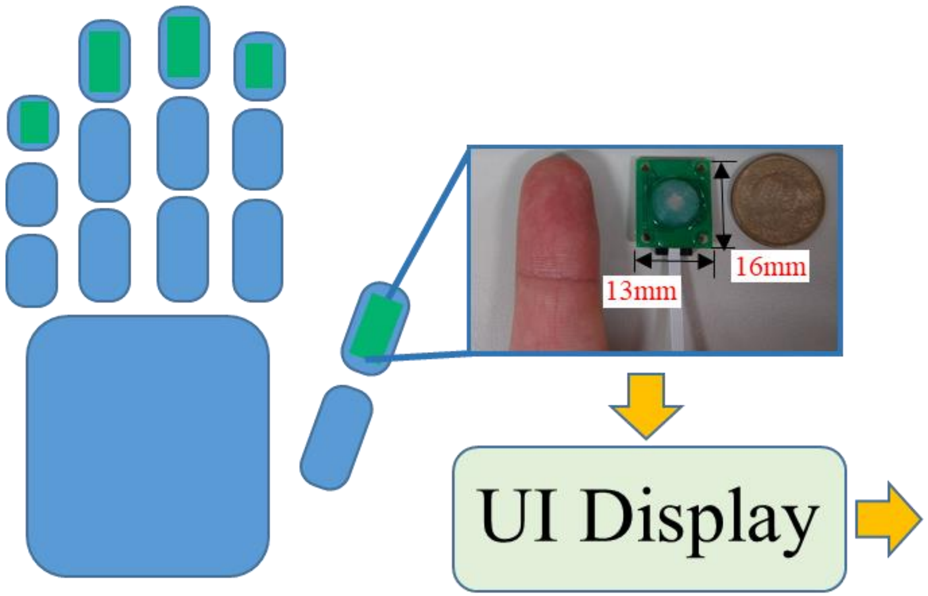
	Shadow Robot – Hand Lite	**NTHU Hand**
DoF	13	7
Weight	2.4 kg	1 kg
Loading	4 kg	1 kg → 3 kg
Driving	Wired (motor fixed)	Wired (motor separated)

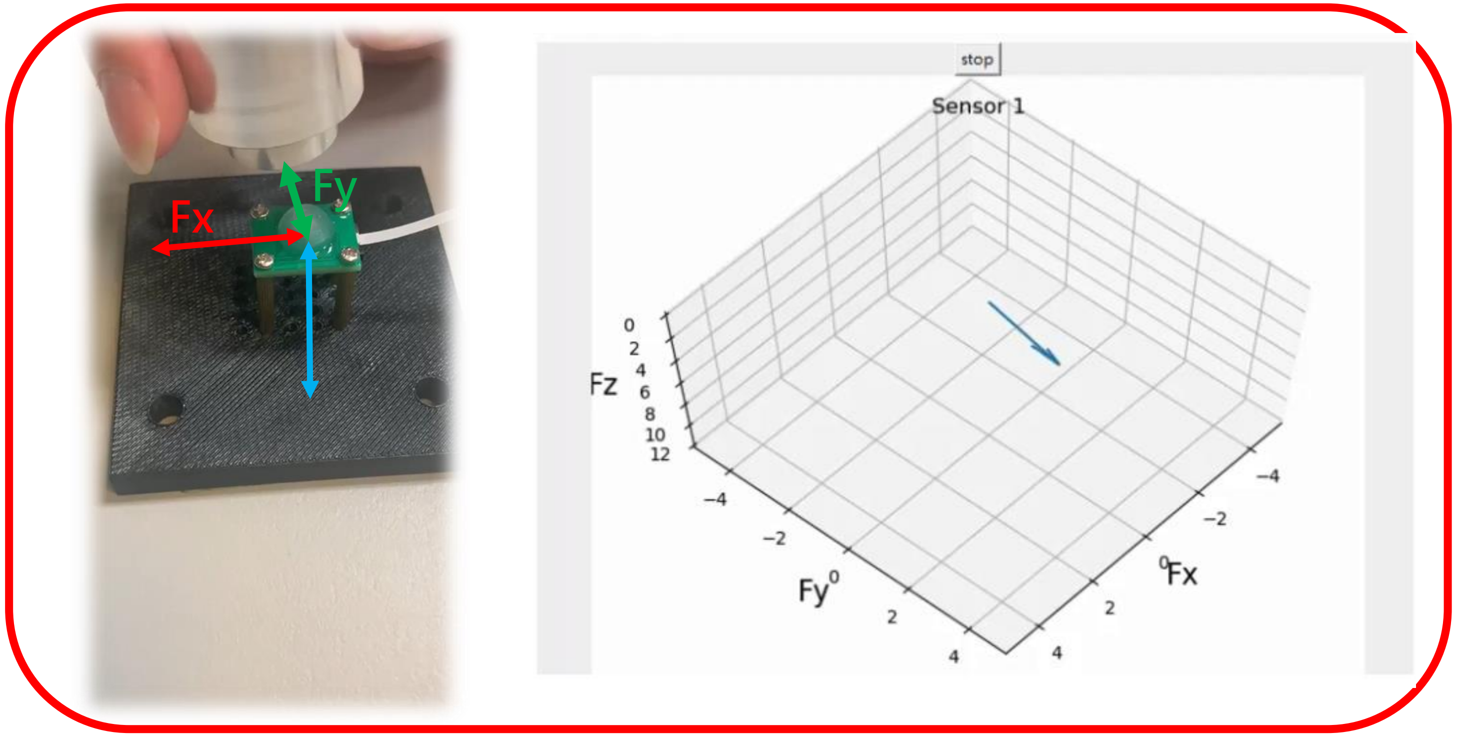


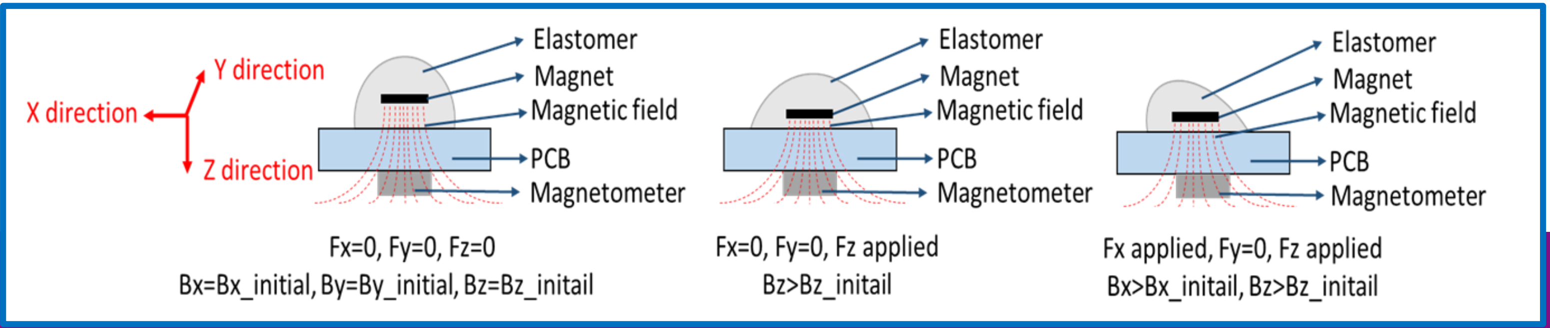
- Underactuated design: 1 actuator to drive two joints
- Better grasps of unknown objects
- Self-adaptive to errors in grasping

Magnetic-based Touch Sensor

(Prof. Jen-Yuan Chang)



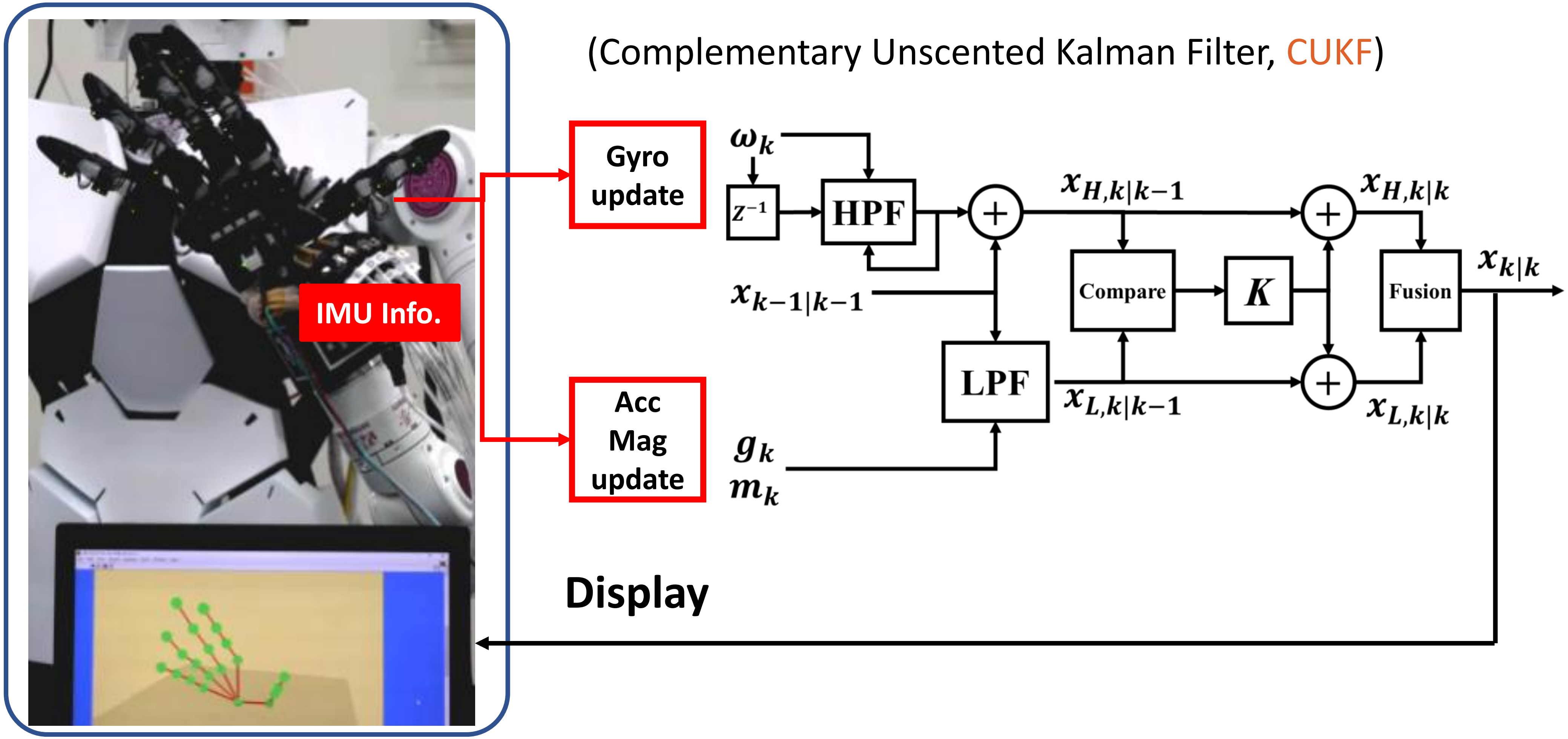




Softhand + IMU

(Prof. Jen-Yuan Chang)

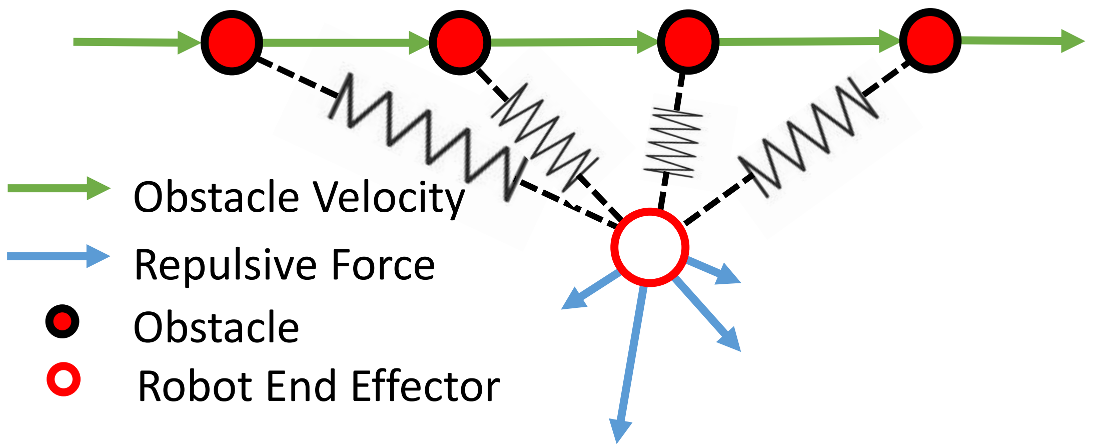
(Complementary Unscented Kalman Filter, CUKF)

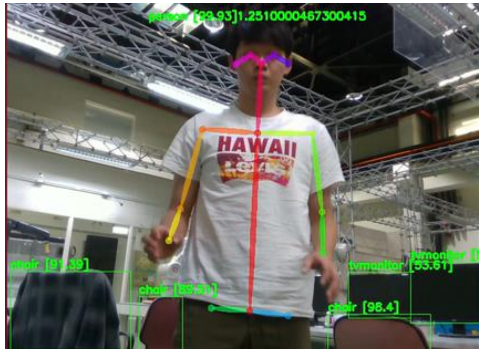


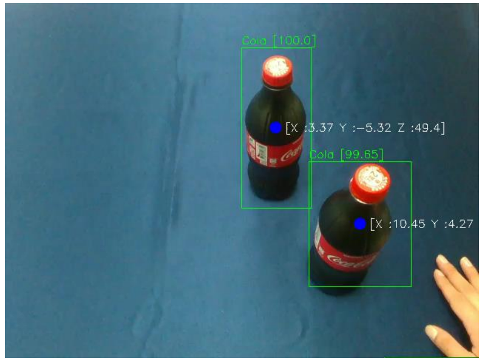
Real-time Obstacle Avoidance

(Prof. Jen-Yuan Chang)

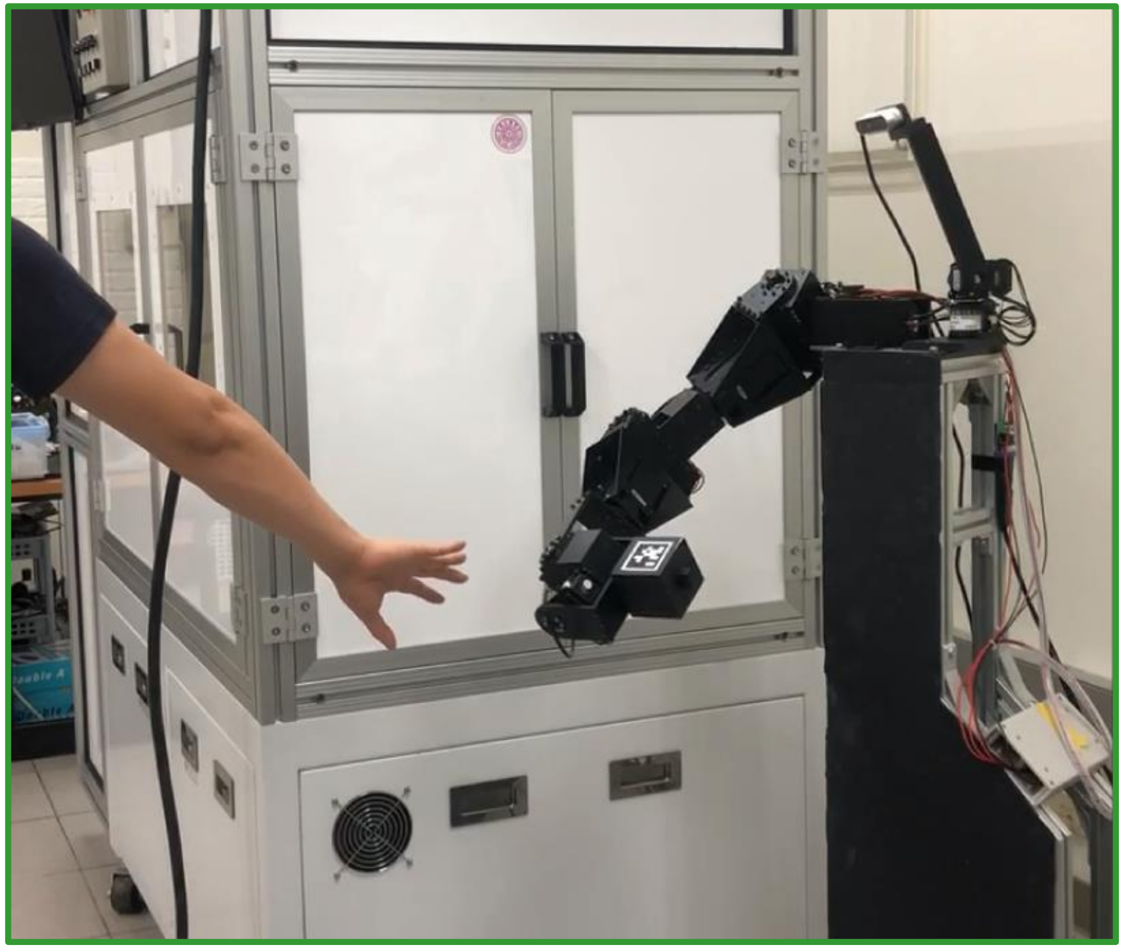
Spring Vector Repulsive Force Algorithm







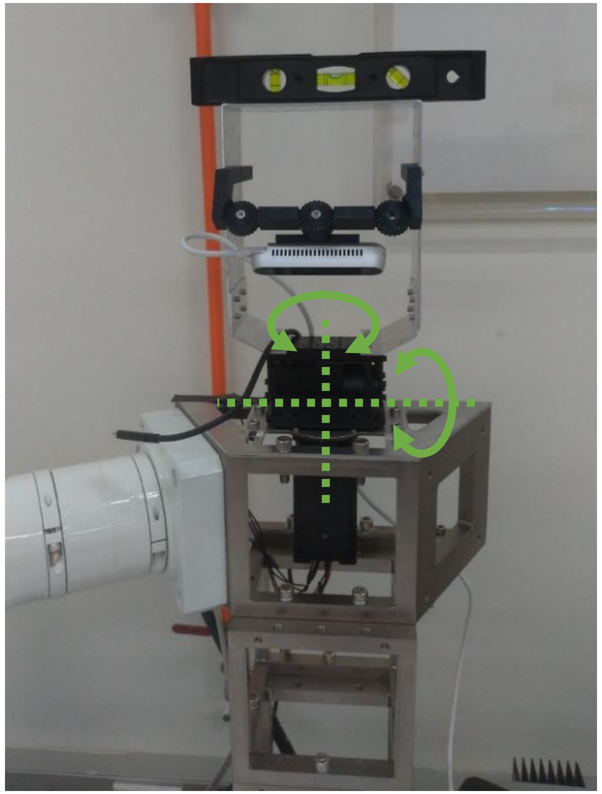
Object and skeleton detection

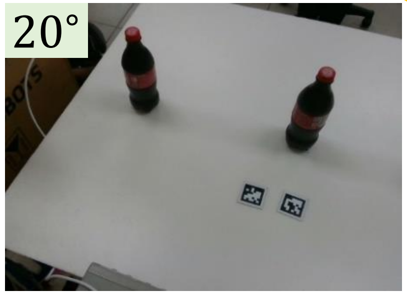


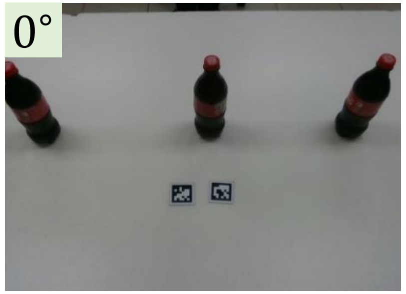
Robot Vision System

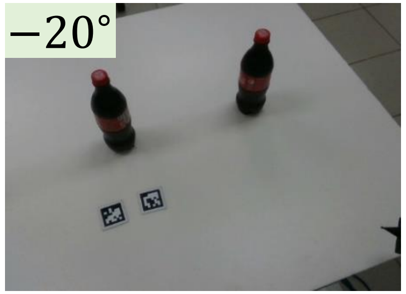
(Prof. Hung-Kuo Chu)

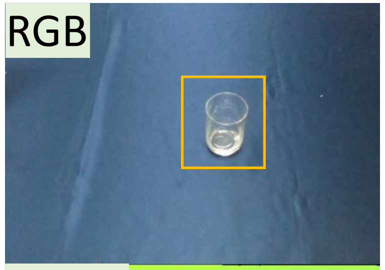
Rotate camera to
increase field-of-view

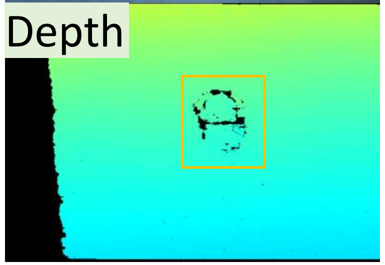




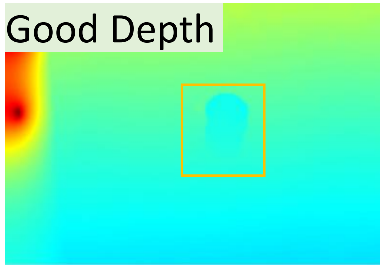








ClearGrasp [1]



Depth of transparent objects

Generate Input image (130ms)

↓

Define problem & equation
& build matrix (170ms)

↓

Use AMGCL library to accelerate
(226ms→8ms)

↓

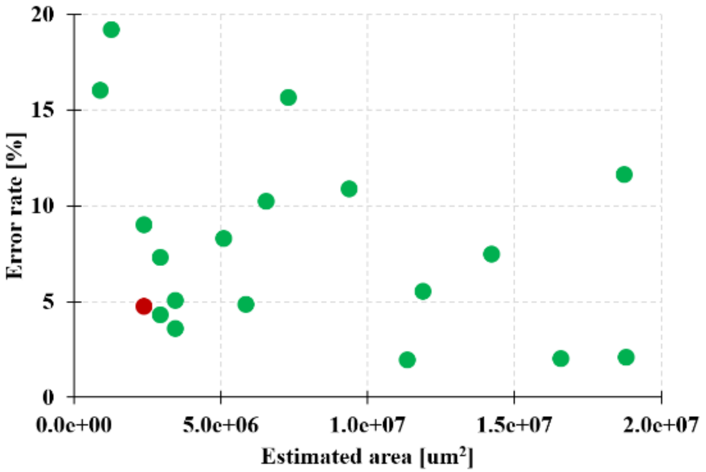
Write image & read image
(100ms)

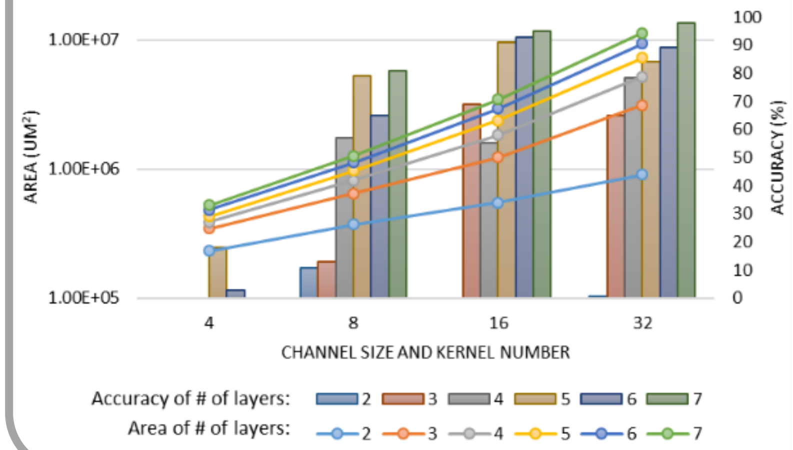
Object Detection SNN AI Chip

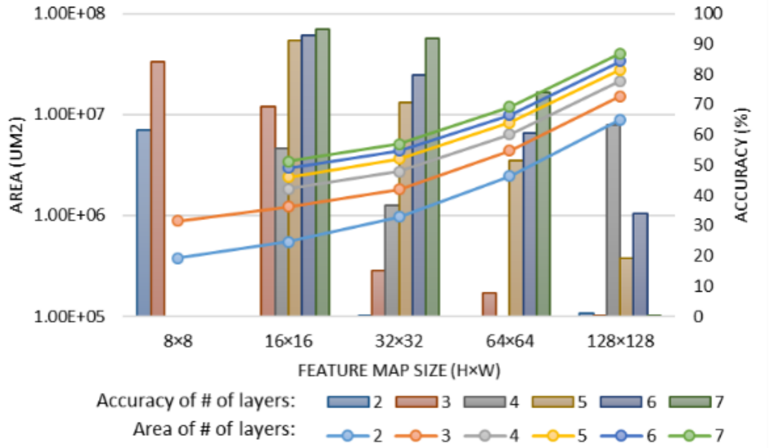
(Prof. Cheng-Wen Wu)

Application-specific NN-based AI accelerator design:

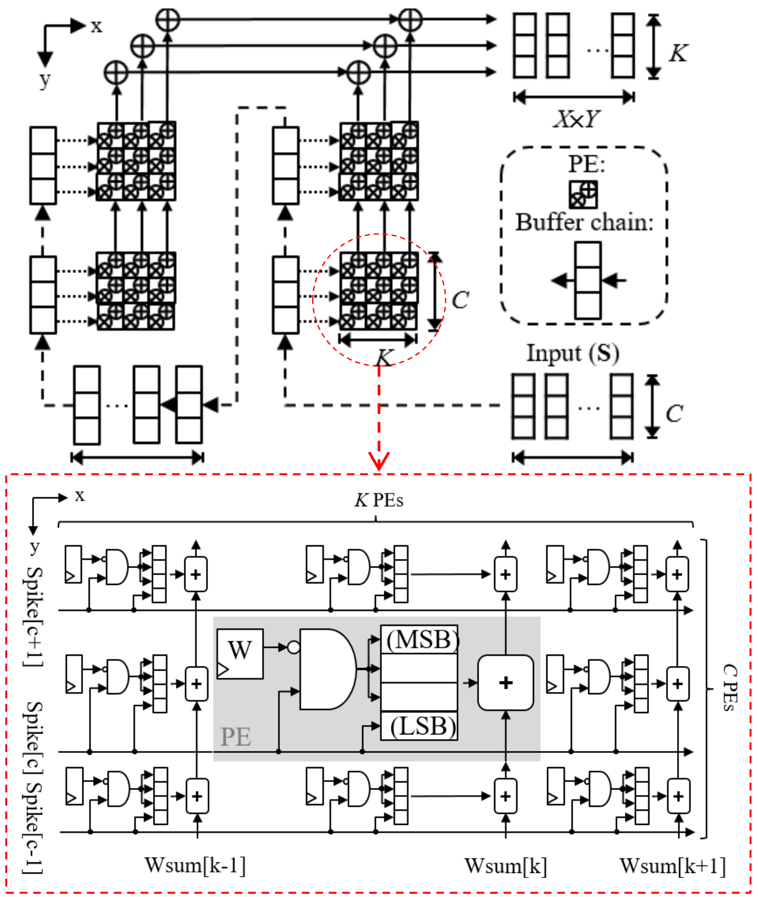
- Choose network parameters to optimize hardware cost and power consumption in terms of accuracy, network throughput, …







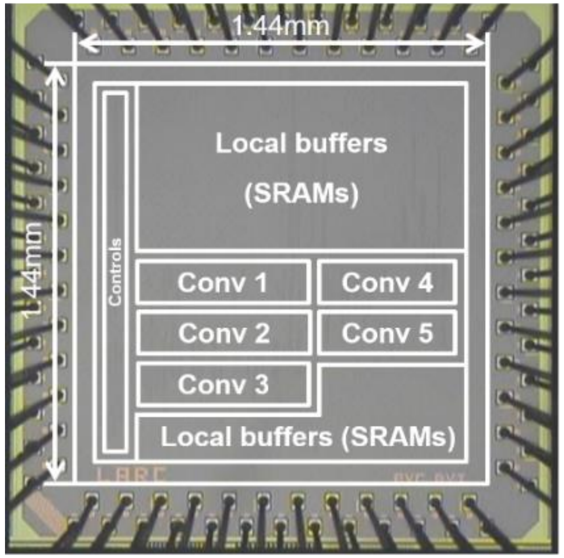
Spiking Neural Network (SNN)



Object Detection SNN AI Chip

(Prof. Cheng-Wen Wu)

- Based on spiking neural network
- Application-specific (label detection)
- Optimized parameters



Design	Ours	BinarEye CICC'18	IBM TrueNorth
Technology	90nm	28nm	28nm
Area [mm^2]	2.07	1.4	430
Algorithm	SNN	DNN	SNN
Voltage [V]	0.6 – 1	0.66 – 0.9	1
Frequency [MHz]	10 – 100	1.5 – 48	-
Efficiency* [TOPS/W]	331 – 45	230 – 145	-
MNIST Energy* [uJ/inference]	0.18 @ 98.01%	0.2 @ 97.4%	6.5 @ 97.5%

*values are scaled to 28nm technology node.

1.4X efficiency, 10% less power

Object Detection SNN AI Chip

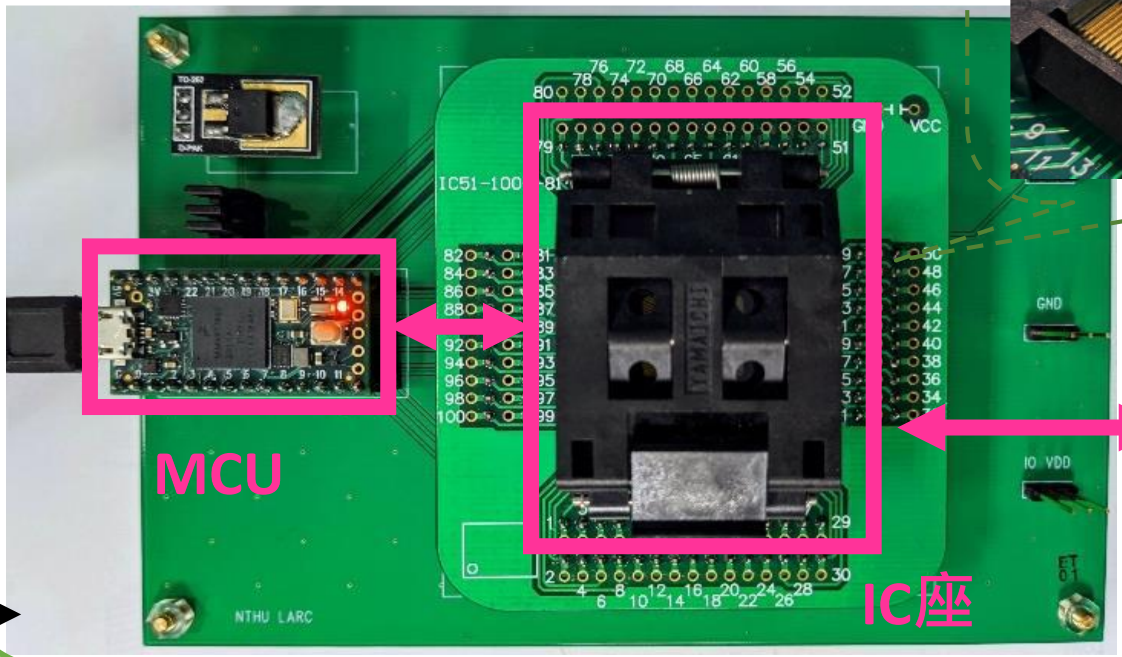
(Prof. Cheng-Wen Wu)

Control API

USB

Chip demo platform

- 300 detections/sec
- 0.6 V operating voltage



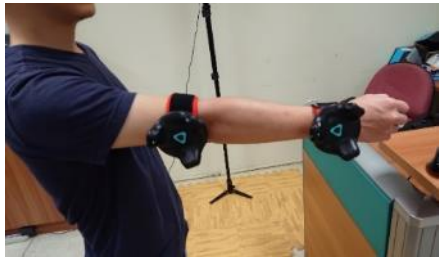
Imitation-based Training Platform (Prof. Chung-Ta King)

Demo data collection

Tracker

HTC Vive

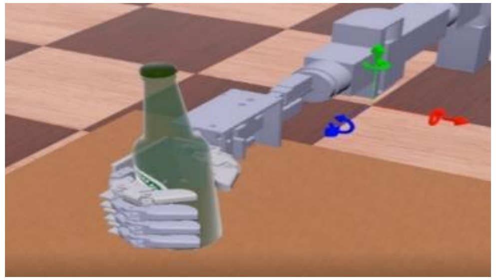
Manus VR



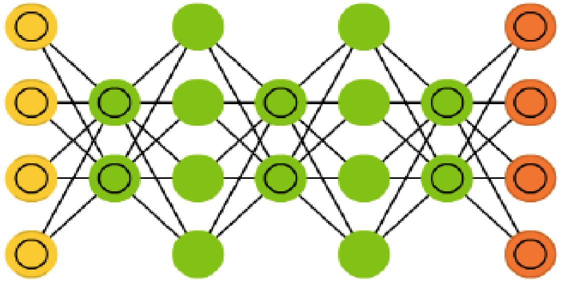
Control

Record

Training in simulator



Webots



Imitation Learning

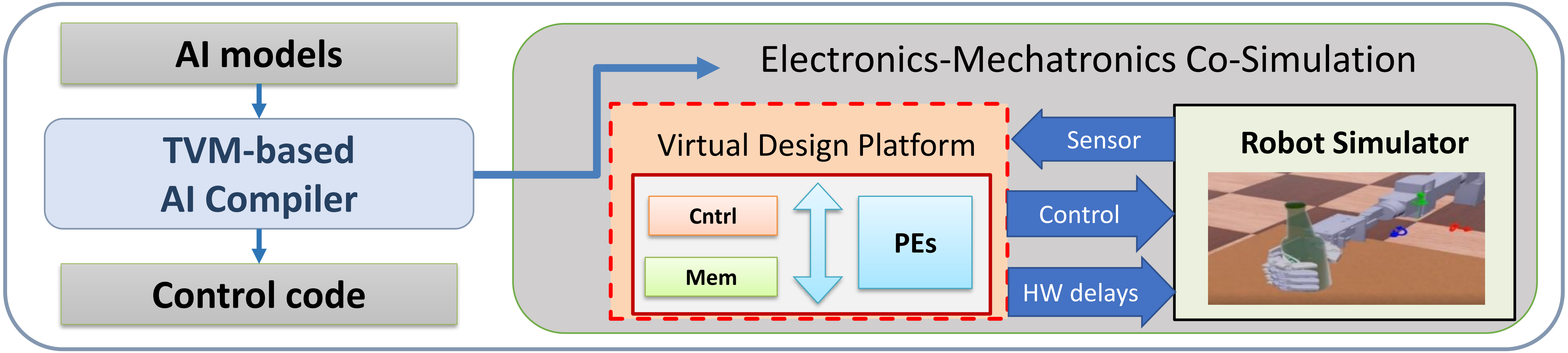
NN model

TVM compiler

Robot control code

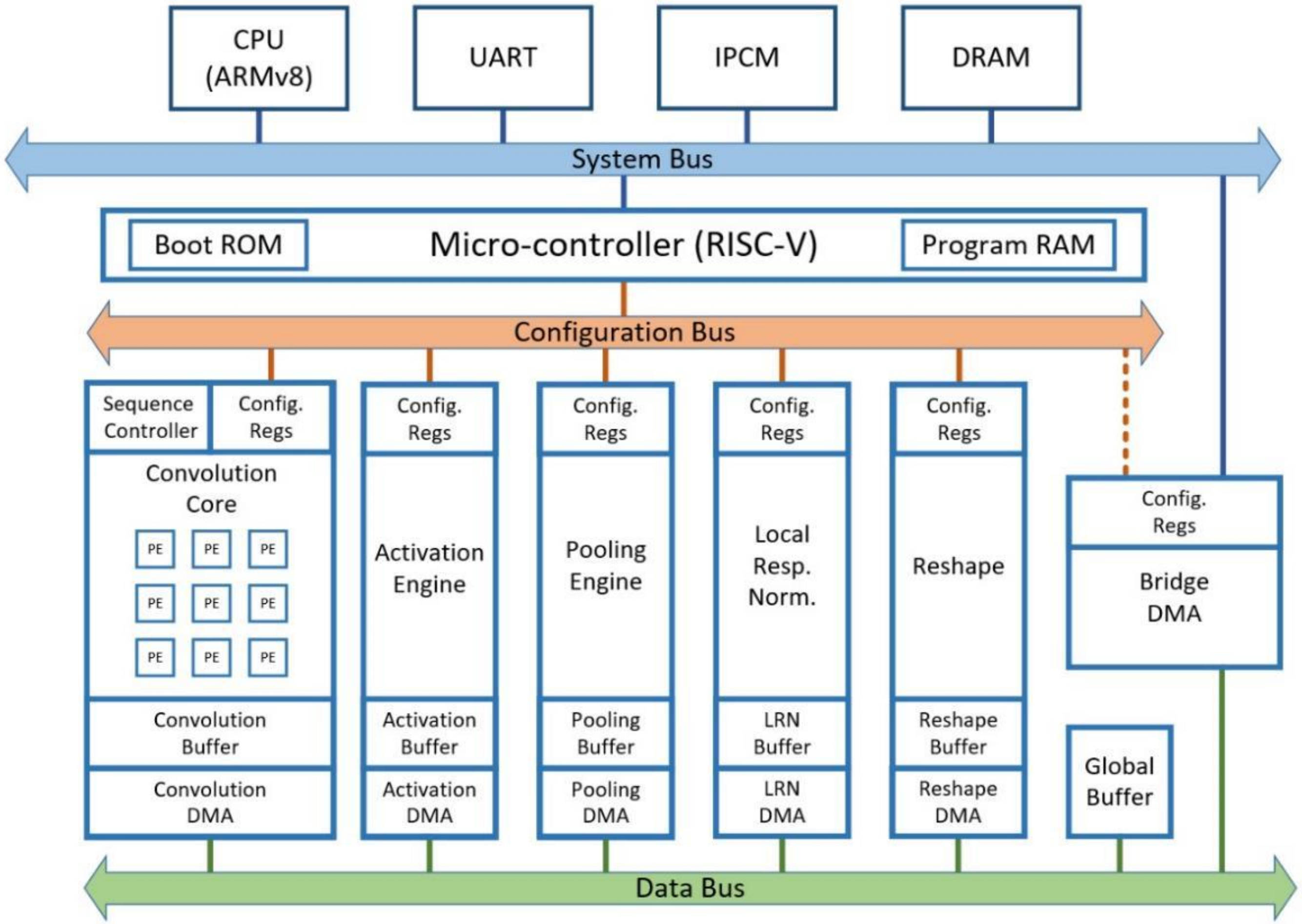
- Demonstrator looked at the virtual scenes to perform actions

Electronics-Mechatronics Co-Simulation (Prof. Jing-Jia Liou)



- SystemC and Webots data exchange interface
- Control hardware delays are feedback to robot arm
- Verify accuracies of AI models on robot arm operations

Architecture of Virtual Design Platform (Prof. Jing-Jia Liou)



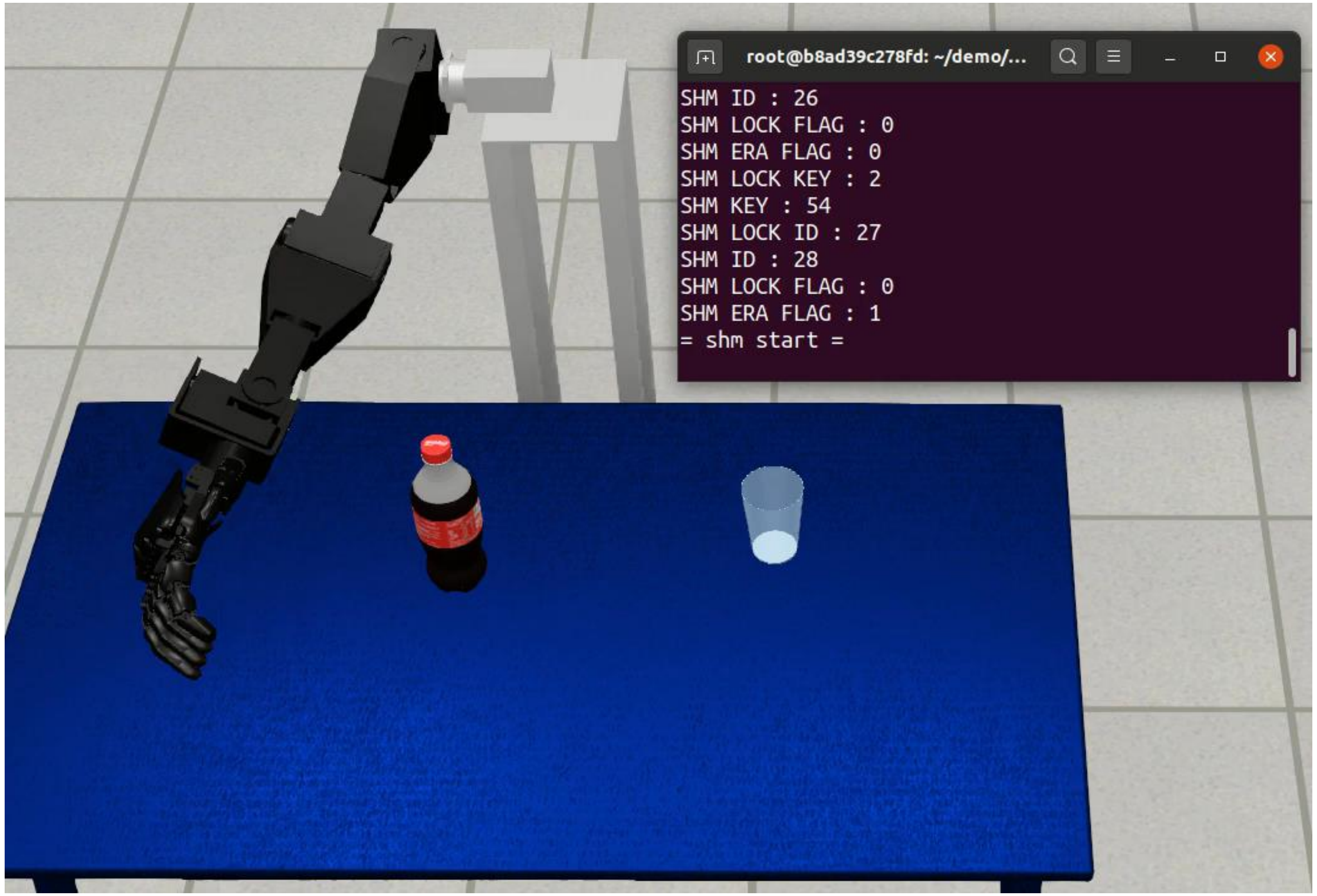
- Hardware components:
  - MCU, DMA
  - AI accelerator modules: Conv, Pool, Relu, etc.
  - Shared memory of AI accelerator modules
- RISC-V MCU for controlling dataflows of various AI accelerator modules
- All in SystemC

Electronics-Mechatronics Co-Simulation (Prof. Jing-Jia Liou)

AI model for robot arm
- Grasping
- 4 FC layers (64, 32, 16, 7)

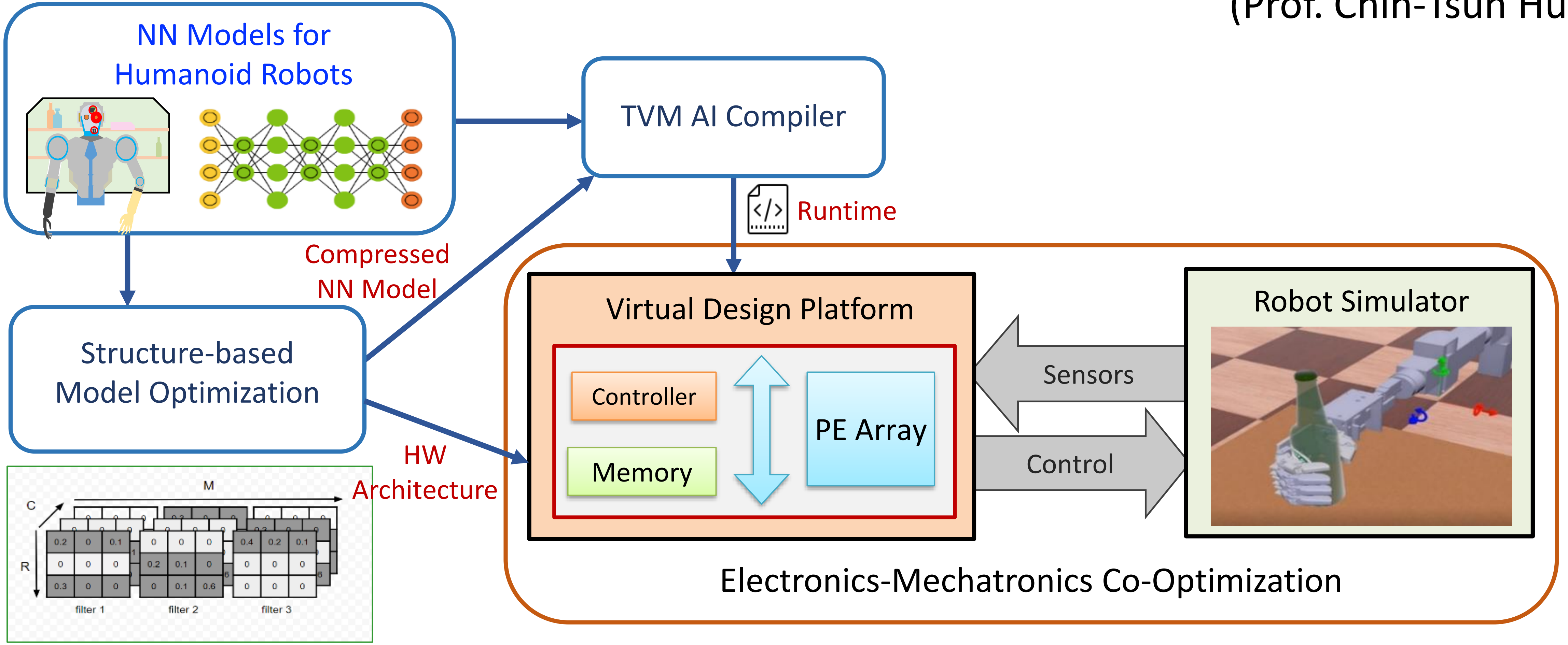
TVM optimization
- Pipelined execution
- Operator fusion
- 14%X speedup

Data exchange
- Delays of accelerator hardware to robot simulator



Electronics-Mechatronics Co-Optimization

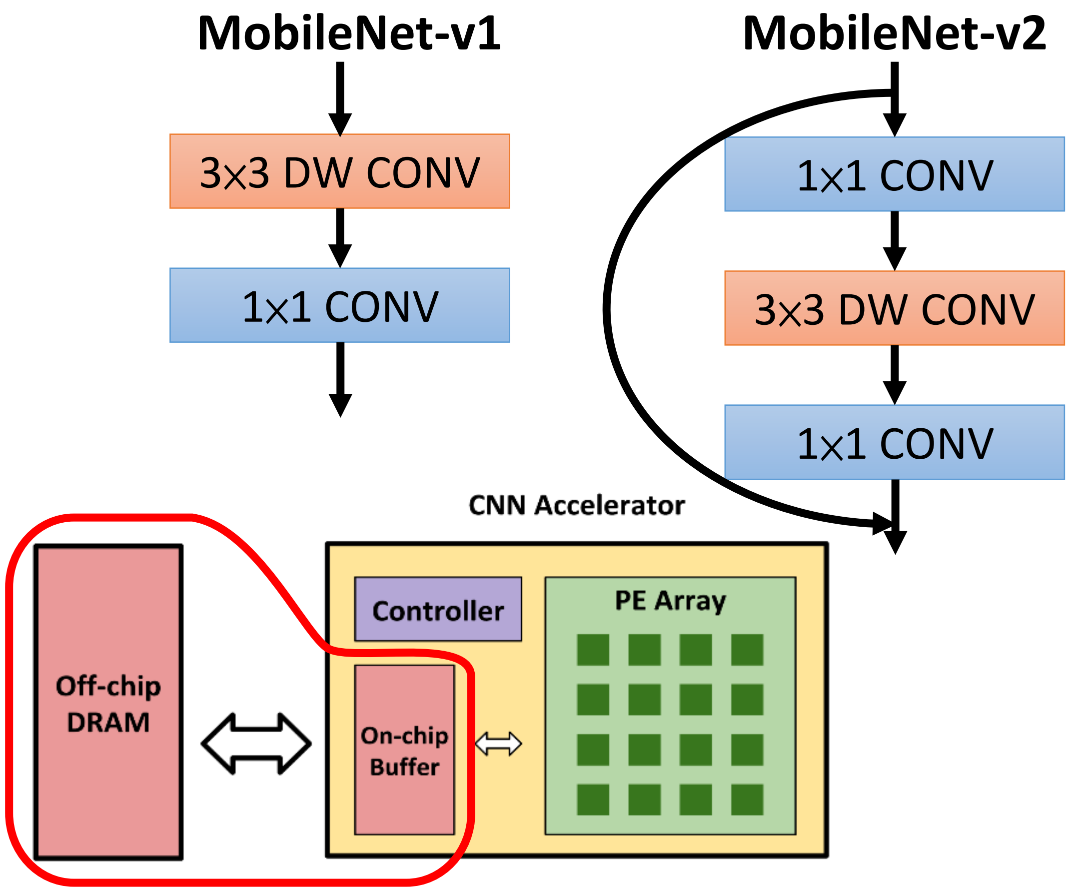
(Prof. Chih-Tsun Huang)



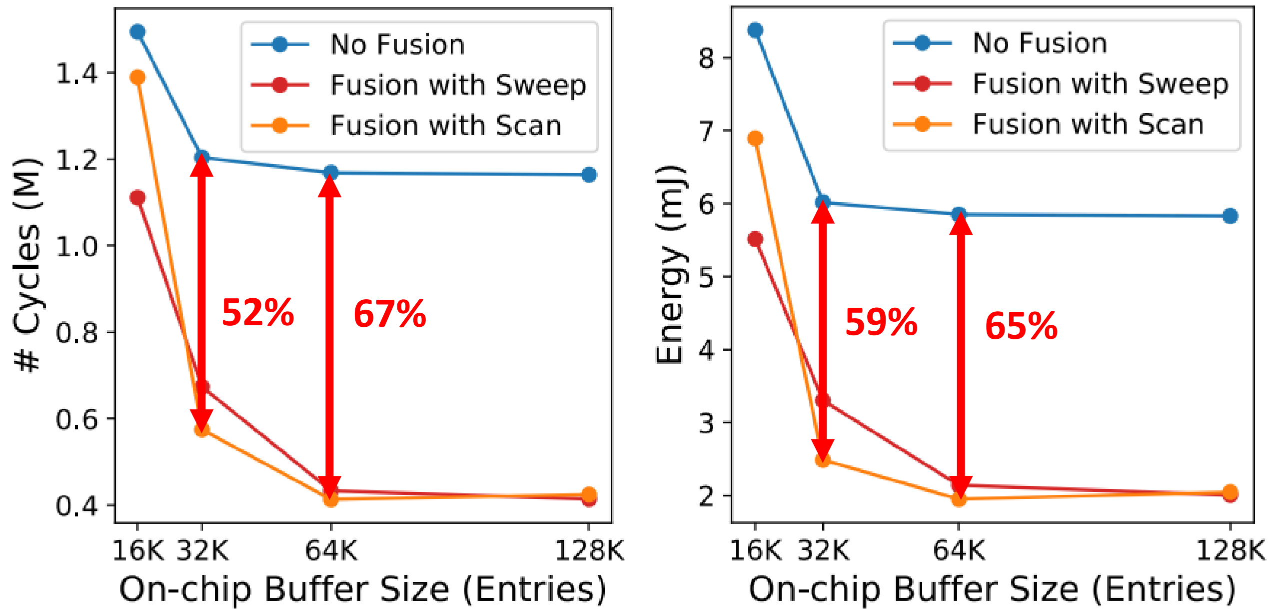
Layer Fusion for Data Reuse in Acc. Memory

- **For Depthwise Separable CNNs**

(Prof. Chih-Tsun Huang)



- **Evaluation for MobileNet-v2**
  - 67% speed improvement
  - 65% power consumption improvement



Conclusions

- An AI chip development environment based on human skill transfer
  - Automate the process and lower the complexity in developing service robot applications
  - Humanoid robot hand with 7-DOF arms, CMOS spiking neural network ASIC, design flows from human skills to AI model and to AI accelerator for robot control
  - Allow more complex robot applications to be more easily developed at a lower cost
- On-going works
  - Robot control with touch sensor feedbacks, memristor-based SNN, long-horizon task planning, hierarchical control models, …

Q&A