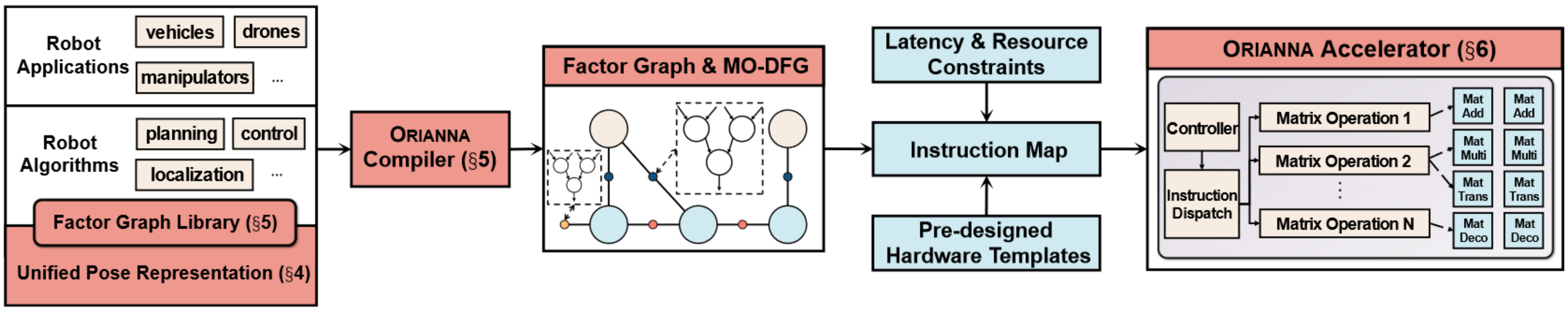
ASPLOS 2024

ORIANNA: An Accelerator Generation Framework for Optimization-based Robotic Applications

Yuhui Hao[1], Yiming Gan[2], Bo Yu[3], Qiang Liu[1], Yinhe Han[2], Zishen Wan[4], Shaoshan Liu[3]

1. Tianjin University
2. Institute of Computing Technology, Chinese Academy of Sciences
3. Shenzhen Institute of Artificial Intelligence and Robotics for Society
4. Georgia Tech

ORIANNA Overview



Robot Applications: vehicles, drones, manipulators ...

Robot Algorithms: planning, control, localization ...

Factor Graph Library (§5)

Unified Pose Representation (§4)

ORIANNA Compiler (§5)

Factor Graph & MO-DFG

Latency & Resource Constraints

Instruction Map

Pre-designed Hardware Templates

ORIANNA Accelerator (§6)

Controller

Instruction Dispatch

Matrix Operation 1

Matrix Operation 2

⋮

Matrix Operation N

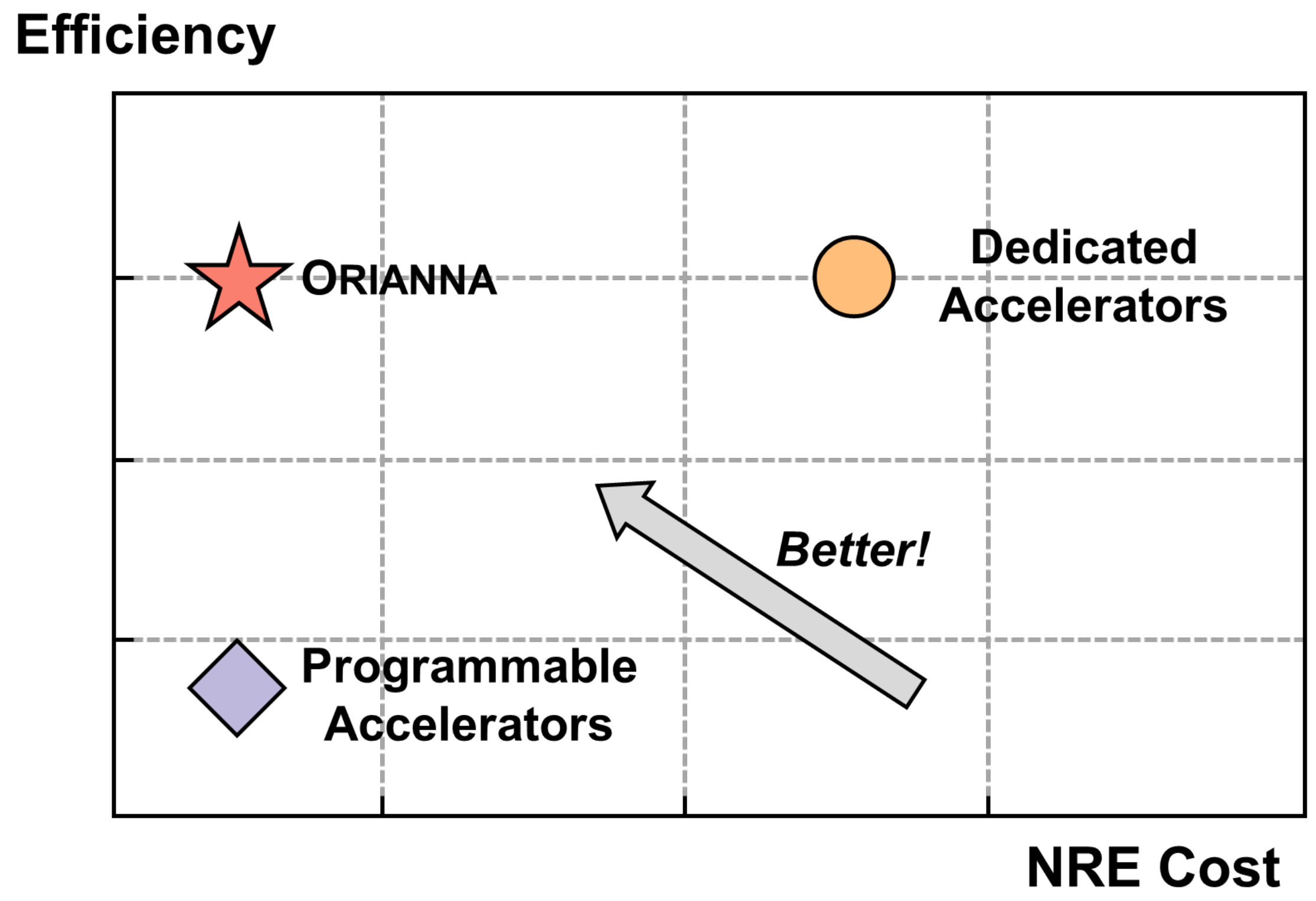
Mat Add | Mat Add

Mat Multi | Mat Multi

Mat Trans | Mat Trans

Mat Deco | Mat Deco

Motivation



Software Framework

Factor Type	Factor	Algorithm
Measurement	LiDAR, Camera, GPS, IMU	Localization
Constraint	Smooth, Collision-free, Prior, Kinematics, Dynamics	Planning, Control

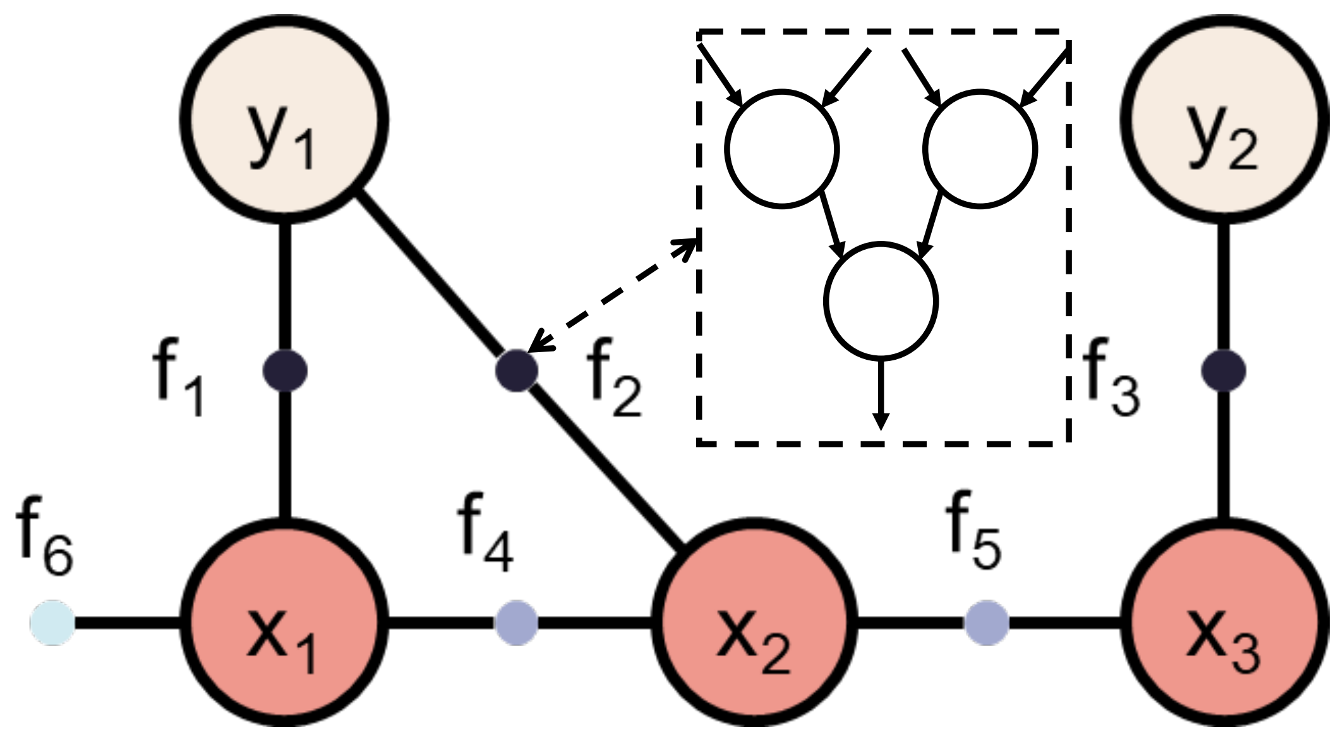
Factors in ORIANNA Factor Graph Library

User High Level Codes

```
# Localization Factor graph
graph.add(CameraFactor(x1, y1, m1))
graph.add(CameraFactor(x2, y1, m2))
graph.add(CameraFactor(x3, y1, m3))
graph.add(IMUFactor(x1, x2, m4))
graph.add(IMUFactor(x2, x3, m5))
graph.add(PriorFactor(x1, p1))
graph.optimize()
```

ORIANNA Compiler

Factor Graph & MO-DFG



Hardware Generation

$$\boxed{p_1^*, p_2^*, \cdots, p_n^*} = \underset{p_1, p_2, \cdots, p_n}{\arg\min} \boxed{L(p_1, p_2, \ldots, p_n)}$$

$$s.t. \quad \boxed{R(p_1, p_2, \ldots, p_n)} \leq \boxed{R^*},$$

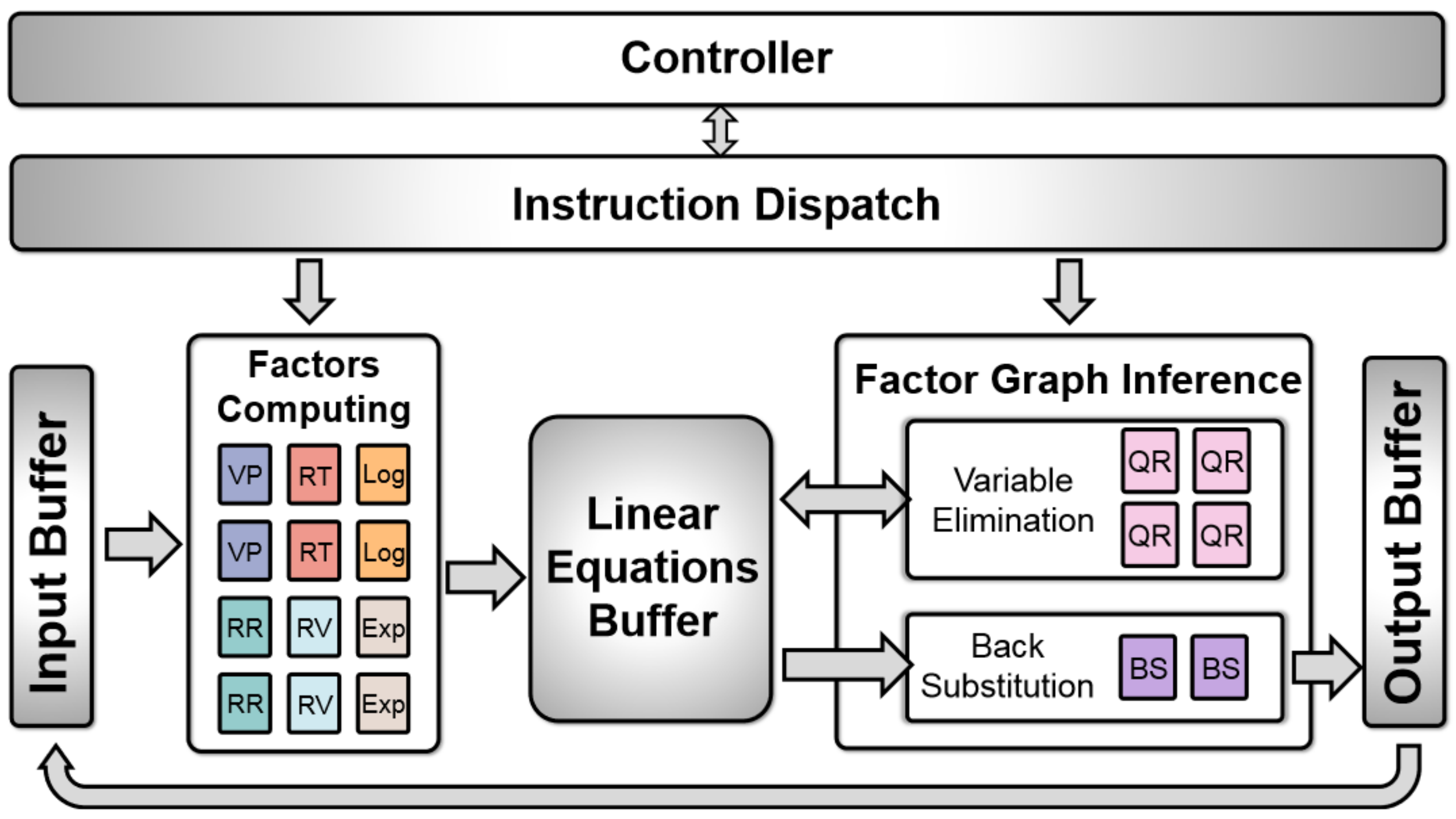
the number of replicated computation units employed for various matrix operations

hardware resource consumption

available on-chip resources

computation latency

Hardware Generation



Controller

Instruction Dispatch

Input Buffer

Factors Computing

VP | RT | Log

VP | RT | Log

RR | RV | Exp

RR | RV | Exp

Linear Equations Buffer

Factor Graph Inference

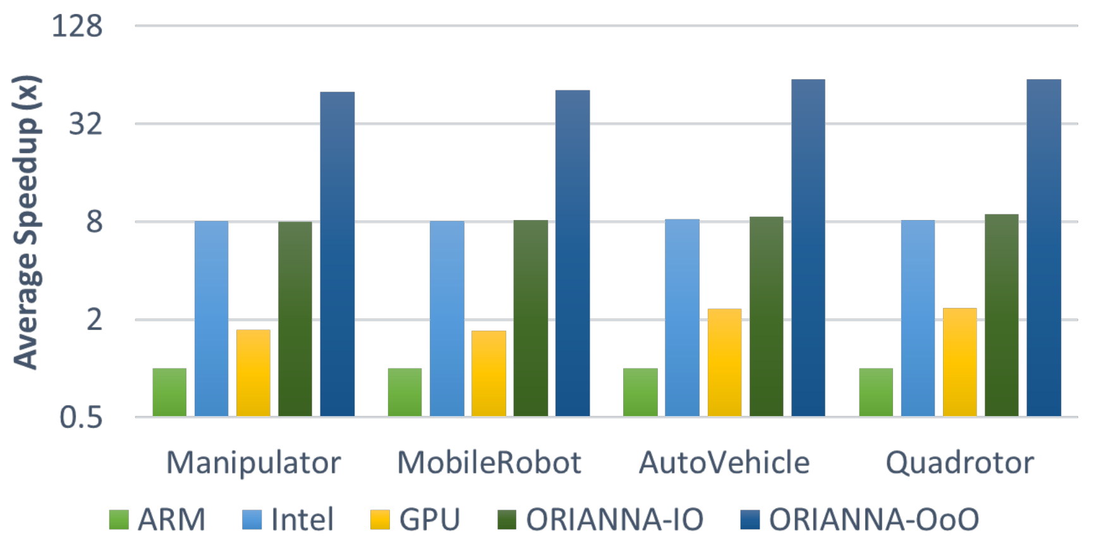
Variable Elimination: QR QR QR QR

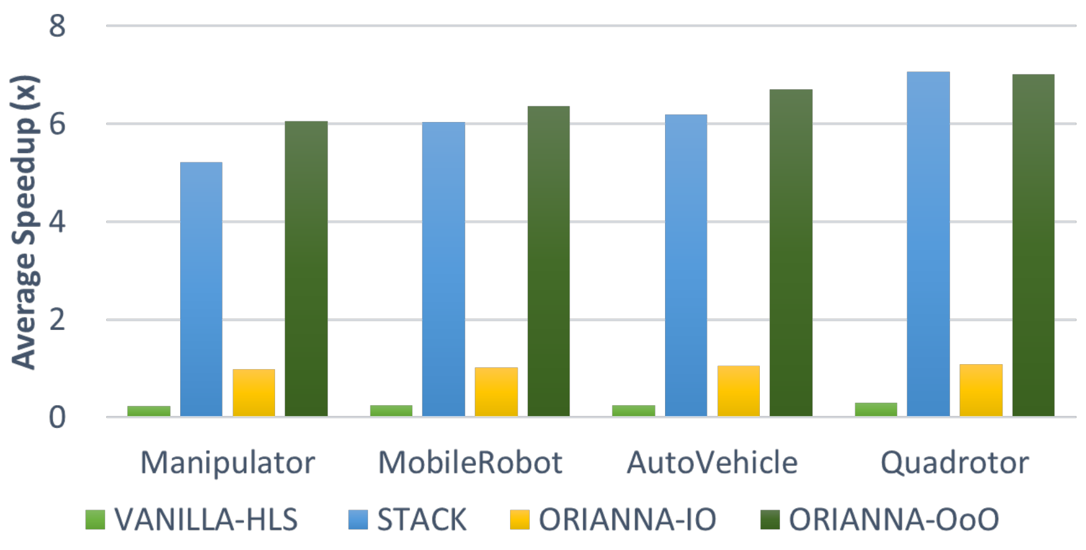
Back Substitution: BS BS

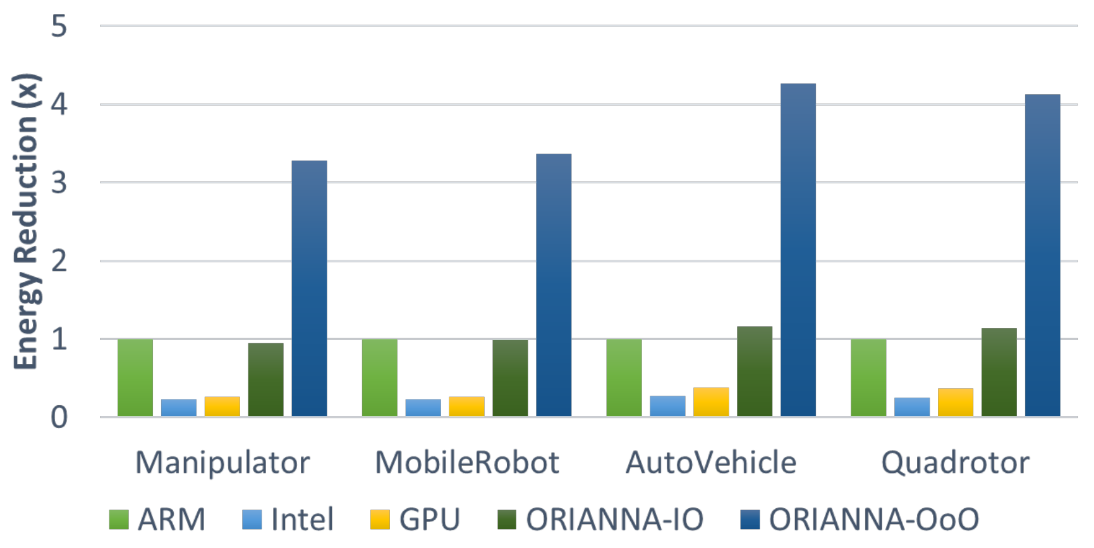
Output Buffer

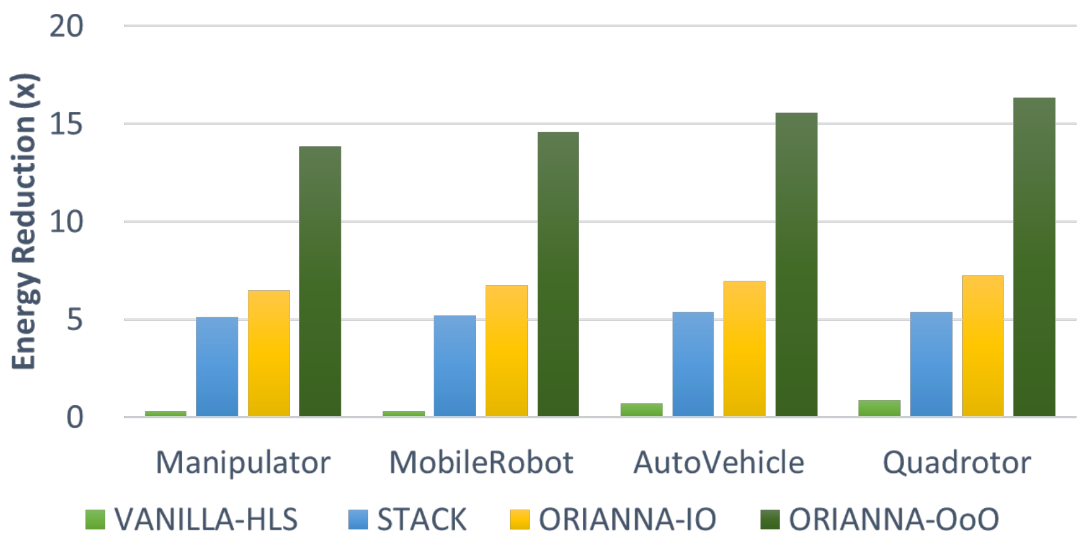
Overall Hardware Architecture

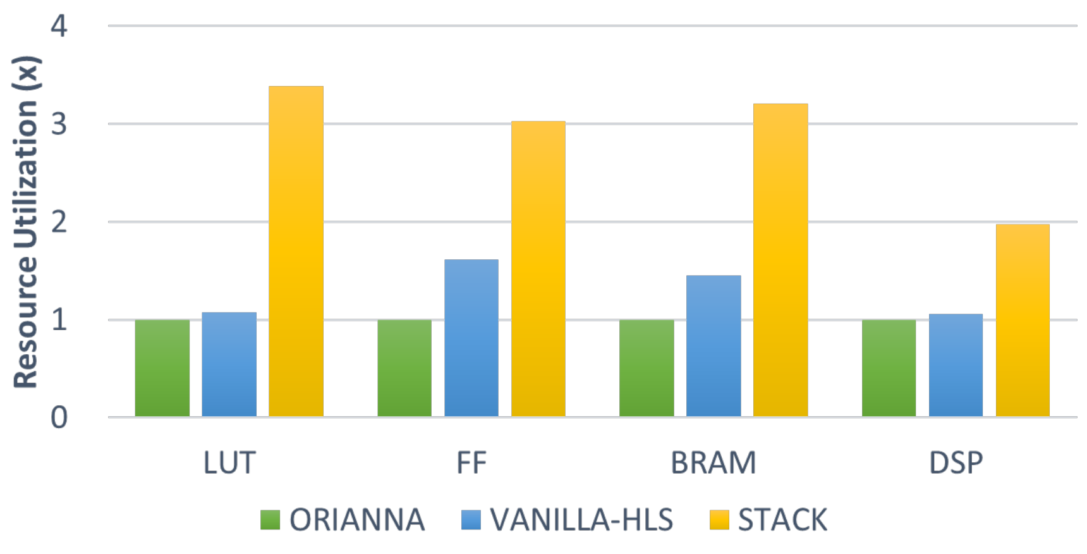
Experimental Results


Average Speedup (x)
Manipulator
MobileRobot
AutoVehicle
Quadrotor
ARM
Intel
GPU
ORIANNA-IO
ORIANNA-OoO


Average Speedup (x)
Manipulator
MobileRobot
AutoVehicle
Quadrotor
VANILLA-HLS
STACK
ORIANNA-IO
ORIANNA-OoO


Energy Reduction (x)
Manipulator
MobileRobot
AutoVehicle
Quadrotor
ARM
Intel
GPU
ORIANNA-IO
ORIANNA-OoO


Energy Reduction (x)
Manipulator
MobileRobot
AutoVehicle
Quadrotor
VANILLA-HLS
STACK
ORIANNA-IO
ORIANNA-OoO


Resource Utilization (x)
LUT
FF
BRAM
DSP
ORIANNA
VANILLA-HLS
STACK