

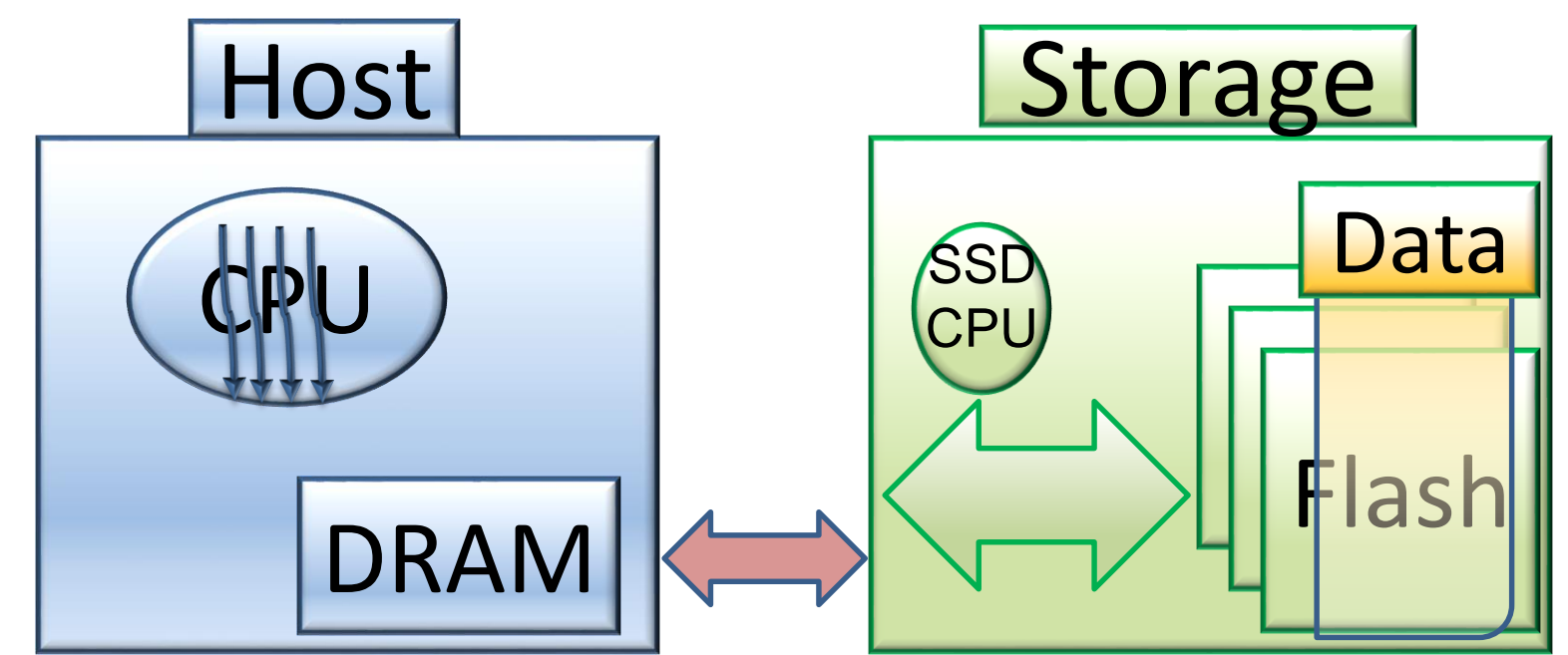
# Summarizer: Trading Communication with Computing Near Storage

## Motivation

- Off-chip interconnection network between the host and the storage becomes the bottleneck for many large-scale applications
- Near-data processing (NDP) improves application performance

## NDP considerations

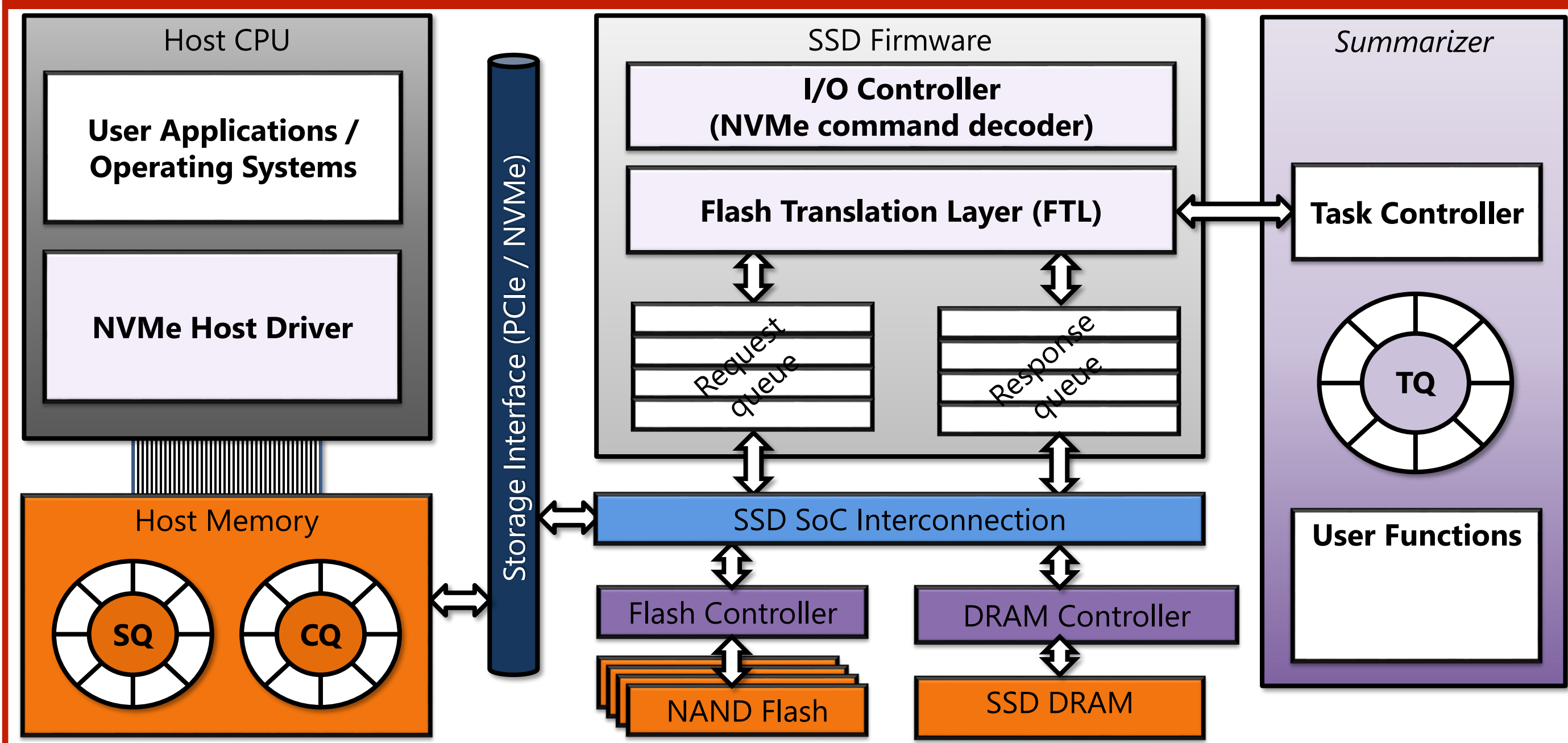
- Wimpy cores
- Dynamic usage of cores
  - Garbage collection, wear levelling, etc.
  - Available computation power to host applications varies dynamically



## Overview

- Summarizer is a framework to support programmer for efficient NDP
  - Dynamically offloads appropriate amount of work
  - Easy to use API
- Key idea: Opportunistically offload work for NDP

## Architecture



## Summarizer API

- Summarizer API
  - Initialization – Transfers in-SSD procedure to SSD memory
  - Computation – Opportunistically executes at SSD
  - Finalization – Gather final in-SSD computation results
- Extended current NVMe commands

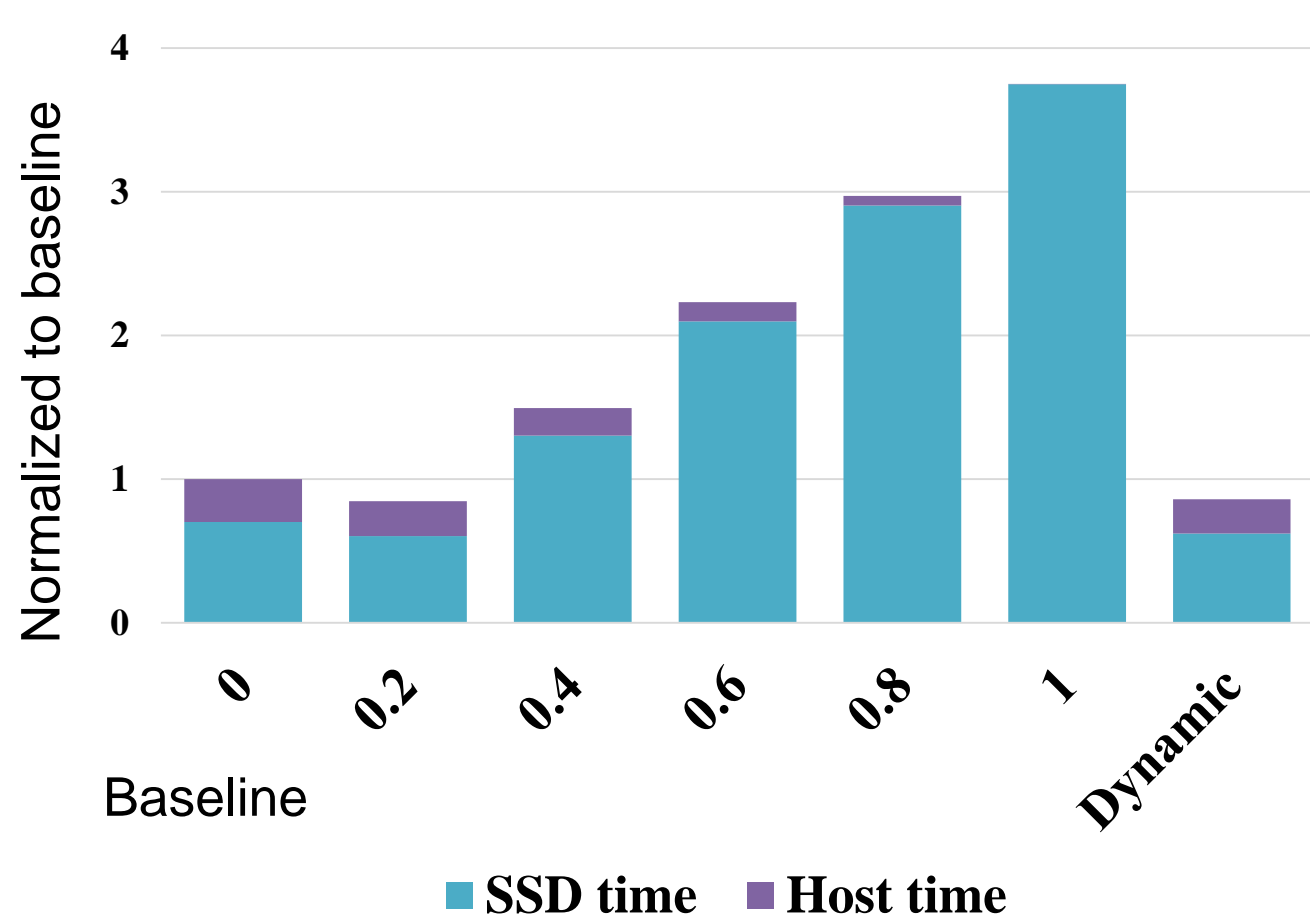
## Evaluation platform

- LS2085a intelligent SSD development platform
- PCIe Gen. 3 4x
- ARM cores – SSD firmware
- FPGA - NAND flash controller



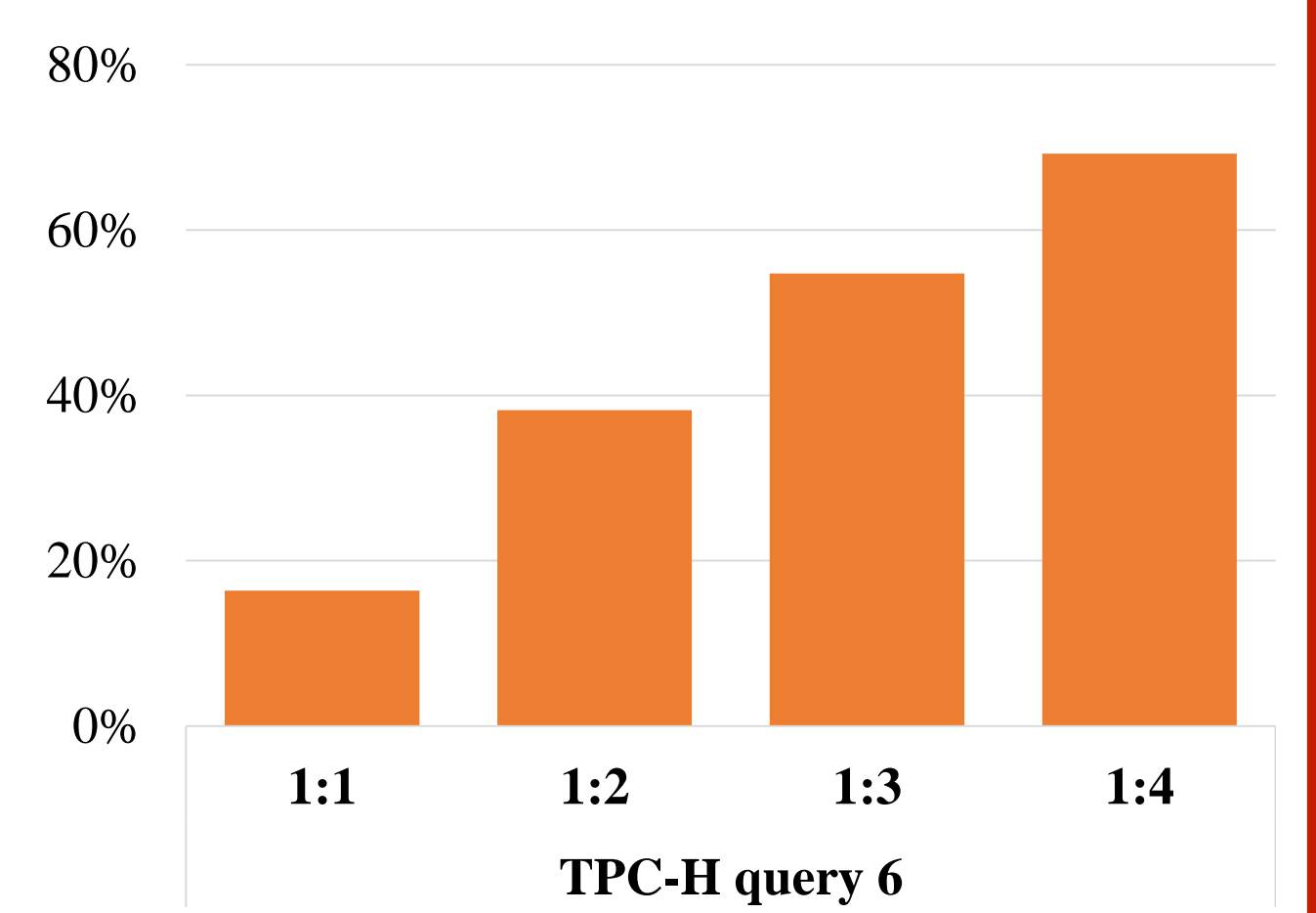
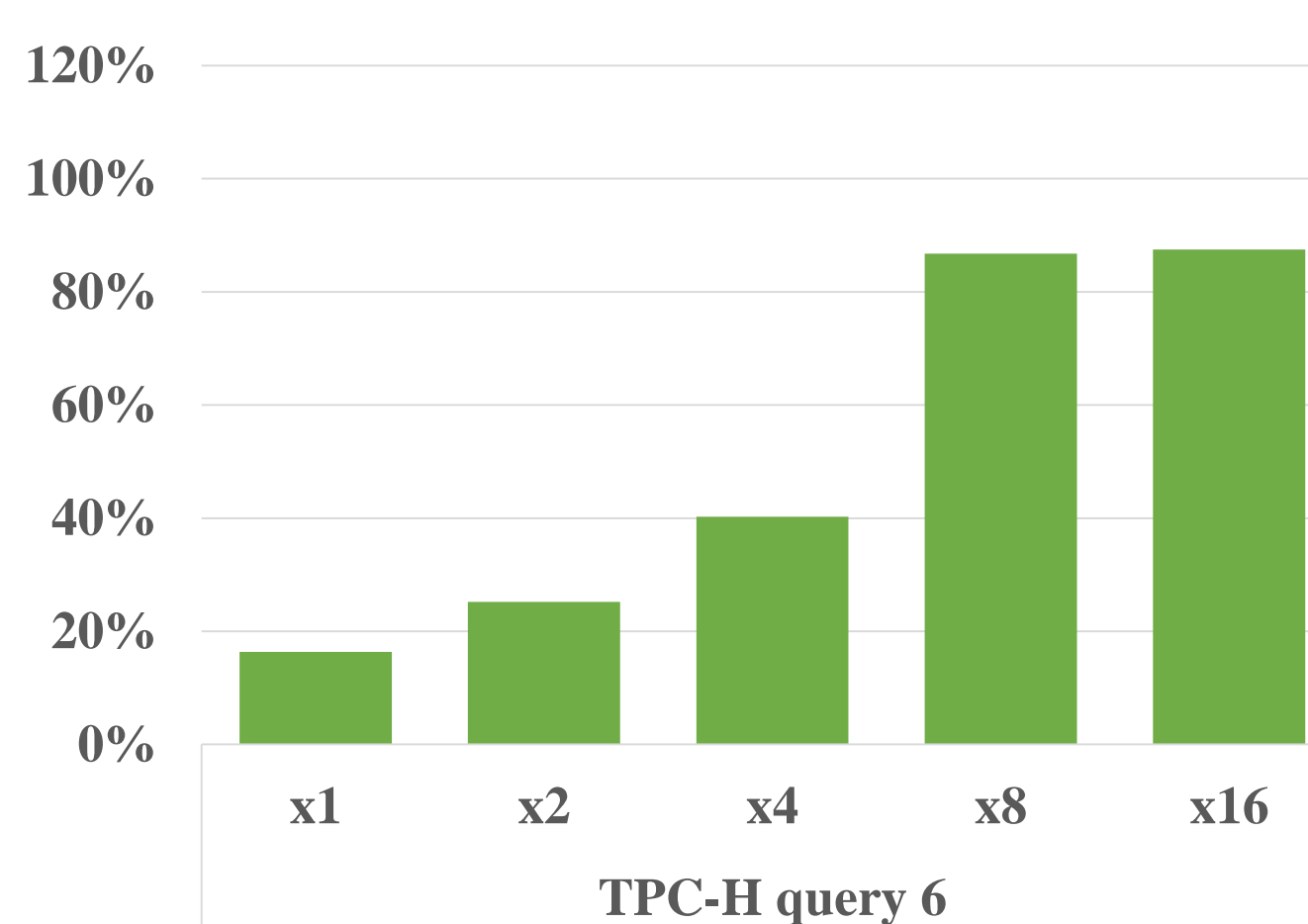
## Results

- 16% improvement for TPC-H query 6

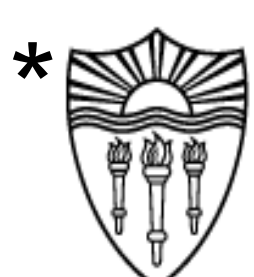


## Design space exploration

- With powerful embedded processors
- With higher internal bandwidth
  - External : Internal bandwidth ratio



Gunjae Koo<sup>\*=</sup>, Kiran Kumar Matam<sup>\*=</sup>, Te It<sup>†</sup>, H.V. Krishina Giri Nara<sup>\*</sup>, Jing Li<sup>‡</sup>, Hung-Wei Tseng<sup>†</sup>, Steven Swanson<sup>‡</sup>, Murali Annavaram<sup>\*</sup>  
<sup>=</sup>Gunjae and Kiran contributed equally to the work



USC

NC STATE UNIVERSITY

UC San Diego