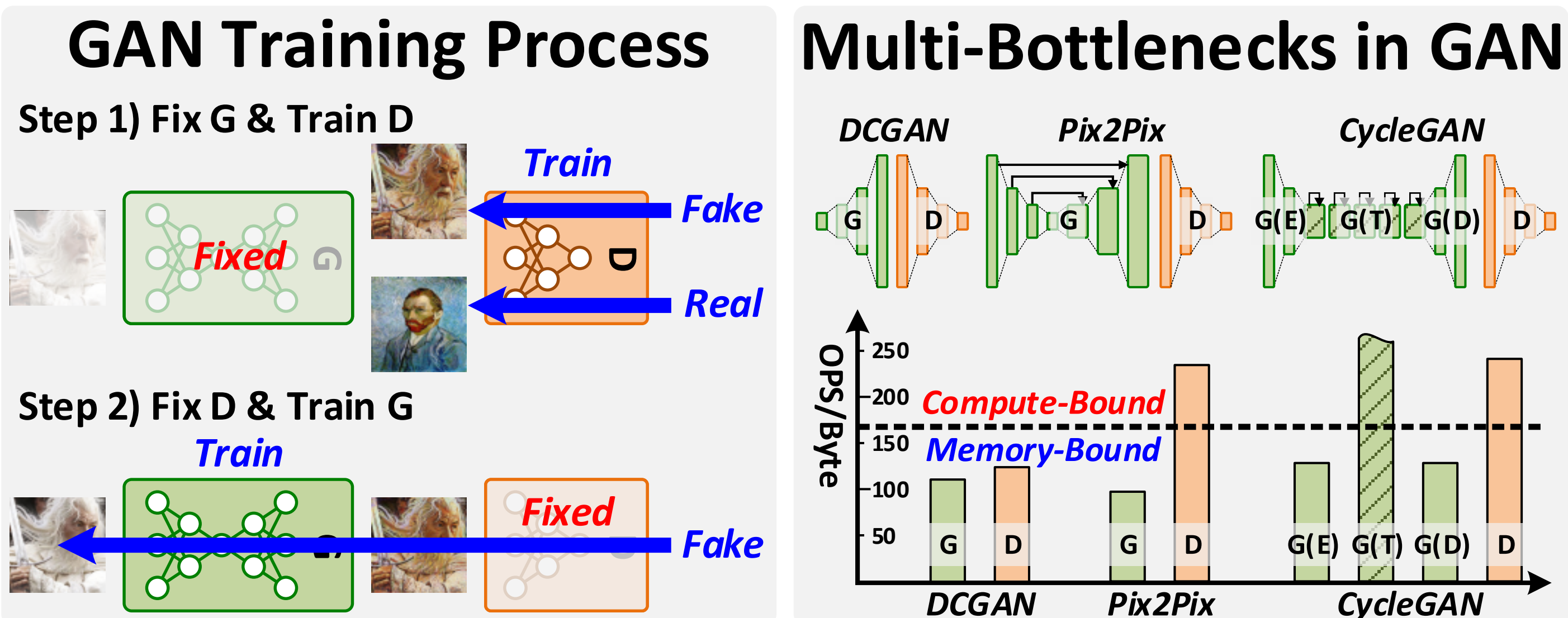


An Energy-Efficient Multi-DNN Training Processor for GANs with Speculative Dual-Sparsity Exploitation

S. Kang, D. Han, J. Lee, D. Im, S. Kim, S. Kim, J. Ryu, and H-J. Yoo (KAIST)

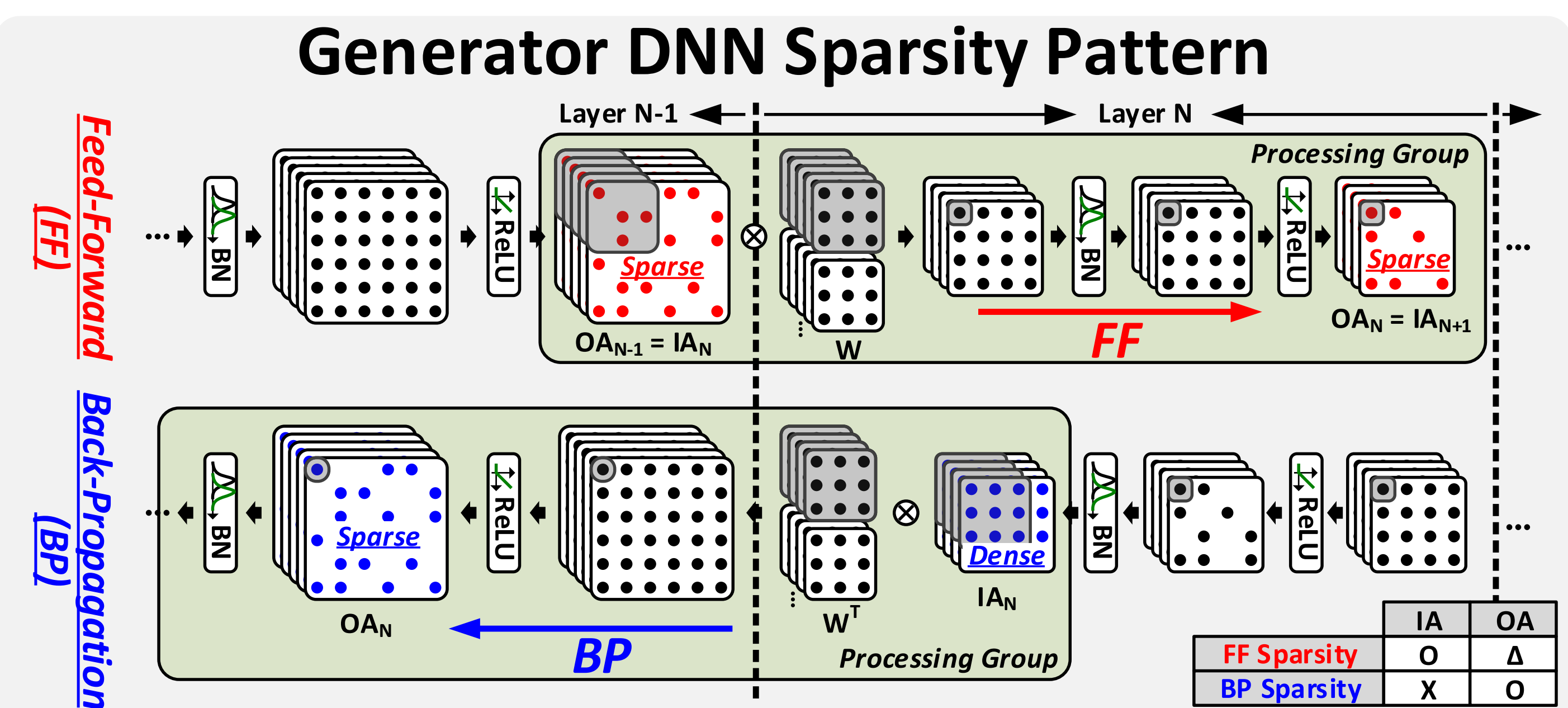
Motivation

1) Multi-DNN Architectures of GANs



- Different bottlenecks co-exist in multi-DNNs (Computation bottleneck & Memory bottleneck)

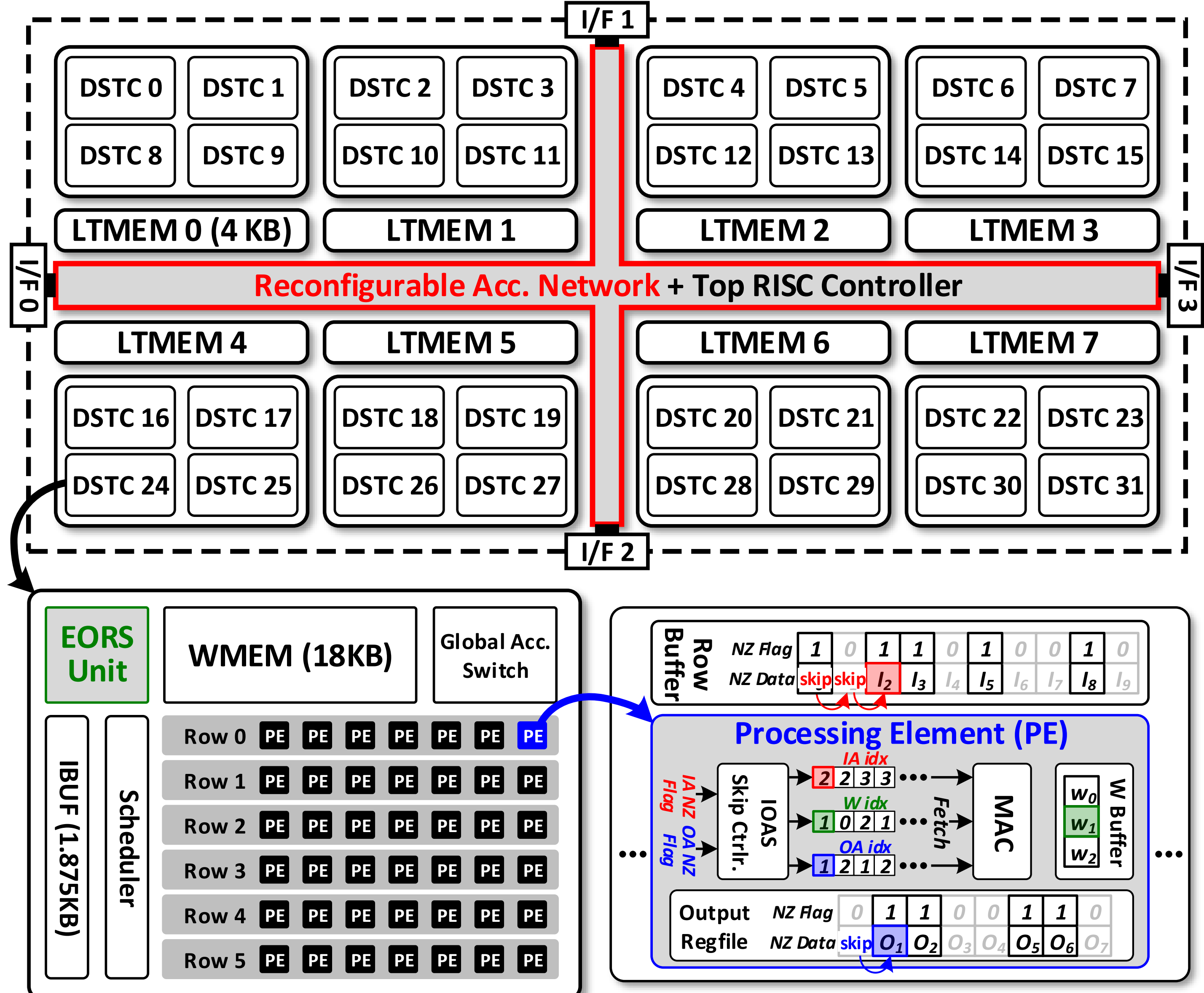
2) Sparsity Pattern of DNN Training



- FF stage:** Both IA sparsity & OA sparsity (unknown)
- BP stage:** Only OA sparsity

Architecture

Overall Architecture



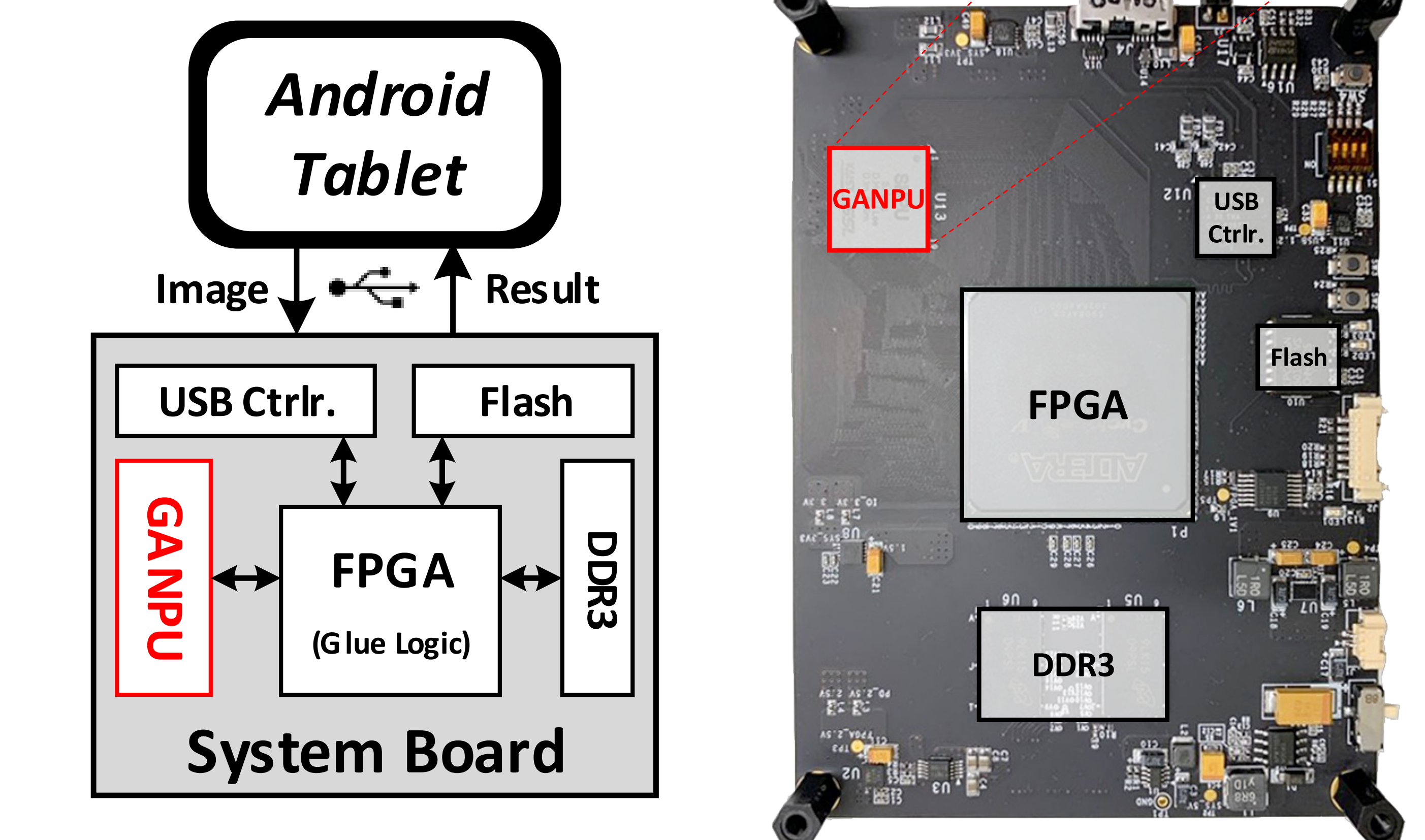
- Total of 32 training cores (DSTC) & Shared memory
- Adaptive spatio-temporal workload multiplexing with reconfigurable accumulation network**
- Dual sparsity (IA & OA) exploitation architecture**
- Exponent-only ReLU speculation (EORS) to utilize OA sparsity**

System Implementation

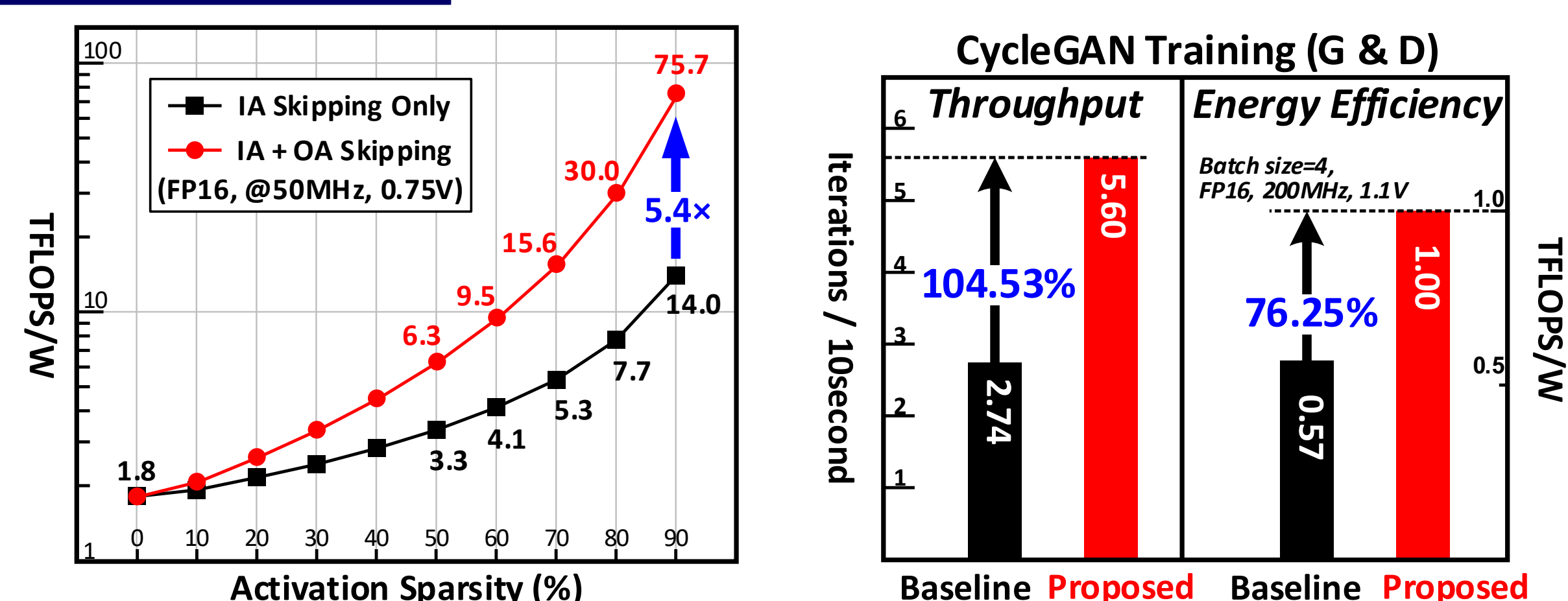
Demonstration System



System Architecture



Verification



Face Modification w/ On-Device Training

