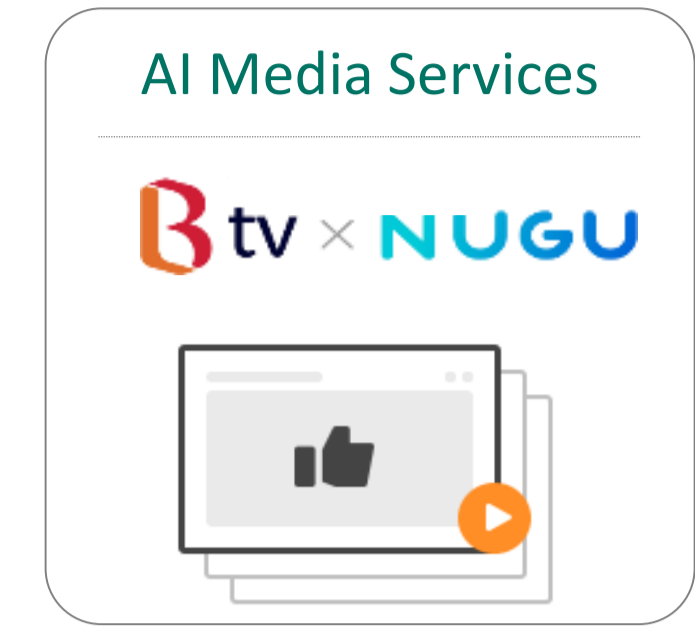
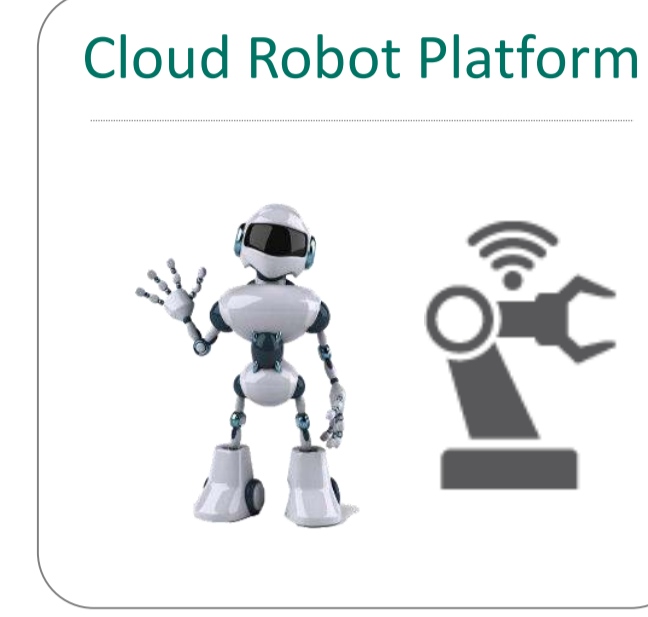


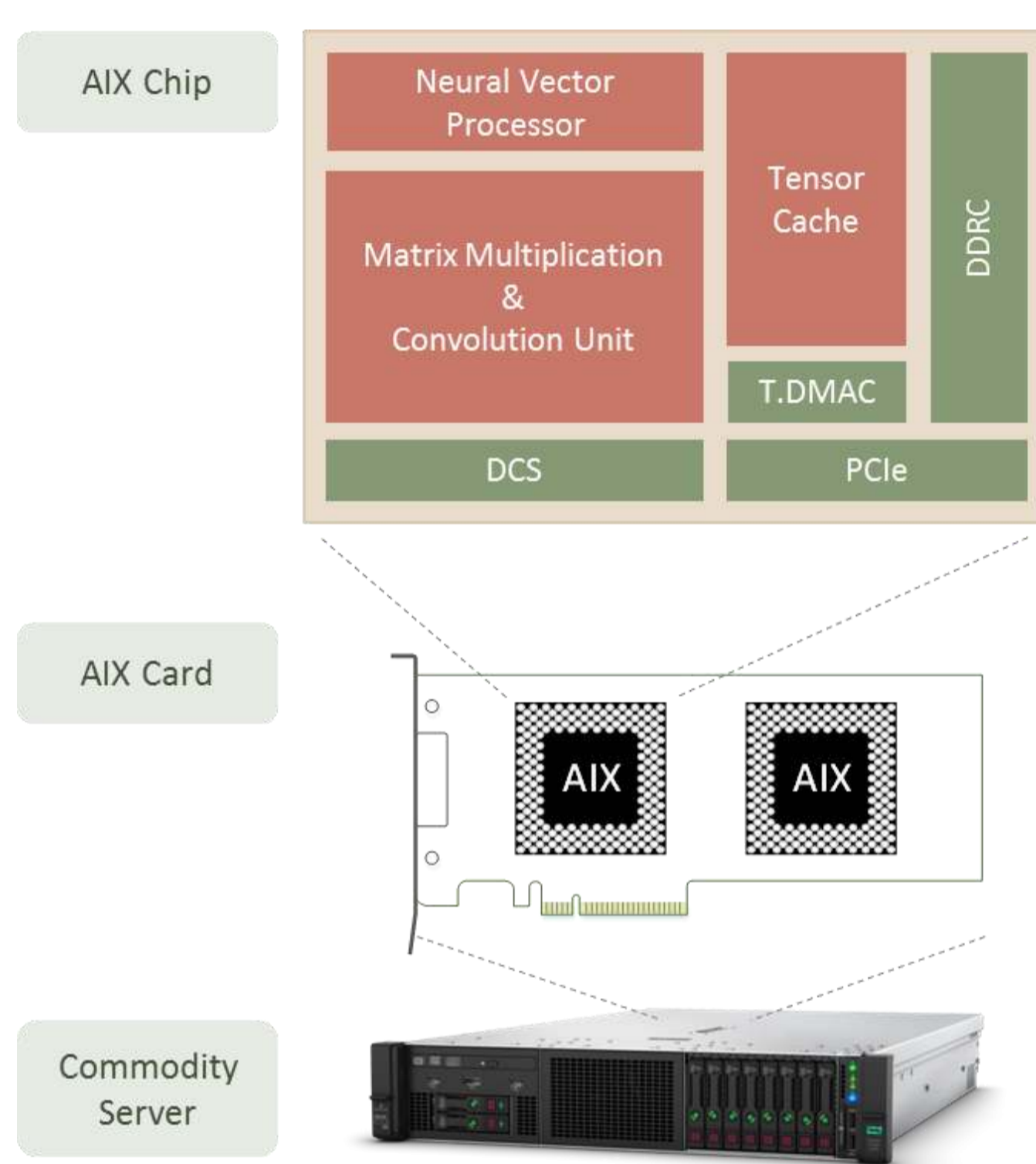
AIX v2: Flexible High Performance AI Inference Accelerator for Datacenters

SKT: from Telecommunication Company to AI Service Company

SK telecom is Korea's largest telecommunications company with more than 29 million mobile subscribers, which accounts for around 50 percent of the total market.



AIX v2 - AI Inference Accelerator for Datacenters



- Static Compute Optimization**
 - Computation Graph Optimization
 - Quantization, Pruning
 - Model Parallel Execution
- Dynamic Compute Optimization**
 - Dynamic Dataflow Graph Scheduling
 - 2xINT8 MACs per DSP
 - Dynamic Precision with FP Scaler
- Ultimate Utilization (Xilinx FPGA)**
 - 775 MHz MAC Frequency
 - 84% of DSPs



AIX w/ ALVEO U250

VU13P (Xilinx VIRTEX UltraSCALE+)

29 TOPS

77 GB/s (DRAM)

200 W



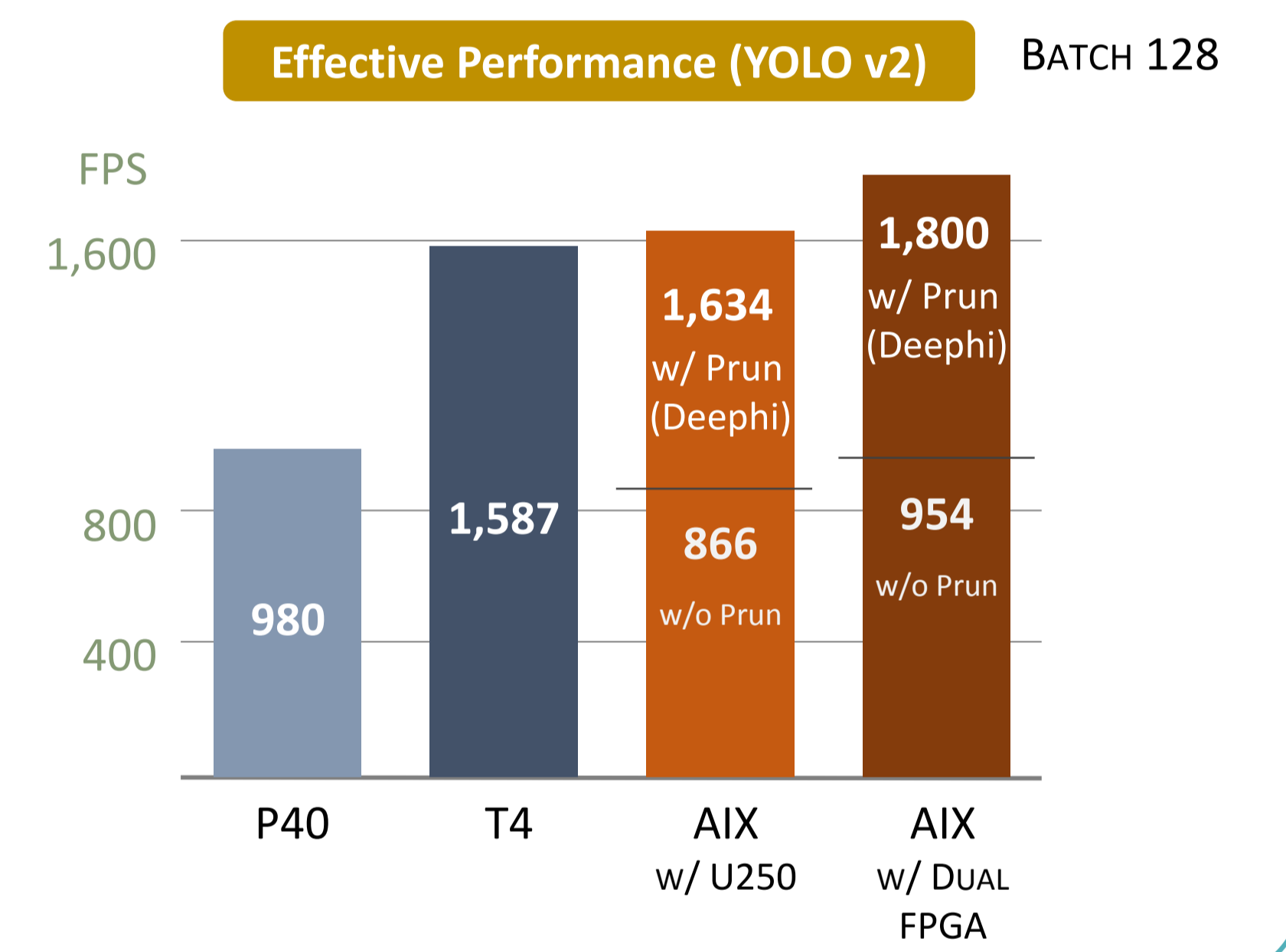
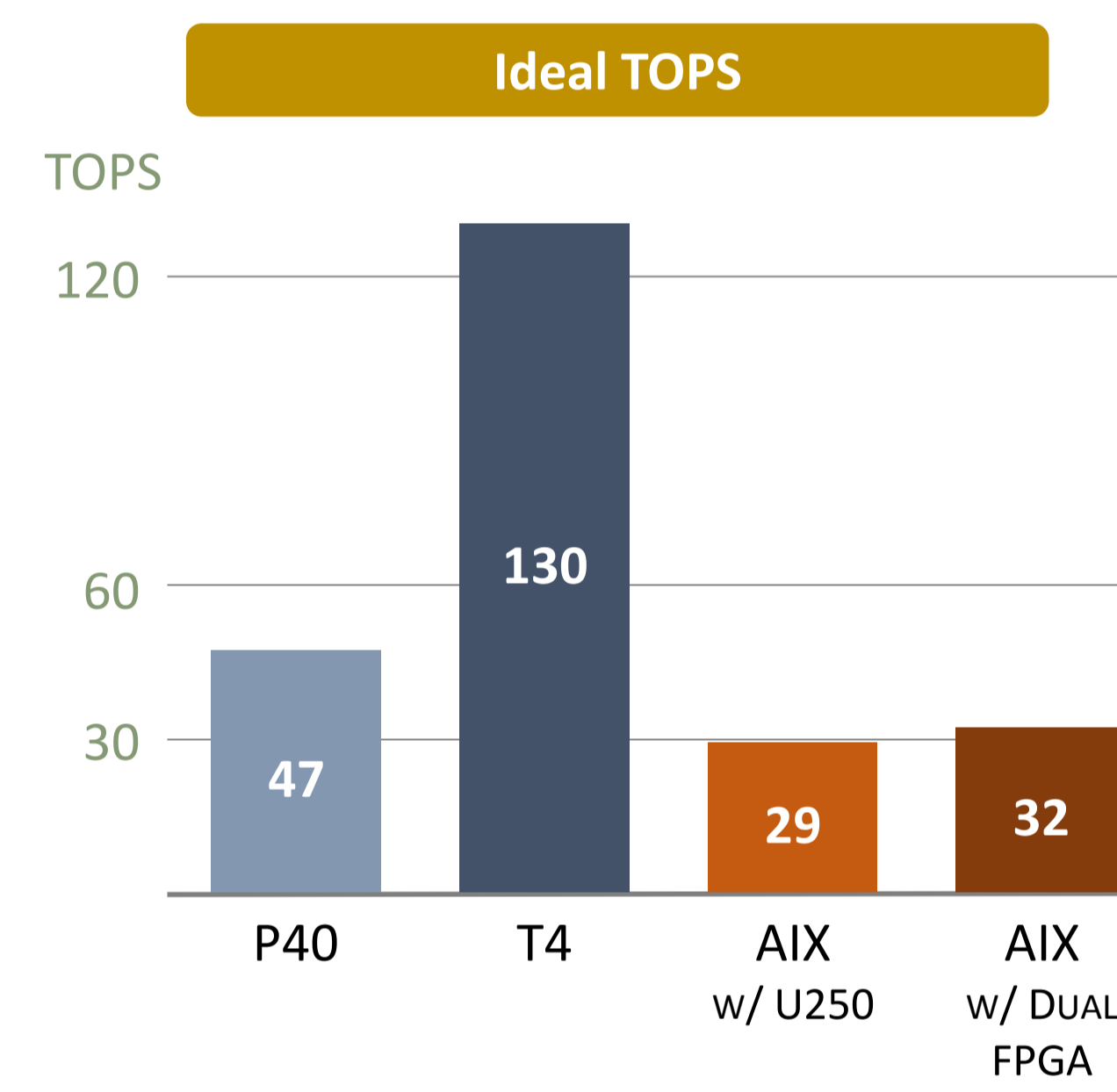
AIX w/ Custom Card (Dual FPGA)

VU9P x2 (Xilinx VIRTEX UltraSCALE+)

32 TOPS

115 GB/s (DRAM)

200 W



AIX v2 Architecture Highlights

