

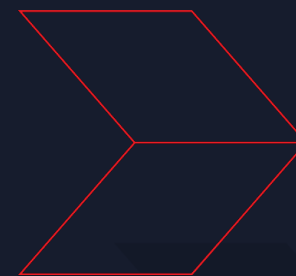


Speech Recognition powered by Xilinx FPGA for Virtual Personal Assistant, NUGU

Presented By



Minwook Ahn
Module Leading Developer
2018.10.02



Contents

- > **Introduction: SK Group & SK Telecom**
- > **NUGU: Virtual Personal Assistant**
- > **SKT AIX: Neural Network Accelerator for NUGU on Xilinx FPGA**
- > **Technical Roadmap**



Introduction: SK Group & SK Telecom

SK Telecom

No.1 Mobile Comm. in Korean

SK Hynix

No.2 Memory Semiconductor Manufacturer in Global Market

SK C&C

Total IT Service Provider

SK planet

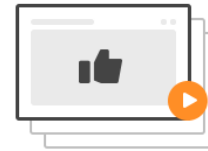
Platform Provider

NUGU AI Speaker



Media Services

B tv x NUGU



Surveillance Cloud

view



Navigation with NUGU



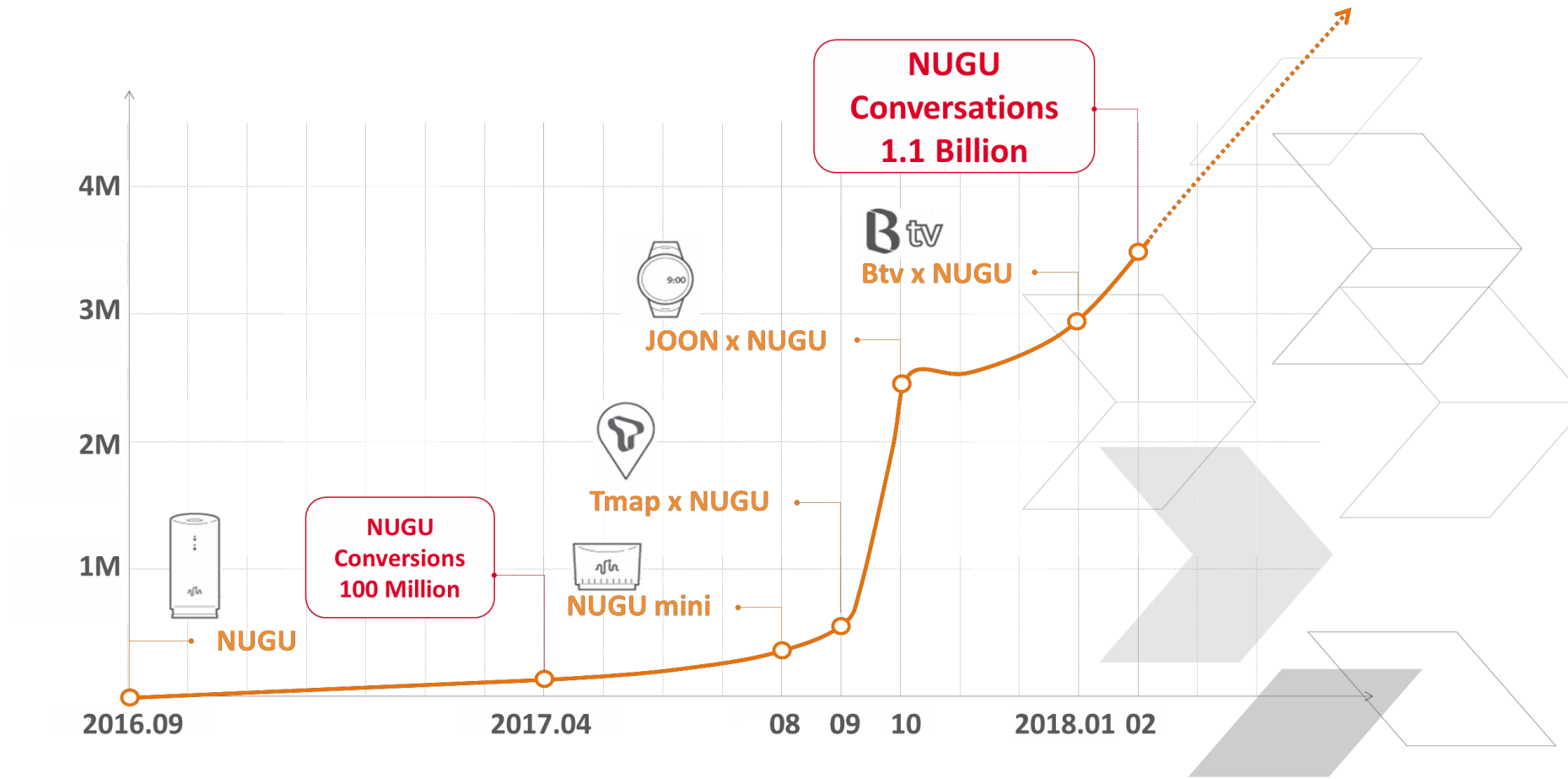
HD Map



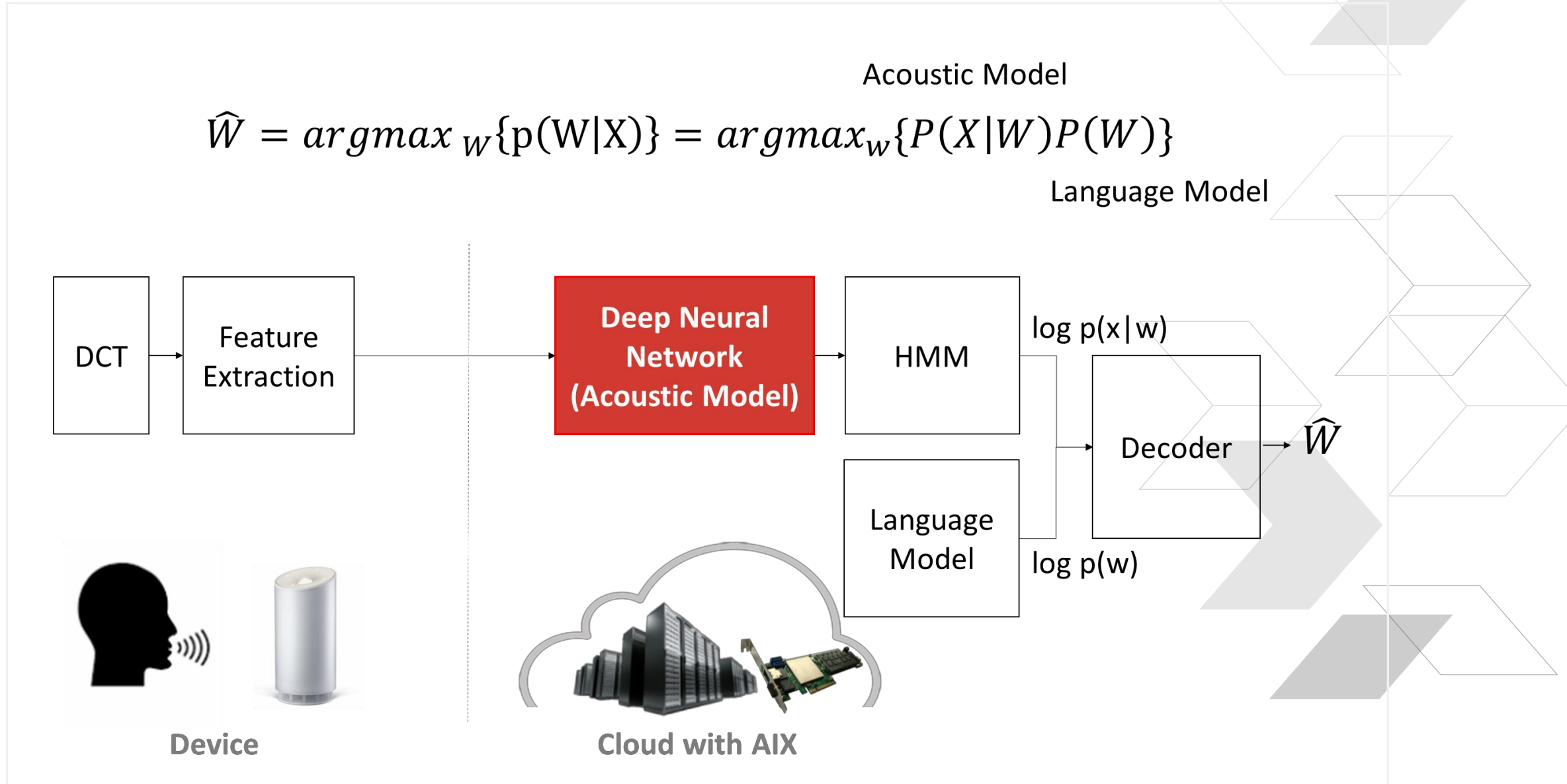
Self-driving Car



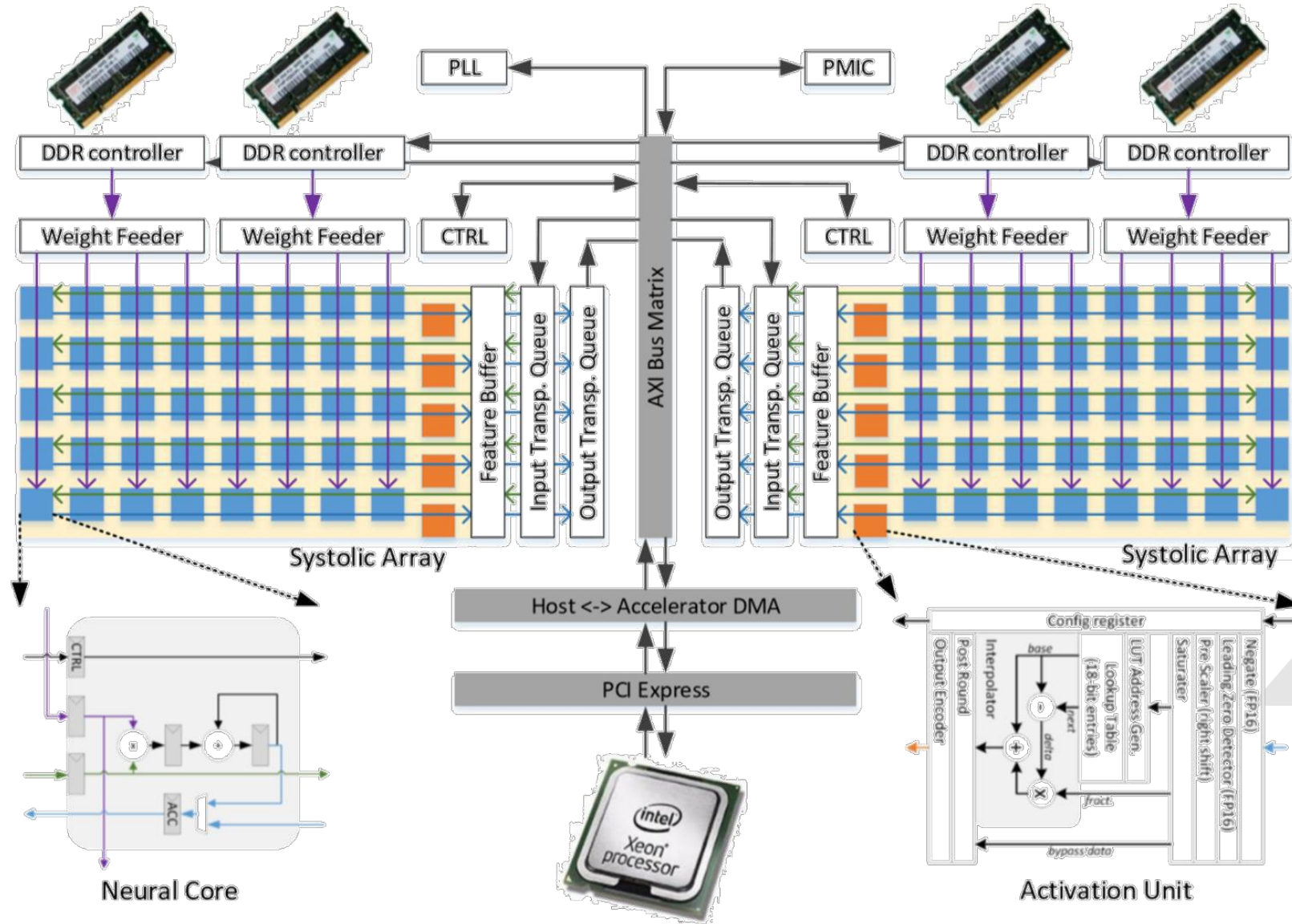
NUGU: Virtual Personal Assistant



NUGU: Virtual Personal Assistant



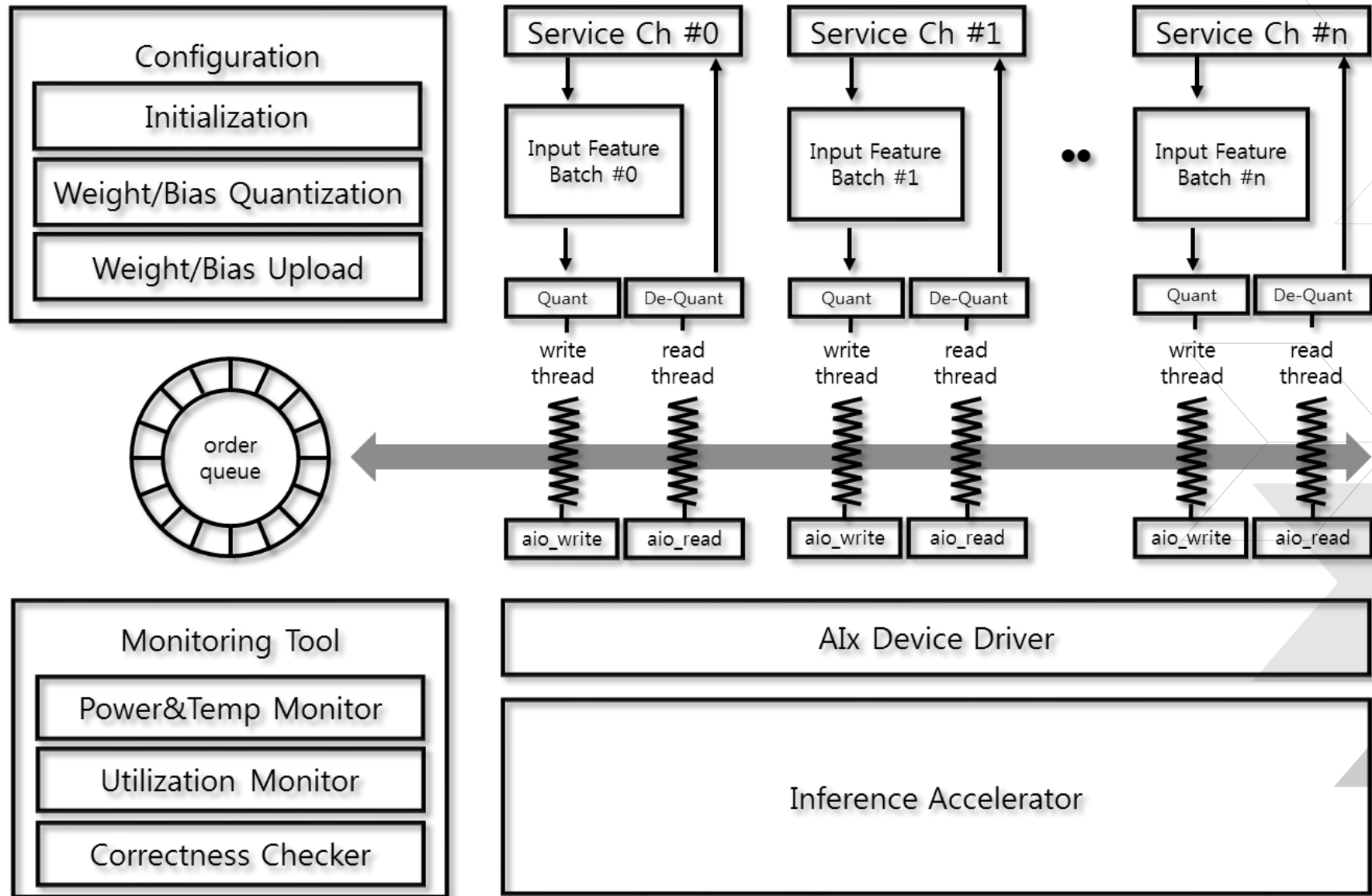
SKT AIX: Neural Network Accelerator for NUGU on Xilinx FPGA



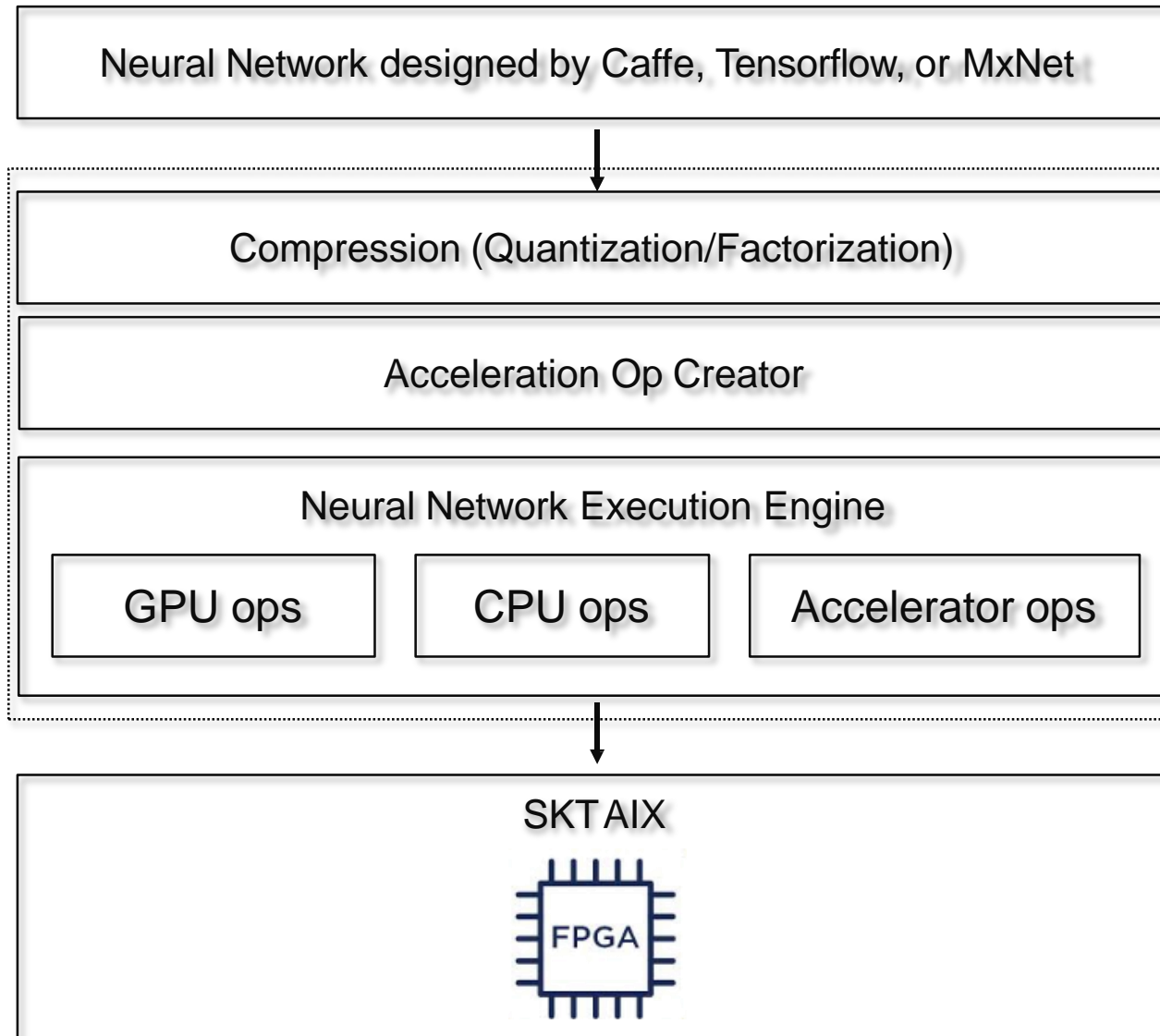
FPGA Resource Utilization for SKT AIX

	LUT as logic	LUT as memory	FlipFlop	BRAM	DSP
Available	663360	293760	1326720	2160	5520
Used	292781	24910	491916	1002	5212
Utilization	44.14%	8.48%	37.08%	46.39%	94.42%

SW stack for NUGU



SDK for SKT AIX



DL Framework

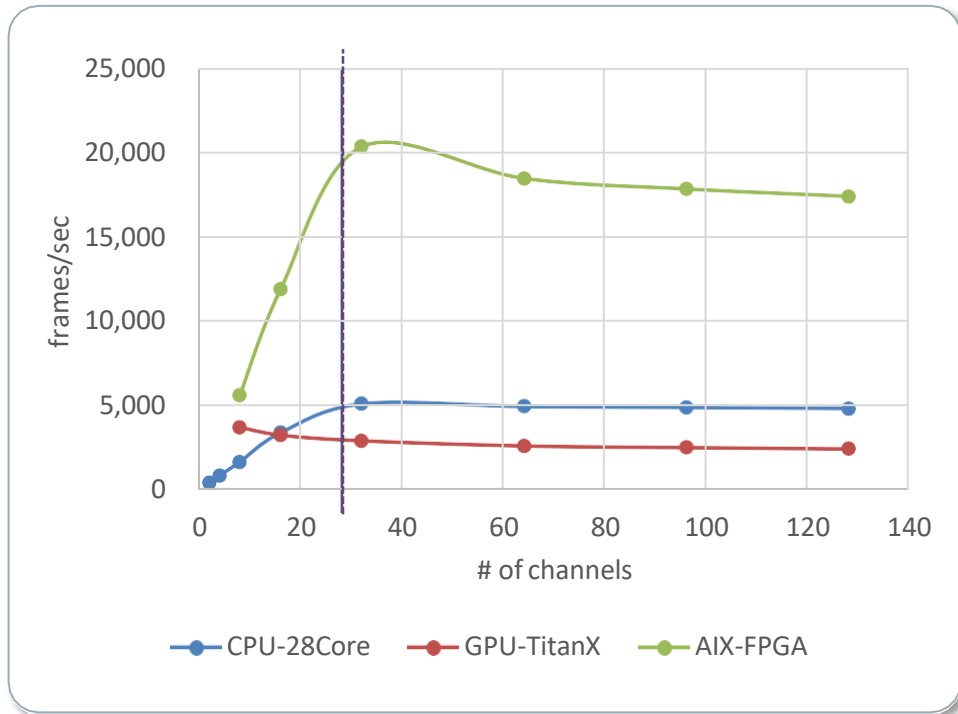
SK AI Inference SDK

SK AI Inference Accelerator

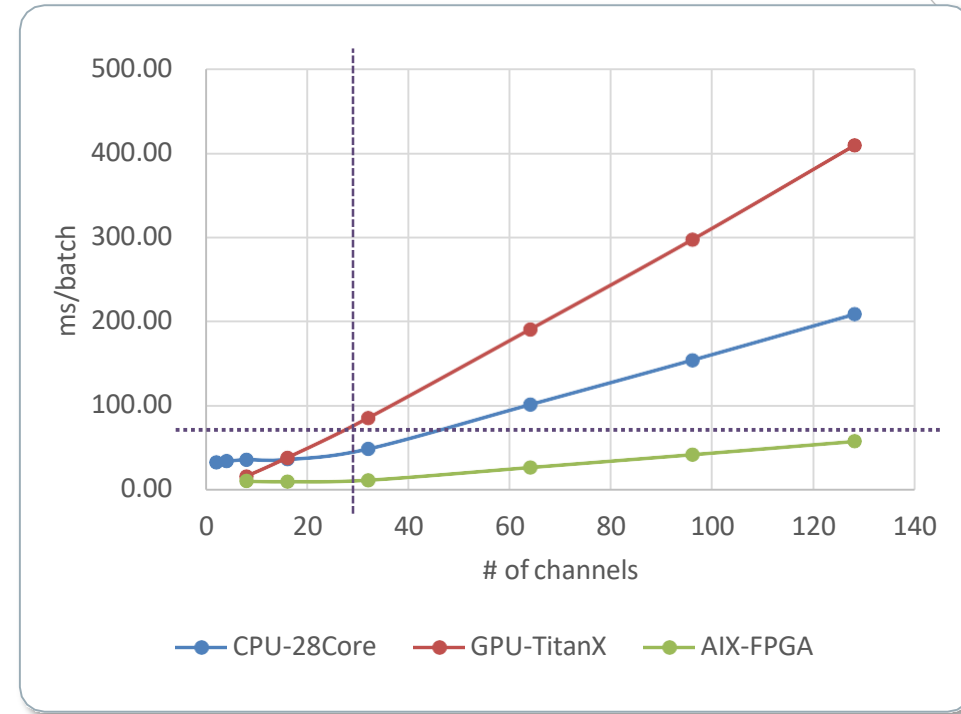
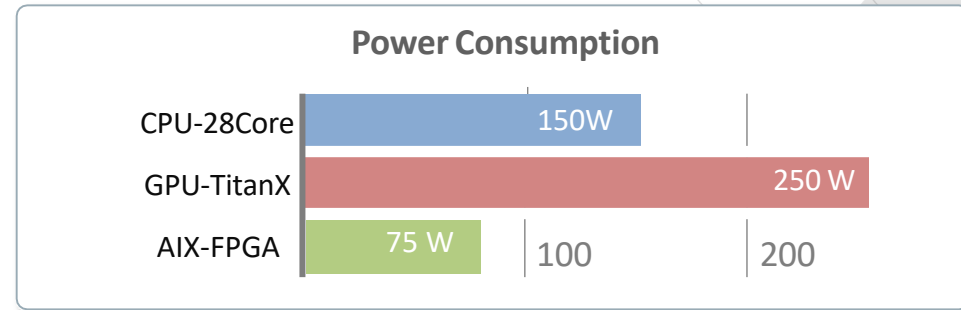
SKT AIX performance

5x Performance
16x Performance-Per-Watt
compared to GPU

Environment: 28-Core CPU, 8-Batch



Throughput according to
number of concurrent channels



Latency according to
number of concurrent channels

Technical Roadmap



AIX-F
LOW-END

AIX-F
HIGH-END

Accelerator Chip	Xilinx FPGA KU115
Compute Capability / Power	3 TOPS _(16bit) / 75 Watt
Neural Networks	MLP
Applications	ASR, Various DL Applications
Launch	2018.04
Accelerator Chip	Xilinx FPGA VU9P x2
Compute Capability / Power	33 TOPS _(16bit) / 250 Watt
Neural Networks	MLP, RNN, CNN
Applications	+ Video, NLP, Speech Syn., etc
Launch	2019.1Q



XILINX
DEVELOPER
FORUM

