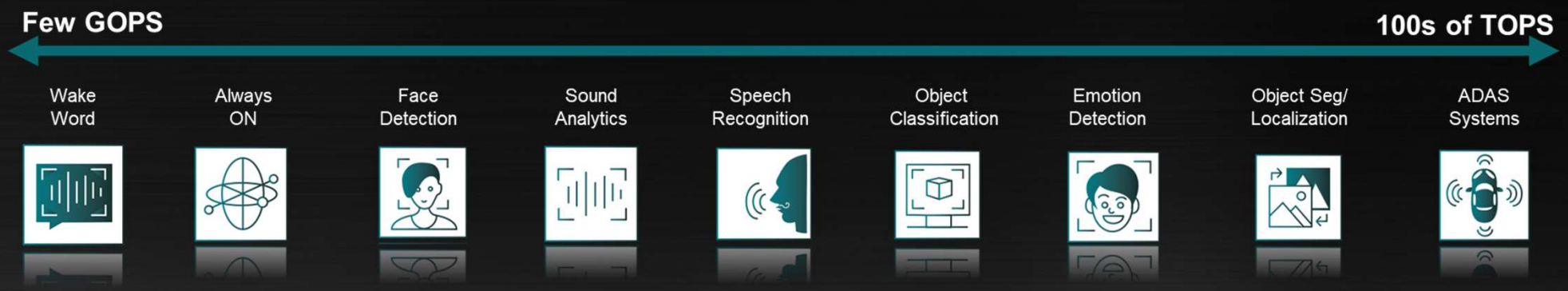
cadence®

The Challenges of Creating Software Development Tools for Diverse AI Hardware Architectures

Eric Stotzer
Sr. Group Director – NeuroWeave™ SDK

AI at the Edge – The Software Development Tools Challenge



AI at the Edge – The Software Development Tools Challenge

Applications

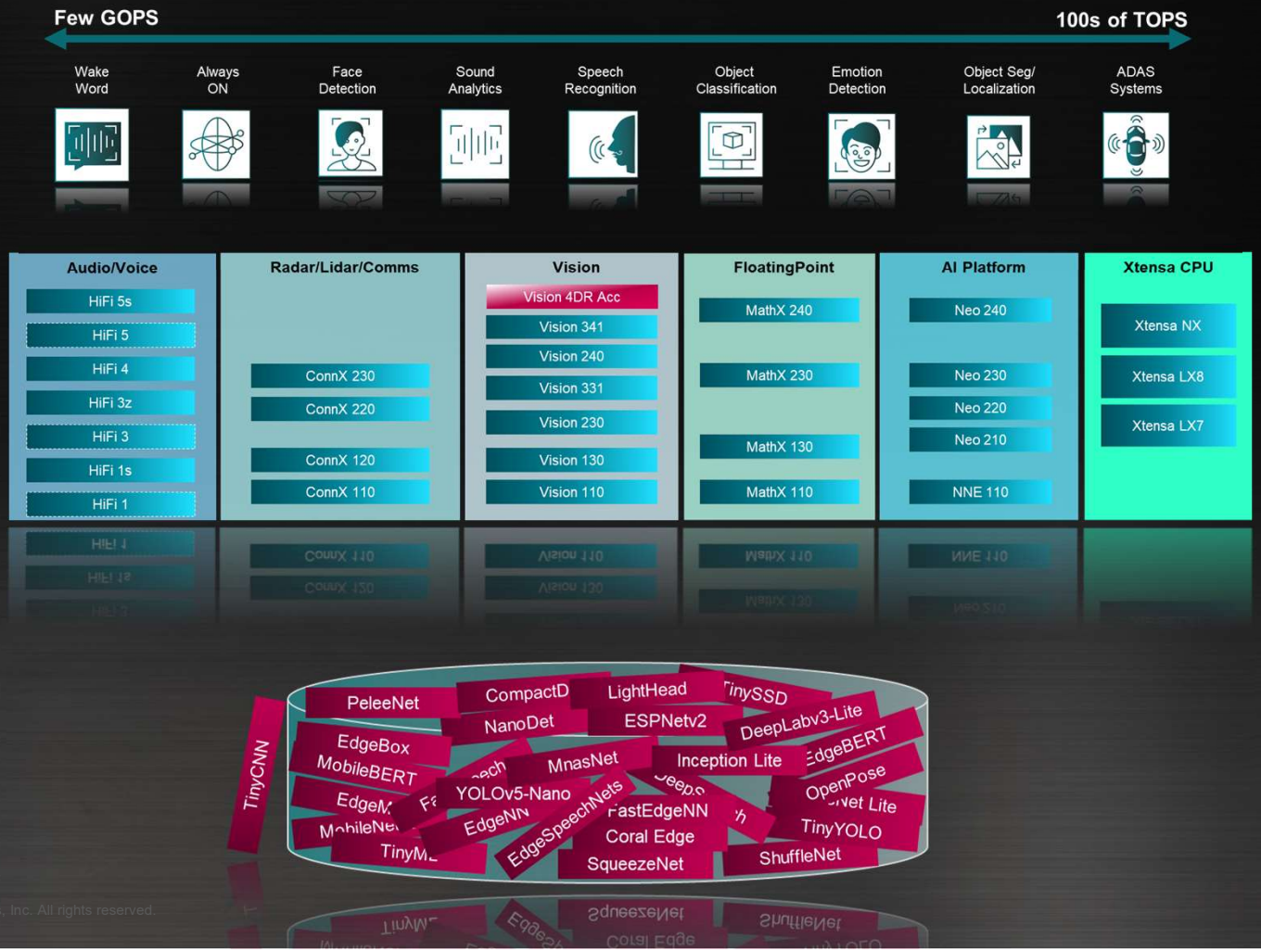
Audio/Voice	Radar/Lidar/Comms	Vision	FloatingPoint	AI Platform	Xtensa CPU
HiFi 5s		Vision 4DR Acc	MathX 240	Neo 240	Xtensa NX
HiFi 5		Vision 341			
HiFi 4	ConnX 230	Vision 240	MathX 230	Neo 230	Xtensa LX8
HiFi 3z	ConnX 220	Vision 331		Neo 220	Xtensa LX7
HiFi 3		Vision 230	MathX 130	Neo 210	
HiFi 1s	ConnX 120	Vision 130			
HiFi 1	ConnX 110	Vision 110	MathX 110	NNE 110	

AI at the Edge – The Software Development Tools Challenge

Applications

Hardware

Networks



AI at the Edge – The Software Development Tools Challenge

The AI Software Development Tools **Challenge**

What challenge?

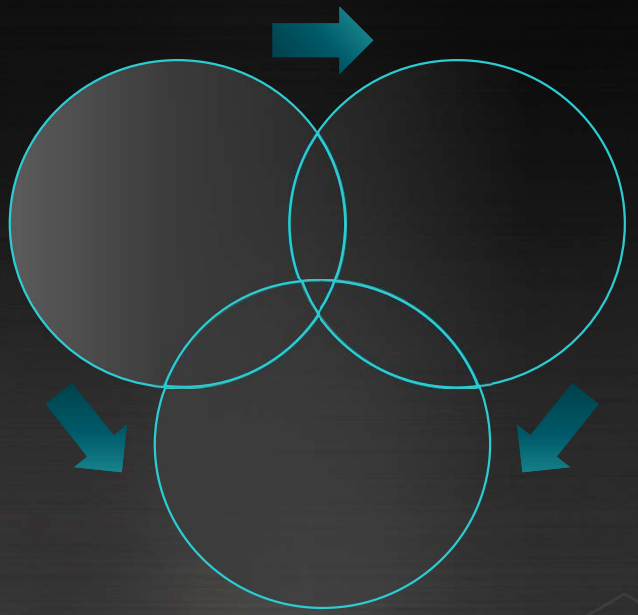
- An intersection of multiple domains
- From AI model in Python to edge quality embedded code
- The Frontier of AI model development

What **Challenge**?

- Opensource AI compilers
  - “XLA, IREE, Glow, TVM, … have already solved this problem.”
- MLIR is a graph compiler toolkit
  - “We solved that with a dialect.”
- Just map NN layers to optimized library functions
  - “We have a kernel for that.”

An Intersection of Multiple Domains

- AI network theory and construction
- High-performance computing
- High Level Language (HLL) compilers and utils
- Embedded system software
- Python and C/C++ and more

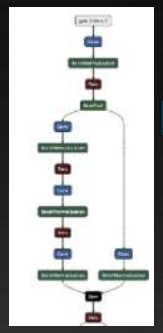


From AI Model in Python to Edge Quality Embedded Code

TensorFlow
PyTorch
Caffe2
ONNX
...

Build AI models

Floating-Point or Quantized Model Graph



NeuroWeave® SDK

Model Analyzer

Model Optimizer/Code Generator

Deploy AI models

Processor

Vision / RLC DSPs

Audio DSPs

AI Accelerators

Wake Word

Always ON

Object Localization

Object Classification

Sound Analytics

Emotion Detection

ADAS Systems

GenAI

One Single Software Toolchain flow for various deployment scenarios

cadence

The Frontier of AI Models and Tools – “The Gold Rush”

- AI models are developed and represented in diverse and evolving frameworks
  - Keras, Caffe, TensorFlow(Lite), Pytorch, ONNX, NNEF, Core ML, MXNet, PaddlePaddle, …
  - A domain specific language expressing a computation graph
- Competing industry platforms
  - Good: Innovation moves quickly
  - Challenge: Trying to keep up
- What is missing compared to more traditional HLLs
  - No ANSI/IEEE specs
  - No Clang/Gcc implementations

How Will Standardization Be Driven in this New World?

How the Cadence Neuroweave SDK Meets the Challenge

- Graph import
  - Civilize the Frontier as best we can via TVM
- Quantization, Optimization, Code Generation
  - A combination of MLIR, LLVM, and Proprietary
- Architecting an open and extensible solution
  - Enable diverse type of AI HW Archiectures
  - Processor + Accelerator + Custom IP

In Closing

AI Applications + Diverse HW + A Heap of Network models

Requires sophisticated software development tools

As a community: converge on a shared AI model representation

Q&A