Other Rogues - Analog, Reconfigurable, and Neuromorphic Accelerators and Related Tools

FPAA Content provided by Dr. Jennifer Hasler,
ECE and CRNCH faculty

CREATING THE NEXT MOORE'S LAW

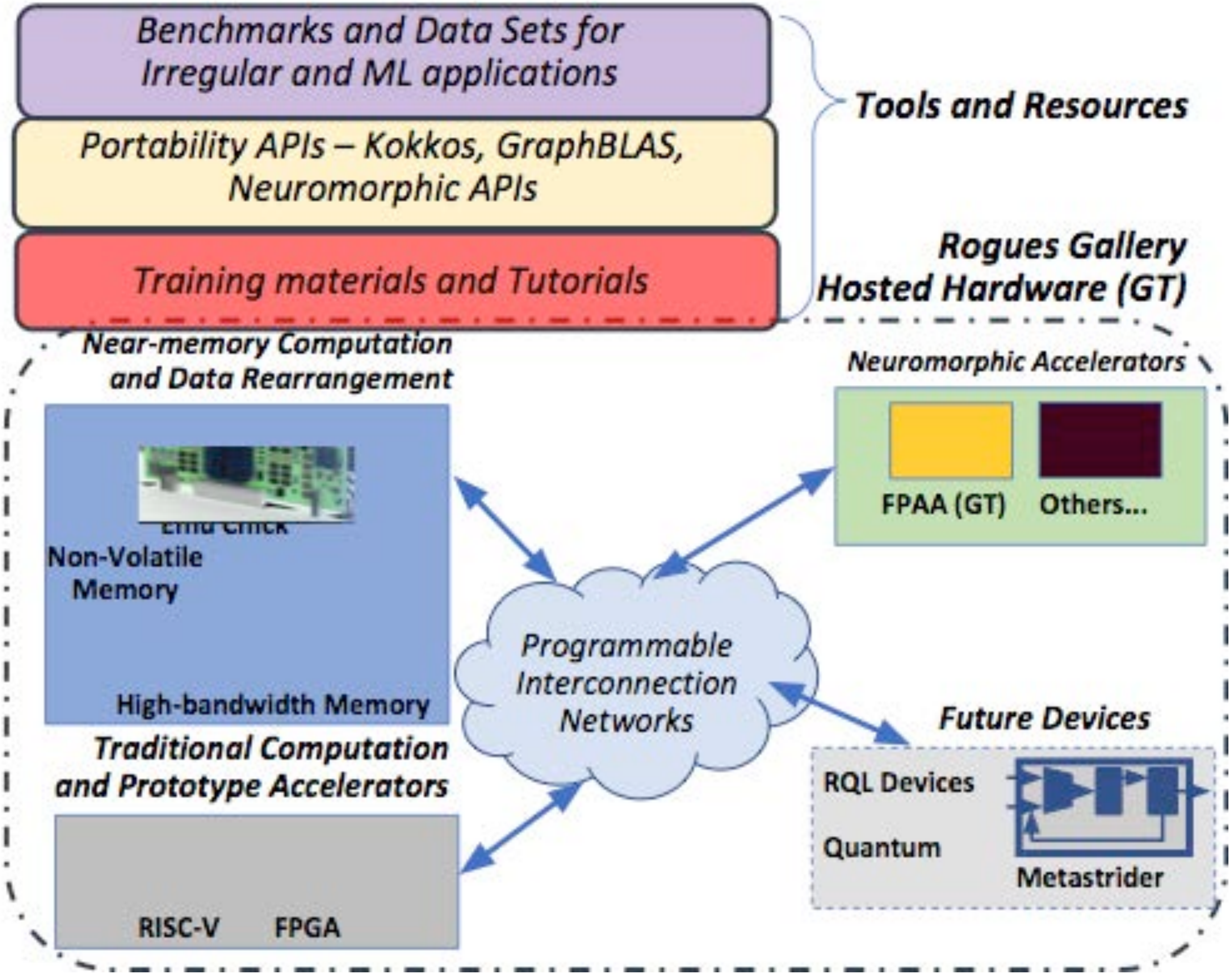
Georgia Tech | Center for Research into Novel Computing Hierarchies

For more information, visit our website at
www.crnch.gatech.edu or send email to
crnch@gatech.edu

Outline

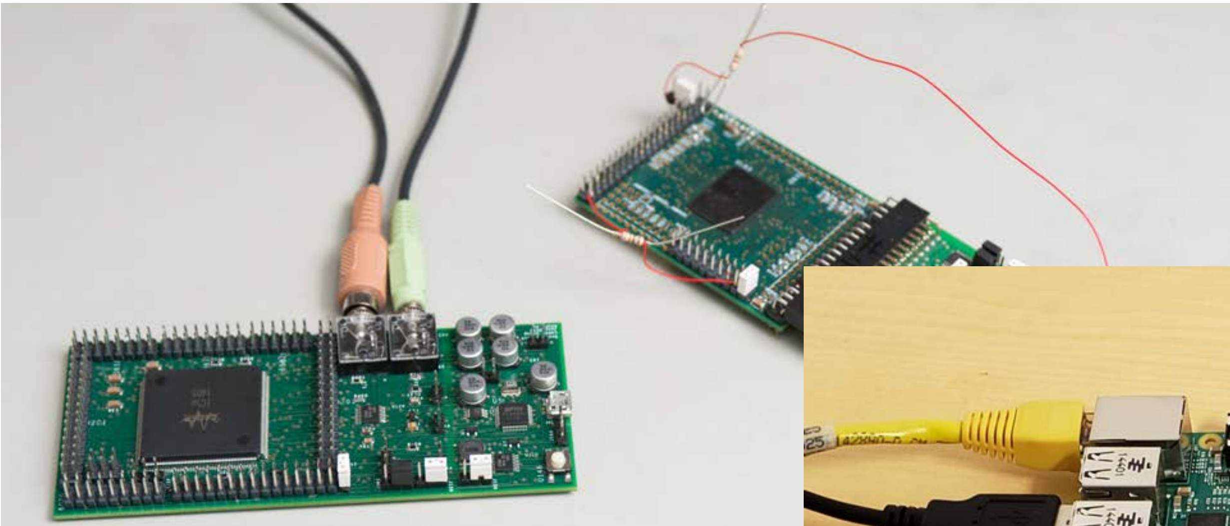
- Neuromorphic Computing
  - Field Programmable Analog Arrays
  - DANNA
  - Other future platforms
- Reconfiguable Computing
- Benchmarking and Tools

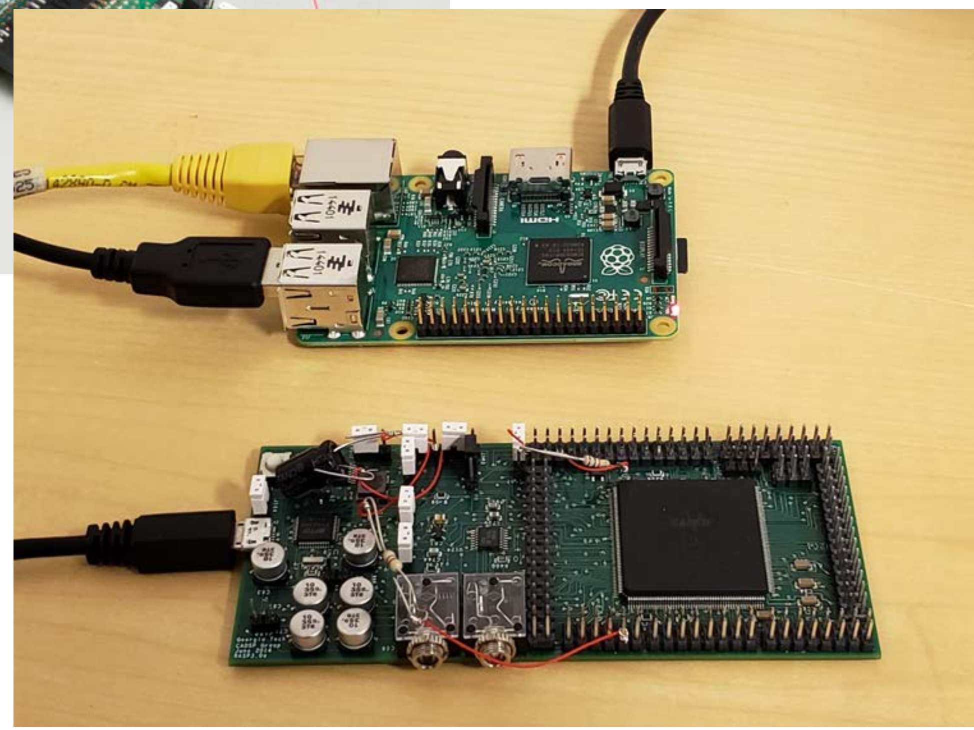
Rogues Gallery Recap



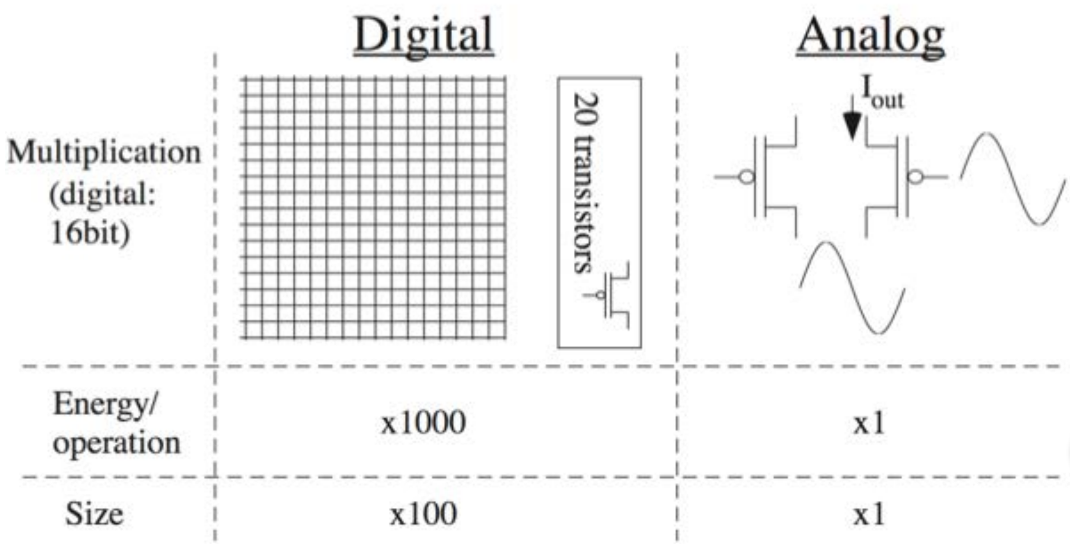
Georgia Tech Center for Research into Novel Computing Hierarchies

Field Programmable Analog Array (FPAA)

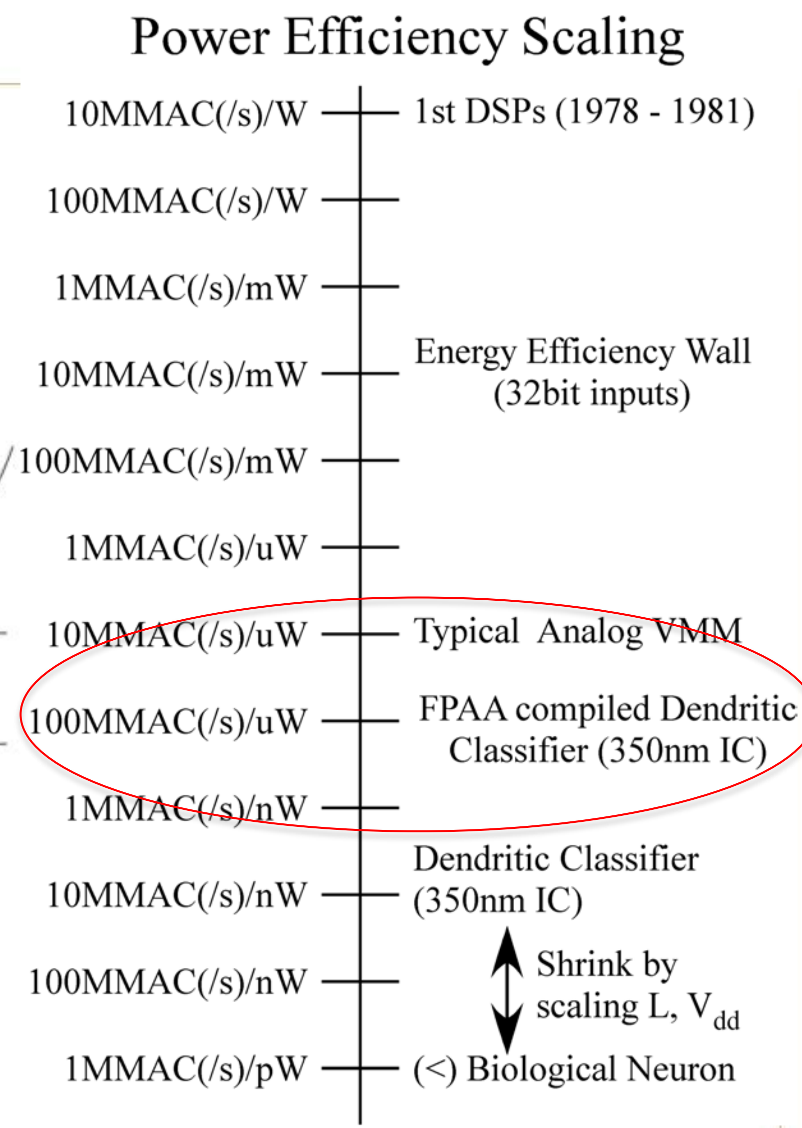




Why Physical / Analog Computing?



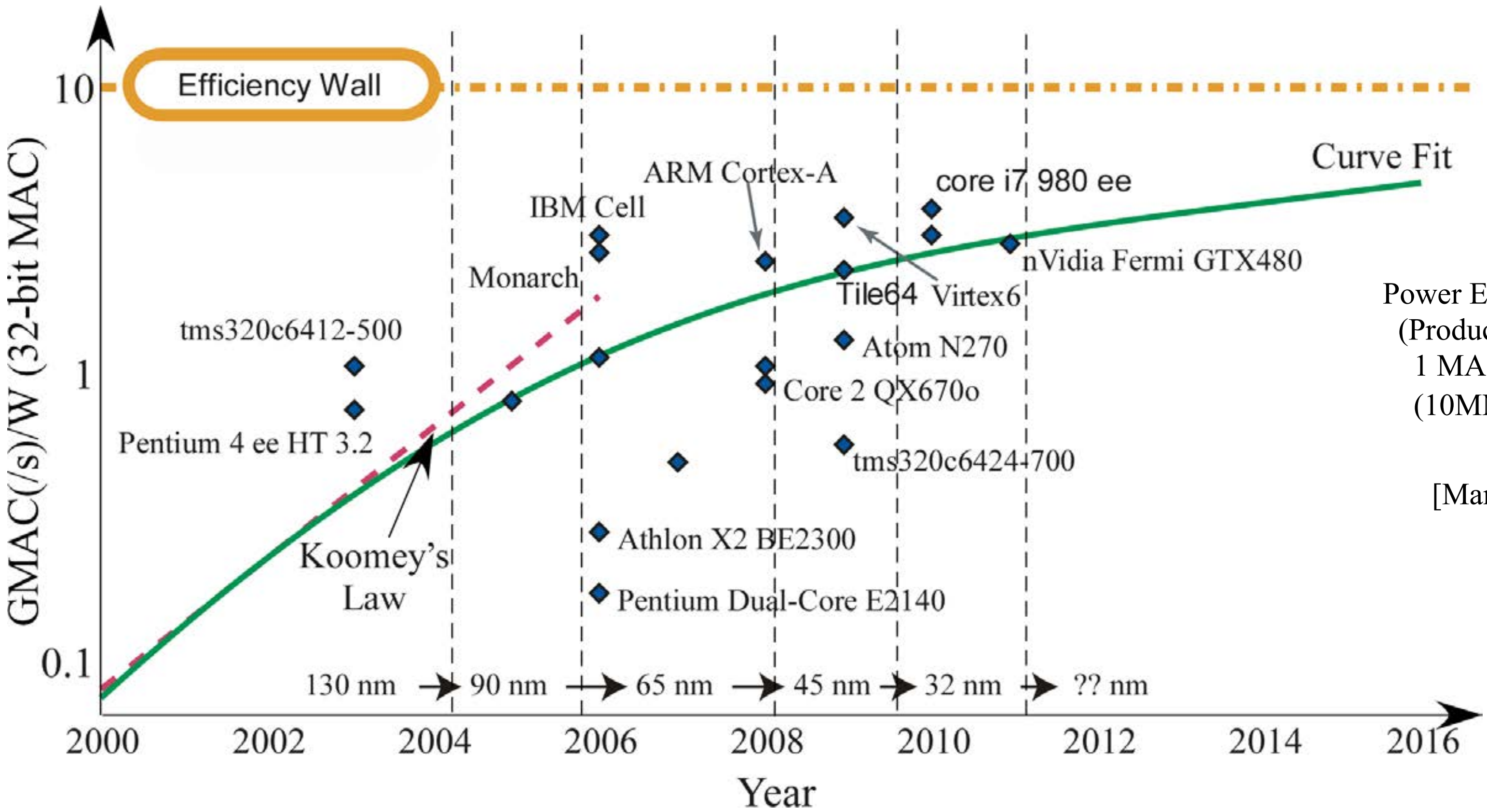
Expansion of C. Mead's original hypothesis, 1990



[Hasler and Marr 2013]

Why Physical / Analog Computing?

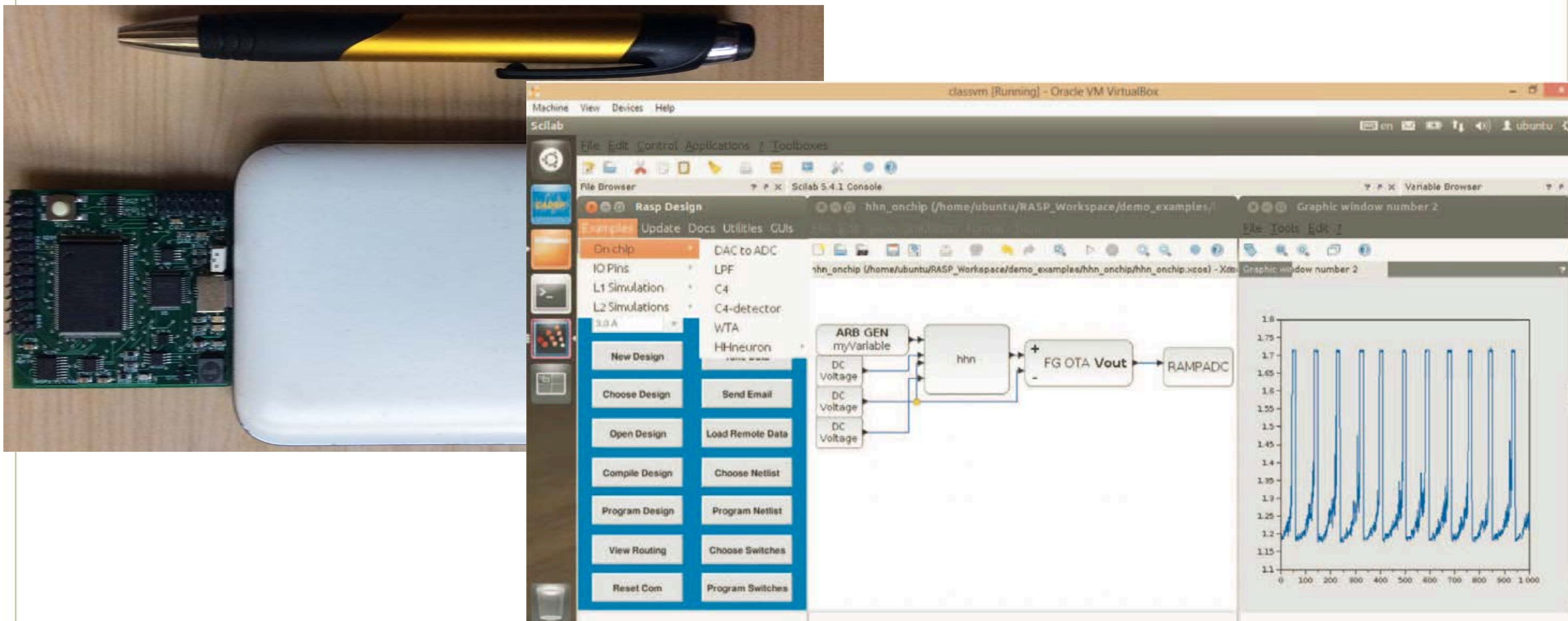
Digital Hitting Limits of Power Efficiency – Koomey's Law



Power Efficiency wall
(Production ICs)
1 MAC in 100pJ
(10MMAC/mW)

[Marr, et. al, 2013]

FPAAs enable Ultra-low Power Computing

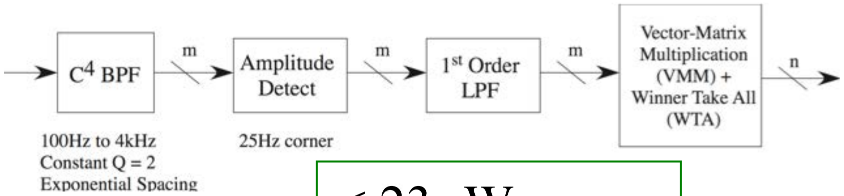


The FPAA provides a mixed analog/digital platform with a general and configurable design that can be used create secure and legacy resistant mixed-signal devices

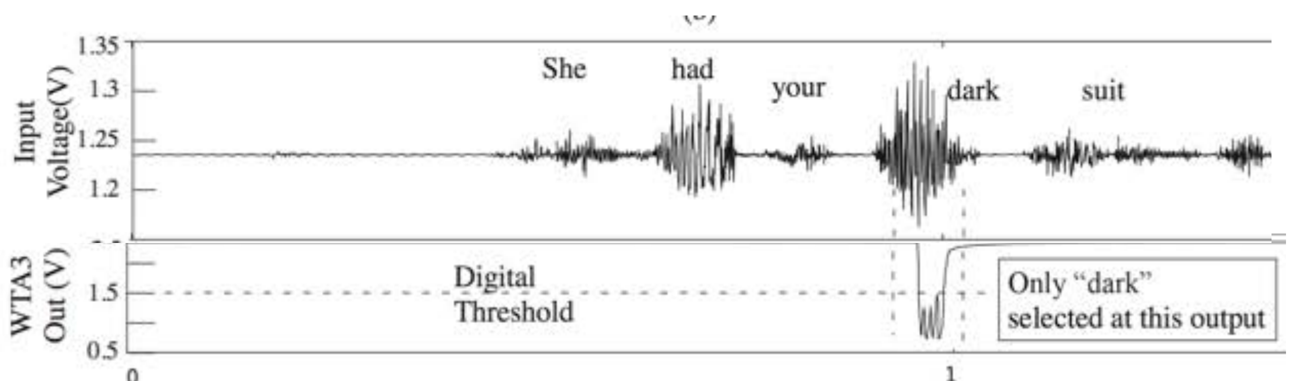
FPAA Classification Task

Accoustic Classification

Ultra-low energy computation



< 23μW power



Embedded learning & classification: 20-30μW on full, 1s Nzero database (GOMAC 2016)

FPAA Classification to Date

Built on Analog Vector-Matrix Multiplication (VMM), including in routing, Winner Take All (WTA), etc.

Initial Command-Word Classification (Hand-tuned Weights) (2015)

Biomedical classifier (Hand-tuned weights) and computational measures
(Knee sounds, Heart monitoring)

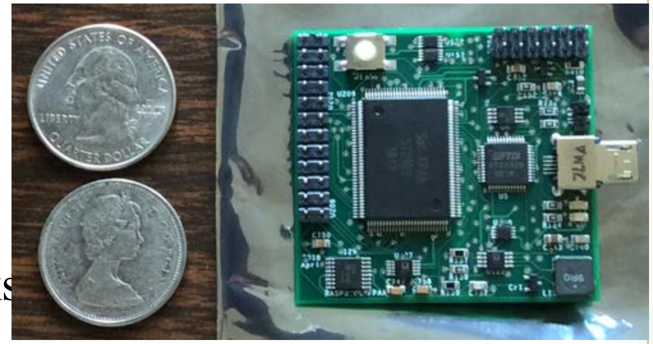
Speech – No-Speech Detector (Hand tuned weights)

Acoustic Classification and Learning:

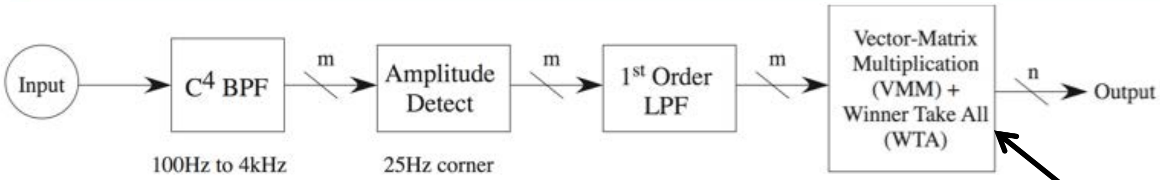
- Developed Theory and Training Algorithm for VMM+WTA Classification
- Trained using modified Lincoln Lab's acoustic data set for Nzero program.
- Yielded correct detection on all sets

Current work on Image classification (2-D sensor inputs)
2-layer NN or DNN are possible…

- VMM+WTA has worked well so far
- Can develop modular spiking (HH neuron) networks

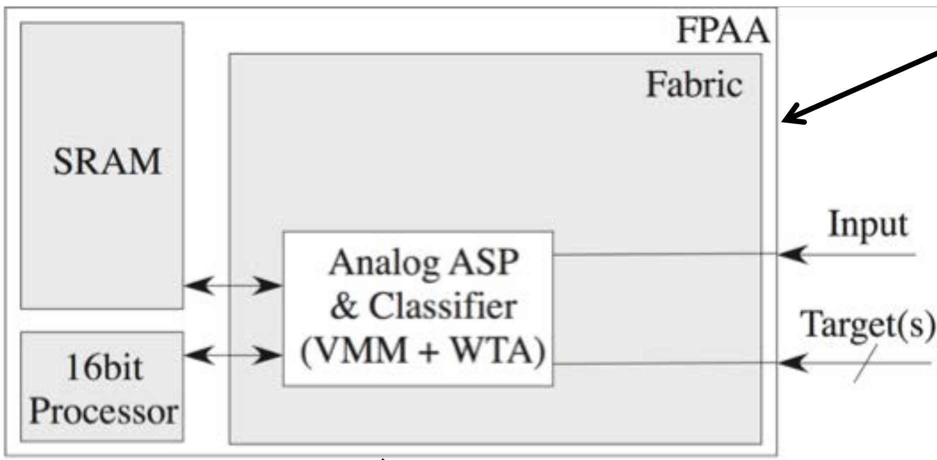


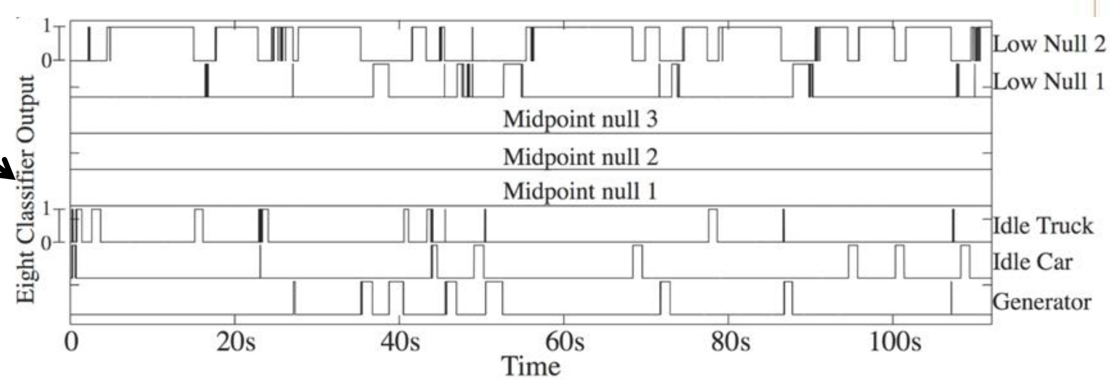
More Analog Classifier (VMM+WTA)



Analog, n-WTA single layer block = universal approximator (2 layer NN)

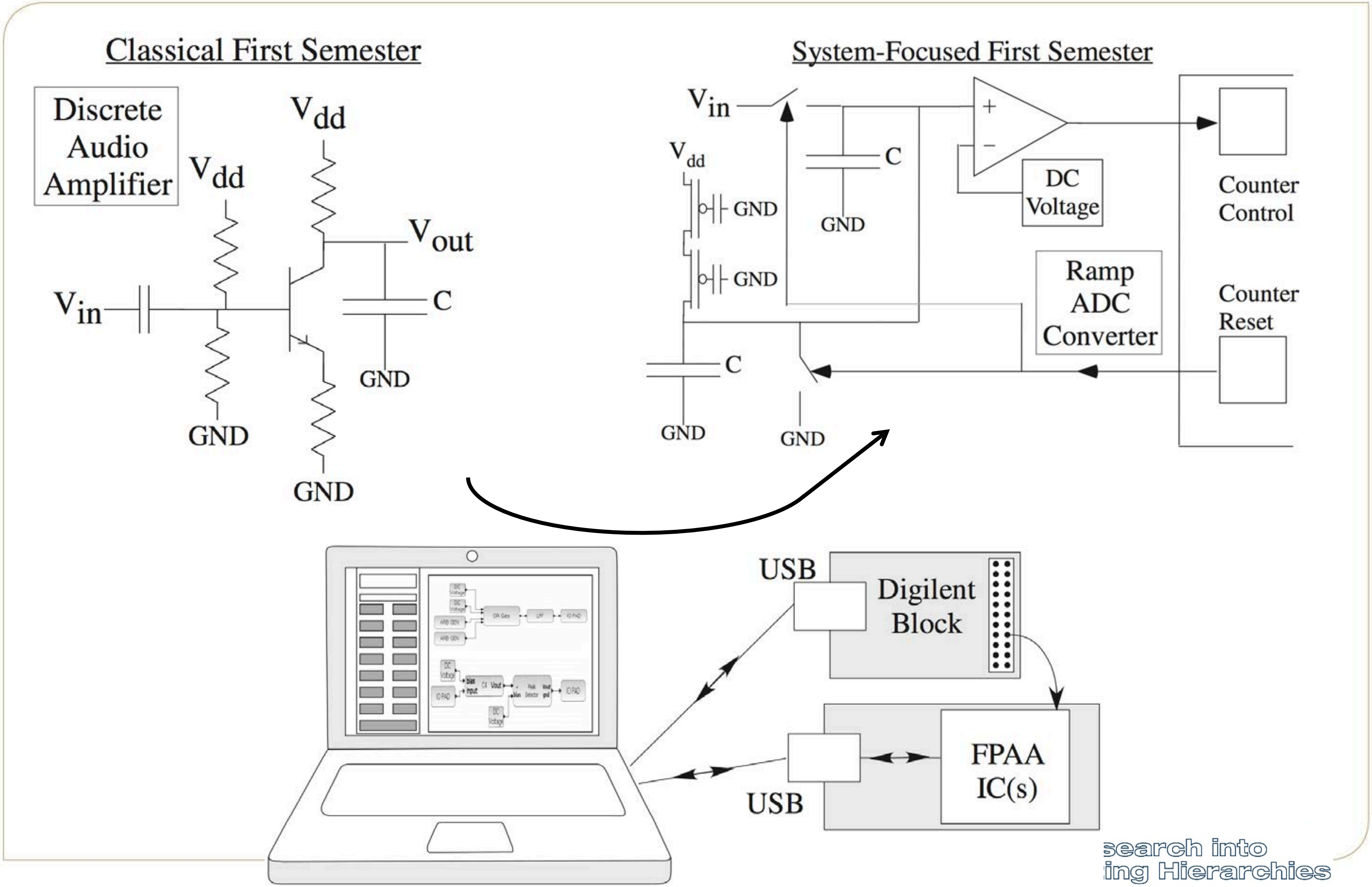
Classification + On-Line Training



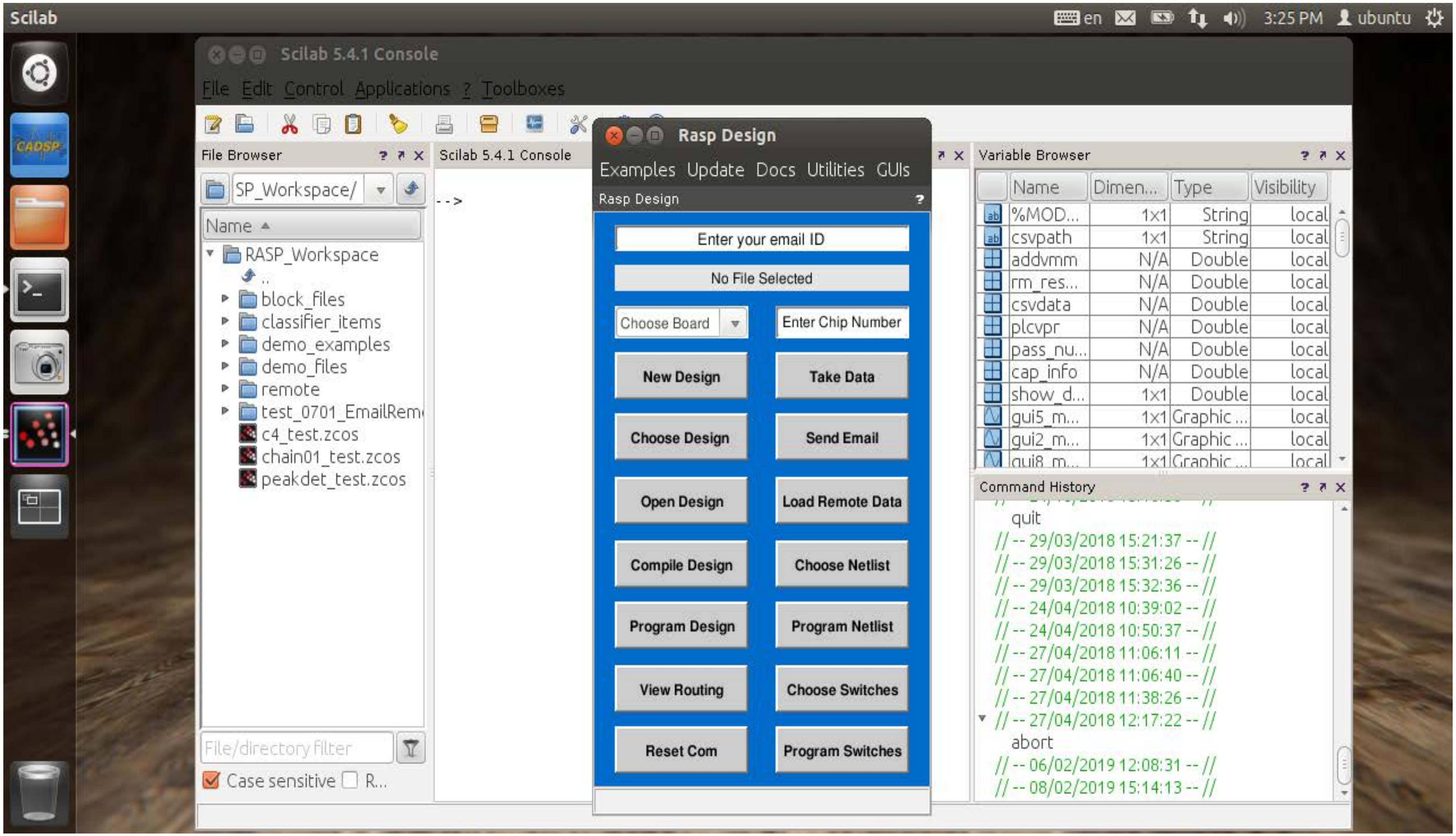


100% classification

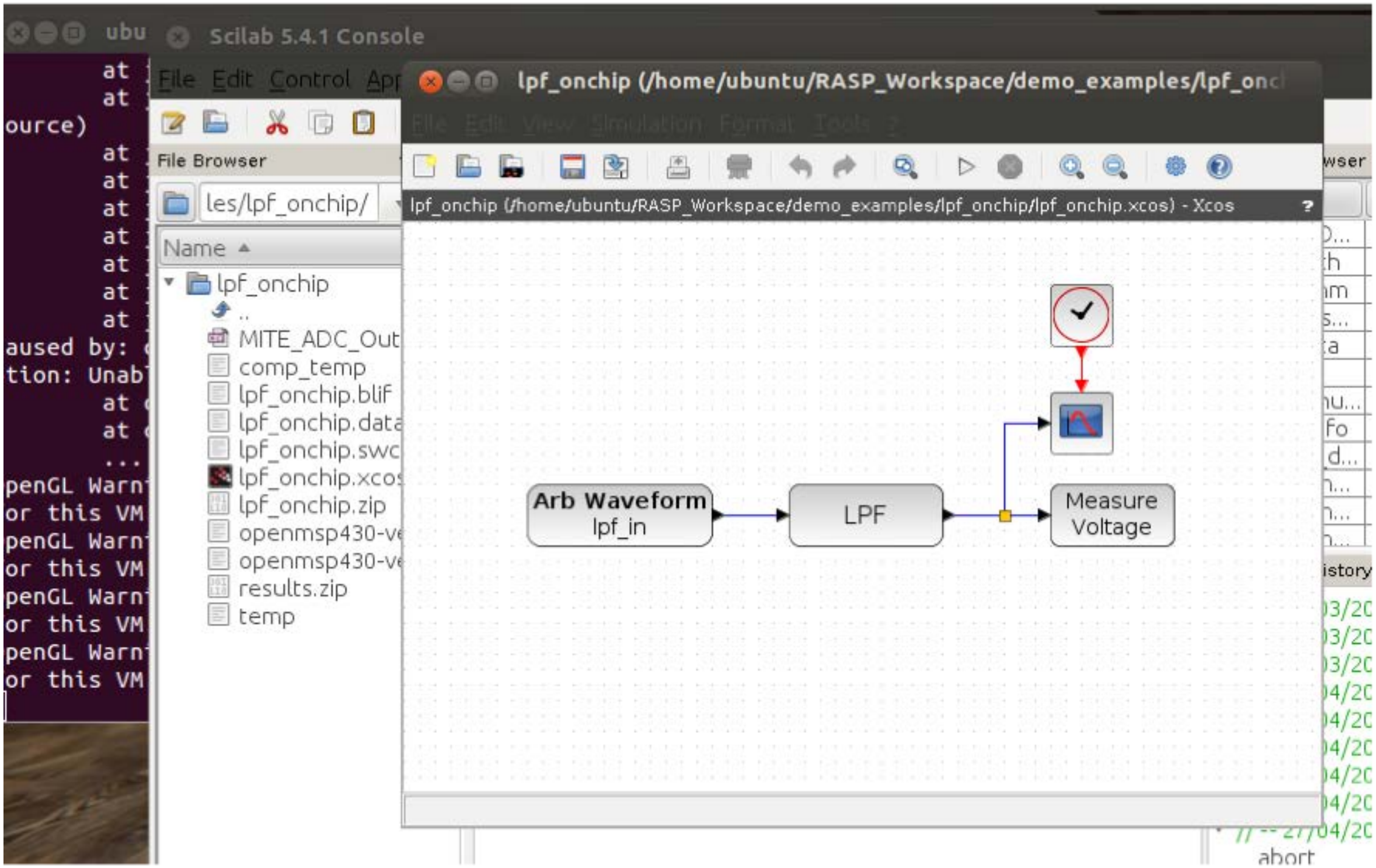
SoC FPAAs in the Classroom



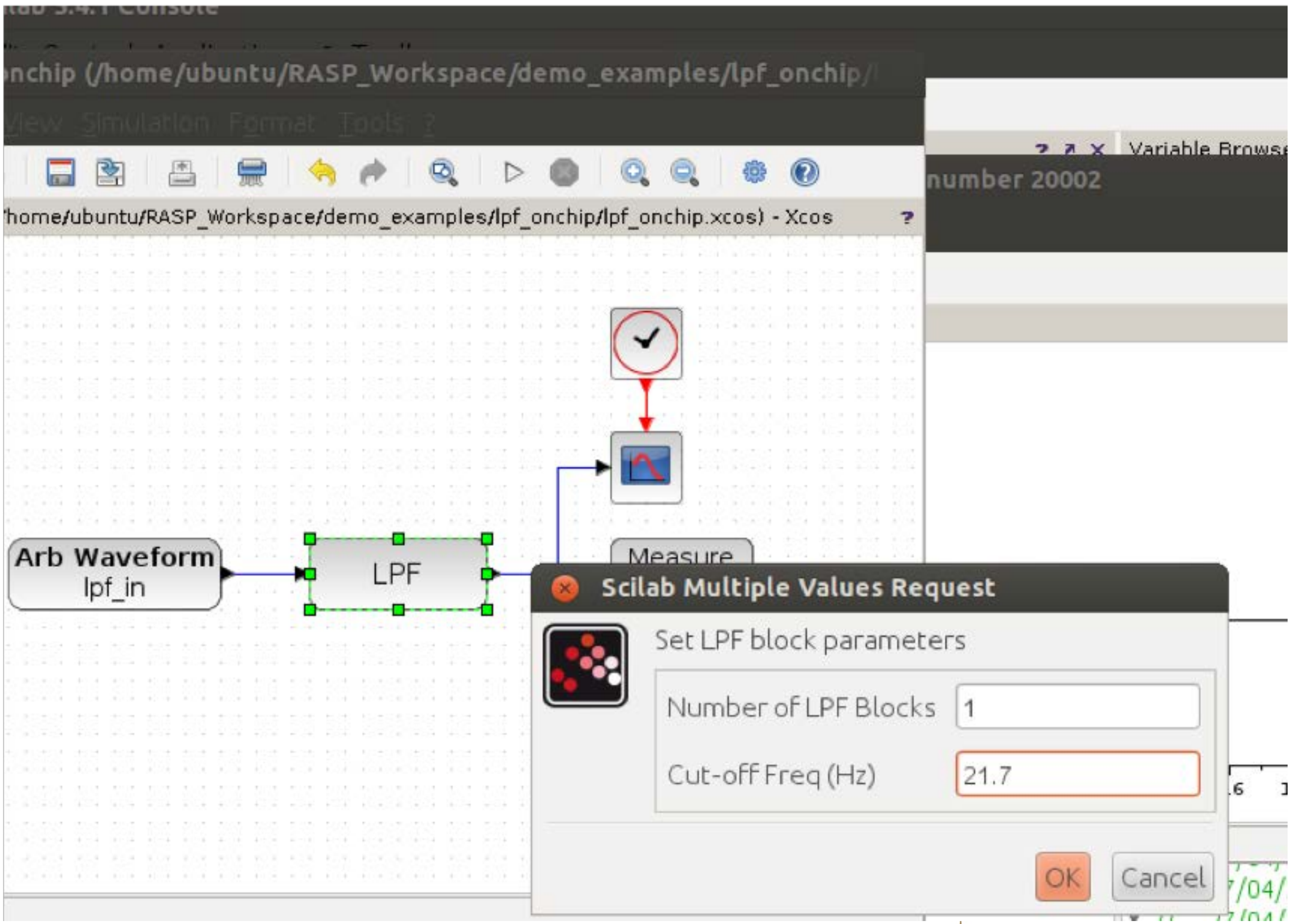
FPAA Toolflow and Examples (1)



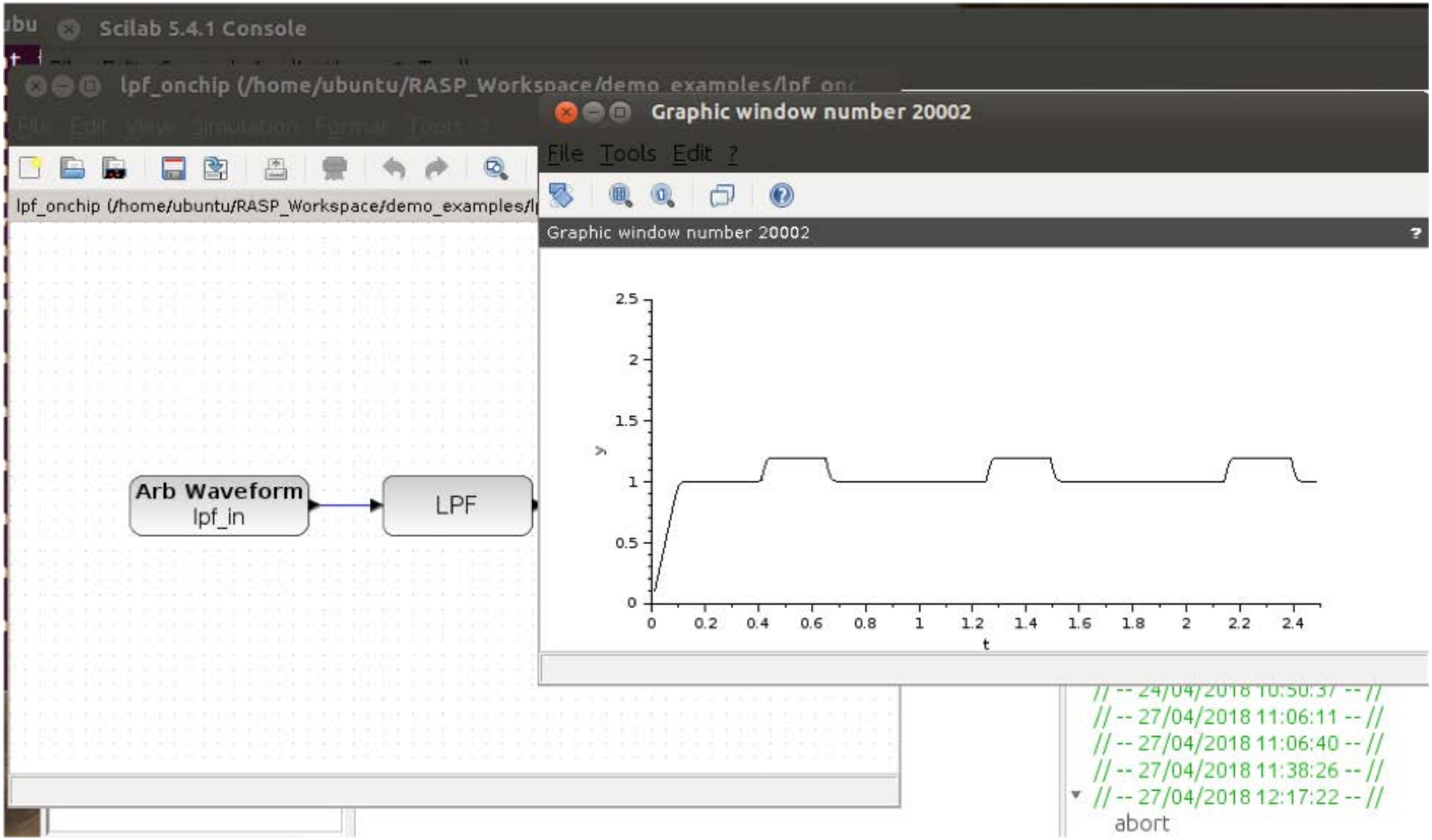
FPAA Toolflow and Examples (2)



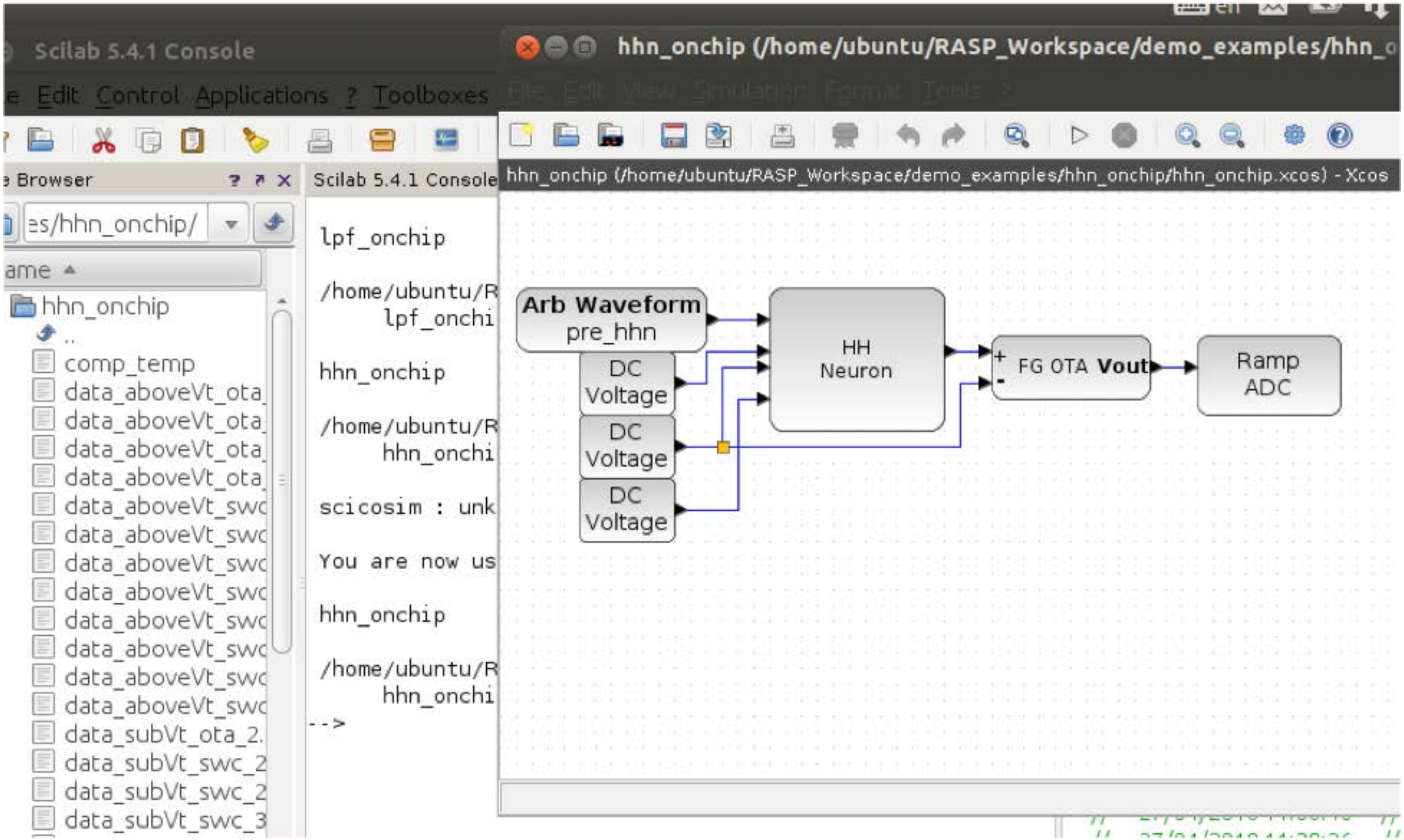
FPAA Toolflow and Examples (3)



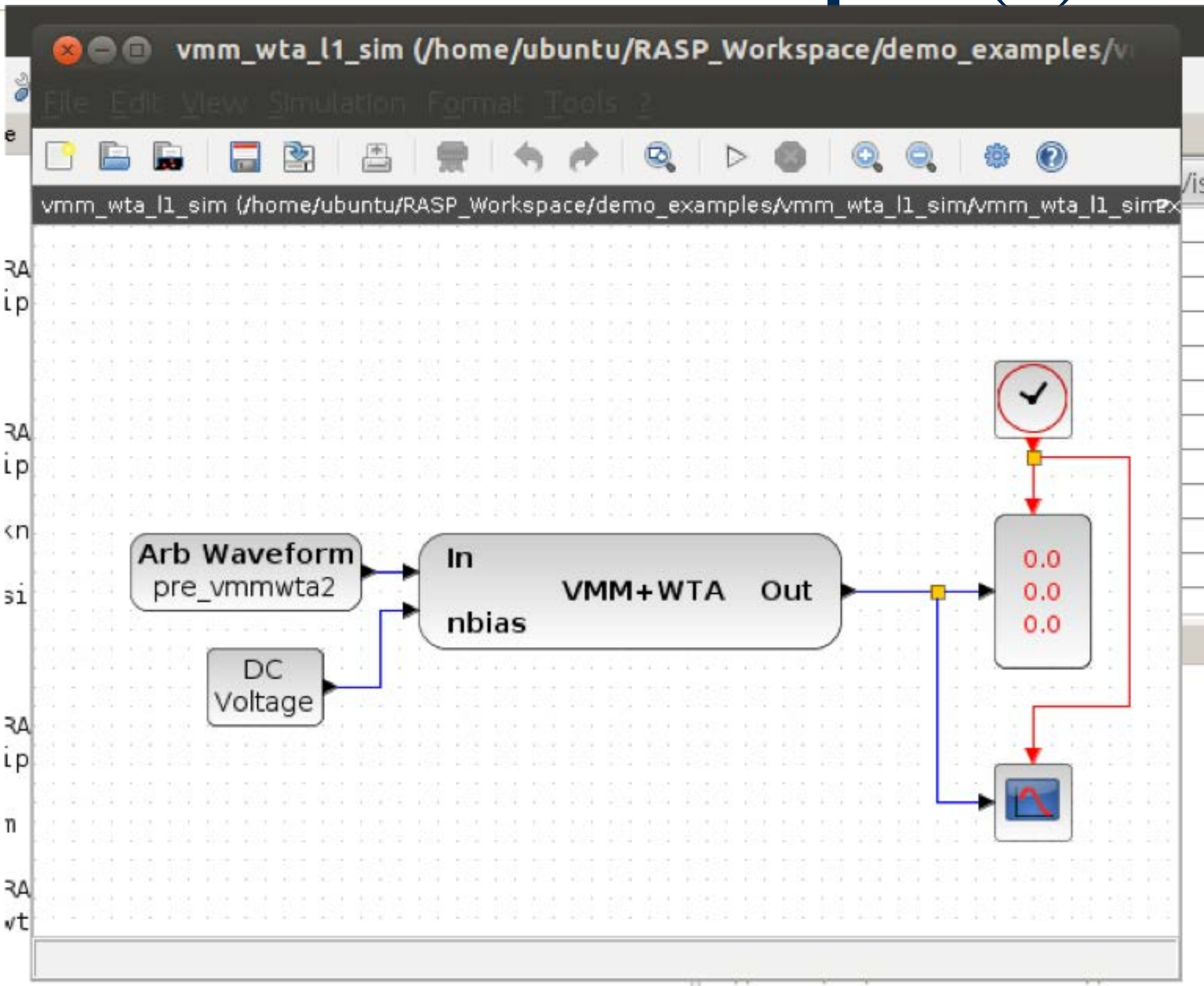
FPAA Toolflow and Examples (4)



FPAA Toolflow and Examples (5)

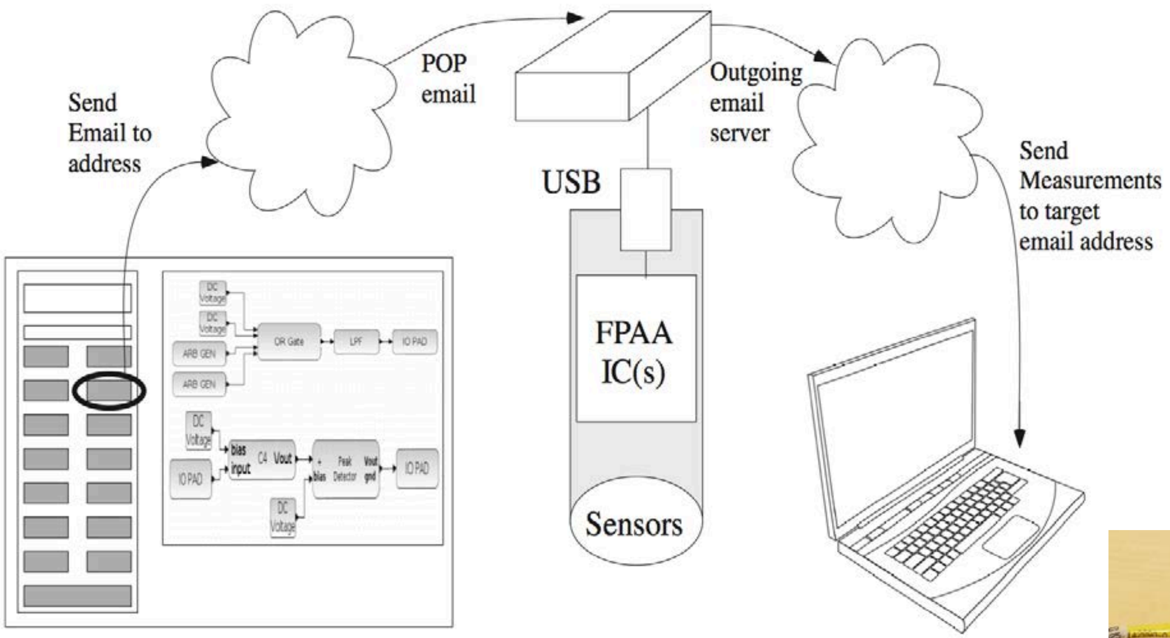


FPAA Toolflow and Examples (6)

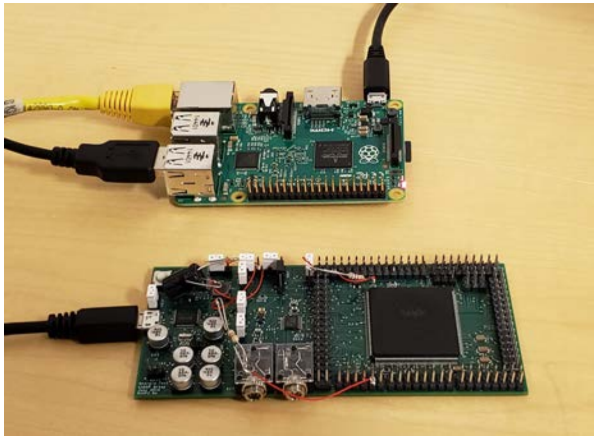


FPAA Infrastructure Details

Remote FPAA System



Pi-hosted FPAAs



FPAA for Neuromorphic Computing

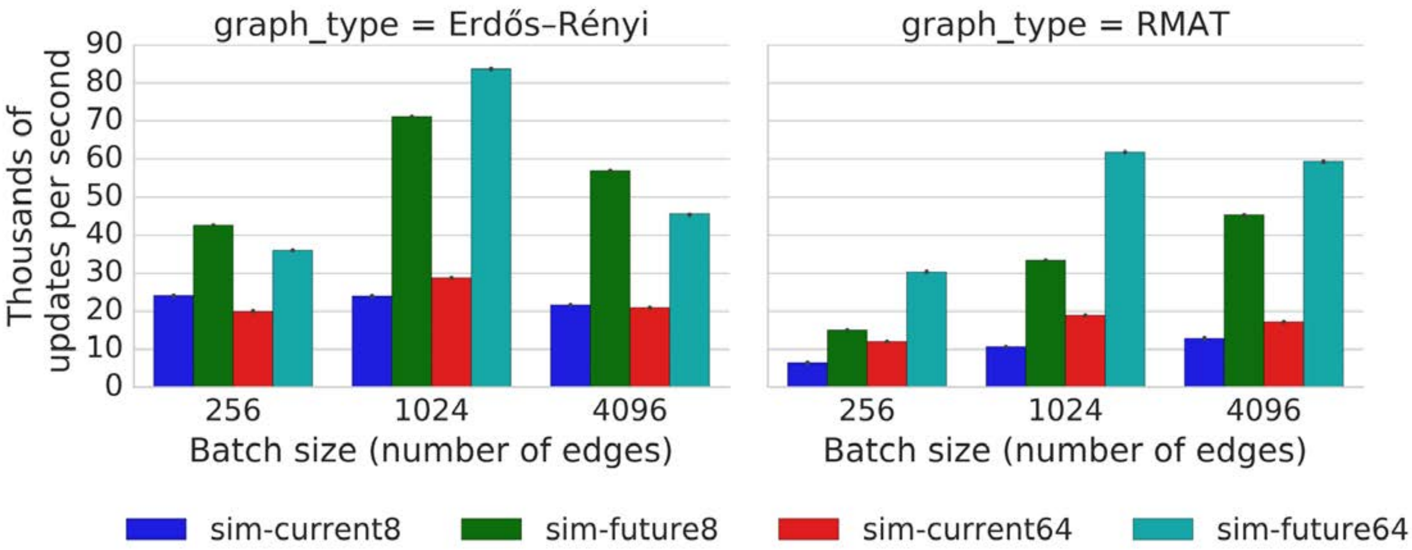
Near-term work is looking at interfacing with other neuromorphic APIs like EONS (Evolutionary Optimized Neuromorphic Systems)

We are also looking to host a prototype of UTK's DANNA architecture, which allows for testing neuromorphic architectures.

FPGAs as Rogue Enablers

Boards with HBM enable high bandwidth for mapping algorithms, machine learning, and neuromorphic computing (Intel Loihi, DANNA)

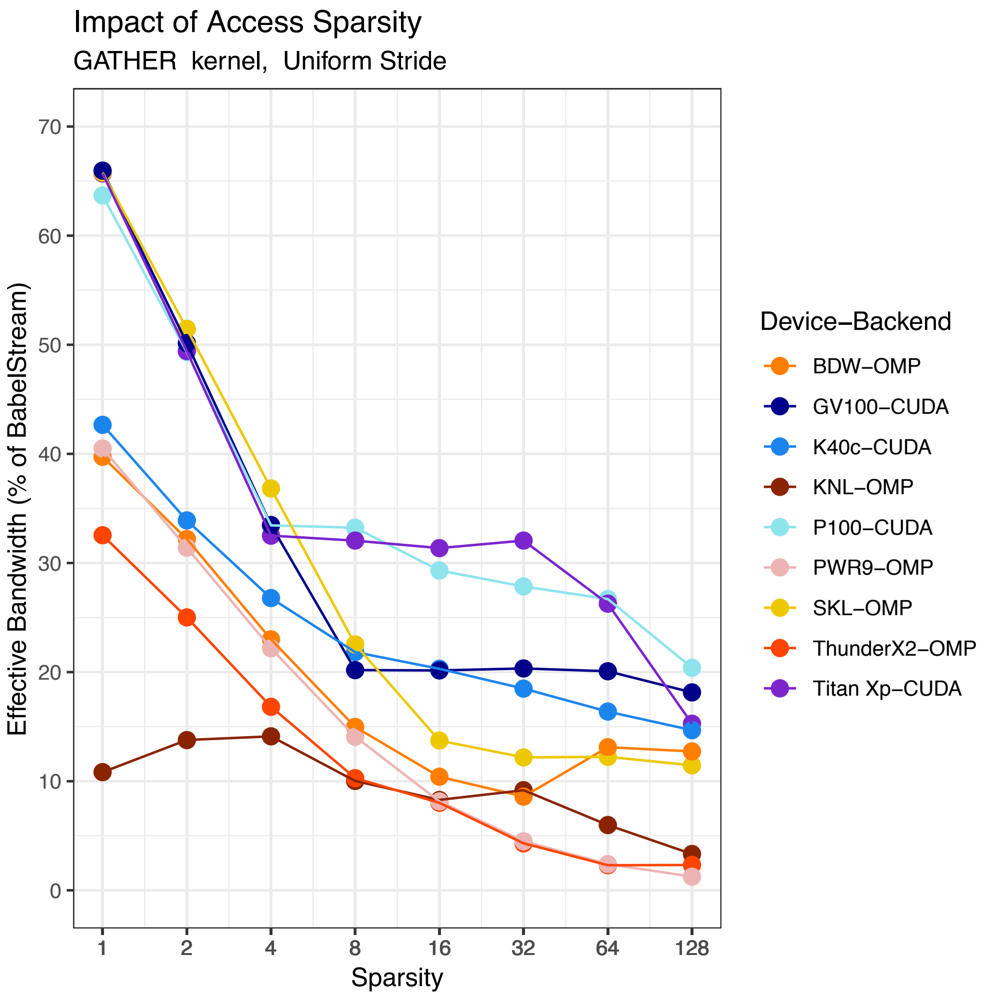
RG Benchmarking - Emu Microbench (STINGER)



Focused on Streaming BFS

Code available at https://github.com/ehein6/emu-microbench and increasingly as part of Emu toolchain releases

RG Benchmarking - Spatter



Code available at Spatter.io

Spatter provides a new benchmark to evaluate gather/scatter accesses across different CPU and GPU platforms.

Emu and FPGA backends are in progress!

RG Libraries – ParTI (Parallel Tensor Infrastructure)

Hardware

Multicore CPUs

GPUs

Distributed Memory

Emu

ParTI! Library

https://github.com/hpcgarage/ParTI

Sparse

Baseline Tensor Routines

Tensor Decomposition

Algorithms

RG Libraries - ParTI

Data Structures/ Platforms	Algorithms	Multicore CPUs	GPUs	Distributed Systems	Emu
COO	CP	✓	✓	✓	✓
	Tucker	✓	✓		
HiCOO	CP	✓	✓	✓	
	Tucker				

RG Tools - Kokkos

- Kokkos is a C++ library that is focused on ***performance portability*** through the mapping of parallel patterns

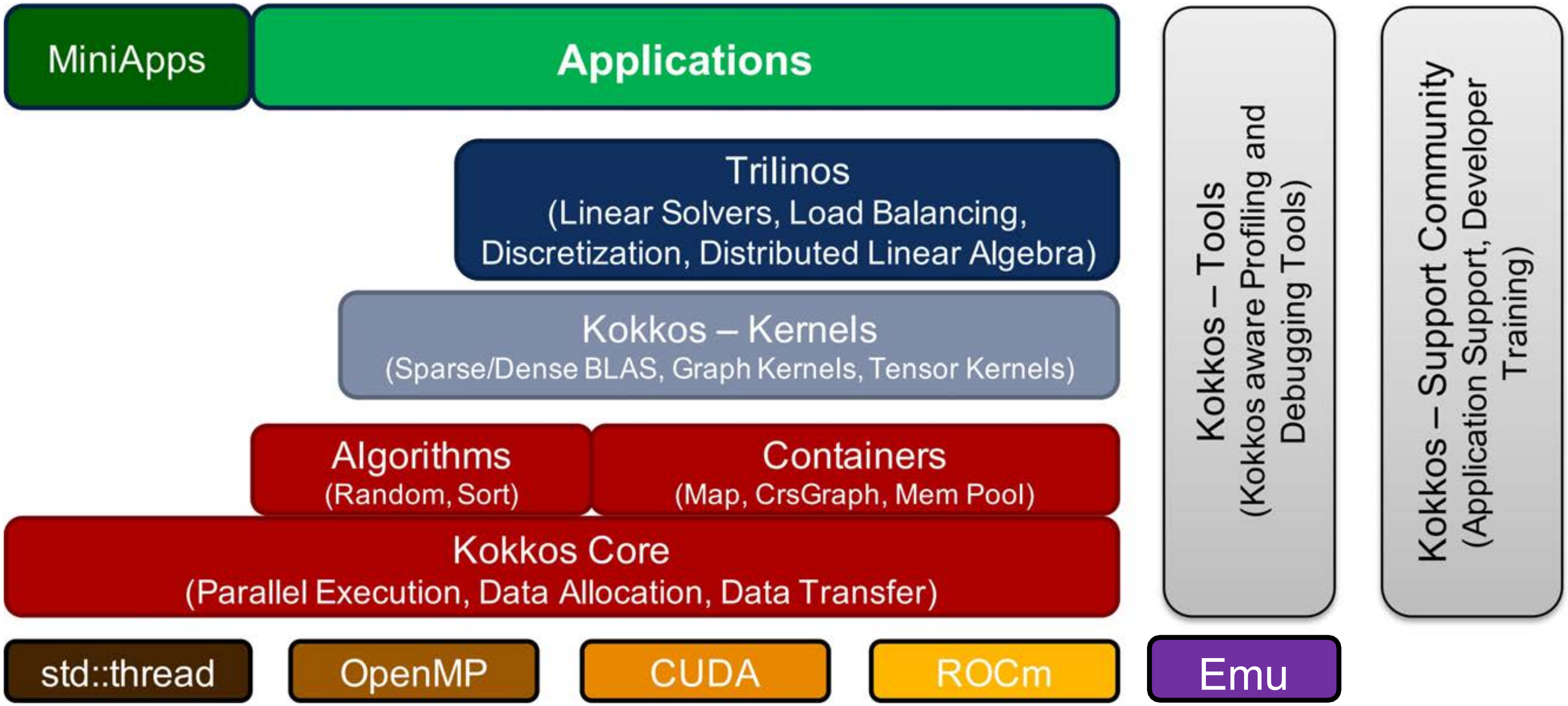
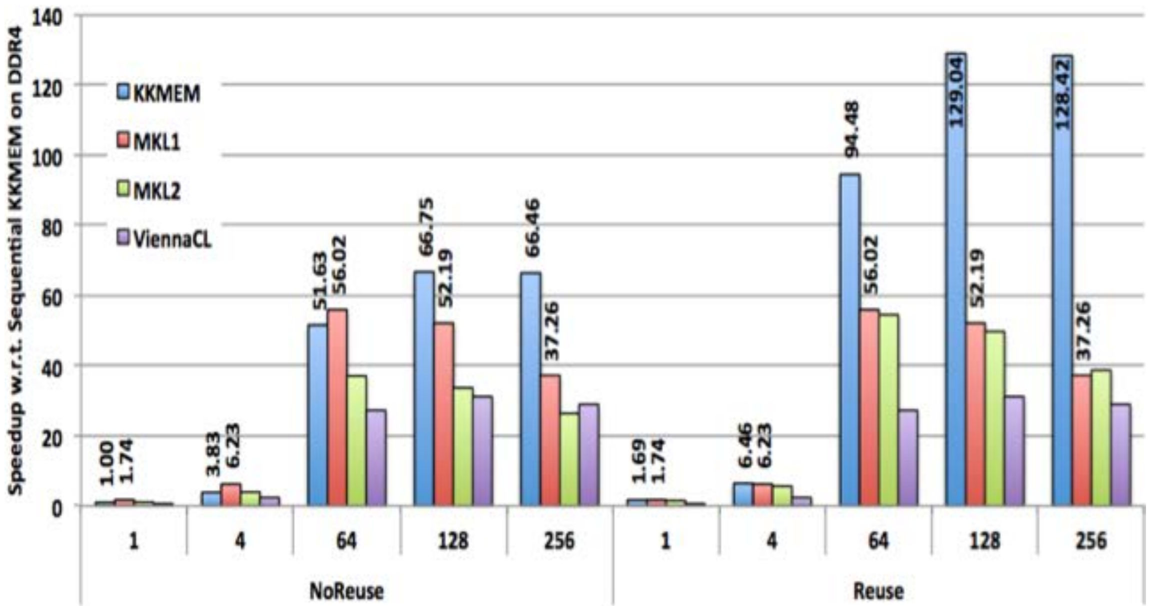


Diagram from Ellingwood, Trott, ANL 2018 Kokkos Tutorial

RG Tools - Kokkos

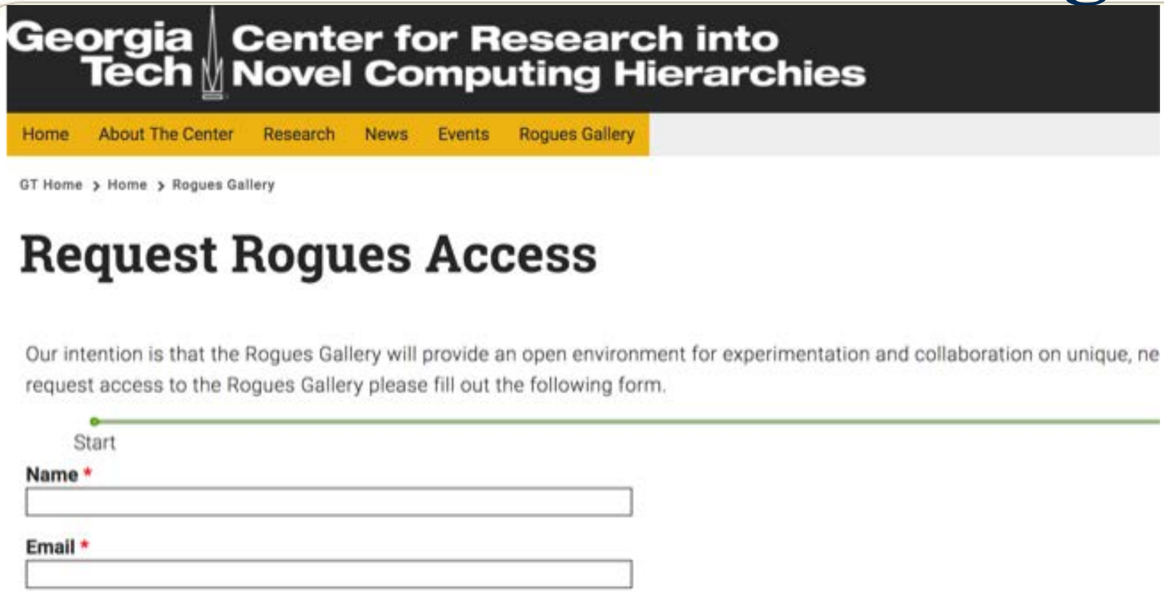
- Kokkos Kernels provide implementations of common sparse, dense linear algebra, and graph-related operations
  - BLAS, SPARSE, AXPY, graph coloring, tensor contraction
  - Common front-end implementations benefit from mapping to Kokkos

SPGEMM Results Comparing Kokkos Kernel implementation with MKL and ViennaCL on Knight's Landing



From Rajamanickam, et al. KokkosKernels: Compact Layouts for Batched Blas and Sparse Matrix-Matrix multiply, Batched BLAS Workshop, 2017

How to Work with the Rogues Gallery



Sign up for an account at crnch.gatech.edu – it's free!

Join our Slack or email lists and look out for CRNCH-sponsored events like tutorials and yearly summits in Atlanta, GA

Acknowledgments

- Emu – Janice McMahon and many others!
- Dr. Jennifer Hasler
- Srinivas Eswar (GT CSE)
- Dr. Eric Hein (GT ECE ⇒ Emu)
- Patrick Lavin (GT CSE)
- Jiajia Li (GT CSE ⇒ PNNL)
- Abdurrahman Yaşar (GT CSE)
- Dr. Ümit Çatalürek (GT CSE)
- Dr. Tom Conte (GT CS/ECE)
- Dr. Vivek Sarkar (GT CS/ECE)
- Dr. Bora Uçar (ENS Lyon CNRS)
- Dr. Rich Vuduc (GT CSE)

Code:

- crnch-rg.gitlab.io – front-end for public resources
- https://github.com/ehein6/emu-microbench
- Spatter.io – Spatter benchmark
- ParTI - https://github.com/hpcgarage/ParTI

Tutorial Feedback

- Feel free to email us:
  - Jason - jason.riedy@cc.gatech.edu
  - Jeff - jyoung9@gatech.edu
- We also have a survey for you to fill out on the event webpage at

crnch-rg.gitlab.io/asplos-2019