The Next Generation of GAP

An IoT Application Processor for Inference at the Very Edge

GREENWAVES TECHNOLOGIES

Presented by:

Martin Croome

VP Marketing
GreenWaves Technologies

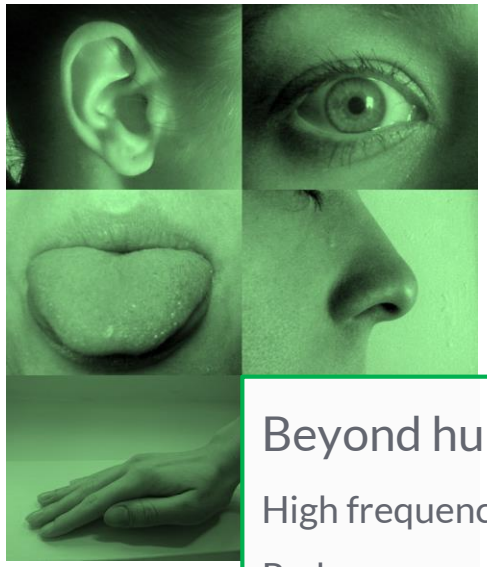
GreenWaves

- French fabless semiconductor startup
- Based in Grenoble, France
- Founded in November 2014
- Focused on designing and selling chips for AI and signal processing on battery operated IoT and wearable devices

Machine learning is becoming an attribute of all connected things

Sense

Human…



Beyond human…

- High frequency vibration
- Radar
- Infrared
- Ultrasound

Interpret & Analyse

Act & communicate

LoRa

NB-IoT

IoT & wearable Devices

require

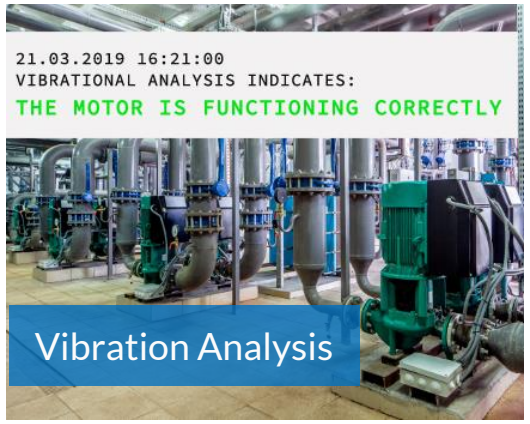
Signal processing and inference capabilities within a highly constrained power budget

GREENWAVES TECHNOLOGIES

Machine learning is becoming an attribute of all connected things

Workspace &
Energy Management

Machine Health
Monitoring

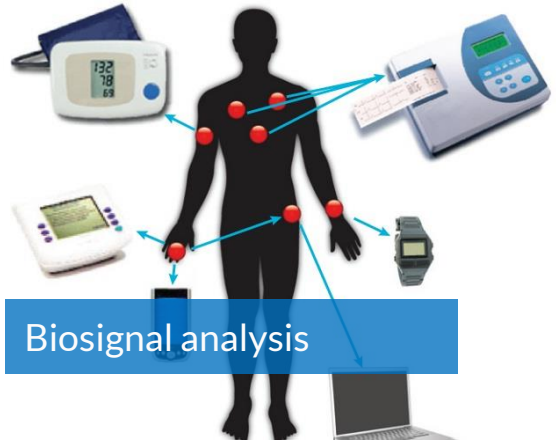


Safety &
Signalling Solutions

Security &
Access Control

Wearables &
Hearables

Medical sensors



Mini Robotics &
Toys

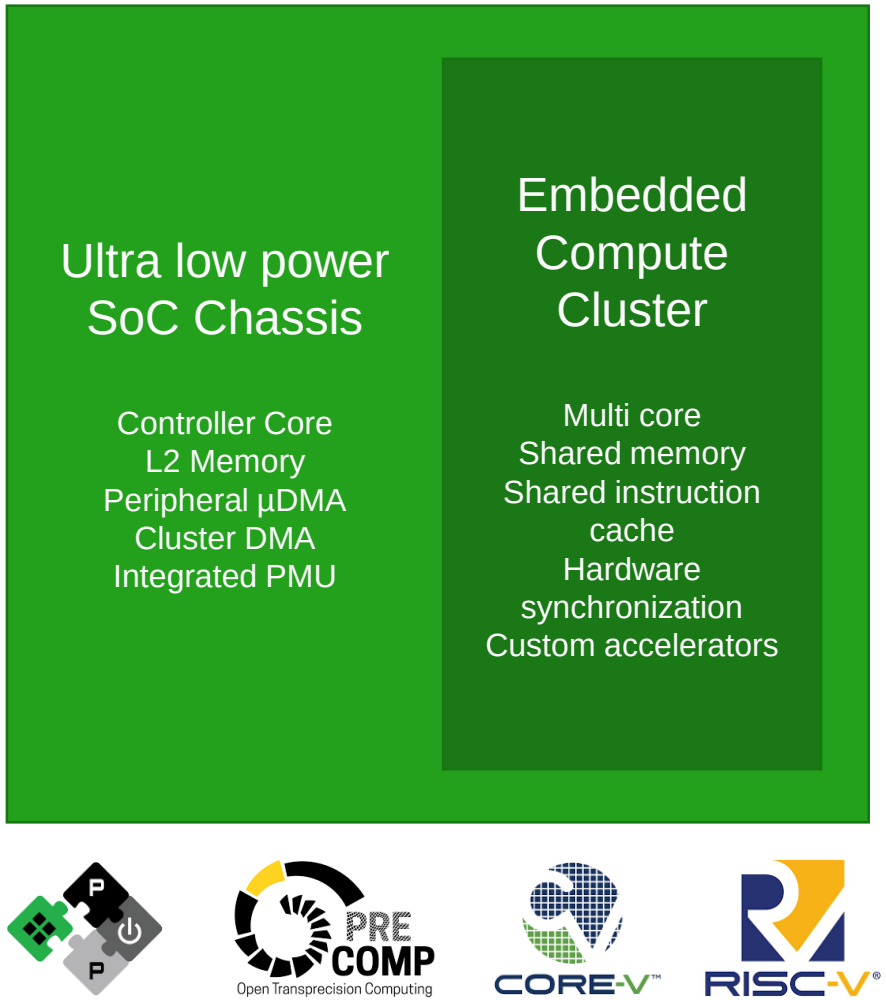
Smart Appliances

GREENWAVES TECHNOLOGIES

The fundamentals of GAP – Intelligence at the very edge

- MCU class energy consumption
  - Highly efficient parallelization
  - Sophisticated architecture (including instruction set architecture extensions)
  - Explicit memory movement
- Agility
  - Fine grained compute / energy scaling
  - Ultra fast state transitions
- Programmability
  - Applicable to many real world problems – not just CNNs
  - Exploits fast evolution of state-of-the-art
  - Single code model across architecture

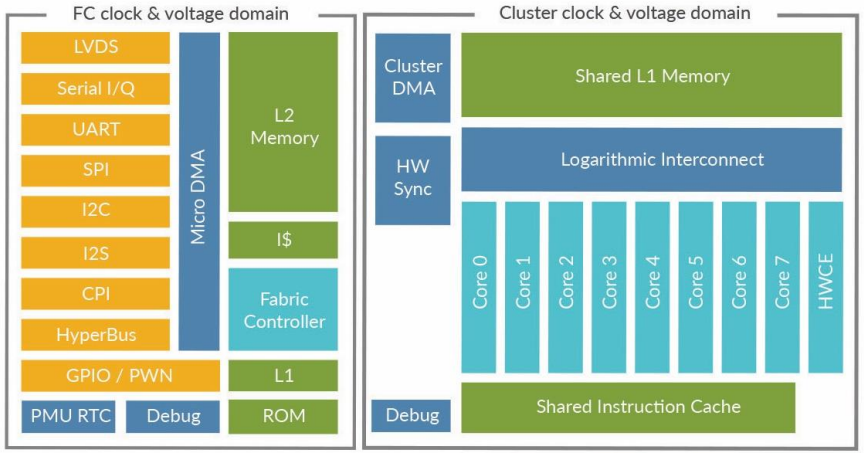
Single chip solution for an intelligent sensor



GREENWAVES TECHNOLOGIES

GAP8

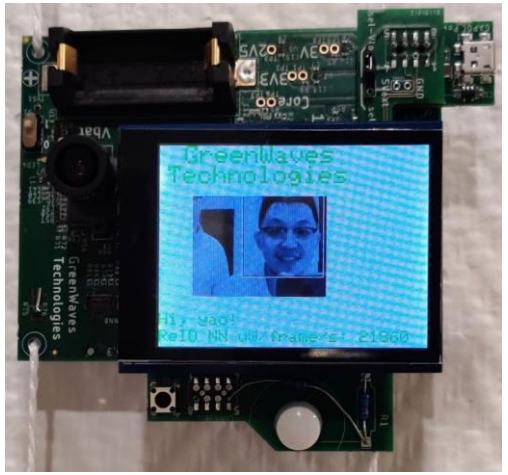
- First generation GAP processor
- Based on 5 years of research in PULP program
- TSMC 55nm process
- 9 Extended ISA RISC-V RV32 IMC cores
- 8 core cluster
- 1 core ‘fabric controller’
- HDKs since May 2018
- Production qualified
- First shipping products Q1 2020



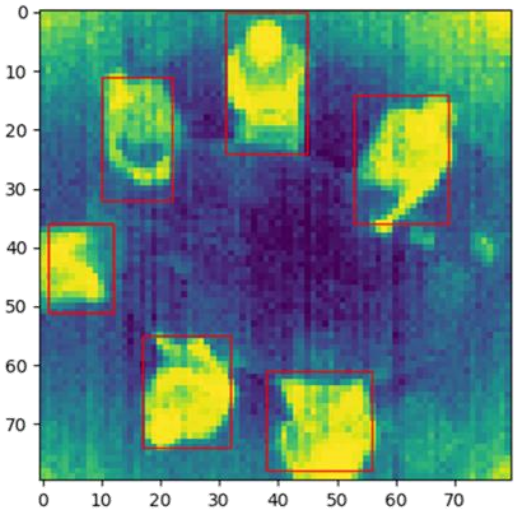
GREENWAVES TECHNOLOGIES

GAP8 has already achieved industry leading performance

- QVGA Face ID
  - Face detection: ~25ms ~1mW / frame / second
  - Face Reidentification: 400ms 22mW / frame / second
  - 93% accuracy on Labelled Faces in the Wild dataset
  - Embeddable owner detection on battery operated devices



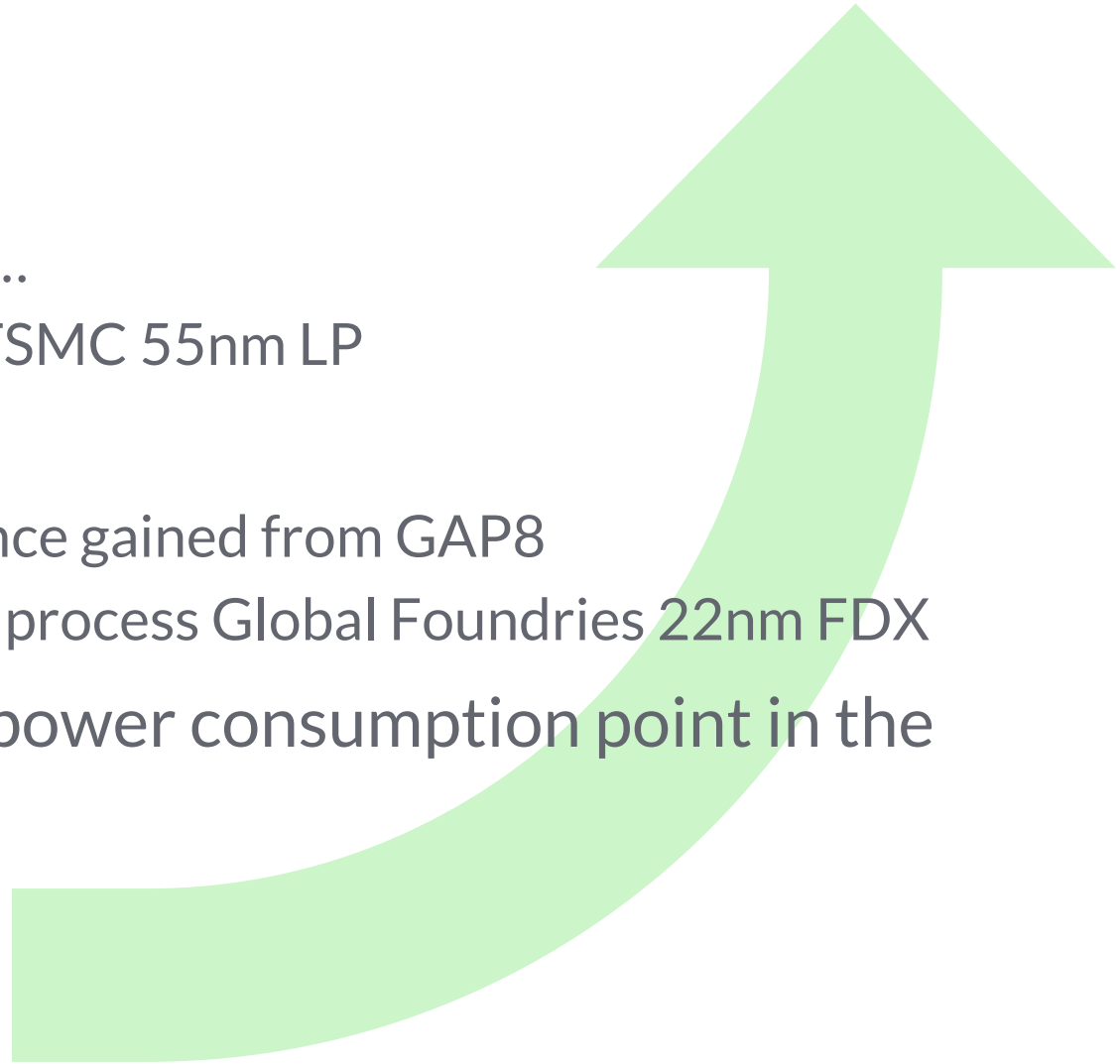
- IR people detection
  - 80 x 80 IR Image - LynRED ThermEye
  - Image preprocessing + human detection
  - 62ms ~4.4mW / frame / second
  - 99% accuracy on internally collected training set.
  - A full solution for people counting / occupancy detection on a battery for > 5 years



GREENWAVES TECHNOLOGIES

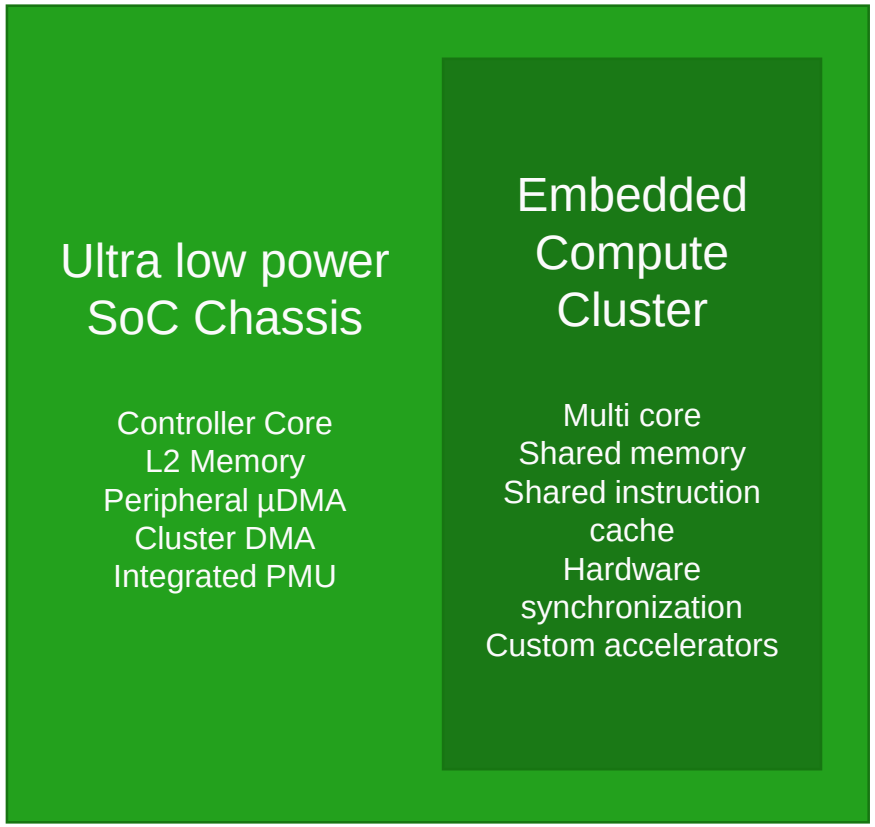
Introducing GAP9

- GAP8
  - Combined market leading architecture…
  - …with mature semiconductor process TSMC 55nm LP
- GAP9
  - Tunes GAP8 architecture with experience gained from GAP8
  - Exploits market leading semiconductor process Global Foundries 22nm FDX
- GAP9 establishes a new capability / power consumption point in the industry
  - 10 times larger problems than GAP8
  - 5 times less power than GAP8
  - Increases agility and programmability



GAP9 – Examples of architectural evolutions

- Increased Capability
  - Larger problems
    - 1.6MB internal RAM
    - Peak cluster L1 bandwidth of 41.6 GB/sec
    - Peak L2 bandwidth of 7.2 GB/s
    - Hardware compression
  - More compute states
    - 400MHz cluster top frequency
    - New power states
  - More flexibility
    - 32 / 16 / 8-bit floating point support
    - New bi-directional multi-channel digital audio interfaces
    - Additional CSI2 camera interface
- Increased security
  - HW AES 256/128 bit
  - HW Programmable Unclonable Function (PUF)



GAP9 vs. Arm M7 on MobileNet v1

Target	Clock (MHz)	Time (ms)	Cycles (M)	FPS	Active Power (mW)	Image Size	Channel Scaling	Top 1 ImageNet Accuracy
STM32 H7	400	162.5	65	6.2	170	160x160	0.25	43%
GAP9	29	162.5	4.77	6.2	5	160x160	0.25	43%
GAP9	400	11.925	4.77	83.9	50	160x160	0.25	43%
GAP9	400	167.5	67	6.0	50	192x192	1	70%

34 x less

14 x more

More accuracy

STM H7 figures - Running MobileNet on STM32 MCUs at the edge, Manuele Rucci
Accuracy estimates from TensorFlow model library
ImageNet performance in 1000 image classes

But architecture is only 50% of the story – tools is the rest

- What are customers expecting (Different things)?

  Expecting a packaged solution

  OR

  Expecting a known network

  OR

  Expecting to revolutionize the world

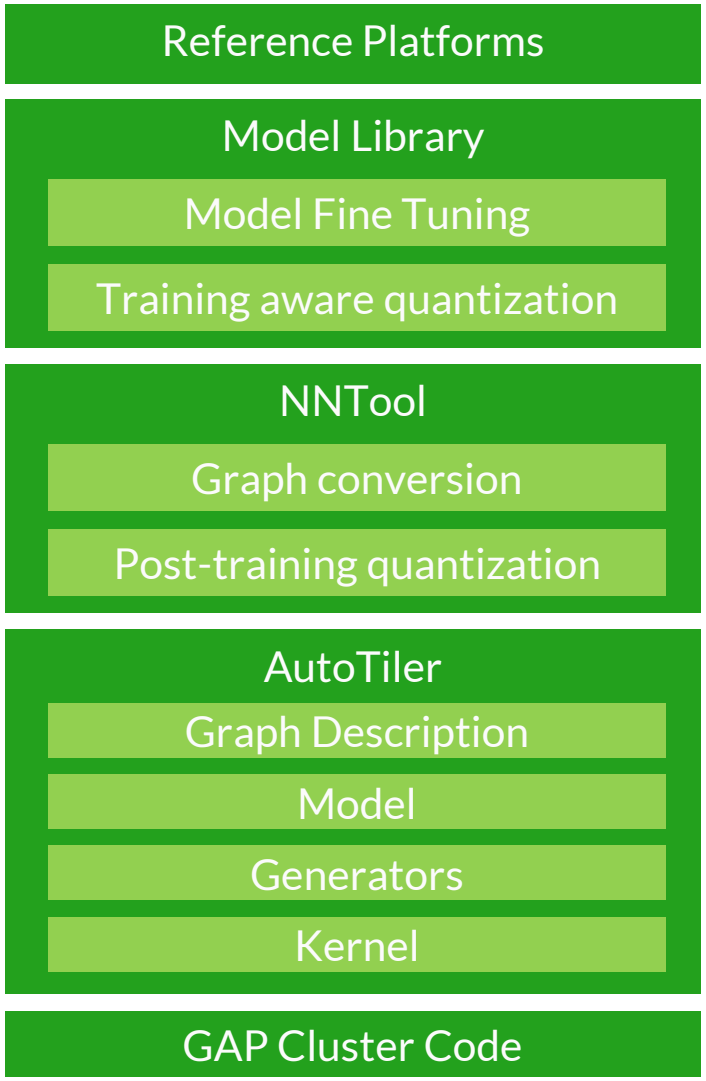
- Each of these customers requires a different tool set

Solutions

Pre-packaged fine-tuneable networks

Expert tools to exploit architecture

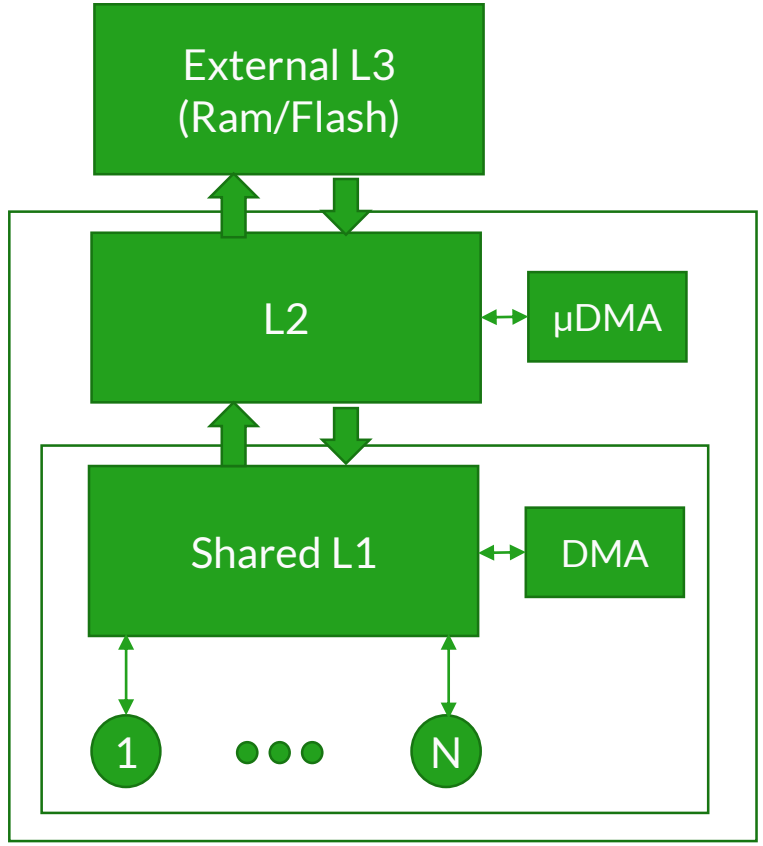
GAP*Flow*

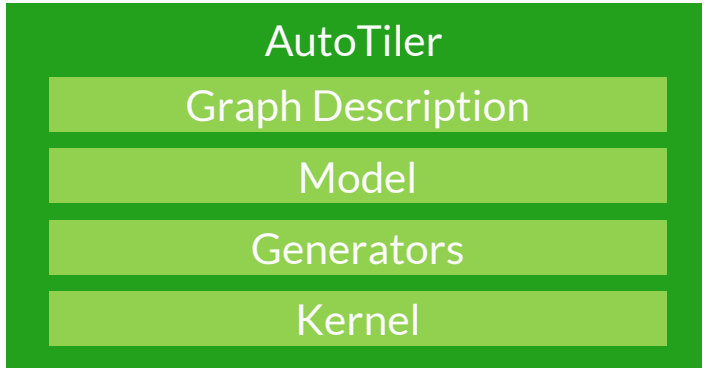


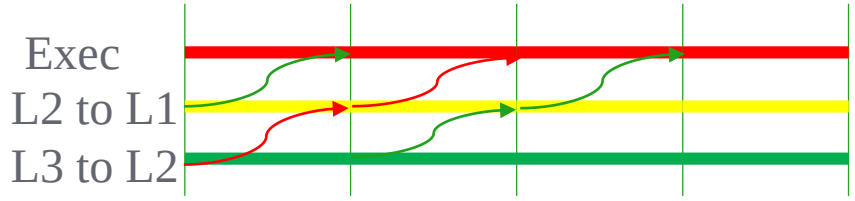
- A series of build system agnostic, modular tools that convert Graphs to GAP code
- Build examples based on Makefiles but usable with any build system
- NN focus but by no means limited to NN
- Use one, use all, use none
- Extendable
- Clear points of failure
- Enhanced with examples networks and full applications that use it

GAP AutoTiler – Explicit memory movement

- Data caches are not good for streamed data – low cache efficiency
- ML / Signal Processing data traffic sizing is known at compile time
- Generate code for automatic data tiling and pipelined memory transfer interleaved with parallel call to compute kernel







GREENWAVES TECHNOLOGIES

GAP family is enabling ground breaking applications at the very edge

- GAP9 development boards available in early 2020 for lead customers
- GAP9 simulator has been in customer hands since May 2019
- GAP8 shipping now in production
- GAP8 development boards and evaluation boards for vision and IR vision shipping now
- GAP SDK available on our GitHub repository
- Come and see our demonstrations on the Open HW Group booth

Real … Now …

GREENWAVES TECHNOLOGIES

Thank you!

Questions?

GREENWAVES TECHNOLOGIES