GAP8 IOT Application Processor

A PULP/RISCV BASED PLATFORM FOR NEAR-SENSOR ANALYTICS

Eric Flamand. CoFounder & CTO
Greenwaves Technologies
ANYTHING THAT BENEFITS FROM NETWORK CONNECTION WILL BE CONNECTED

Source: Ericsson
Cost of transporting these data over the air?

Serial short reach link: best results around 0.5 pJ/bit
LTE: between 300 and 600 uJ/bit

Even assuming distributed computing is marginally more efficient than centralized we win big if data volume to be exchanged over the air is shrunk by several order of magnitude moving from quantitative data to qualitative data!
So in practice ....
If we want data reach sensors

Move from (raw) data to meta data (abstract/pertinent)
Perform this transformation close to sensor
While fitting in a tight power and cost budget
And being seamlessly integrated to the Internet over the air
Three main sources of intensive data

- **Image**: Raw input in the order of 100KB/s for a small sensor
  - Scene classification
  - Posture analysis
  - Identification
  - Output is a single index

- **Voice/Sound**: Raw input in the order of 10KB/s per mic
  - Recognition
  - Identification
  - Signature analysis
  - Output is a single index

- **Vibrations**: Raw input in the order of 10KB/s
  - Preventive maintenance
  - Monitoring
  - Output is a single index or an alarm

Once properly processed, common denominator is: extremely compact output (single index, alarm, …)

Bandwidth is reduced by several order of magnitude.
What we want to achieve

Giga/Mega Bytes per second of *incoming* raw data from sensors

Few (Kilo) Bytes per second of *outgoing*, heavily processed data @ minimum Joule per operation
System level view

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered

Sense

Analyze and Classify, Manage radio diversity

Transmit

Short range, BW

Low rate (periodic) data

SW update, commands

Long range, low BW

Resolution just enough for the job, order of QVGA/VGA

Battery powered

U Controller

L2 Memory

IOs

Parallel Processor

ULP

Resolution just enough for the job, order of QVGA/VGA

Battery powered
General pattern for content understanding

- Extract descriptors from raw data
  - 2D: Corners, blobs, HOG, DOG, …
  - 1D: LPC coefficients, Cepstral coeffs, …

Usually highly parallel

- Use descriptors to classify data among representative families
  - Machine learning (CNN, SVM, Boost), Bayesian, ….

Also highly parallel
GAP8: Ultra Low Power IoT Processor

Performances
- up to 12GOPS
- up to 0.4GOPS @ 1mW,
- up to 40MOPS @ 300uW
- 3 uWatt stand-by power consumption

Architecture efficiency
- Extended Risc-V ISA
- Low contention shared memory 8 +1 core clustered architecture
- Tight synchronization
- CNN based pattern matching engine (HWCE)

HW features
- Smart IOs
- Voltage regulator/DVFS
- RTC
- Secured execution

Performances
- up to 12GOPS
- up to 0.4GOPS @ 1mW,
- up to 40MOPS @ 300uW
- 3 uWatt stand-by power consumption

Application affinity
- Dominant signal processing part
- Limited memory requirement
- Limited SW legacy

GAP8 has a unique energy efficiency across a very large range of computing power

Leveraging open source projects
- Risc-V (Berkeley)
- PULP (ETHZ, UniBo)

Low cost processor
- 55nm LP
- 0.5MB L2
- aQFN 84

L2 Memory
- PMU
- DC/DC
- RTC
- LVDS
- UART
- SPI
- I2S
- I2C
- // 10b
- GPIOs
- Jtag
- Micro DMA
- I$
- Fabric Ctrlr
- L1
- Rom
- Ddbg
- Clk

Cluster
- DMA
- HWCE
- HW Sync
- Ddbg Unit

Shared L1 Memory
- Logarithmic Interconnect
- Shared Instruction Cache

Shared L2 Memory

Leveraging open source projects
- Risc-V (Berkeley)
- PULP (ETHZ, UniBo)

Low cost processor
- 55nm LP
- 0.5MB L2
- aQFN 84

Application affinity
- Dominant signal processing part
- Limited memory requirement
- Limited SW legacy

GAP8 has a unique energy efficiency across a very large range of computing power
## GAP8 Hierarchical Architecture

<table>
<thead>
<tr>
<th><strong>monitoring</strong></th>
<th><strong>event qualification, protocol stack, system control</strong></th>
<th><strong>data analysis &amp; classification, SW modem</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Smart I/Os, voltage regulator &amp; RTC, SRAM in retentive mode</td>
<td>extended RISC-V</td>
<td>extended RISC-V</td>
</tr>
<tr>
<td>quasi stand-by</td>
<td>low computing power</td>
<td>efficient 8 core parallelization</td>
</tr>
<tr>
<td>uWs</td>
<td>mWs</td>
<td>HW synchronization</td>
</tr>
<tr>
<td>primary energy consumption</td>
<td></td>
<td>shared instruction cache</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CNN HW engine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>high computing power</td>
</tr>
<tr>
<td></td>
<td>10 to 20 mWs</td>
<td>primary energy consumption</td>
</tr>
</tbody>
</table>

### Diagram Elements
- **L2 Memory**
- **Cluster**
  - DMA
  - CNN-HWE
  - HW Sync
  - I$\$ (L1)
  - Rom
- **Dbg Unit**
- **Shared L1 Memory**
- **HyperBus**
- **GAP8 Hierarchical Architecture**
  - **monitoring**
    - Smart I/Os, voltage regulator & RTC, SRAM in retentive mode
  - **event qualification, protocol stack, system control**
    - extended RISC-V
    - low computing power
    - mWs
  - **data analysis & classification, SW modem**
    - extended RISC-V
    - efficient 8 core parallelization
    - HW synchronization
    - shared instruction cache
    - CNN HW engine
    - high computing power
    - 10 to 20 mWs
    - primary energy consumption

### Energy Consumption
- **uWs**
- **mWs**
- **10 to 20 mWs**

### Features
- **SW modem**
- **primary energy consumption**
- **extended RISC-V**
- **efficient 8 core parallelization**
- **HW synchronization**
GAP8 architectural energy efficiency gains

- Data analysis & classification, SW modem
- Extended RISC-V
- Efficient 8 core parallelization
- HW synchronization
- Shared instruction cache
- CNN HW engine

Overall, in practice on targeted algorithms, typically 20x

3-5x
1.4x
1.5x
5x
**GAP8 Advanced Power Management**

### MCU sleep mode
- Embedded DC/DC, low current
- Real Time Clock 32KHz only
- L2 Memory partially retentive

### MCU active mode
- Embedded DC/DC, high current
- Voltage can dynamically change
- One clock gen active, frequency can dynamically change
- Systematic clock gating

### MCU + Parallel processor active mode
- Embedded DC/DC, high current
- Voltage can dynamically change
- Two clock gen active, frequencies can dynamically change
- Systematic Clock Gating

Ultra fast switching time from one mode to another
Highly optimized system level power consumption

Ultra fast voltage and frequency change time
Qualitative data from real life applications
The work horse for radio, sound and vibration: FFT

Key operations for performance:

- Complex Multiplications
- Complex Rotations
- Post modified accesses
- Vectorial operations

All these butterflies are evaluated in parallel.
The work horse for radio, sound and vibration: FFT

Number of operations (\(*,+,\gg,\text{Ld/St}\) vs. Number of Cycles running on 8 cores

<table>
<thead>
<tr>
<th>FFT 256</th>
<th>FFT 1024</th>
<th>FFT 4096</th>
</tr>
</thead>
<tbody>
<tr>
<td>11264</td>
<td>56320</td>
<td>225280</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FFT 256</th>
<th>FFT 1024</th>
<th>FFT 4096</th>
</tr>
</thead>
<tbody>
<tr>
<td>1167</td>
<td>4842</td>
<td>22710</td>
</tr>
</tbody>
</table>

Ideal means perfect scaling, perf on N cores = Perf on 1 core / N
The work horse for radio, sound and vibration: FFT

ARM FFT1024 Q15 Data are with CMSIS optimized library
Visual Localization: FFT2D + HOG

FFT2D 256x256. Radix 4

<table>
<thead>
<tr>
<th>Number of Cores</th>
<th>Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4000000</td>
</tr>
<tr>
<td>2</td>
<td>3000000</td>
</tr>
<tr>
<td>4</td>
<td>2000000</td>
</tr>
<tr>
<td>8</td>
<td>1000000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of Cores</th>
<th>Cycles Per Pixel</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2000</td>
</tr>
<tr>
<td>2</td>
<td>1000</td>
</tr>
<tr>
<td>4</td>
<td>500</td>
</tr>
<tr>
<td>8</td>
<td>250</td>
</tr>
</tbody>
</table>

Histogram Of Gradients
8x8 Cells, Blocks: [2x2] Cells, 50% Overlap. 640x480 BW Image

<table>
<thead>
<tr>
<th>Number of Cores</th>
<th>Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1000000</td>
</tr>
<tr>
<td>2</td>
<td>500000</td>
</tr>
<tr>
<td>4</td>
<td>250000</td>
</tr>
<tr>
<td>8</td>
<td>125000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of Cores</th>
<th>Cycles Per Pixel</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>25</td>
</tr>
</tbody>
</table>

1 384 000 cycles per image
589 000 cycles per image
We need only 2 MHz per image
CNN based Image Classification

**CIFAR10**

- 32x32 → 28x28 → 14x14 → 10x10 → 5x5 → Linear
- 5x5 Conv
- MaxPool 2x2:1

**MNIST**

- 28x28 → 24x24 → 12x12 → 8x8 → 4x4 → Linear
- 5x5 Conv
- MaxPool 2x2:1

1 core to 8 cores + HWCE: 10.9 speedup
1 core to 8 cores + HWCE: 9.3 speedup
CNN based Image Classification

CNN 13 Layers, 128x128 Input, 14 Outputs

Trainable Par: 421 263
Neurons: 1 511 904

33ms per image
People Counting

- Filtering + Difference of Gradient + SVM-RBF
- Open Space. Accuracy: approx 90%
- 1 Image every 3 minutes => 10 years on a battery
Audio Processing

N sub bands, one triangle filter per sub band, mel scale

3N Coefficients (Coeff, Coeff', Coeff'') at each sample time

MFCC, 64 Bands, 16KHz/PCM16, 30ms Frame, 20ms Overlap

0.65 MHz on 8 Cores
Hierarchical Power Processing?

Gradually turn on resources

Sound ? ➞ Phonem ? ➞ Word ?

Cluster
- DMA
- HWCE
- HW Sync
- Dkg Unit

Shared L1 Memory

Logarithmic Interconnect

Shared L1 Memory

L2 Memory
- Fabric Ctrler
- L1
- Rom

Dbg Unit
- Dkg
- Clk

PMU
- DC/DC
- RTC

LVDS
- I2C
- SPI
- // 10b
- GPIOs
- Jtag

Micro DMA

Frequency bands

Static, A, AA

Convolution layer feature maps

max pooling layer nodes

Share same weights

Conclusion

• GAP8 bridges the gap between ultra low power MCU and multi-core processor:
  – Smart sensing on data rich sensors achievable within tinny power budget: uW in Idle, mW in micro controller mode, 5-20mW in number crunching mode: Few Mops to up to 12Gops
  – Low cost bill of material

• GAP8 agile power management architecture combined with IOT low duty cycling is a perfect fit for FDSOI process