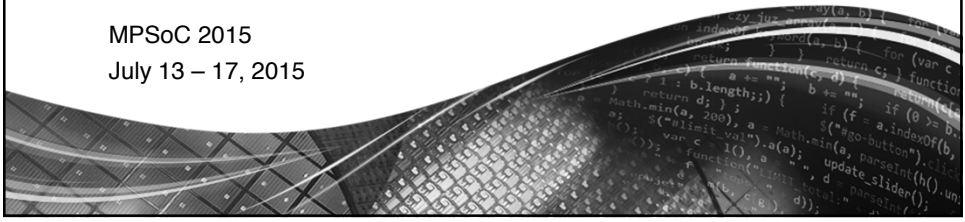


Scalable Processor Solutions for IoT Applications

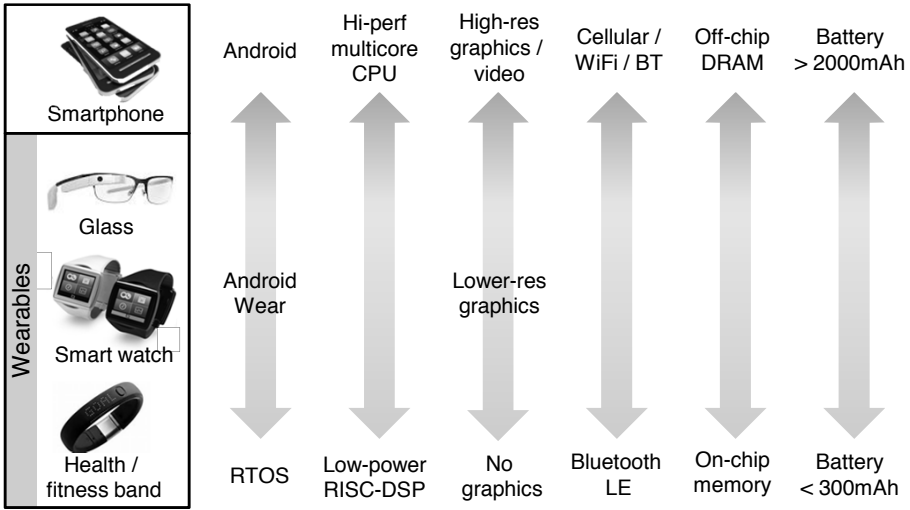
Pieter van der Wolf



MPSoC 2015
July 13 – 17, 2015

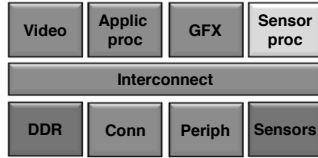


Always-on Mobile Devices *Devices with different characteristics*



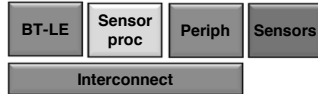
Always-on Mobile Devices

Processors for always-on processing



Processor for always-on

- Operate in different SoC contexts
- Fmax typically < 100 MHz
- Lowest power in each mode
- Mixed control and DSP



- **Separate core for always-on processing**
 - Wake-up application processor only when needed
- **Low power**
 - > 10x lower power than application processor
 - Battery in wearable needs to last weeks
- **Multiple modes**
 - E.g. voice activation:
 - Standby / detection mode
 - Recognition mode
- **Mixed control and DSP**
 - DSP for processing of sensor inputs

Processors for IoT Applications

No 'one-size-fits-all'

- Diverse functionalities → diverse requirements
 - Great variety of IoT products with different mixes of functionalities
 - Data rates and computational demands vary widely

Connectivity	Bluetooth LE Bluetooth WiFi ZigBee GPS ...
Security	Secure boot, authentication & authorization, encryption, ...
Sensor	Motion sensing, voice activation, face detection, ...
Infotainment	Audio (codecs, pre-/post-processing), voice (ITU-T), ...

- Scalable
 - To offer right performance level
 - Both control and DSP
- Lowest energy consumption
 - Low power consumption (μ W/MHz)
 - High cycle efficiency (for lowest MHz)

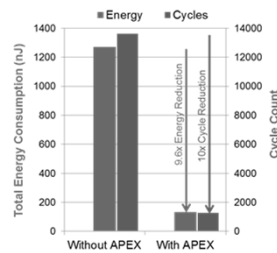
No 'one-size-fits-all' processor that can offer:

- Required features
- Required performance
- At lowest energy consumption for different IoT products

Processors for IoT Applications

Satisfy diverse requirements with processor IP

- Product family
 - Set of processors with different features and performance levels
 - IoT: too much diversity for small set of “fixed” processors
- Configurability
 - Allow tuning by (de-)configuring features for specific IoT application
 - E.g. bitstream parsing, ITU-T operations, FFT butterfly, crypto accelerators, ...
- Extensibility
 - Allow customer-specific extensions
 - New instructions in processor
 - E.g. trig-accelerator with Sin, Cos, ...



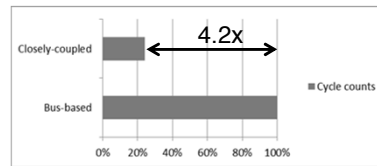
© 2015 Synopsys, Inc. 5

SYNOPSYS

Processors for IoT Applications

Memory architecture

- Low latency memory access
 - Closely coupled memories
 - Rather than access over interconnect
 - Fewer stall cycles and lower energy
- Smaller memories with concurrent access
 - Use multiple smaller memories to reduce energy per access
 - Allow concurrent access to reduce stalls
 - X/Y memories, interleaving
 - Fewer stall cycles and lower energy
- Fewer accesses
 - Fetch less instruction data
 - Zero-overhead loops, address generation units, implicit load/store, ...
 - Wide instruction memory to fetch multiple instructions at once
 - Fewer cycles, smaller code size and lower energy



Memories	Rd power (μW/MHz)	Wr power (μW/MHz)	Area (μm ²)
4kB (1024x32)	1.755	1.765	9394
8kB (2048x32)	2.586	2.633	16755
32kB (8192x32)	4.301	4.512	57899
64kB (16384x32)	6.387	6.319	111856

© 2015 Synopsys, Inc. 6

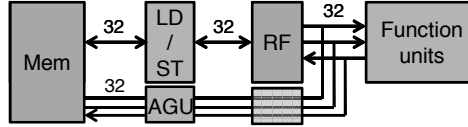
SYNOPSYS

Processors for IoT Applications

XY memory architecture

```

C source code
q31_t foo(q31_t *b, q31_t *c) {
    q31_t s = 0;
    for (i = 0; i < N; i++)
        s += b[i] * c[i]
    return s;
}
    
```



```

Non-XY assembly
// loop, unrolled 2x
LP lpend
LD r2, [r0, 4]
LD r3, [r1, 4]
LD r4, [r0, 4]
LD r5, [r1, 4]
MAC 0, r2, r3
MAC 0, r4, r5
lpend: // epilogue for odd sized loops
...
    
```

	Non-XY	XY
Performance [MAC/cycle]	0.3	1.0
I-power [B/MAC]	12	4

```

XY assembly
// prologue, set-up address gen
SR ...
SR ...
LP lpend
MAC 0, r32, r33
lpend: // no epilogue
...
    
```

3x performance →

Processors for IoT Applications

Multi-issue architecture

```

C source code
q31_t foo(q31_t *b, q31_t *c) {
    q31_t s = 0;
    for (i = 0; i < N; i++)
        s += b[i] * c[i]
    return s;
}
    
```

	Non-XY	XY	Multi-issue
Performance [MAC/cycle]	0.3	1.0	1.0
I-power [B/MAC]	12	4	8

```

Multi-issue assembly
// Prologue SW pipelined
LDD r2, [r0, 8] ; 64b vector load
LDD r4, [r1, 8] ; 64b vector load
LDD r6, [r0, 8] ; 64b vector load
// 4x unrolled loop
LP lpend
{MAC 0, r2, r4; LDD r8, [r0, 8]} ; 32b MAC and 64b load
{MAC 0, r3, r5; LDD r2, [r1, 8]} ; 32b MAC and 64b load
{MAC 0, r6, r8; LDD r4, [r0, 8]} ; 32b MAC and 64b load
{MAC 0, r7, r9; LDD r6, [r1, 8]} ; 32b MAC and 64b load
lpend: // epilogue if loop count not multiple of 4
...
    
```

Conclusions

- IoT applications pose diverse requirements
 - Different feature mixes and performance levels
 - Lowest energy consumption
- Configurability and extensibility are essential
 - To provide right features & performance at lowest energy
- Memory architecture is key
 - For high performance and low energy consumption
 - Must be configurable to allow area – performance – energy trade-offs

Thank You

