



Compound circuits in multi-accelerator architectures: How to balance between flexibility and specialization

IWLS 2010 – 4th June 2011
Sami Yehia

IWLS / 4th June 2011

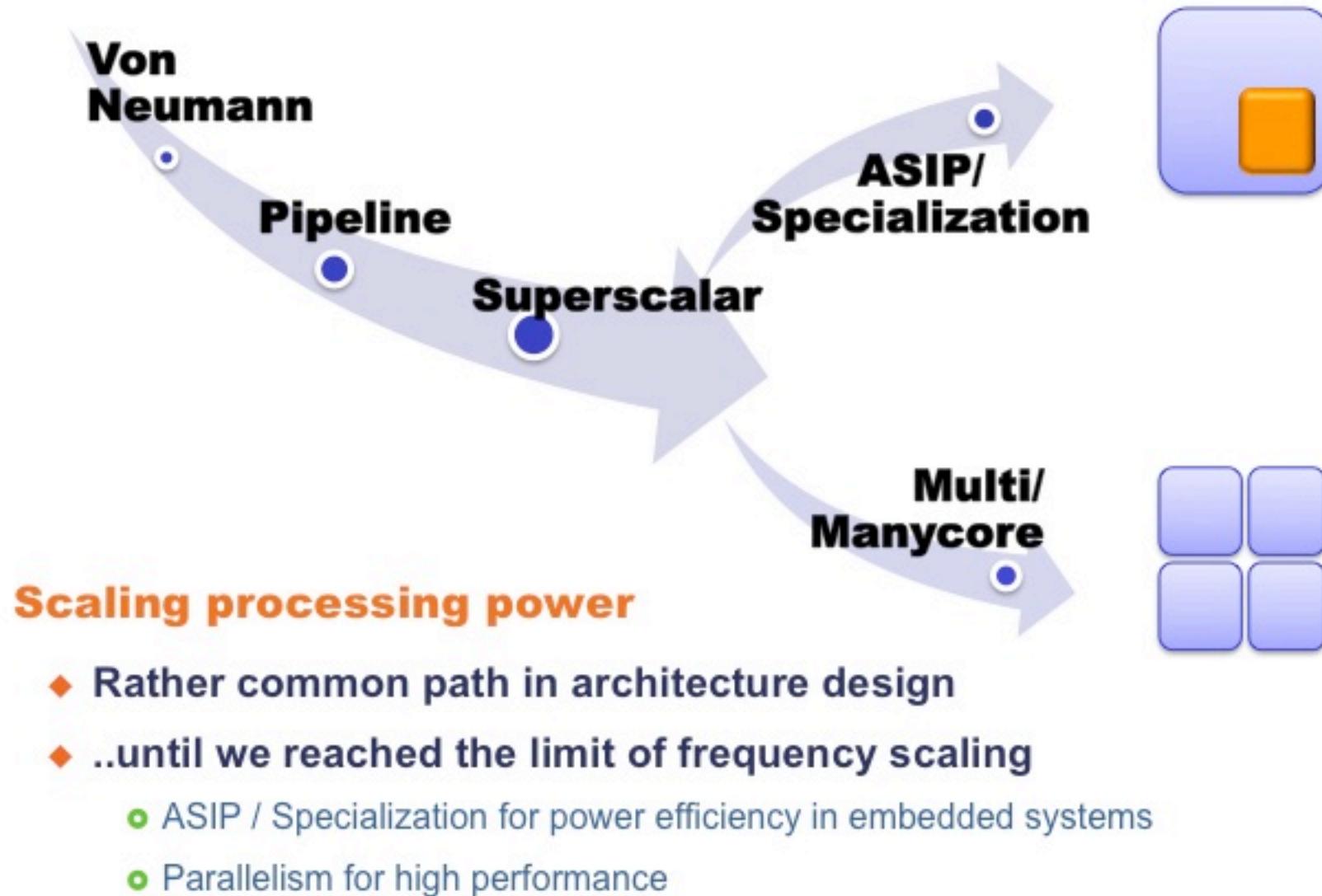


Research & Technology

SMECY
Smart multicore
embedded systems



THALES



About the future of Many cores

- “We need to consider how much to move into domain-specific architectures rather than having a generic C engine,” said Flautner. “The vision of the future as some people express it in the post-frequency scaling environment is the move from one to two to four to eight to thirty-two all the way to a billion cores. I wondered what these guys are smoking. Will it be another trend where you just turn the knob? And live on it for 20 years?” Kris Flautner (ARM senior VP) on the future of homogeneous multicore.

Parallel pain, Chris Edward, Institution of engineering technology, July 2008

15th century



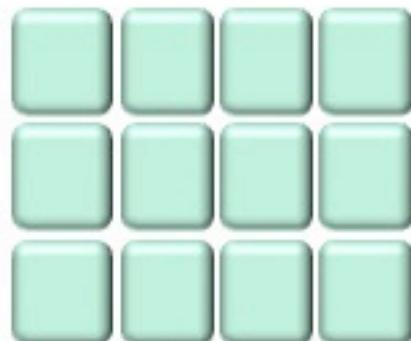
scientist, mathematician,
engineer, inventor, anatomist,
painter, sculptor, architect,
botanist, musician and writer

16th century

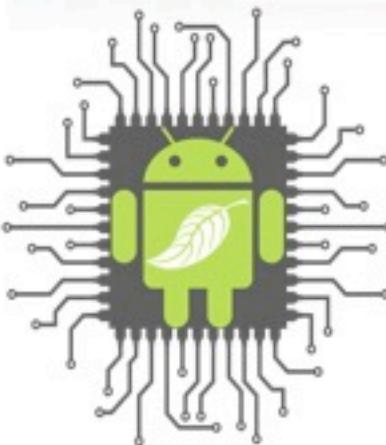
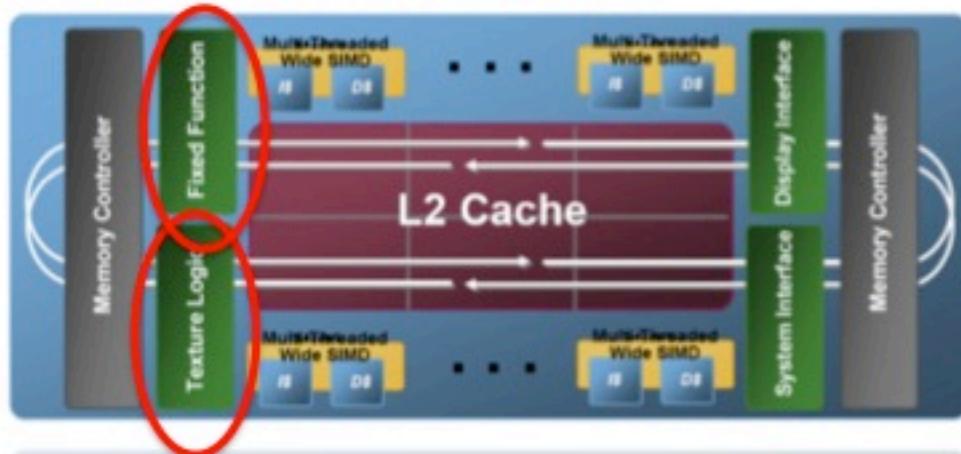


physicist, mathematician,
astronomer, and philosopher

20th century



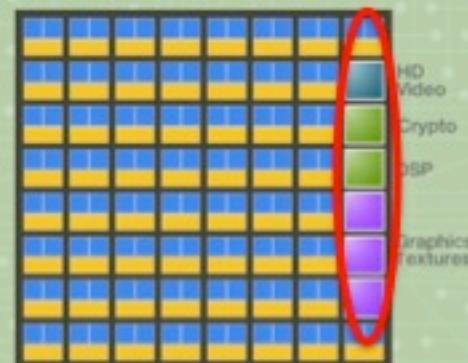
Larrabee Block Diagram



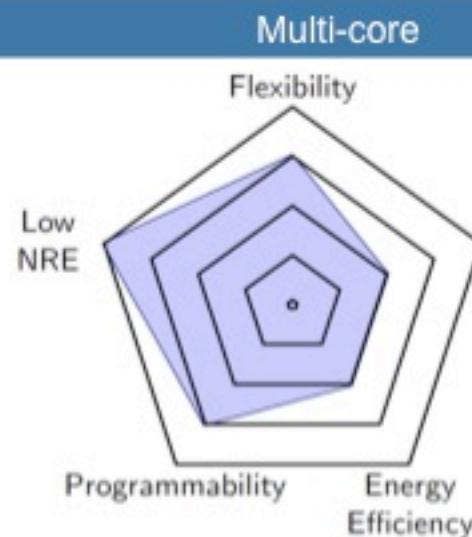
GREENDROID

New Capabilities

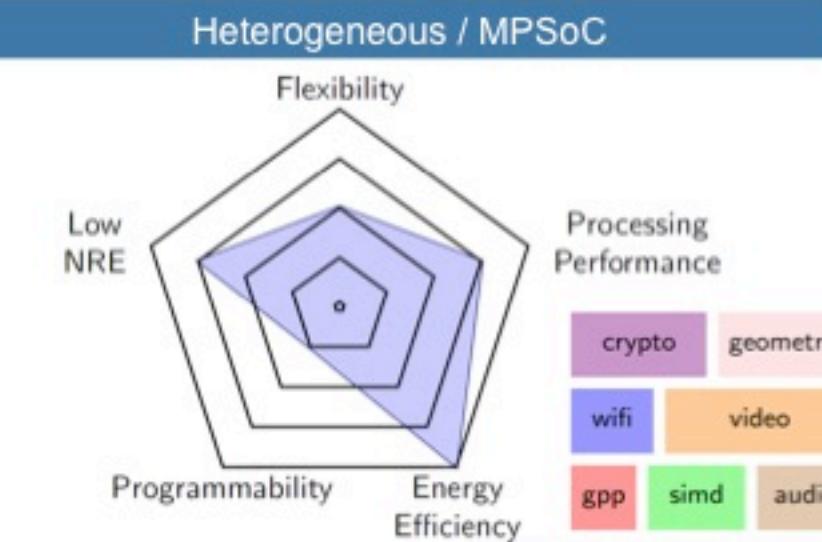
To expand functionality, some dedicated hardware engines could be integrated to accelerate multimedia, networking, security, or other tasks. The number and type would vary by target platform



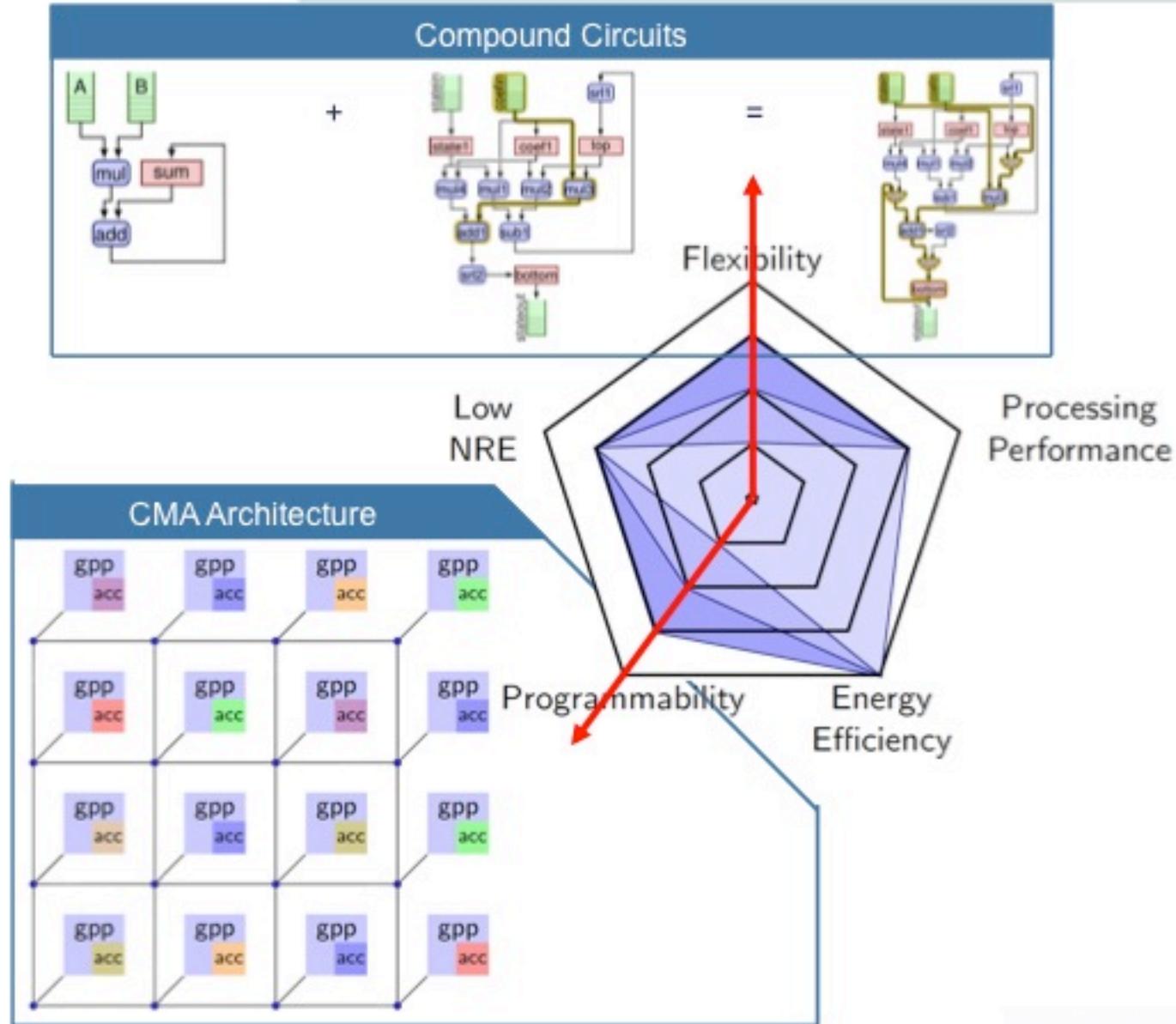
THALES



Processing Performance			
gpp	epp	gpp	gpp
gpp	epp	gpp	gpp
gpp	epp	gpp	gpp
gpp	epp	gpp	gpp



CMA : Chip Multi-accelerator



◆ Introduction & Concept of CMA

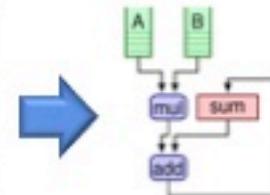
◆ Compound Circuits 

◆ SDR Case Study

◆ Conclusion

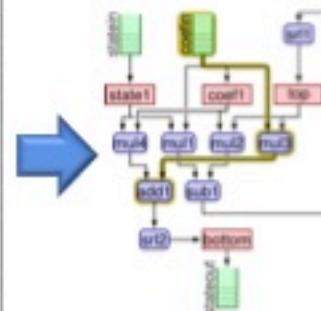
Compound Circuits: Overview

```
for (i = 0; i < AROW; i++)
  for (j = 0; j < BCOL; j++)
    sum = 0;
    for (k = 0; k < BROW; ++k)
      sum += A[i][k] * B[k][j];
    C[i][j] = sum;
```



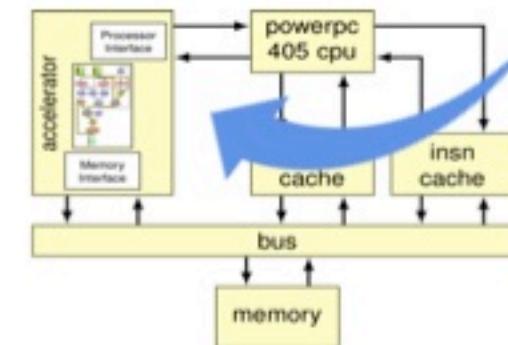
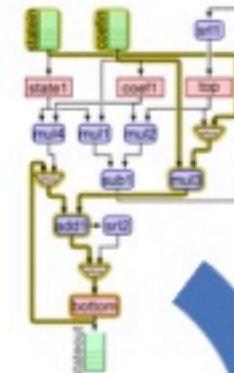
- ▶ Custom loop accelerators (circuits) for performance, power and area **efficiency**

```
for (i = 0; i < NPOINTS; i++)
  top = data[i];
  for (j = 1; j < ORDER; j++)
    left = top;
    right = state[j];
    state[j] = botm;
    top = (coef[j-1] * left)
      - (coef[j] * right)
      ) >> prec;
    botm = ( coef[j-1] * right)
      + (coef[j] * left)
      ) >> prec;
    state[ORDER] = botm;
    state[ORDER+1] = top;
  ...
```



Explore

- ▶ Exploratory combination of circuits for greater **flexibility**

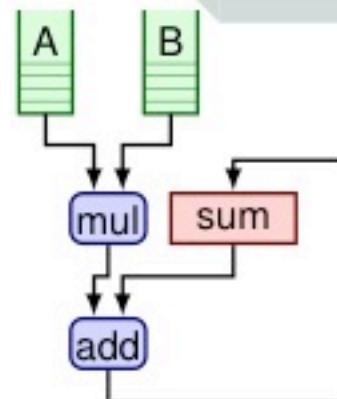


- ▶ Reuse of memory and processor interface for **lower NRE costs**.

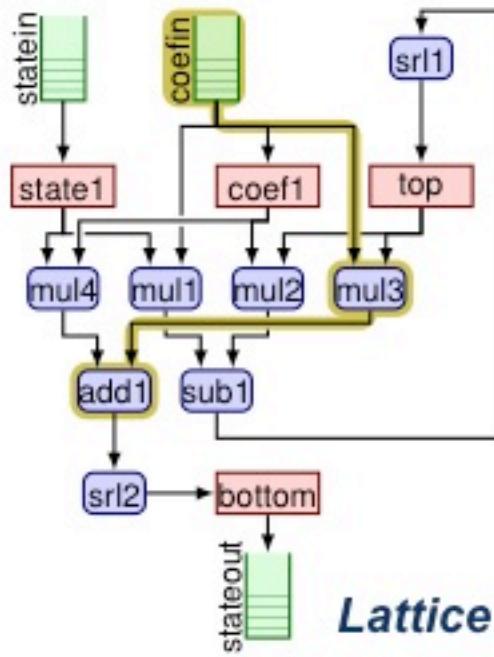
[Yehia et al. HPCA09]

THALES

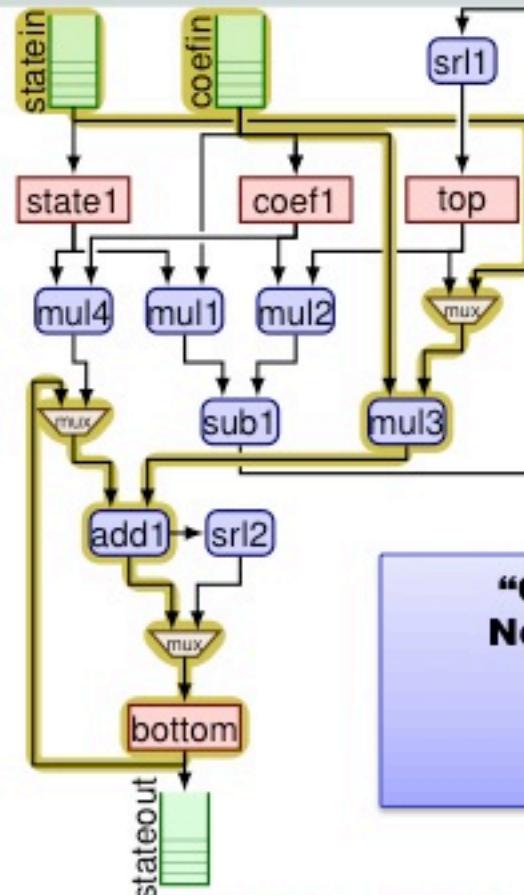
Merging the circuits: principles



Dot product



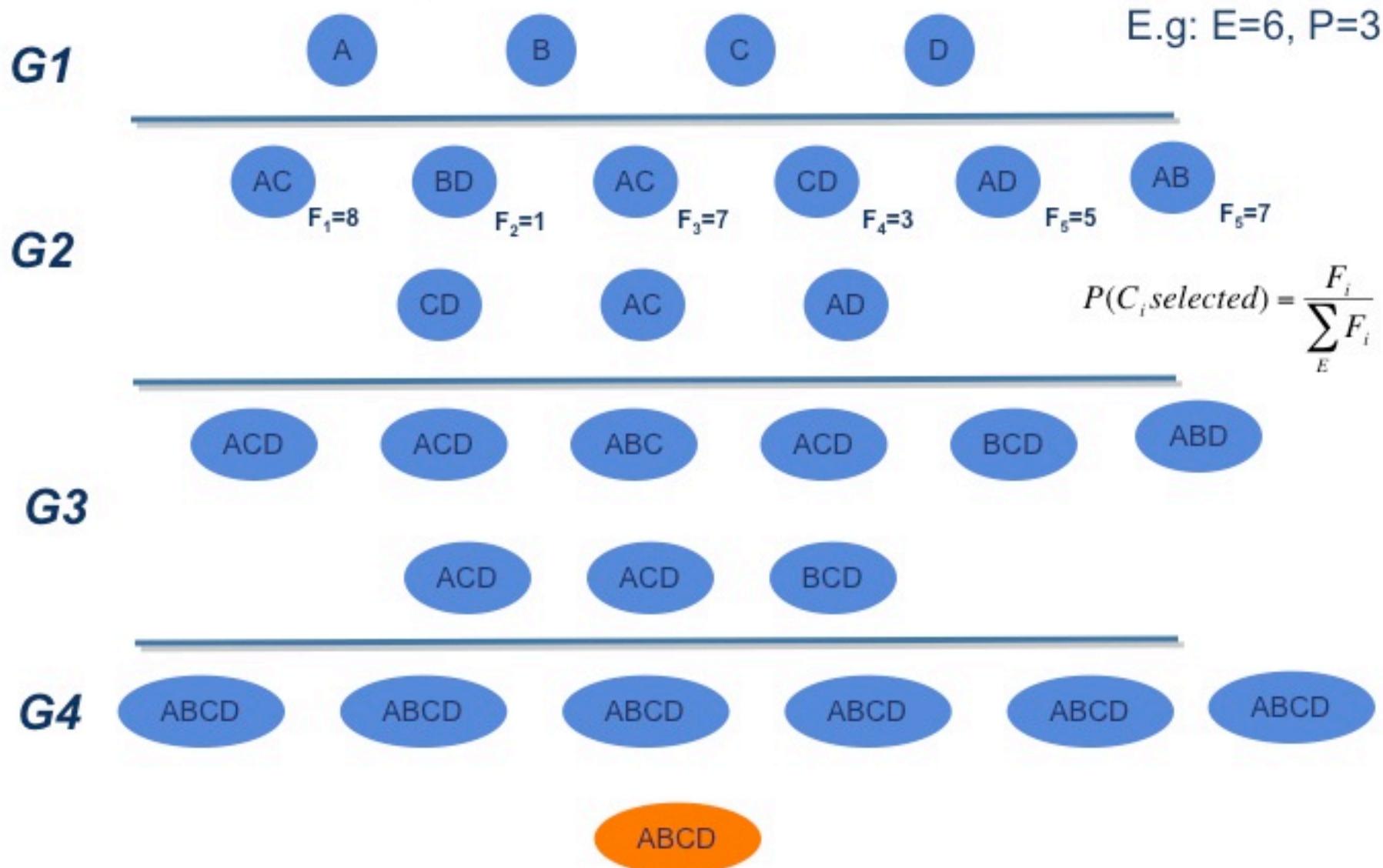
Lattice Filter



Cost function

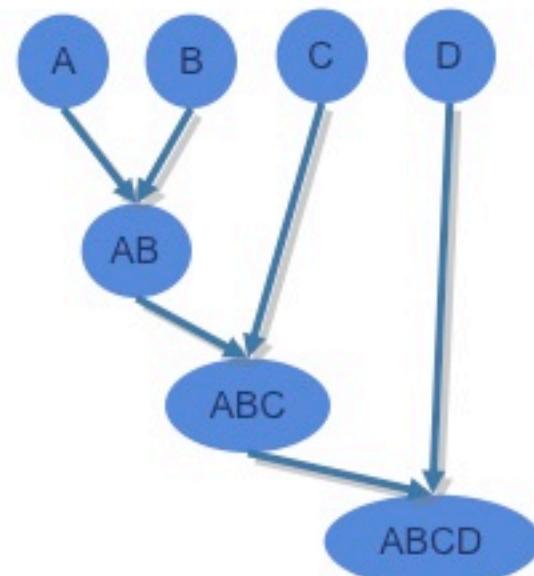
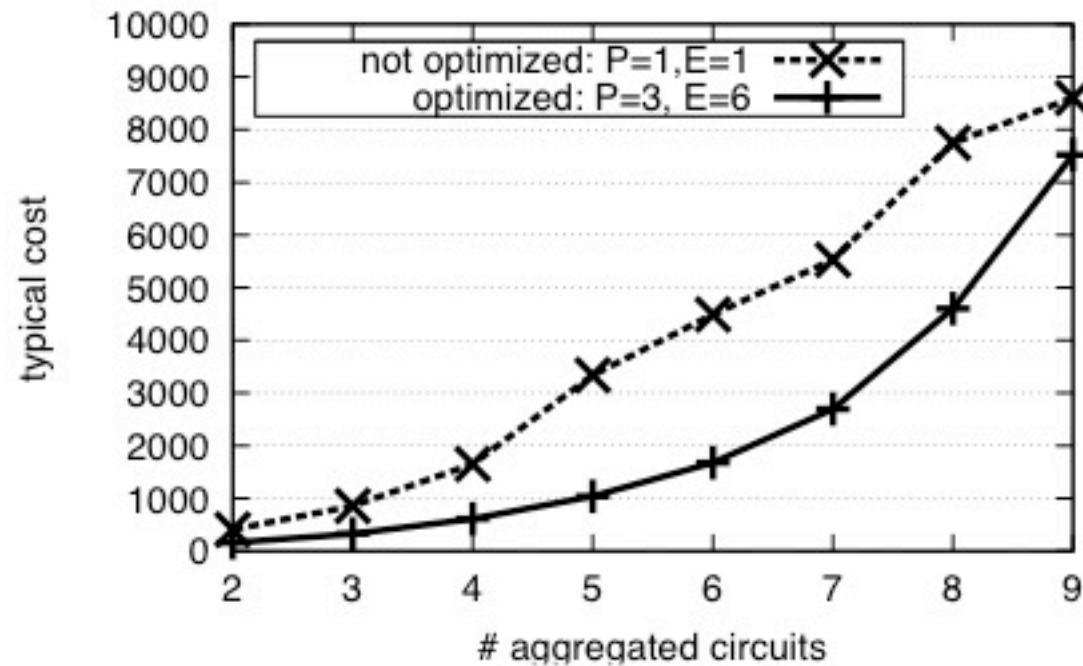
**“One solution”
Not necessarily
the best**

- Synthesis too costly for each explored compound
- Overhead cost = cost of muxes = \sum number of inputs * bit width

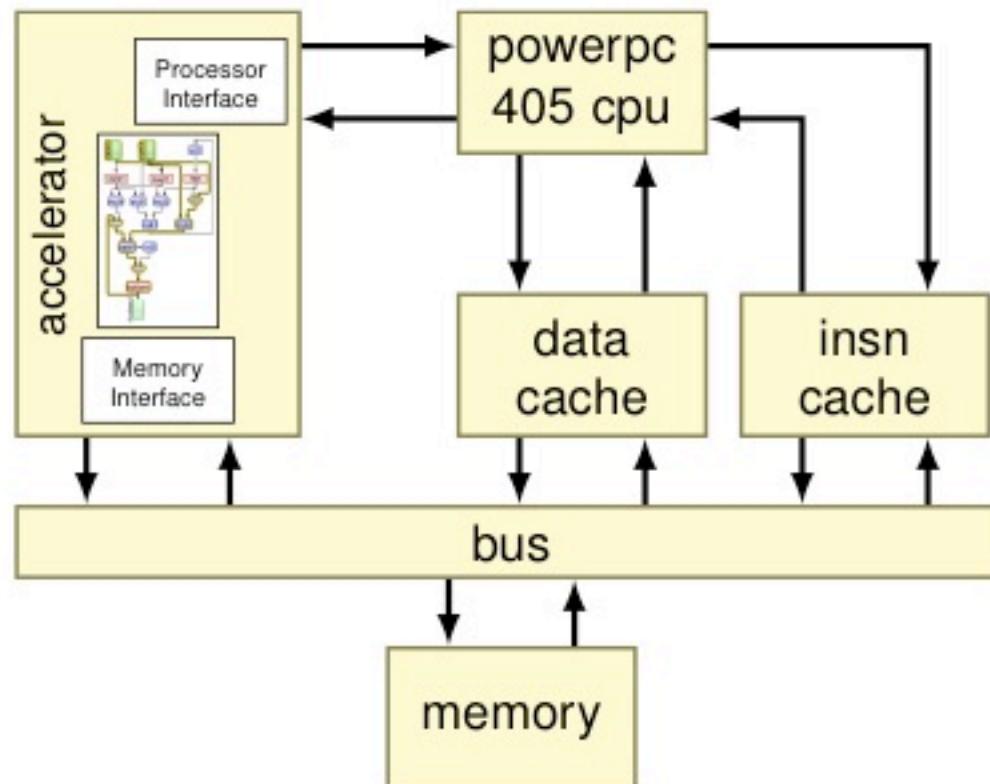


Multiple-circuit compound circuits

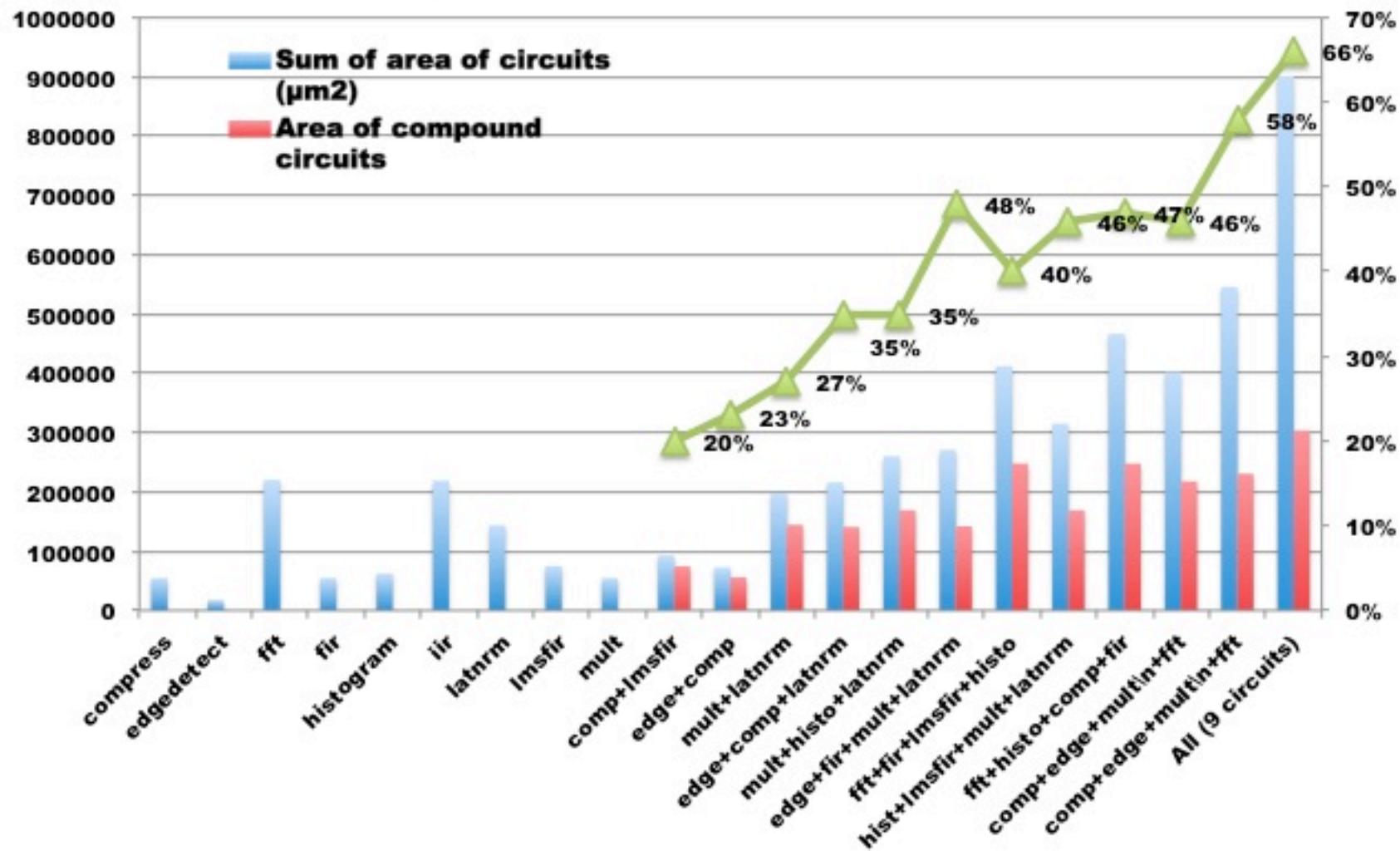
$E=1, P=1$

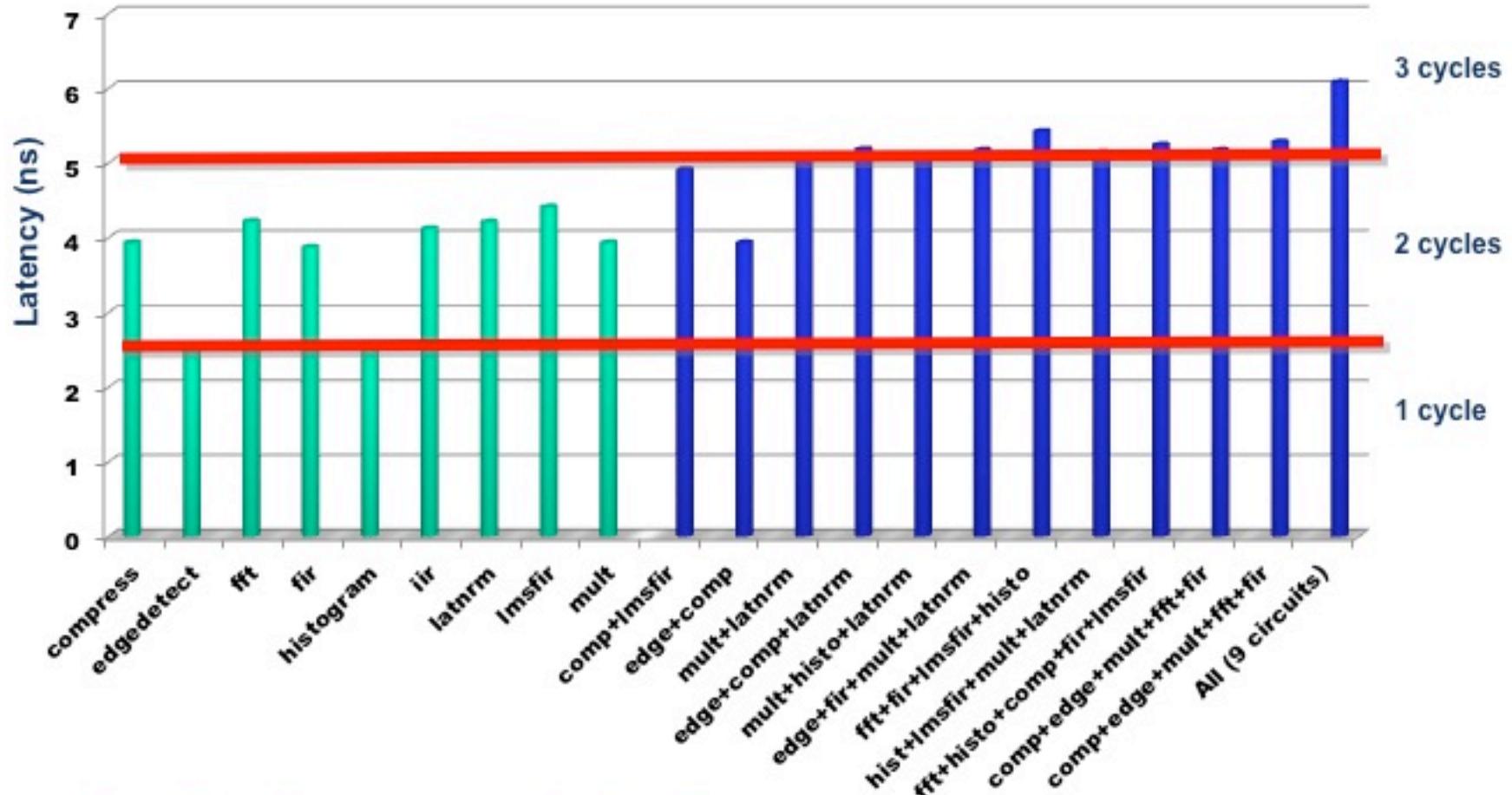


- ▶ 400-MHz (90nm) PPC 405 (5-stage pipeline, 32 registers)
 - ▶ Area evaluation using Synopsys DC, TSMC 90nm
 - ▶ 9 UTDSP benchmarks

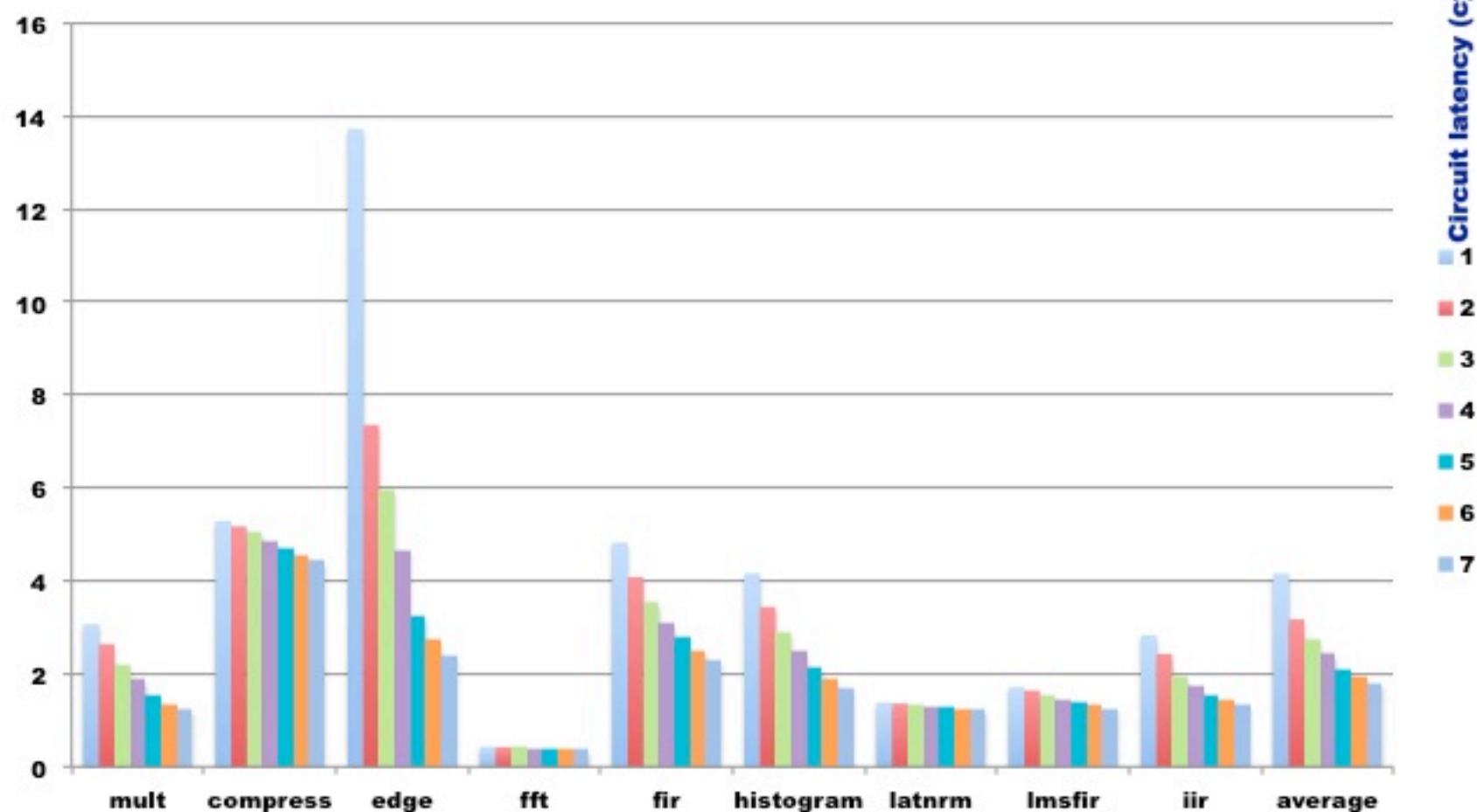


► 66% area saving





► ~2-cycle compound circuits



The information contained in this document and any attachments are the property of THALES. You are hereby notified that any reading, dissemination, distribution, copying or otherwise use of this document is strictly prohibited without the prior written agreement of THALES. ©THALES 2011. Teradata 3GP version 7.0.7

- ◆ Introduction & Concept of CMA

- ◆ Compound Circuits

- ◆ SDR : Case Study 

- ◆ Conclusion



Quite fast evolution of wireless standards



HSDPA
UMTS



Enabling technology:

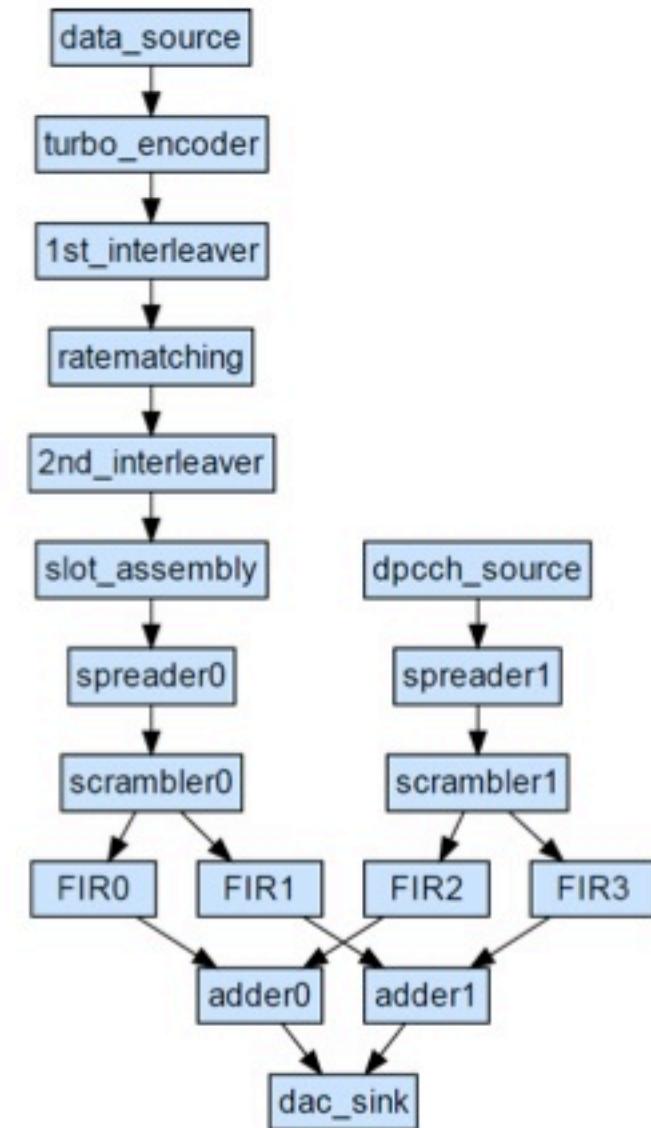
- ◆ Multi-Mode Radios
- ◆ Cognitive Radios

**SDR: keep up with the pace
Through flexibility**

THALES

Typical: waveform description

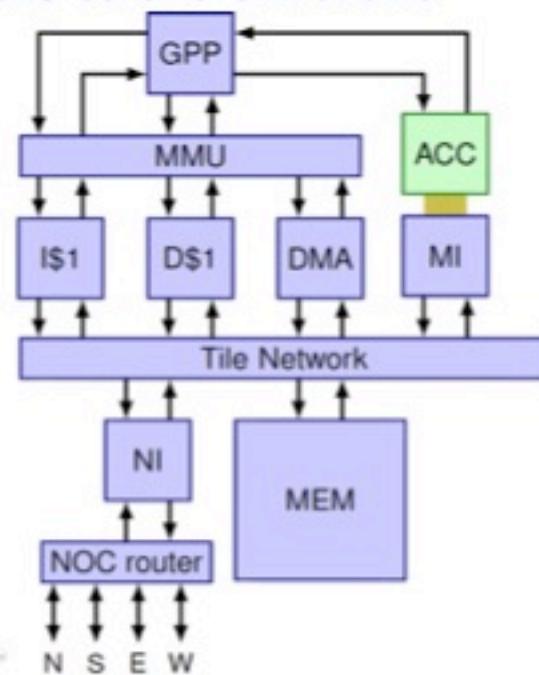
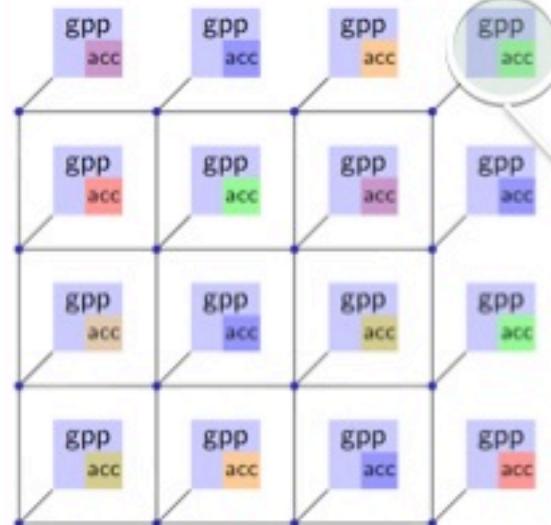
- ◆ Task flow graph
- ◆ Interconnected sequential tasks
- ◆ Low inter-task throughput
- ◆ High intra-task throughput
- ◆ Hard real-time requirement at Task level
- ◆ E.g WCDMA TX
- ◆ In this case study:
 - 802.11a and WCDMA on single CMA
 - 4 Waveforms : 802.11a TX, 802.11a RX, WCDMA TX, WCDMA RX

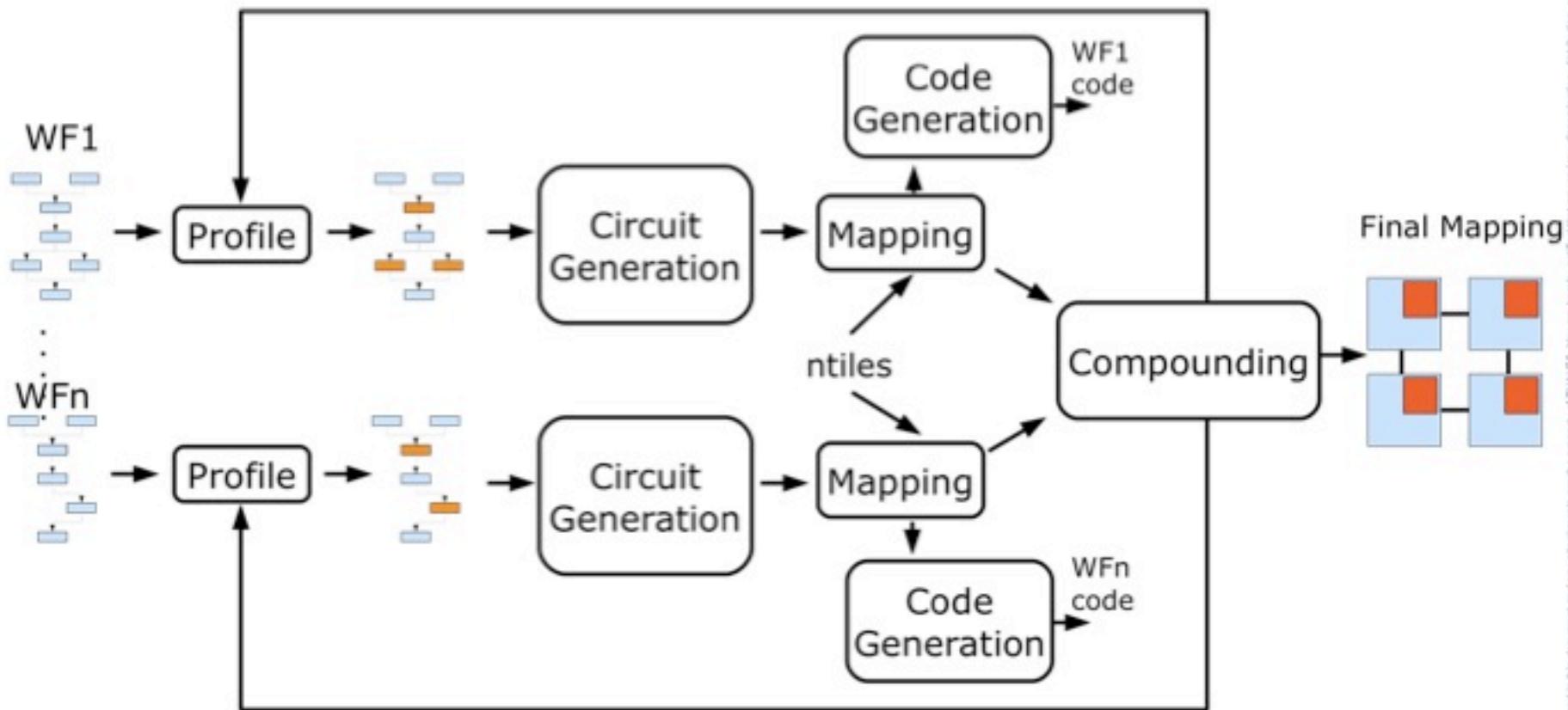


CMA Architecture

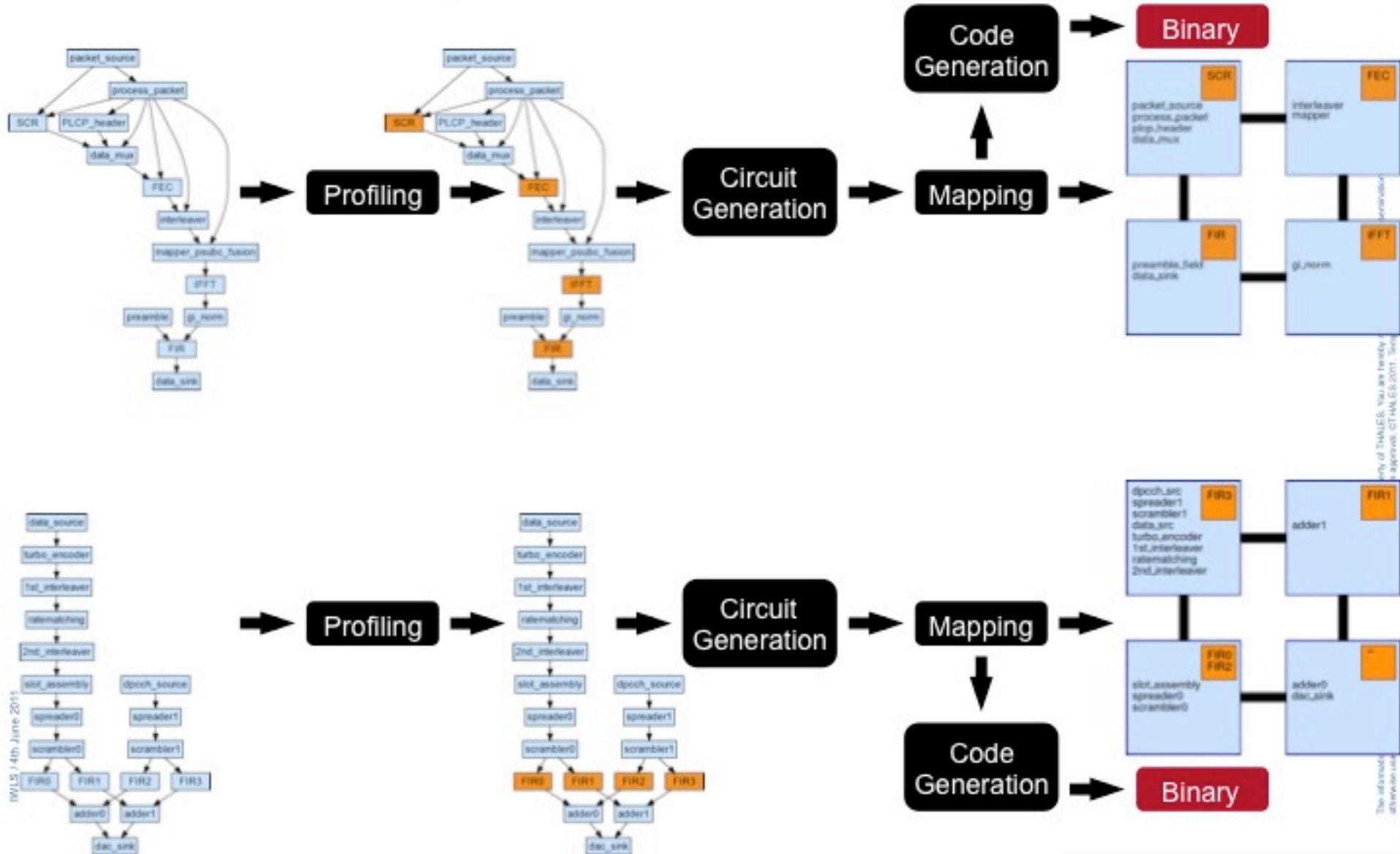
Heterogeneous but regular

- ◆ Regular infrastructure (GPP, NOC, ..) for programmability
- ◆ Heterogeneous processing thanks to the different accelerators.

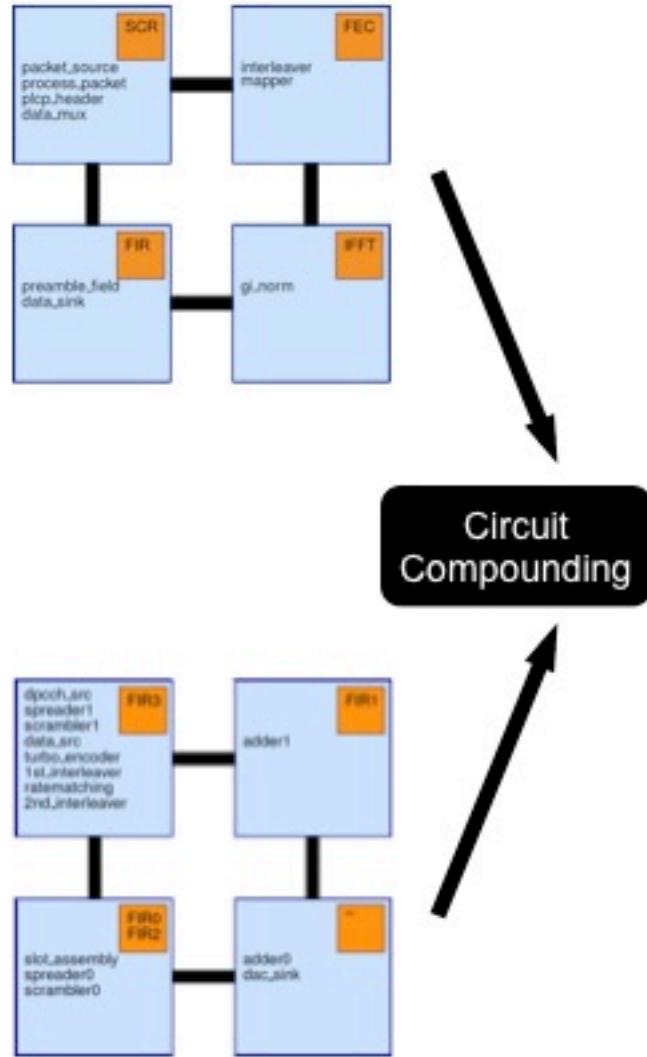
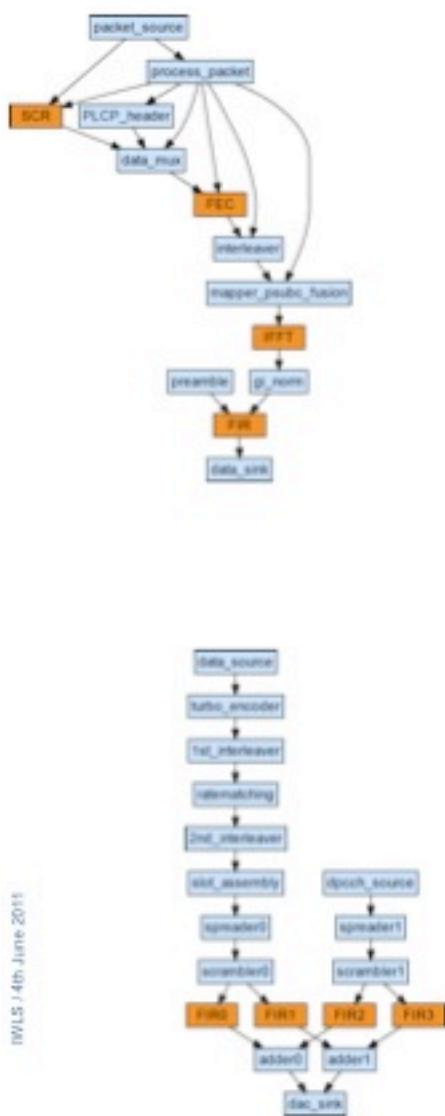




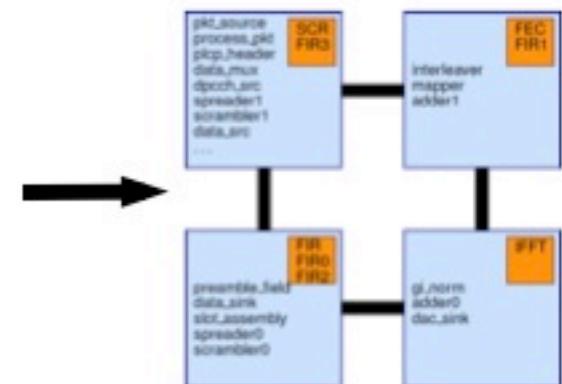
Mapping Individual Waveforms



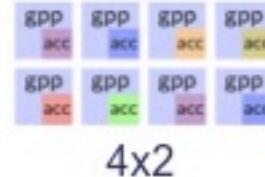
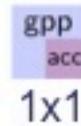
Mapping Individual Waveforms



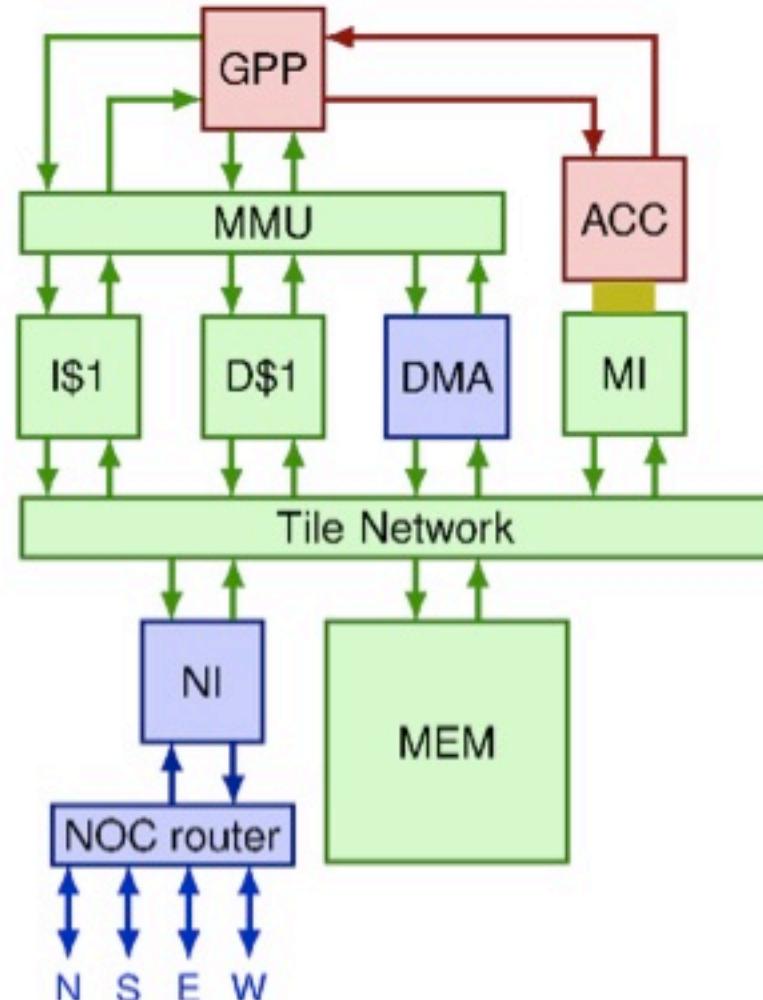
Circuit
Compounding



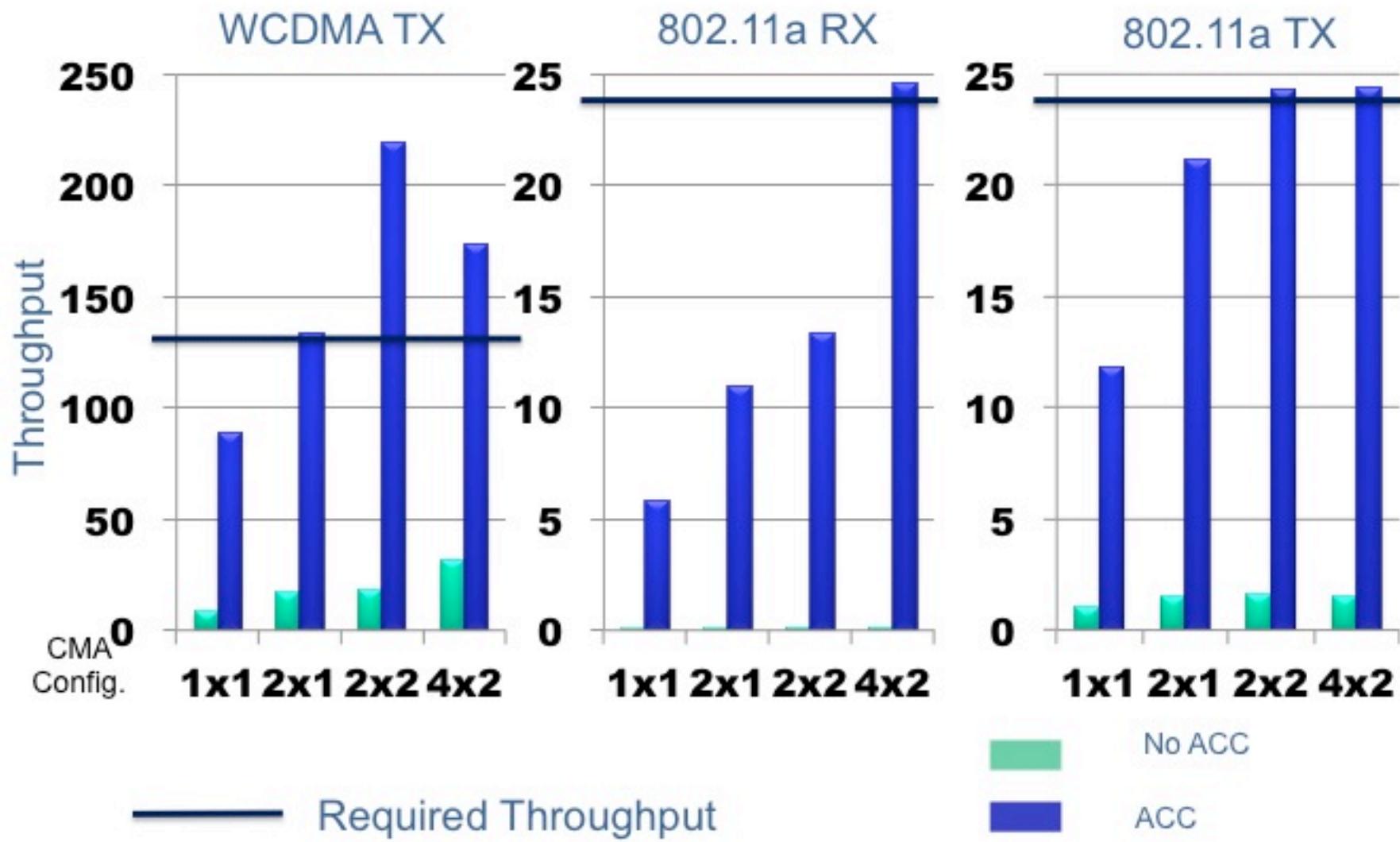
- ▶ PowerPC405 @650 MHz
- ▶ 3-cycle circuit latency
~4.6ns
- ▶ Network-on-Chip
- ▶ Augmented ISA for ACC
- ▶ UNISIM Infrastructure
- ▶ 4 Configurations



- ▶ Synopsys Design Compiler
- ▶ TSMC 90nm Standard Library



The information contained in this document and any attachments are the property of THALES. You are hereby notified that any reading, dissemination, distribution, copying or otherwise use of this document is strictly prohibited without the prior written agreement of THALES. Version 7.0 - 2011. Teradata - 950 version 7.0.7



- ◆ Introduction & Concept of CMA

- ◆ Compound Circuits

- ◆ SDR : Case Study

- ◆ Conclusion



CMA

=

Ease of programming (homogenous structure)

+

Efficiency of custom acceleration

- ◆ Heterogeneity made more accessible
- ◆ Use of available on chip transistors by using many accelerators
 - Potentially, activate needed accelerators on demand
- ◆ Flexibility through compounding
 - Many tasks in few tiles and area saving
 - Or widen the scope of application(s)

Main Contributors

- ◆ Olivier Temam – INRIA
- ◆ Sylvain Girbal – Thales Research and Technology
- ◆ Hugues Berry – INRIA
- ◆ Dominik Auras – RWTH

Funded by

- ◆ The research leading to these results has received funding from the ARTEMIS Joint Undertaking under grant agreement number 100230 and from the national programmes / funding authorities for the SMECY project (www.smecy.eu).
- ◆ The HiPEAC Network of Excellence