Reconfigurable Computing
可重构计算

Introduction

The entire system operates in a configuration described as the "Fixed-Plus-Variable" Structure Computer such that the same elements used for the special computer may be reorganized for other problem applications.

– Gerald Estrin (UCLA) 1962

Philip Leong 梁恆惠 (philip.leong@sydney.edu.au)
School of Electrical and Information Engineering

http://phwl.org/talks

THE UNIVERSITY OF SYDNEY

Course Details

Course Objectives

- Prerequisites
  - Computer programming in C
  - Basic digital systems (combinatorial circuits, sequential circuits, finite state machines, data paths)
  - Experience using a hardware description language (Verilog or VHDL)
- Objectives
  - An introduction to the field of reconfigurable computing
  - Advance digital design skills by developing a reconfigurable computing application
  - An introduction to research methodology

Topics

Lecture Schedule

1. Introduction （简介）
2. FPGA architecture（FPGA结构）
3. Trends and Exploration（趋势与探索）
4. Parallelism（并行性）
5. Precision（精度）
6. Interface（接口）
7. Customisation（定制）
8. Case studies（案例）

› Reconfigurable Computing
  - EPIC approach（EPIC方法）
  - Computer architecture（计算机体系结构）
  - Computer arithmetic（计算机算术）
  - VLSI design（大规模集成电路设计）
  - Trends in semiconductor technology （半导体技术的趋势）

› Case studies
  - Examples from research

Labs

- A major part of this course are the labs concerning FPGA implementation of machine learning (long short-term memory neural network，长短时记忆神经网络)
  - Lab1 – Familiarisation & Testbench
  - Lab2 – Parallelism
  - Lab3 - Precision
  - Lab4 – Exploration
  - Lab5 – Interface I
  - Lab6 – Interface II
- Report
  - Write a 4 page paper describing your design

Introduction to Reconfigurable Computing

- FPGAs
- Reconfigurable computing
- Applications

FPGAs

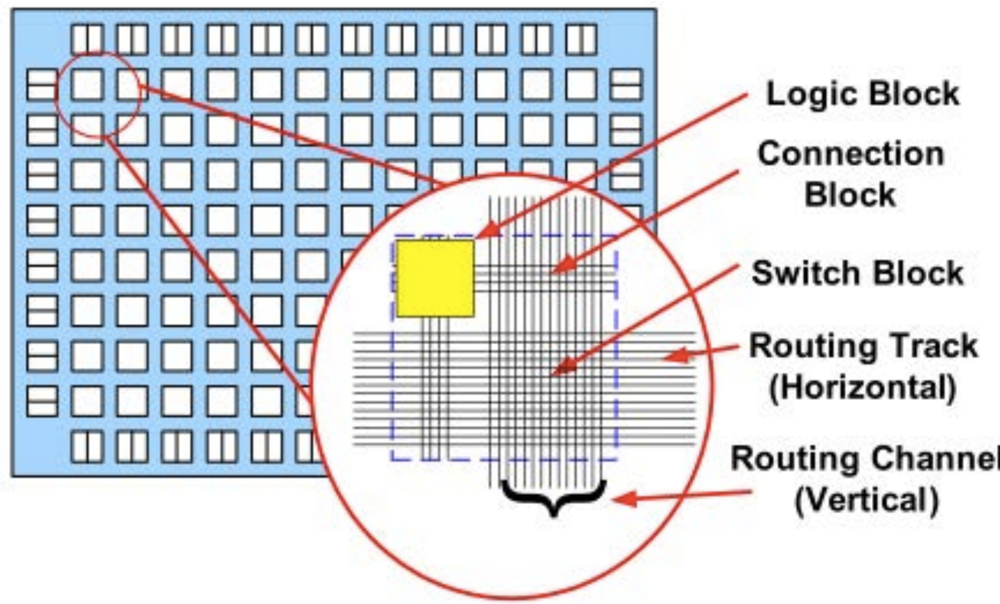
Integrated Circuits

Most electronics rely on application-specific ICs (ASICs) for perf, cost and P

Source: Arvind MIT

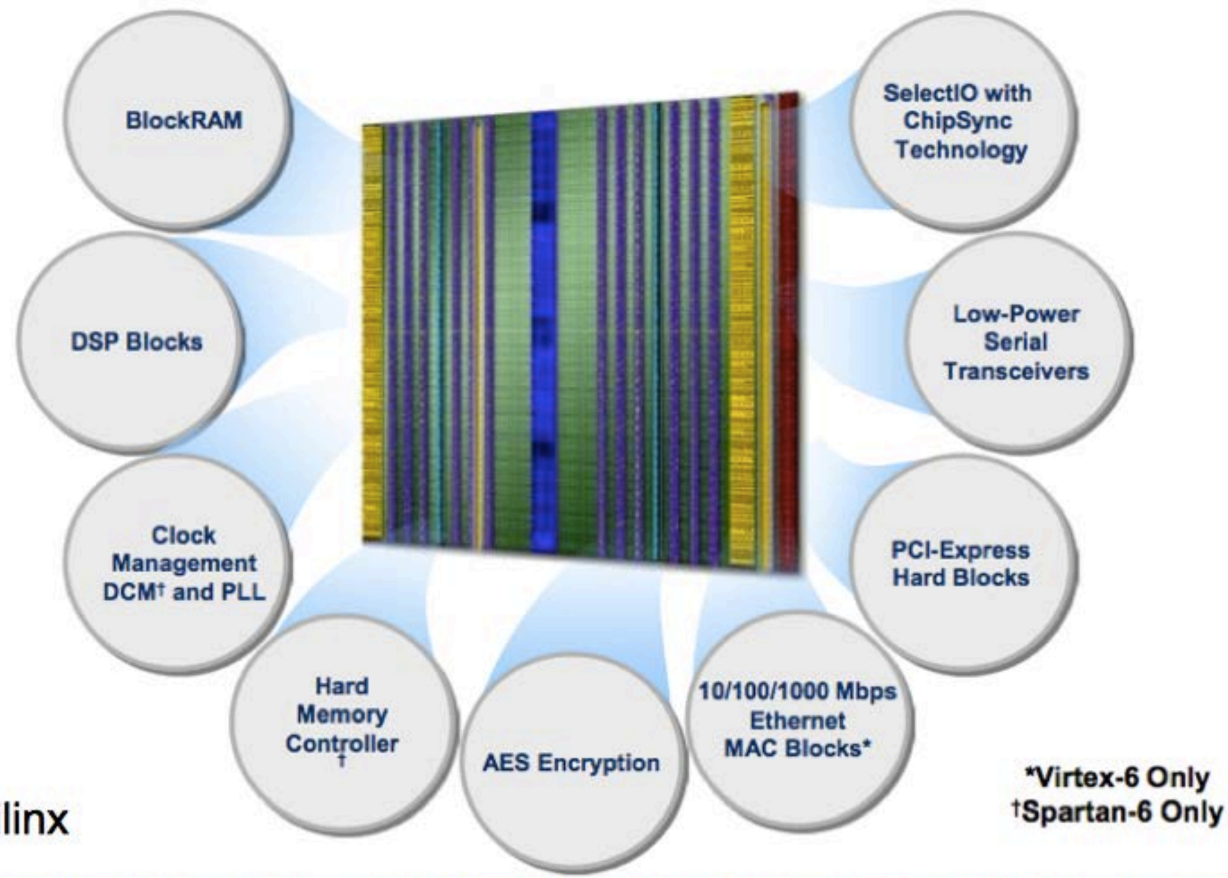
FPGA

- A generalised integrated circuit
  - Logic blocks for digital operations
  - Programmable interconnect for routing
- Arbitrary digital circuits can be implemented
- Functionality downloaded to FPGA memory (in seconds)



Source: Steve Wilton (UBC)

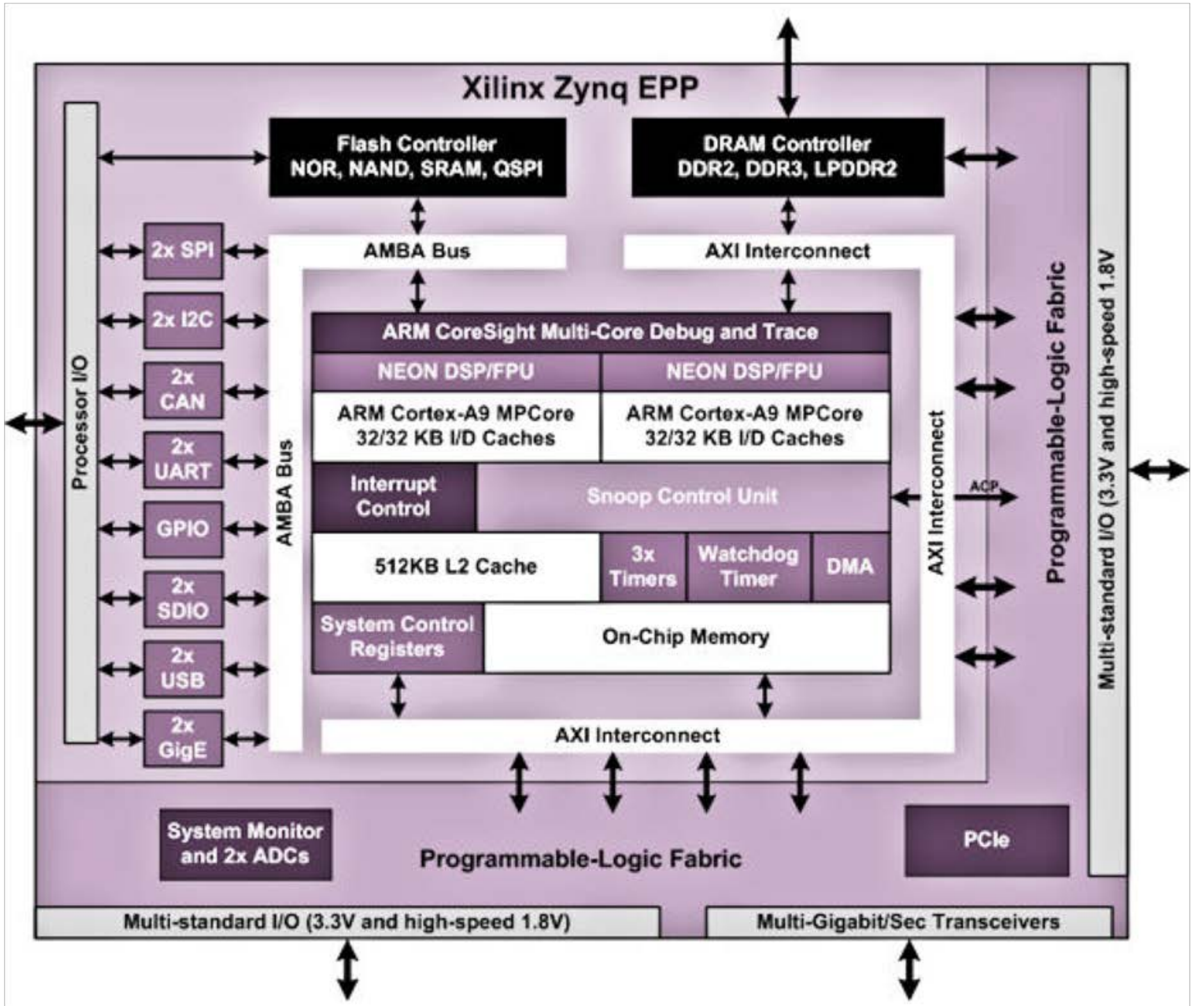
FPGA Embedded Blocks



Source: Xilinx

Hard IP blocks for widely-used functions: faster, more efficient, lower power
Careful choice: every user must pay for these functions, whether used or not

Zynq (ARM+ Reconfigurable Fabric)



Source: Xilinx

THE UNIVERSITY OF SYDNEY

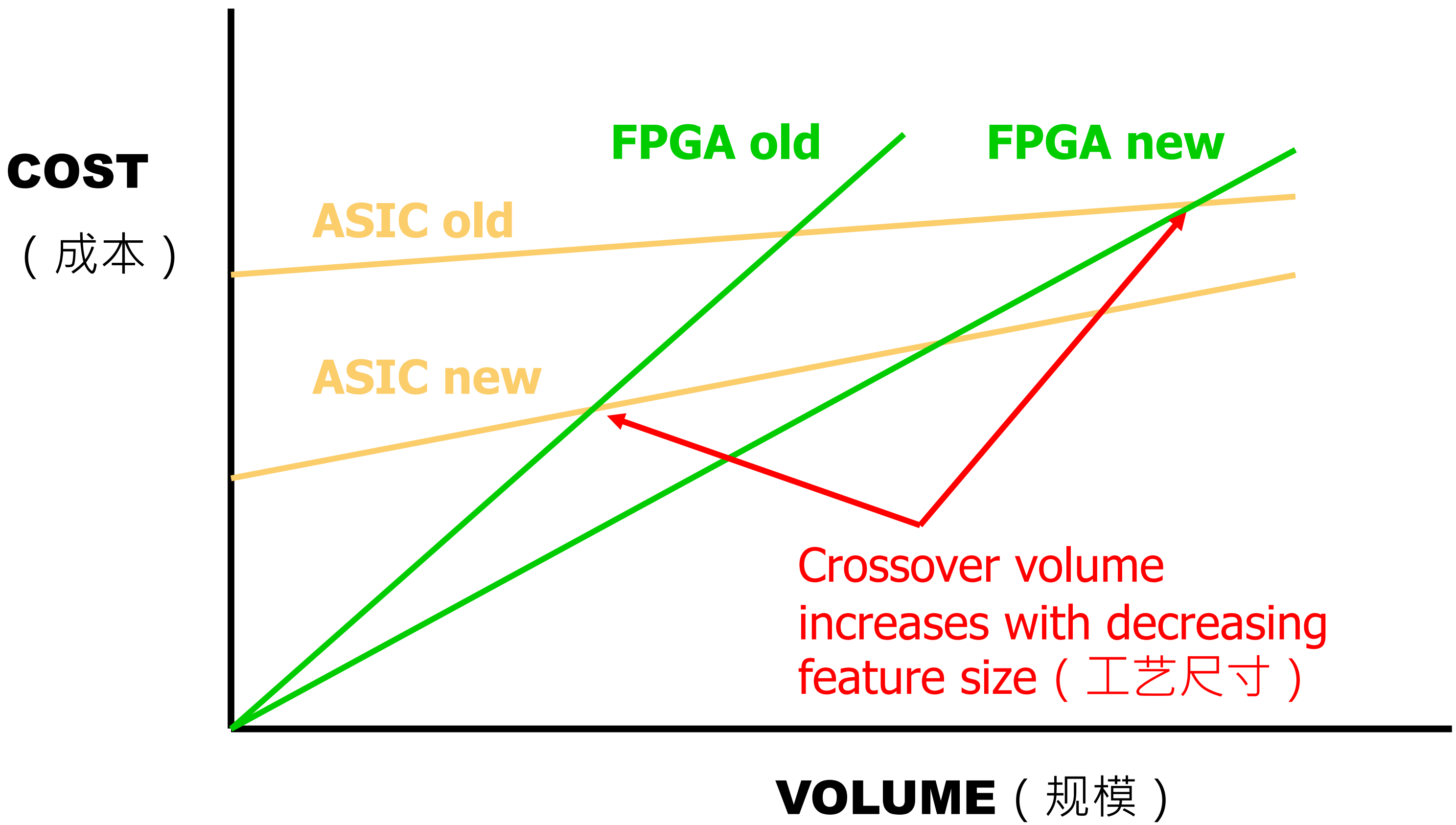
FPGA Families

Xilinx 7-series FPGAs, 28nm

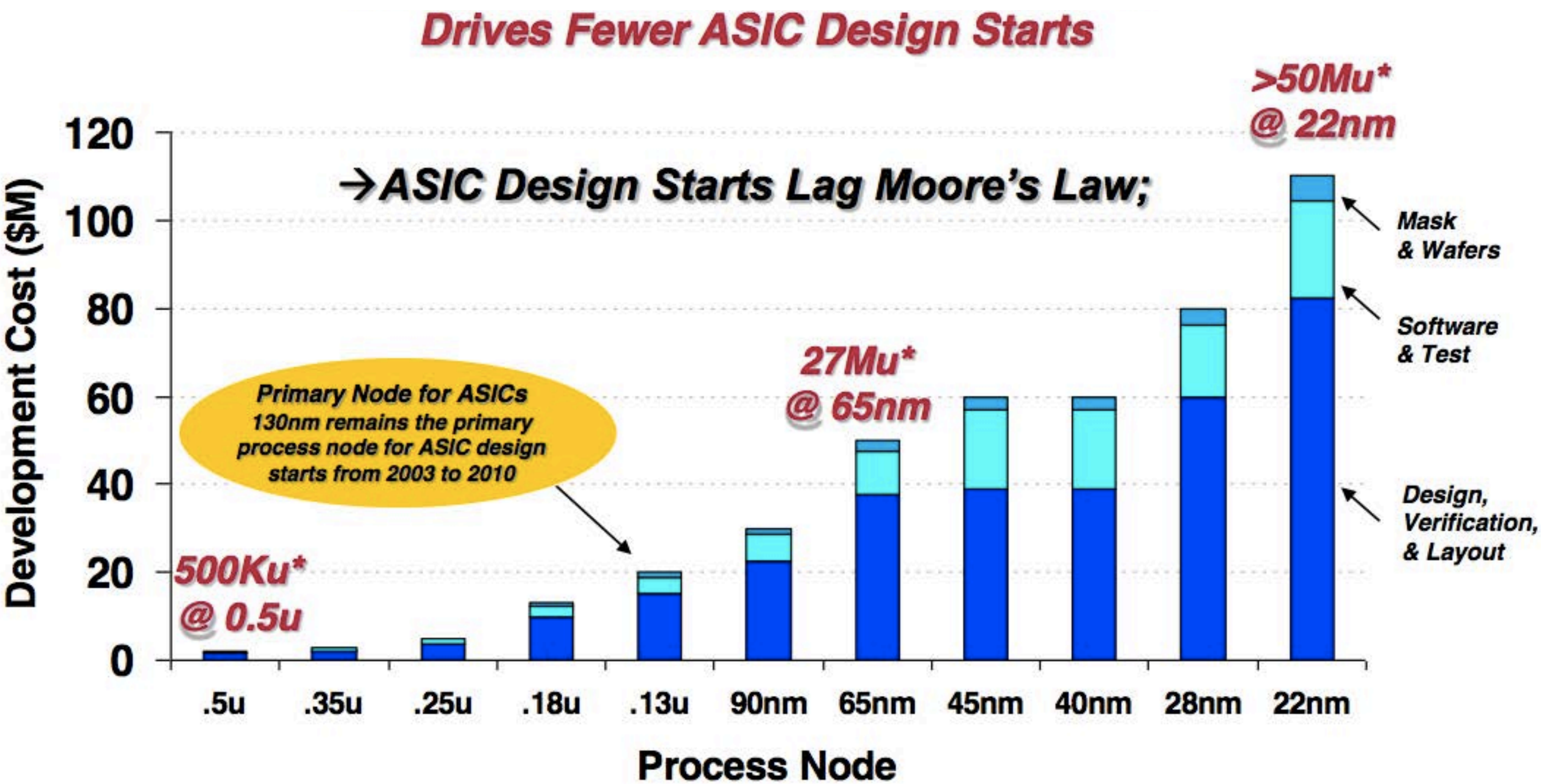
DSP Slices	740	1,920	3,600
DSP Performance (symmetric FIR)	930GMACs	2,845GMACs	5,335GMACs
Transceiver Count	16	32	96
Transceiver Speed	6.6Gb/s	12.5Gb/s	28.05Gb/s
Total Transceiver Bandwidth (full duplex)	211Gb/s	800Gb/s	2,784Gb/s
Memory Interface (DDR3)	1,066Mb/s	1,866Mb/s	1,866Mb/s
PCI Express® Interface	x4 Gen2	x8 Gen2	x8 Gen3
Analog Mixed Signal (AMS)/XADC	Yes	Yes	Yes
Configuration AES	Yes	Yes	Yes
I/O Pins	500	500	1,200

Source: Xilinx

ASIC vs FPGA Cost



ASIC Development Costs



Source: Altera

Return on Investment Analysis

ASIC / ASSP (65nm) Development Cost — **$55M**

20% of Revenue on R&D

Revenue Target — **$275M**

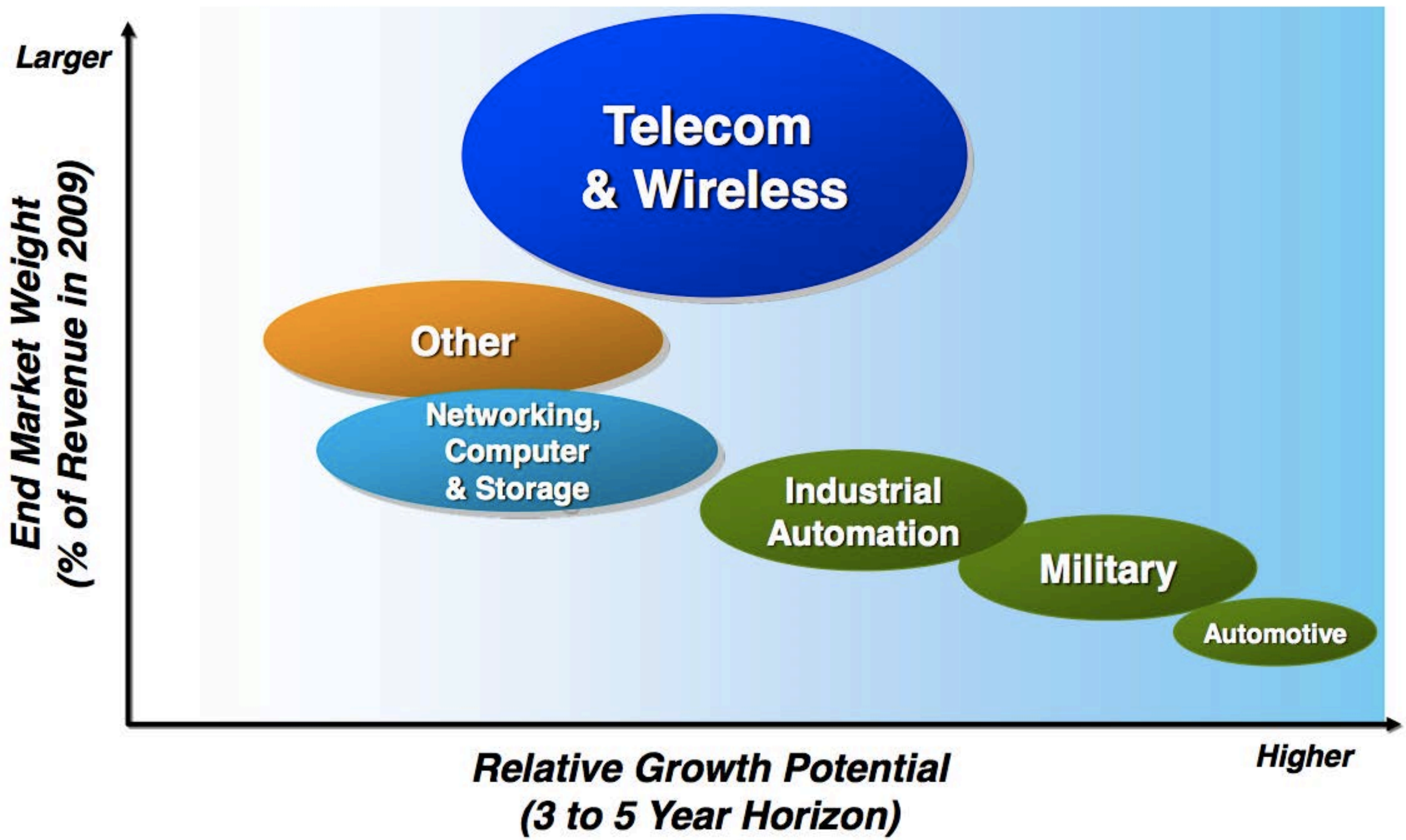
$10 to $50 Device ASP

Unit Volume Req'd — ***5Mu to 27Mu***

Very Few High Volume Applications Justify ASIC / ASSP Development

Source: Altera

Application Domains



Source: Altera

Typical High Performance Commercial Applications

Application

Optical Transport
OTU Transponder

40GbE/100GbE Switch

Radar

Requirements

- >350 MHz performance
- 28 Gbps transceivers
- 10GBASE-KR backplane support
- High-performance on-chip memory
- High-performance and flexible memory controller
- Hard system-level IP for bandwidth
- High precision DSP

Solution

ALTERA Stratix V

Process: 28HP

- >350 MHz performance
- Lowest power in its class
- Up to 1.1M LEs on a monolithic die

Transceiver: 14.1 Gbps/28 Gbps

Product Architecture:

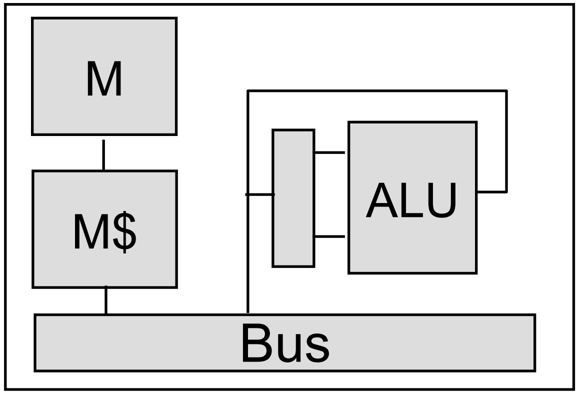
- Soft memory controller supports 800MHz DDR3 DIMM
- 2,560 M20K memory blocks
- 54x54 variable precision DSP

System IP:

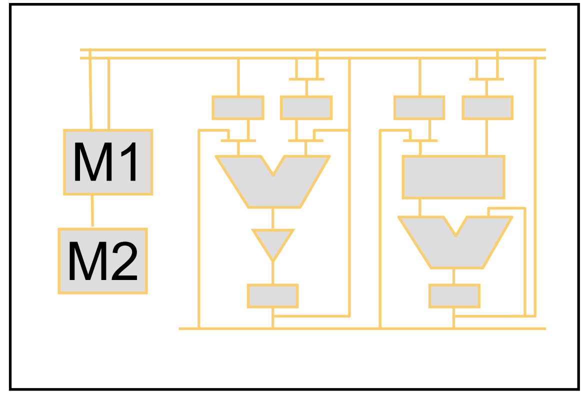
- PCIe Gen3 x8, 40 GbE/100 GbE, Interlaken

Source: Altera

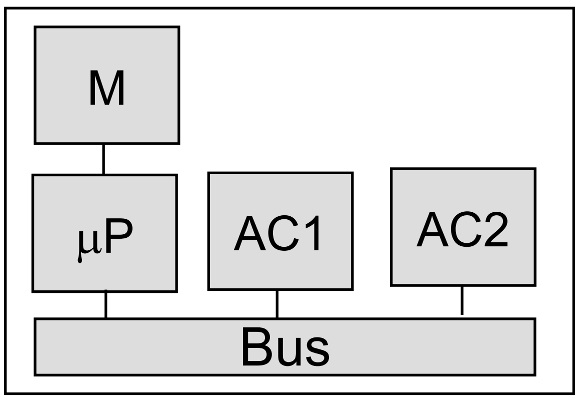
Architectural Choices



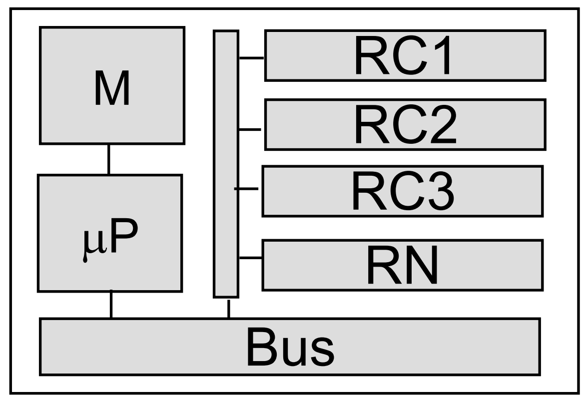
General-purpose processor



Application-specific processor



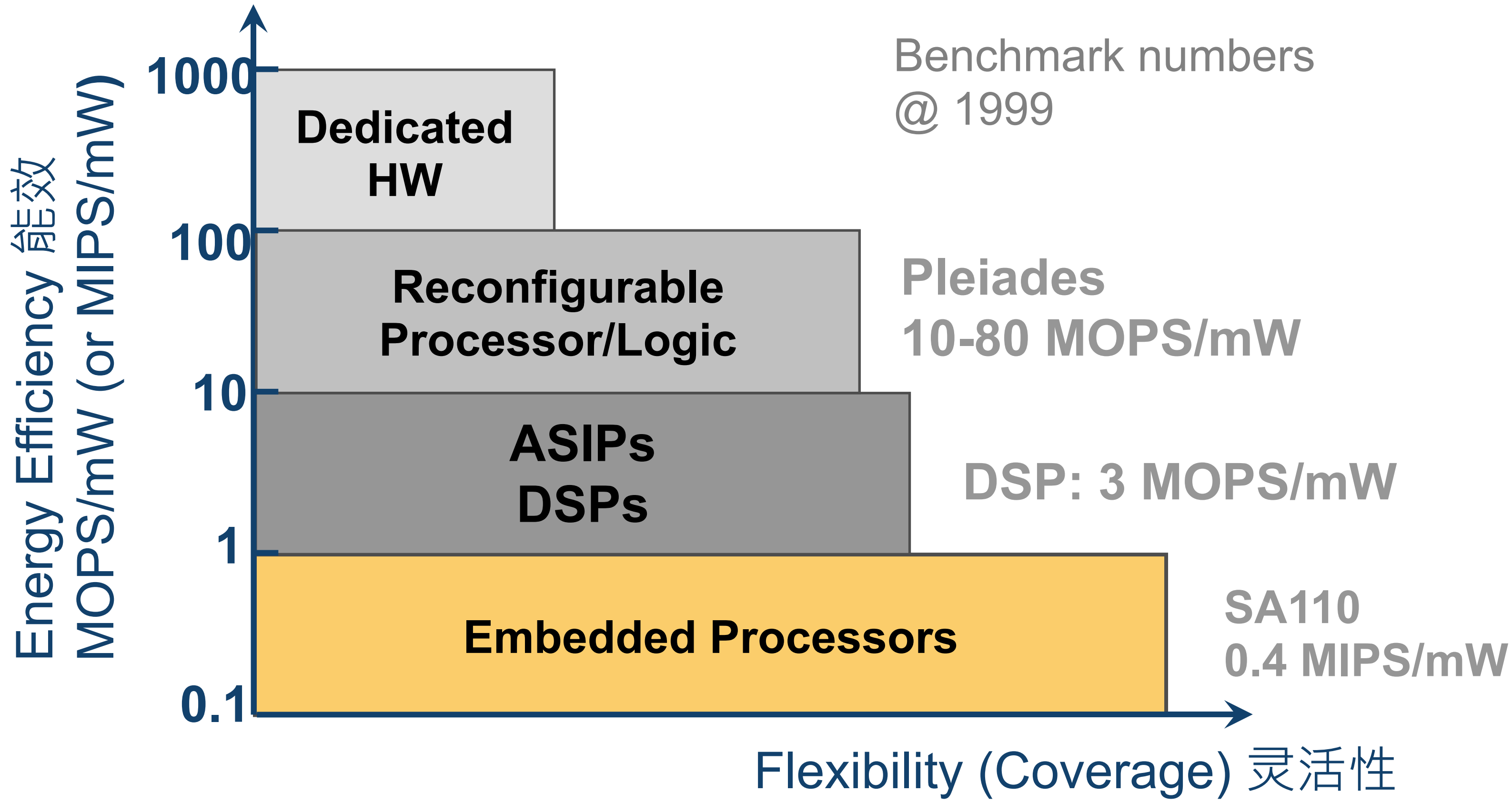
Dedicated accelarators



Reconfigurable processor

Source: Rabaey UCB

Flexibility vs Energy

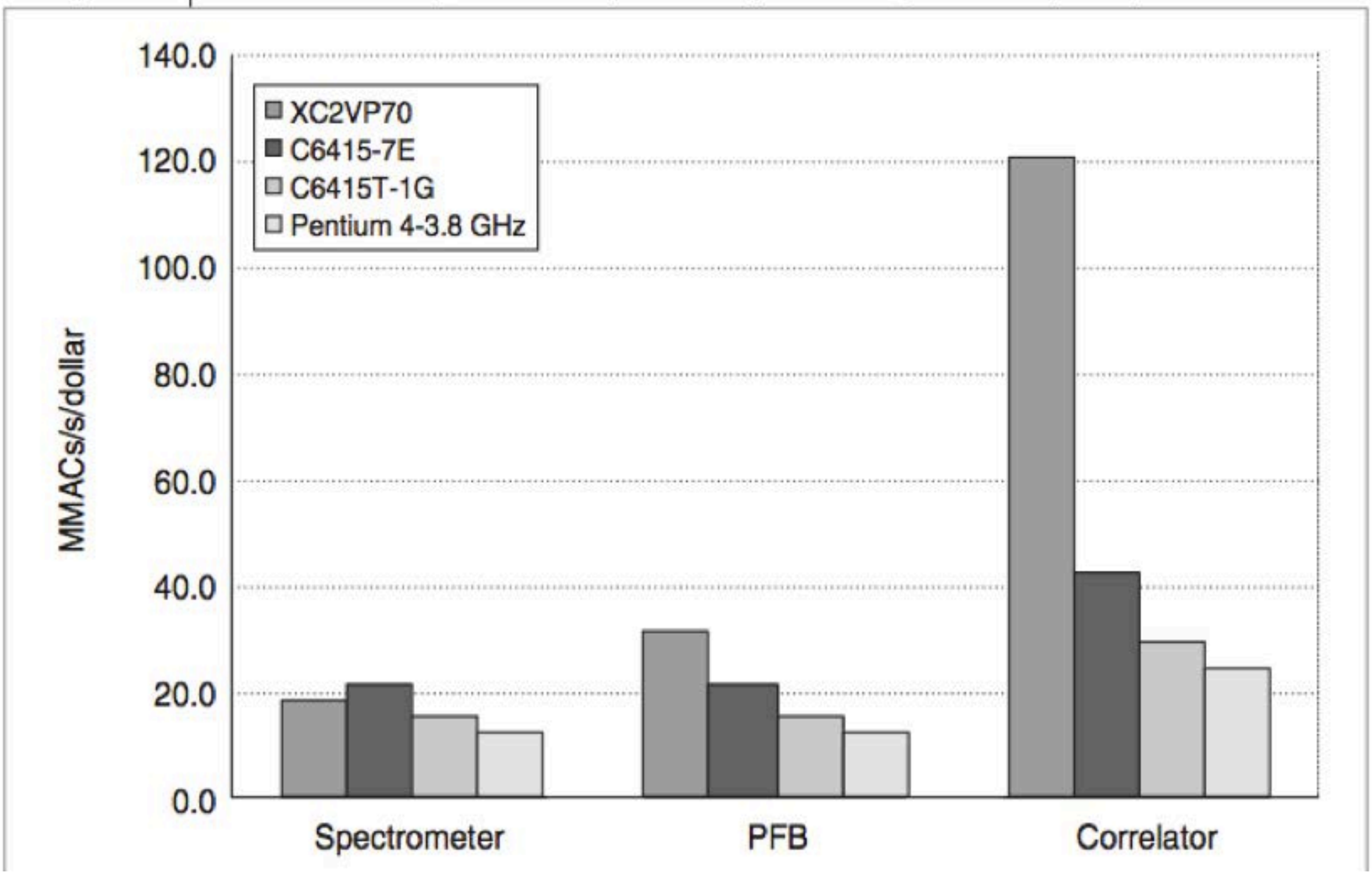


Approximately three orders of magnitude in inefficiency from general-purpose to dedicated!

Source: Rabaey UCB

FPGA vs DSP and CPU Cost Comparison

Berkeley BEE2 cost comparison (FPGA, DSP1, DSP2, uP)



THE UNIVERSITY OF SYDNEY

Tools

- Traditionally designed using ASIC development tools
  - VHDL/Verilog very low level
  - Chisel is a recent tool which is higher level
- Recent advances
  - Vivado HLS
  - OpenCL
- Extensive module generators and libraries e.g. filters, fft, floating-point, maths coprocessors, soft processors, network controllers, memory controllers, I/O controllers …
- Still an active research topic

	Hand-coded VHDL	Vivado HLS C
Design Time (weeks)	12	1
Latency (ms)	37	21
Memory (RAMB18E1)	134 (16%)	10 (1%)
Memory (RAMB36E1)	273 (65%)	138 (33%)
Registers	29686 (9%)	14263 (4%)
LUTs	28152 (18%)	24257 (16%)

Resource utilization example: hand coded versus Vivado HLS.

Comparison of FPGAs with uP and ASIC

- Compared with uP and DSP
  - higher speed, lower power, smaller variance in execution time
  - Longer development times, higher cost per unit
- Compared with ASICs
  - Lower initial cost
- Rides Moore's Law, development costs amortised（分摊） over users
  - Faster time to market, lower risk
  - Can be customised to problem in ways not possible with ASICs

Reconfigurable Computing

THE UNIVERSITY OF SYDNEY

Reconfigurable Computing

- Application of FPGA devices to computing problems

FPGA Design

- Good reconfigurable designs are EPIC
  - Exploration
  - Parallelism
  - Precision
  - Integration
  - Customisation

Exploration

- All equally bad
  - Making a brilliant design of the wrong algorithm
  - Making a poor design choice the right algorithm
  - Making a really fast core without adequate interface
- Need to make sure we understand
  - Algorithm
  - Previous work, “An afternoon in the library can save a year in the lab.”
  - System-level issues

Parallelism

- Do what would take many cycles on uP in fewer cycles (instruction level parallelism)
- Do many independent tasks/threads/processes in parallel (multiprocessor)
- Tradeoff latency with throughput by doing things in stages (pipelining，流水线)

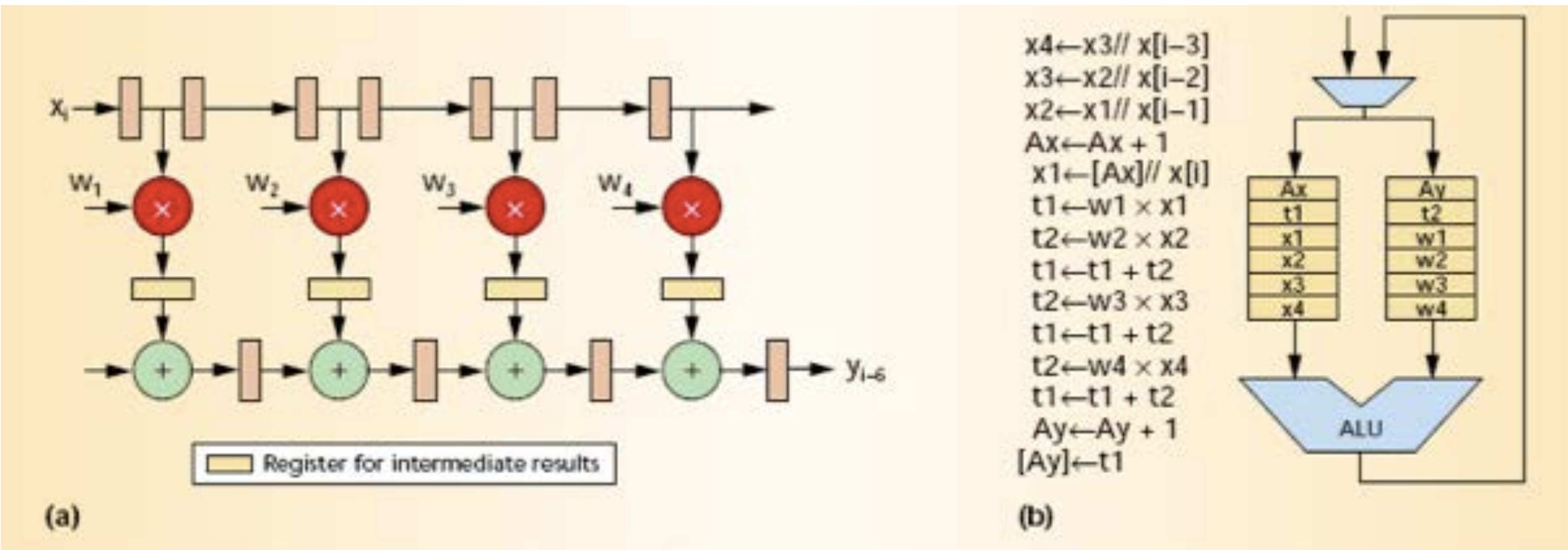
http://fernandoexperiences.blogspot.com.au/

Parallelism (Reality)

Unfortunately this is the reality (but FPGAs allow better control of this)

Parallelism Example

- Microprocessor: data passed sequentially to computing unit
- FPGA & ASIC: spatial composition of parallel computing units (multiple muls, pipelining)
- E.g. 4-tap FIR filter, FPGA 1 output per cycle, uP takes multiple cycles
- Lower power and higher speed



Source: DeHon "The Density Advantage of Configurable Computing"

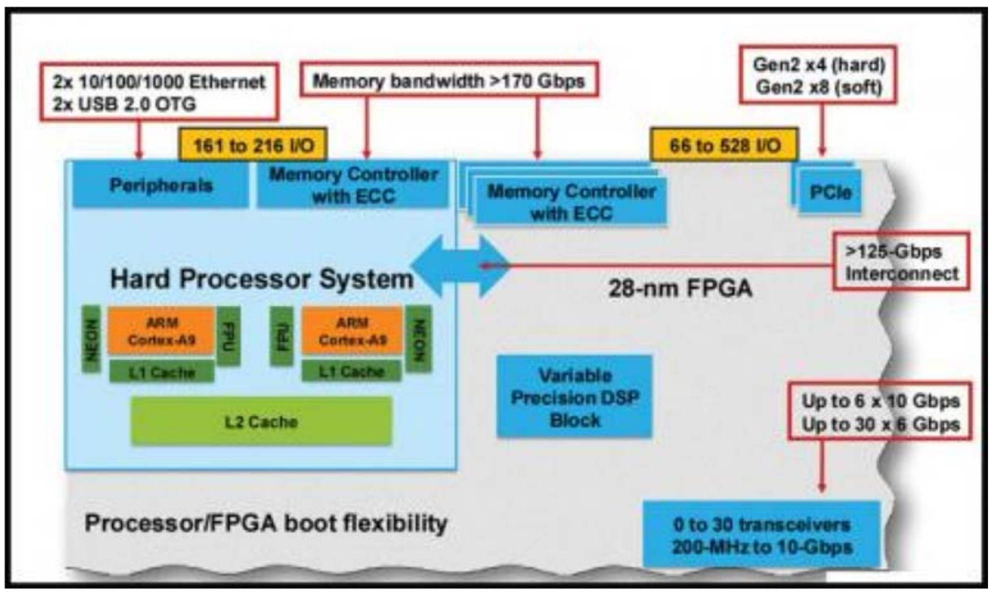
THE UNIVERSITY OF SYDNEY

Parallelism (Precision)

- Microprocessors are really good at single and double precision
- Often overkill for many applications
- Can use reduced precision which is efficient on FPGAs
- Can used mixed precision where high accuracy is achieved using mainly low-precision operations
- There are also different algorithms which have different cost, performance and accuracy tradeoffs e.g. CORDIC vs polynomial approximation for sin()
- Reduced precision -> reduced area so more spatial parallelism can be realised

Integration

› Networking, chip IO and computation on same device
- Reduction of buffering can help latency
- Single chip operation massive interconnect within chip exploited
- Multiple (small) memories within FPGA offer enormous memory bandwidth



Customisation

- More specific functions can be implemented more efficiently
- Too expensive to design ASIC to perform very specialised function
- FPGAs can be heavily customised due to their programmability i.e. only do one thing efficiently
  - Tradeoffs between speed and accuracy can be exploited, on uP, only get single or double; char, short or long
  - General operators can be replaced with specific ones
- E.g. Chip which only encrypts for a specific password

Applications

BMW Williams Formula 1 Team 2003

- Vehicle Control Module uses Virtex-II devices
  - gearbox, differential, traction control, launch control and telemetry
- High speed real-time control and DSP application

Source: BMW Williams

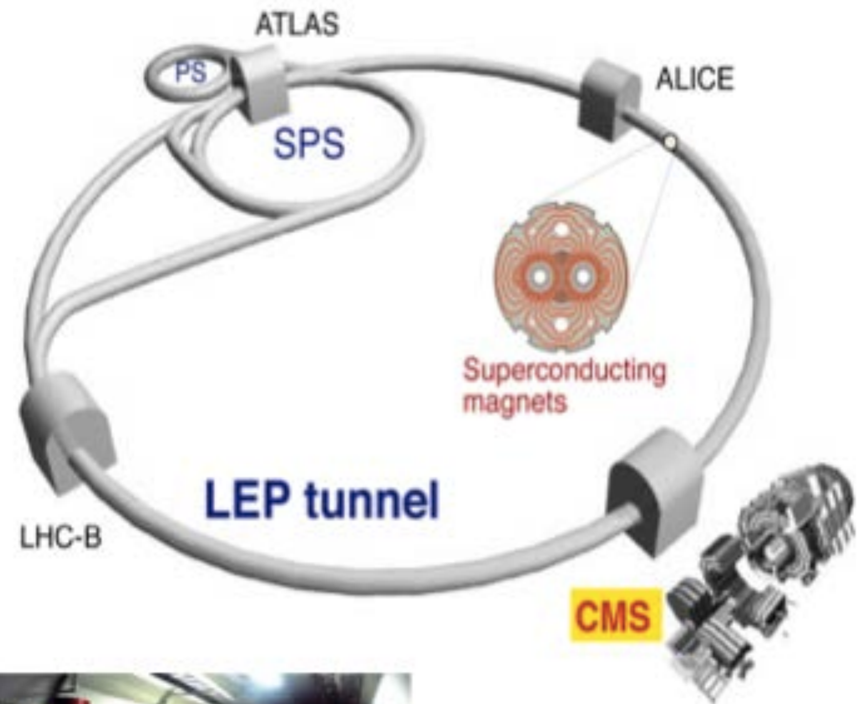
CERN Large Hadron Collider

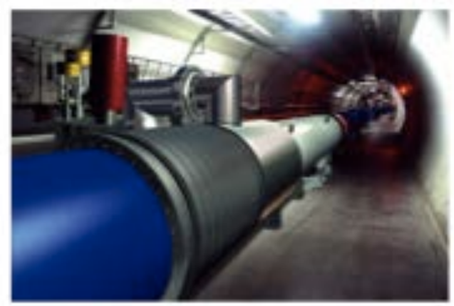
› Compact Muon Solenoid

（紧密介子绕线圈）

- 10^{15} collisions per second
- Few interesting events ~ 100 Higgs events per year
- 1.5Tb/s real-time DSP problem

- More than 500 Virtex and Spartan FPGAs used in real-time trigger





Source: Geoff Hall, Imperial College

Square Kilometer Array

- Square Kilometre Array (SKA) will be one of the largest and most ambitious international science projects ever devised (€1.5 billion).
- CSIRO Developing Australian SKA Pathfinder (ASKAP), a $150M next- generation radio telescope using FPGA technology for the data collection & processing

Source: John Bunton CSIRO

Other RC Applications

- Applications suited to acceleration
  - seismic processing astrophysics FFT
  - adaptive optics (transforming to frequency domain and removing telescope image noise)
  - biotech applications such as BLAST, Smith Waterman and HMM
  - computational finance

- Functions well suited to FPGA acceleration
  - searching & sorting
  - signal processing (audio/video/image manipulation)
  - encryption（加密）
  - error correction
  - coding/decoding
  - packet processing
  - random-number generation for Monte Carlo simulations

Source: cray.com

THE UNIVERSITY OF SYDNEY

Conclusion

- uPs are the most flexible technology but performance (speed and power) is relatively low
- FPGAs provide
  - Easy interfacing with hardware (tighter coupling than GPUs)
  - Parallelism
  - Have become large enough to implement DSP and ML algorithms
  - Very interesting research area: architectures, tools, applications
- ASICs becoming only be suitable for highest volume, highest performance applications, FPGAs will do the rest
- Many of the highest performance accelerators, particularly for real-time problems, are FPGA-based