System-on-a-Chip: A Case for Heterogeneous Architectures

Jan M. Rabaey
BWRC
University of California @ Berkeley
http://bwrc.eecs.berkeley.edu

With contributions from Richard Newton and many others

Berkeley Wireless Research Center

Design at a Crossroad

Silicon technology tracking Moore's Law

Silicon in 2010

Die Area: 2.5x2.5 cm

Voltage: 0.6 - 0.9 V

Technology: 0.07 μm

	Density (Gbits/cm2)	Access Time (ns)
DRAM	8.5	10
	[illegible]	10
	[illegible]	1.5

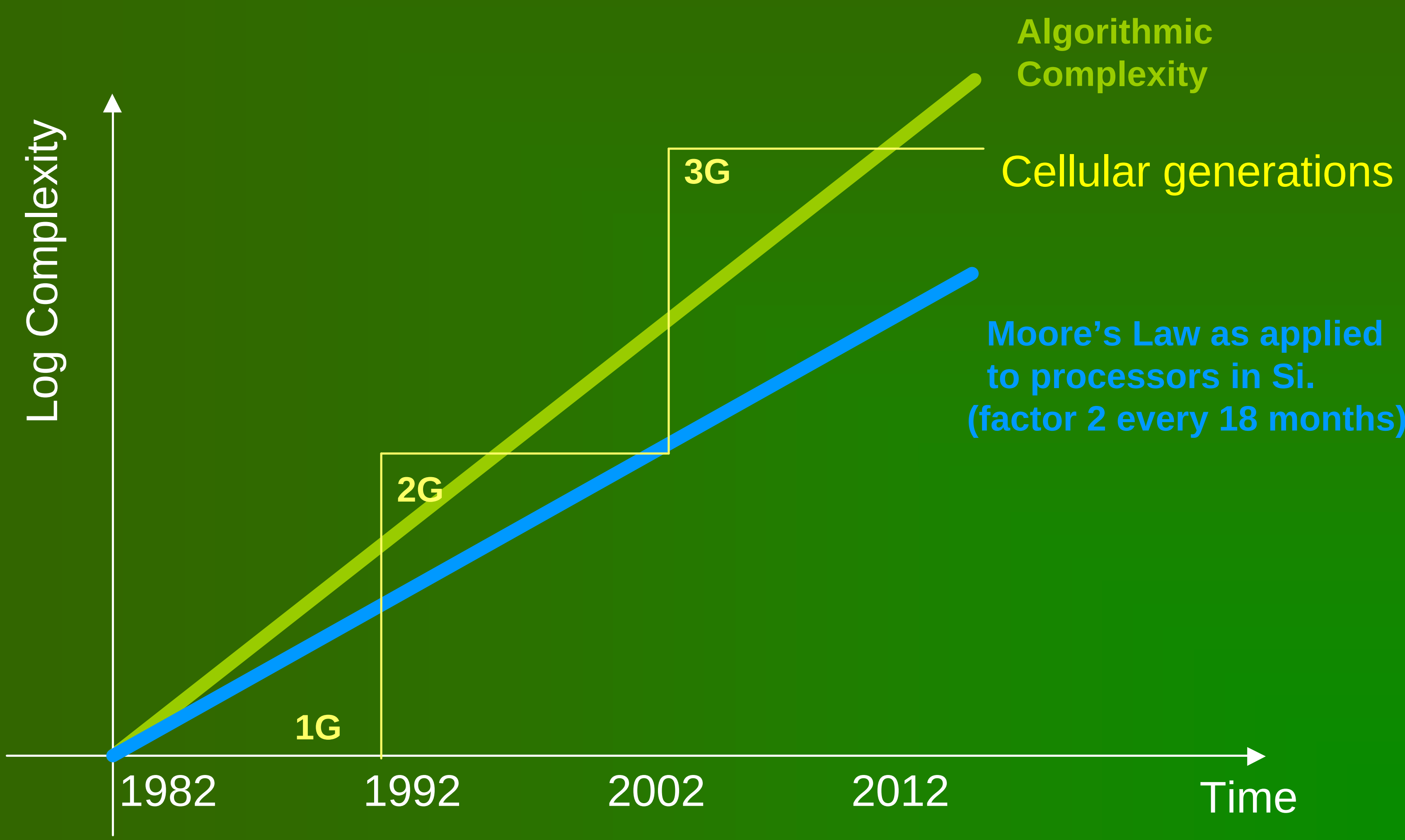
2.5 times power

5 times clock rate

	Density (Mgates/cm2)	Max. Ave. Power (W/cm2)	Clock Rate (GHz)
Custom	25	54	3
Std. Cell	10	27	1.5
Gate Array	5	18	1
Single-Mask GA	2.5	12.5	0.7
FPGA	0.4	4.5	0.25

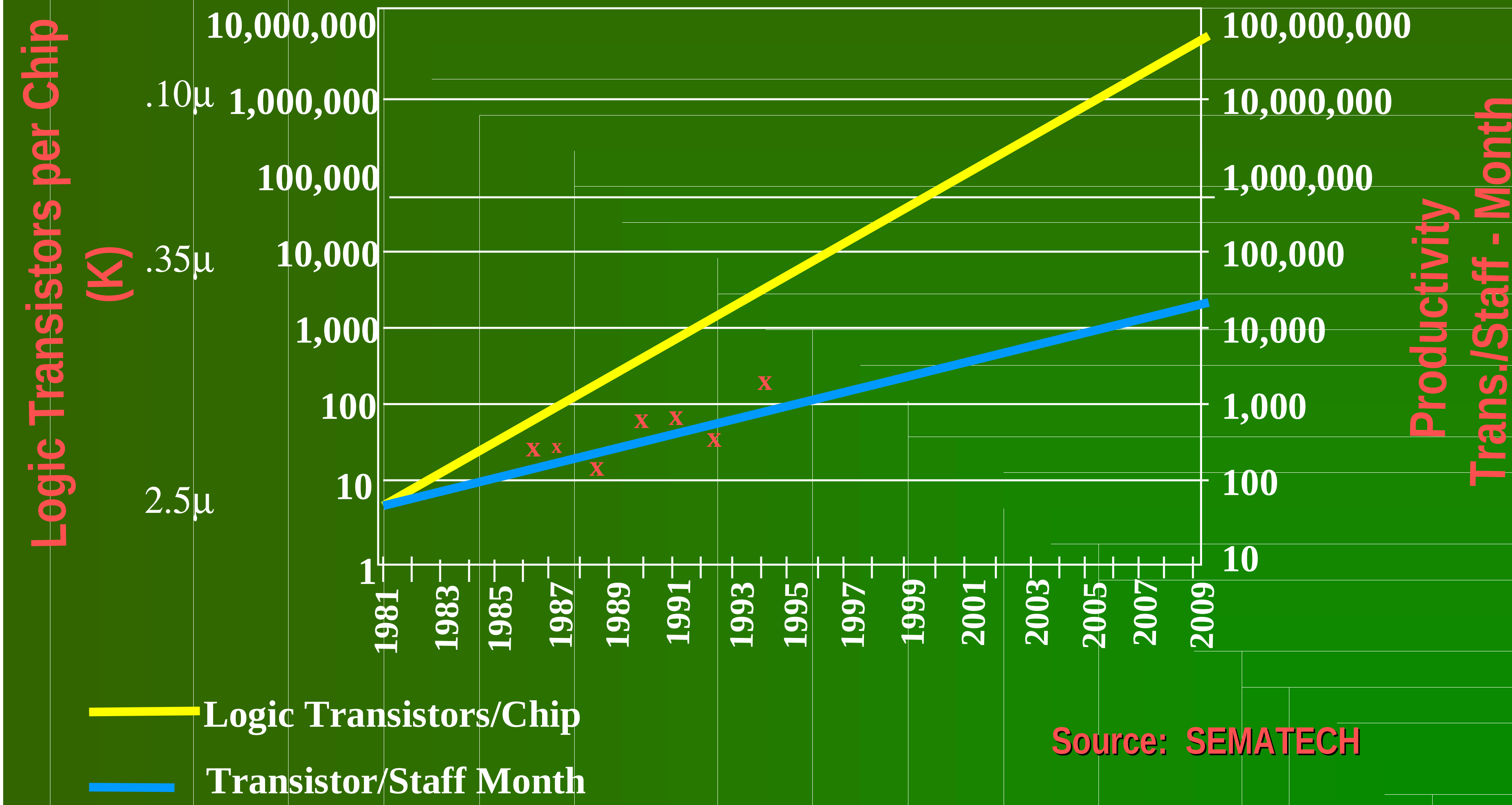
Design at a Crossroad

Applications beat Moore's Law

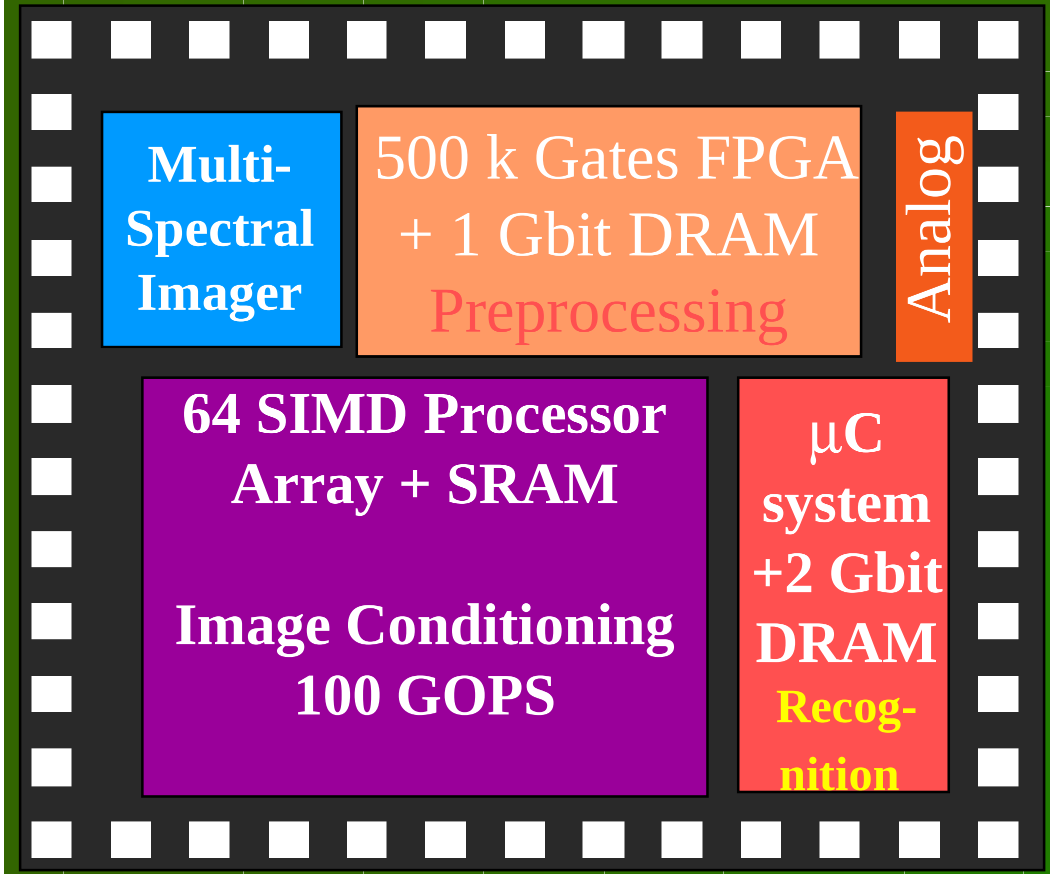


Source: R. Subramanian, Mophics Tech. Inc

Design at a Crossroad
The Productivity Gap



Design at a crossroad
System-on-a-Chip



- Embedded applications where cost, performance, and energy are the real issues!
- DSP and control intensive
- Mixed-mode
- Combines programmable and application-specific modules
- Software plays crucial role

The Distributed Approach to Information Processing

The Changing Metrics

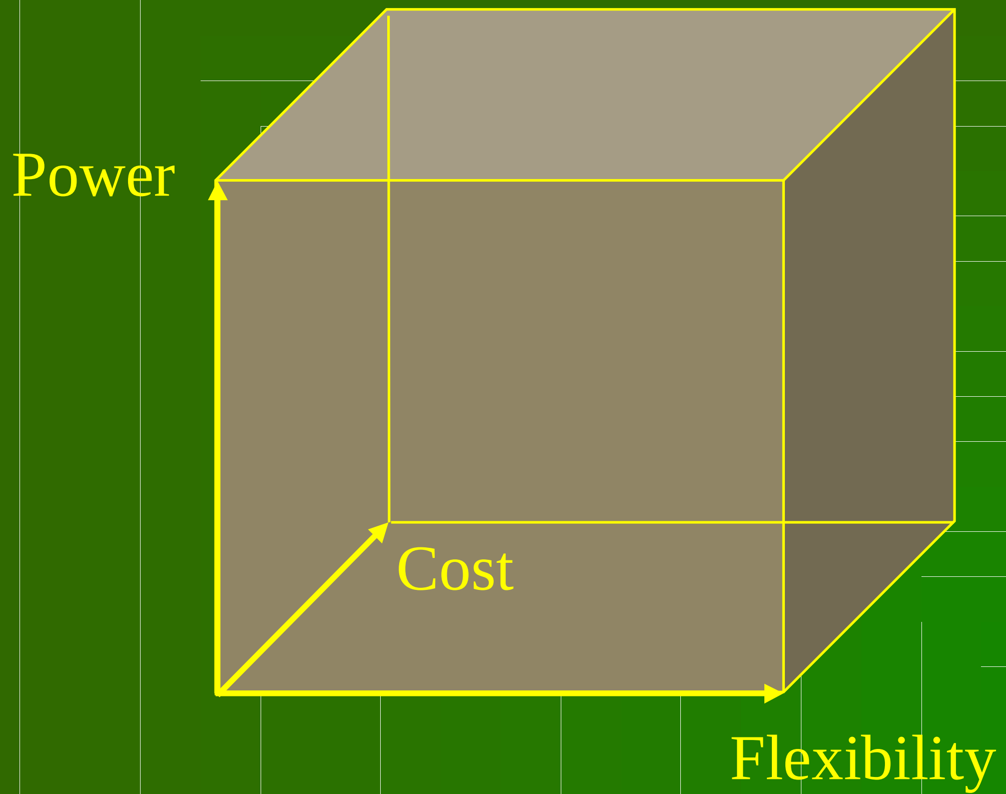
- Power and/or Energy have become dominant drivers
  – Limiting factor for performance and reliability in wall-plugged applications
  – Enabler for wide-spread use of distributed computing and data access
- Energy reduction requires joint optimization process between application and implementation

The Changing Metrics

- Cost of fabrication facilities and mask making has increased significantly
  - NRE cost of new design has increased significantly
- Physical effects (parasitics, reliability issues, power management) are increasingly significant in the design process
  - These must now be considered explicitly at the circuit level
- Design complexity, and “context complexity” is sufficiently high that design verification is a major limitation on time-to-market

Towards Fewer, but more Flexible and Reusable Silicon Platforms

The Changing Metrics

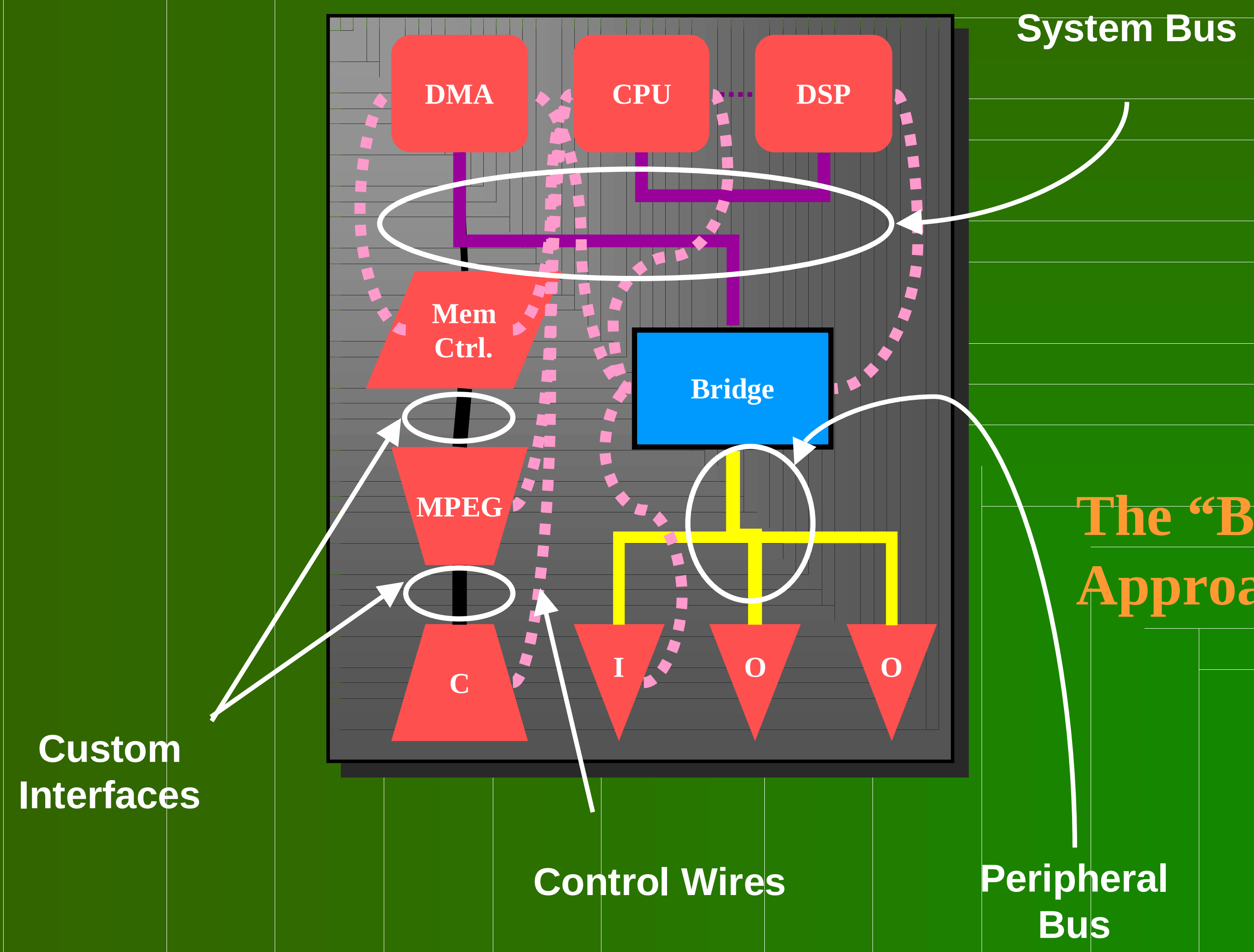


Performance as a Functionality Constraint
("Just-in-Time Computing")

The System-on-a-Chip Nightmare

"Femme se coiffant"
Pablo Ruiz Picasso
1940

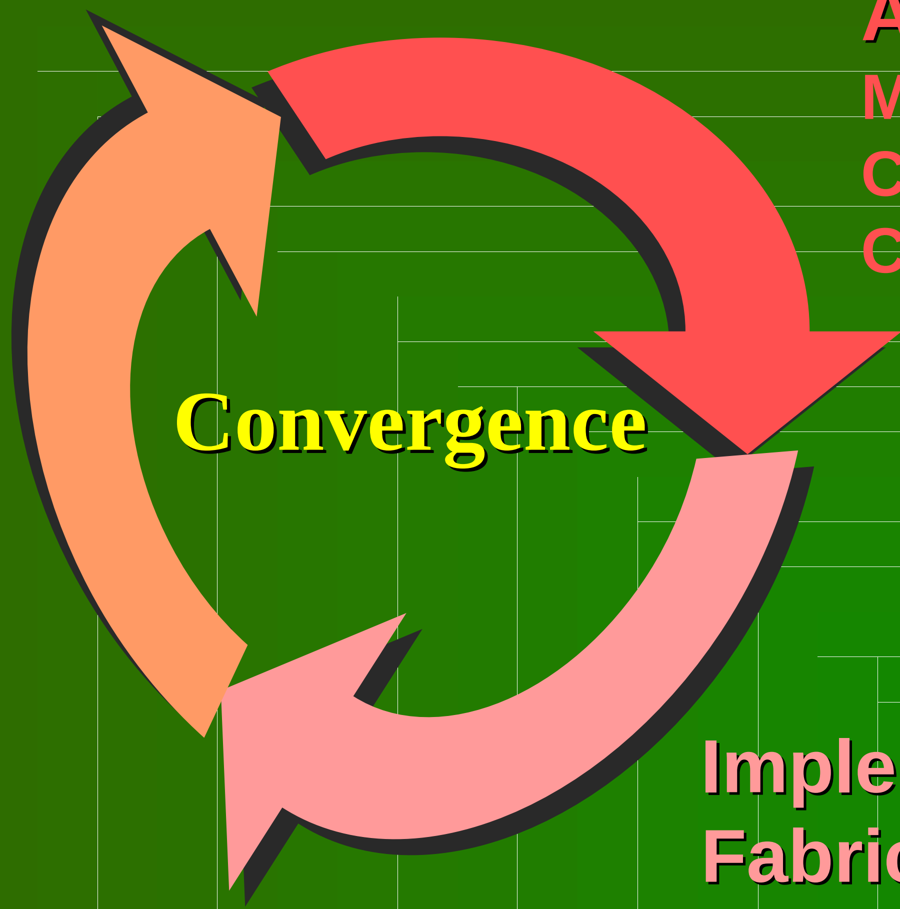
The System-on-a-Chip Nightmare



The "Board-on-a-Chip" Approach

System-on-a-Chip
A Renaissance in Design

Design Methodology
Hard+Soft

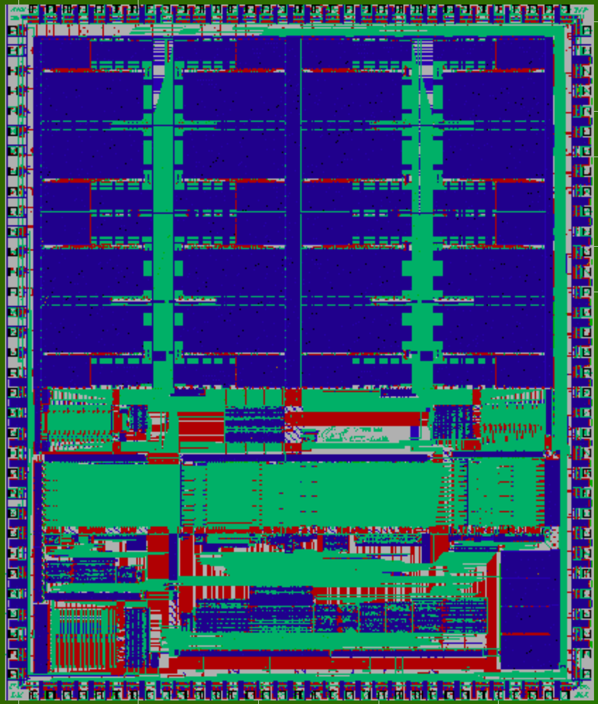


Applications
Multimedia
Consumer
Communications

Implementation Fabrics
Silicon substrate
Silicon fabrics

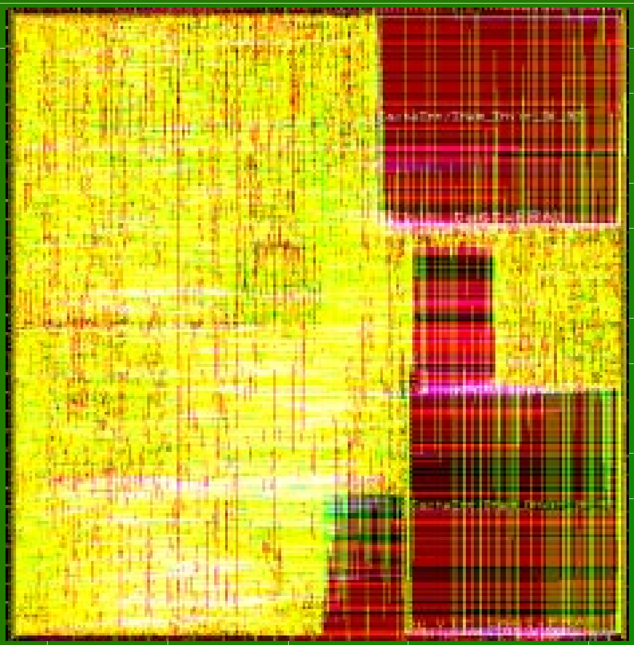
Aart De Geus
DAC'99

An Architectural Renaissance



Embedded ARM-8
Microprocessor
(Hard IP)

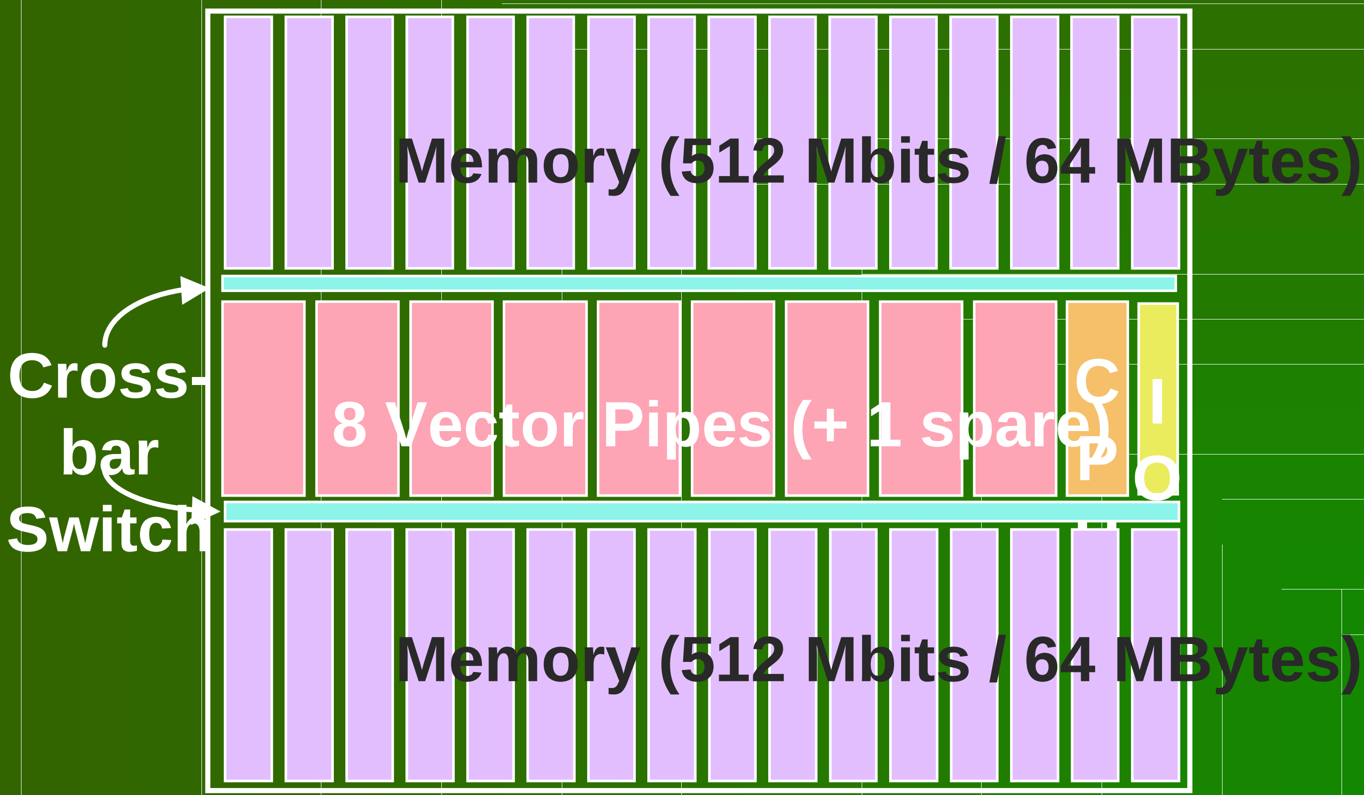
Tensilica Synthesized and
Configurable μProcessor
(Soft IP)



Courtesy of ARM, Tensilica Inc

An Architectural Renaissance

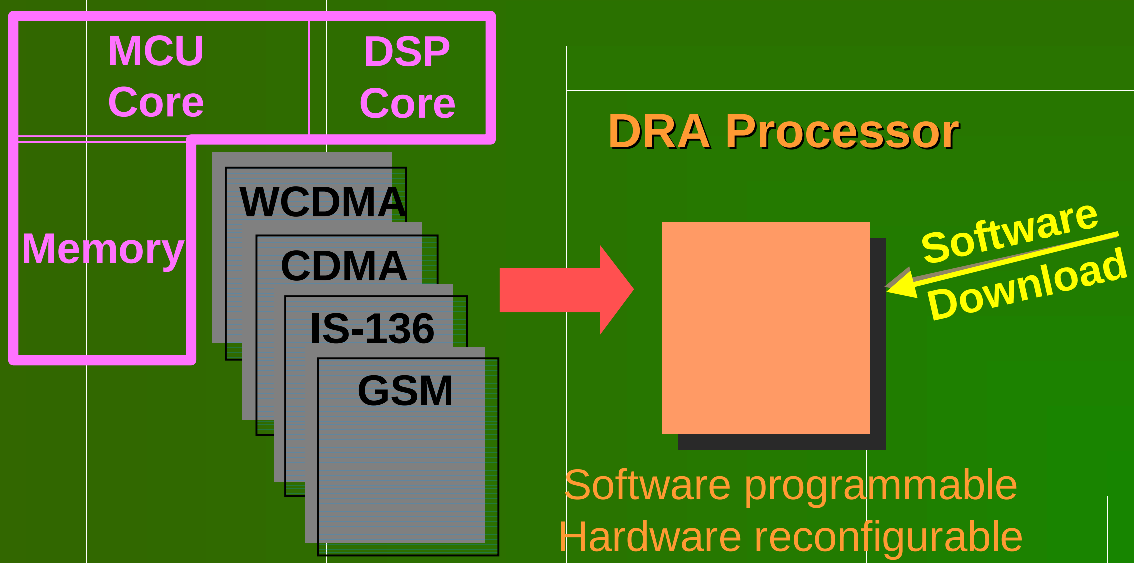
“Very-Short Instruction Word” Processors



0.13 µm CMOS
1 GHz
16 GFLOPS
64 GOPS
(projected)

V-IRAM: An integrated Vector Processor for Media Processing
[Patterson et all]

An Architectural Renaissance



Fixed logic…

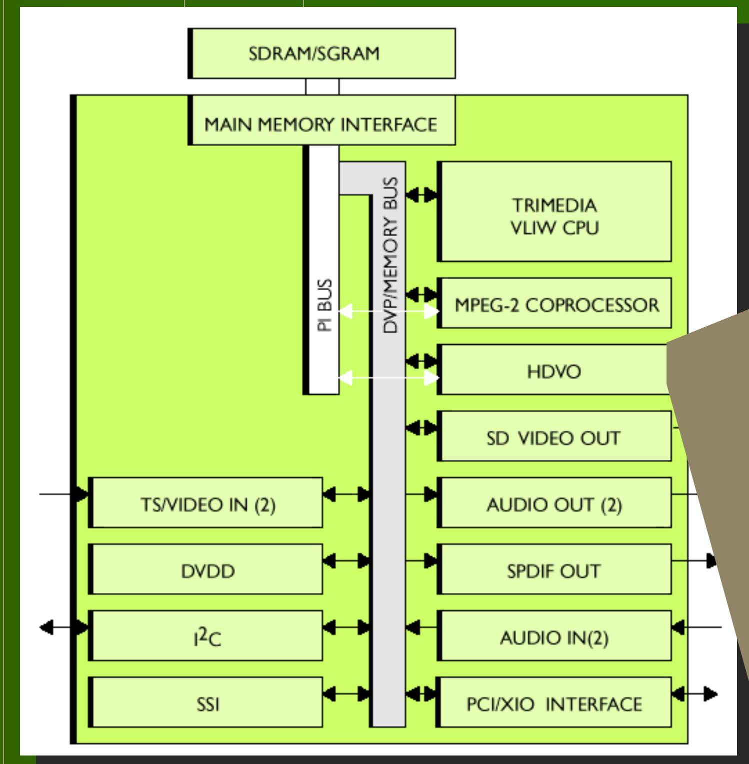
WCDMA	(mode, param)
CDMA	(mode, param)
WTDMA	(mode, param)
TDMA	(mode, param)

- SIM Card
- Handset Memory
- POS Programming
- Network Download
- OTA Download

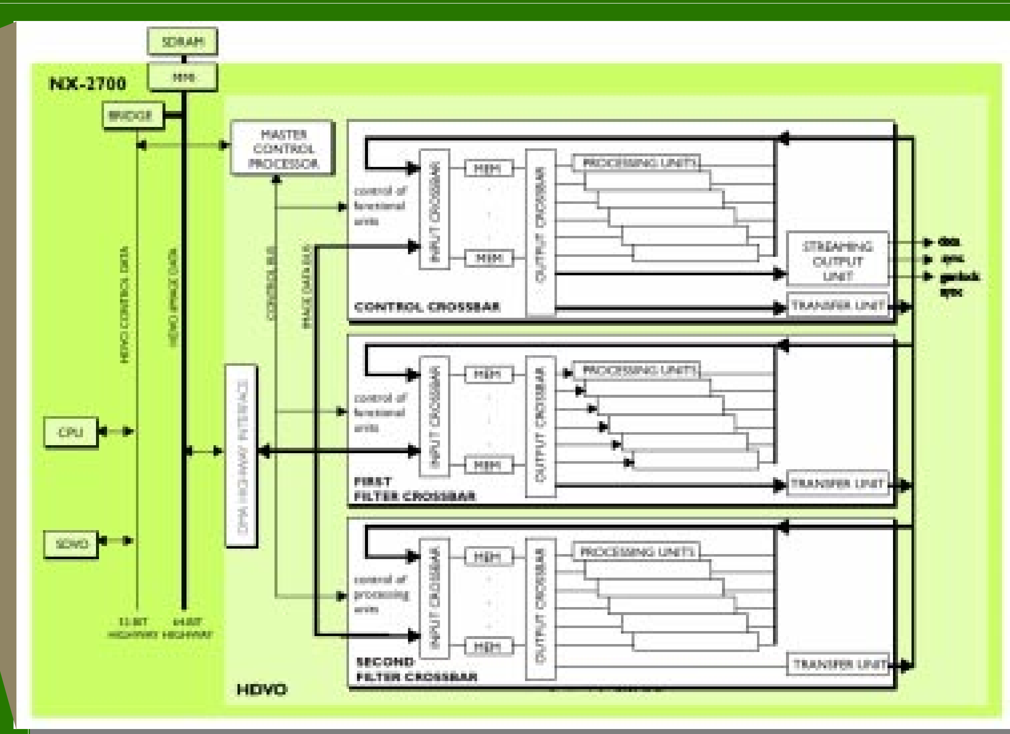
Realizes cost, size and power targets similar to traditional core+hardwired

MorphICs Dynamically Reconfigurable Architecture (DRA) Processor

An Architectural Renaissance



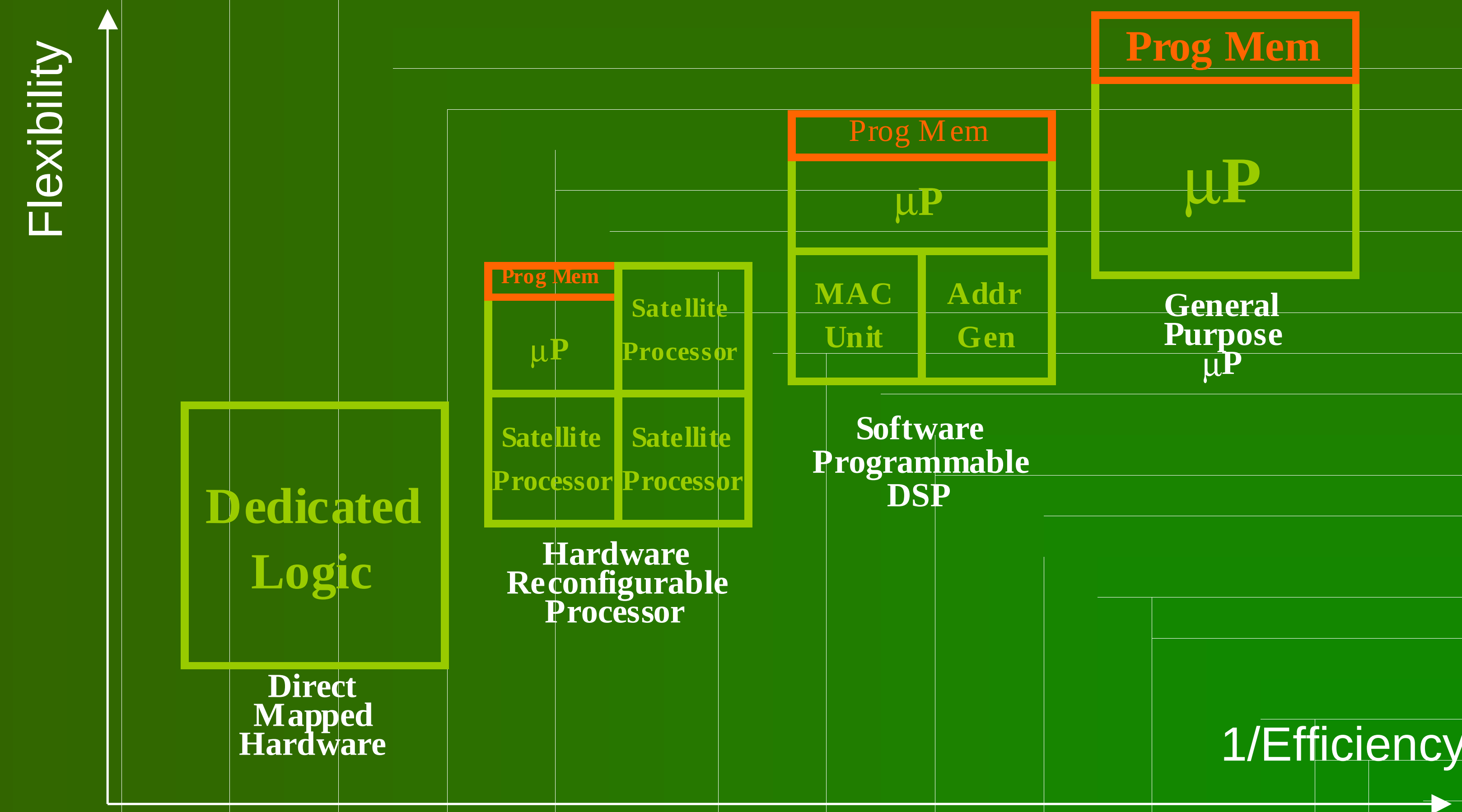
Combines Trimedia VLIW with Configurable media co-processors



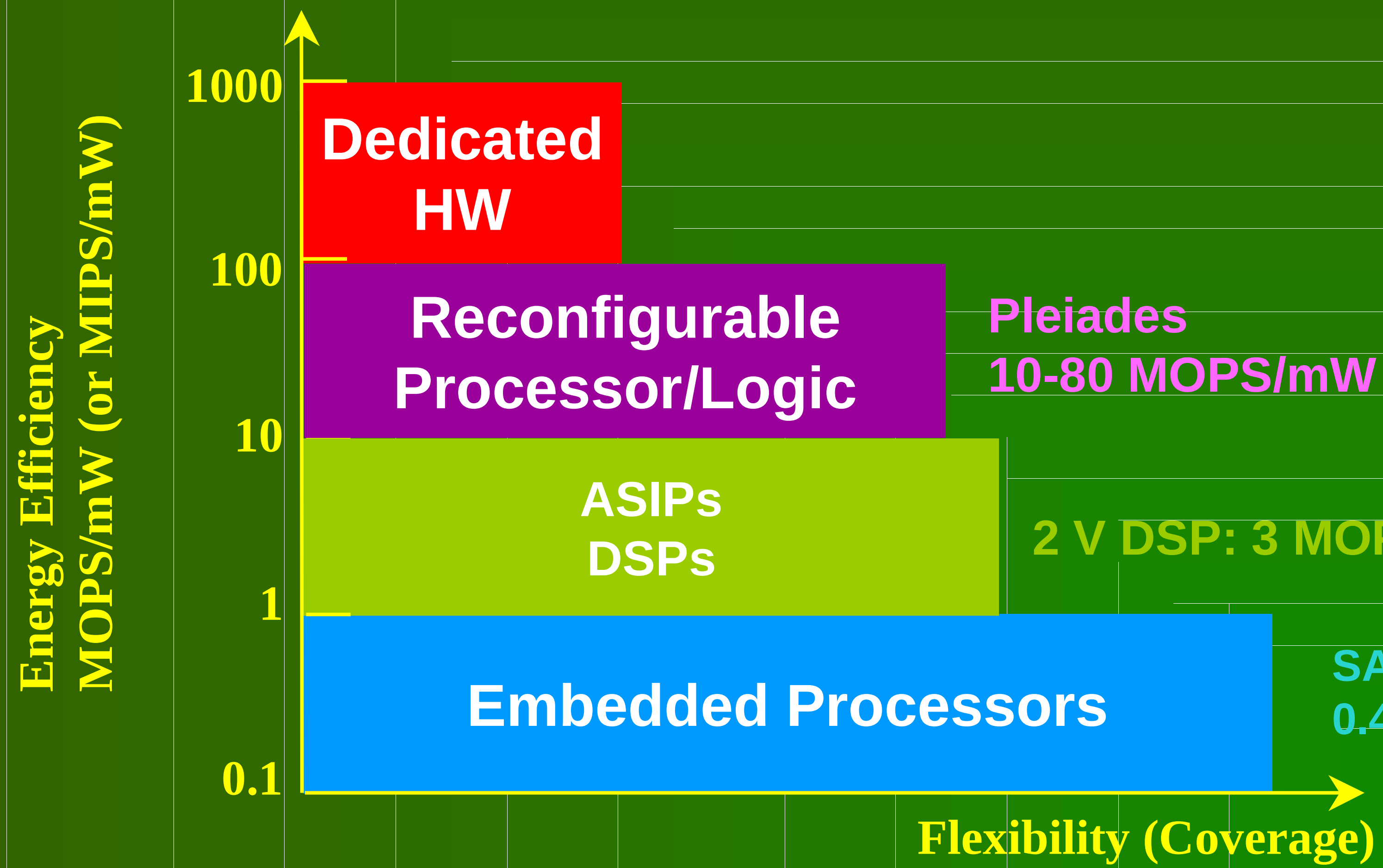
Philips Nexperia NX-2700

A programmable HDTV media processor

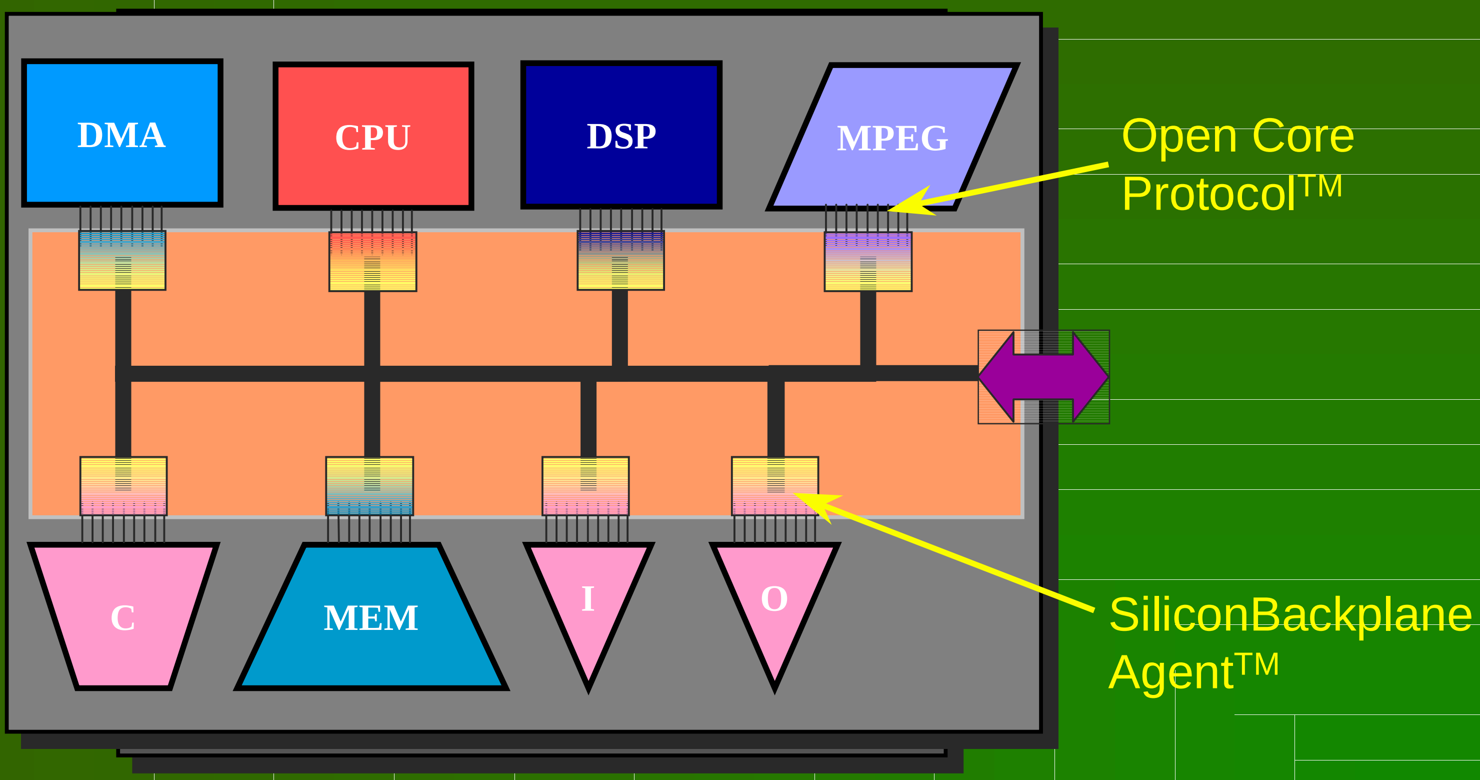
Architectural Choices

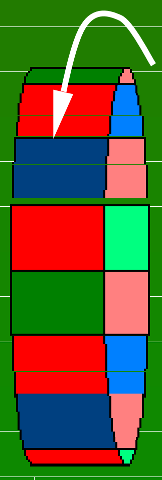


The Energy-Flexibility Gap



An Architectural Renaissance

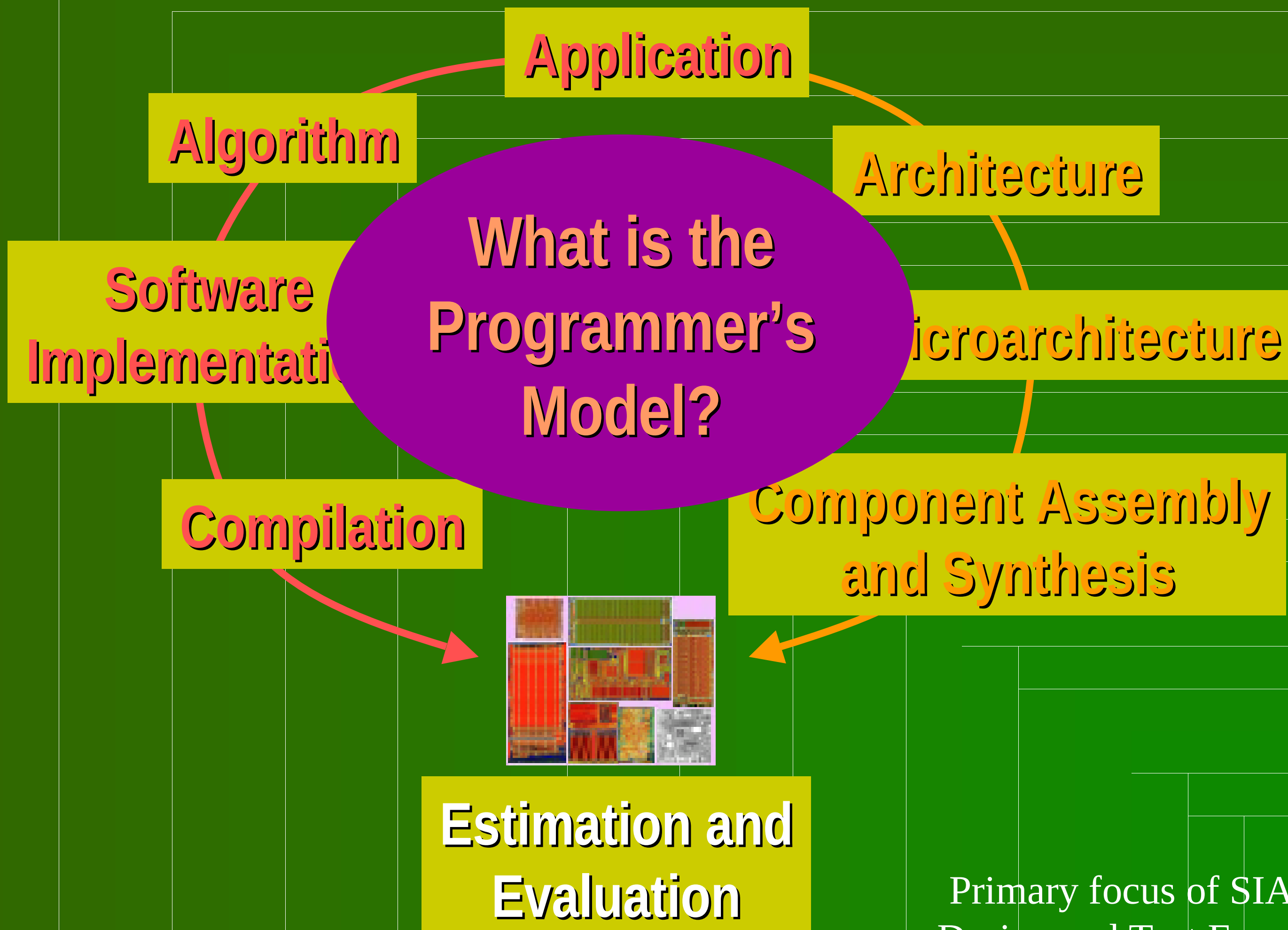




Communications-based Design

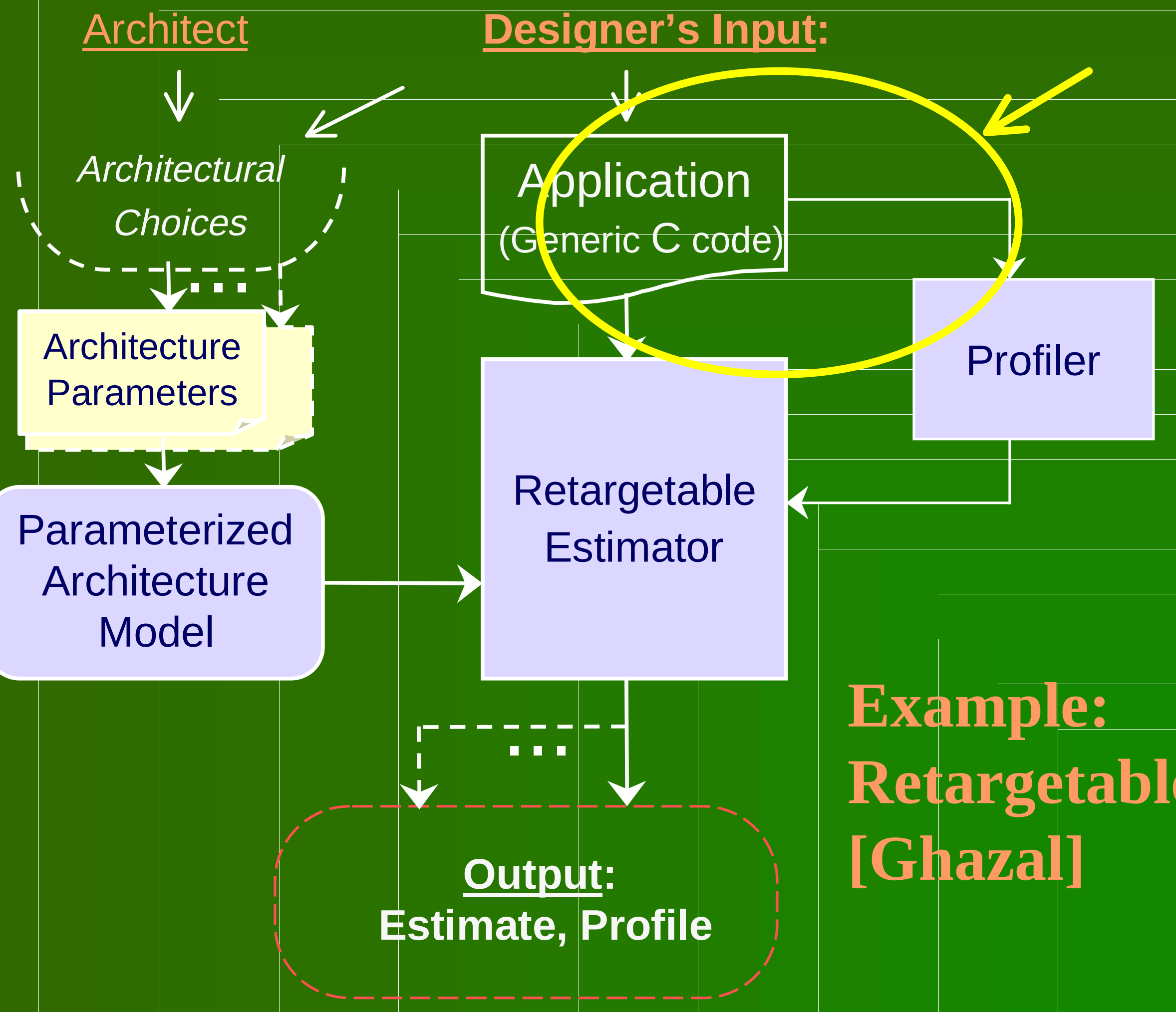
Example: “The Silicon Backplane” (Sonics, Inc)

Programming the Platform



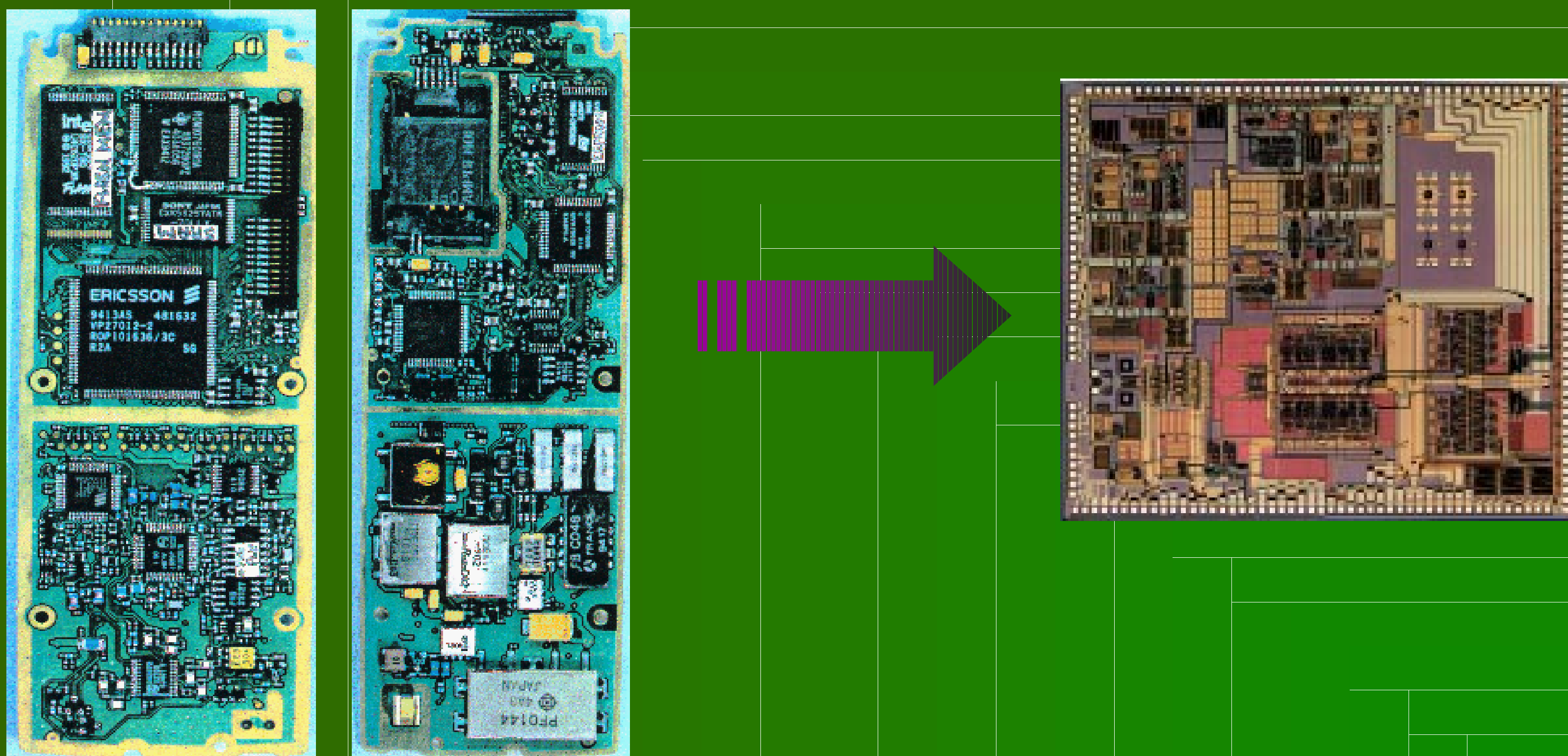
Primary focus of SIA GSRC
Design and Test Focus Center

Fast Design Space Exploration



Example:
Retargetable estimation [Ghazal]

A Case Study
The Integrated CMOS Radio



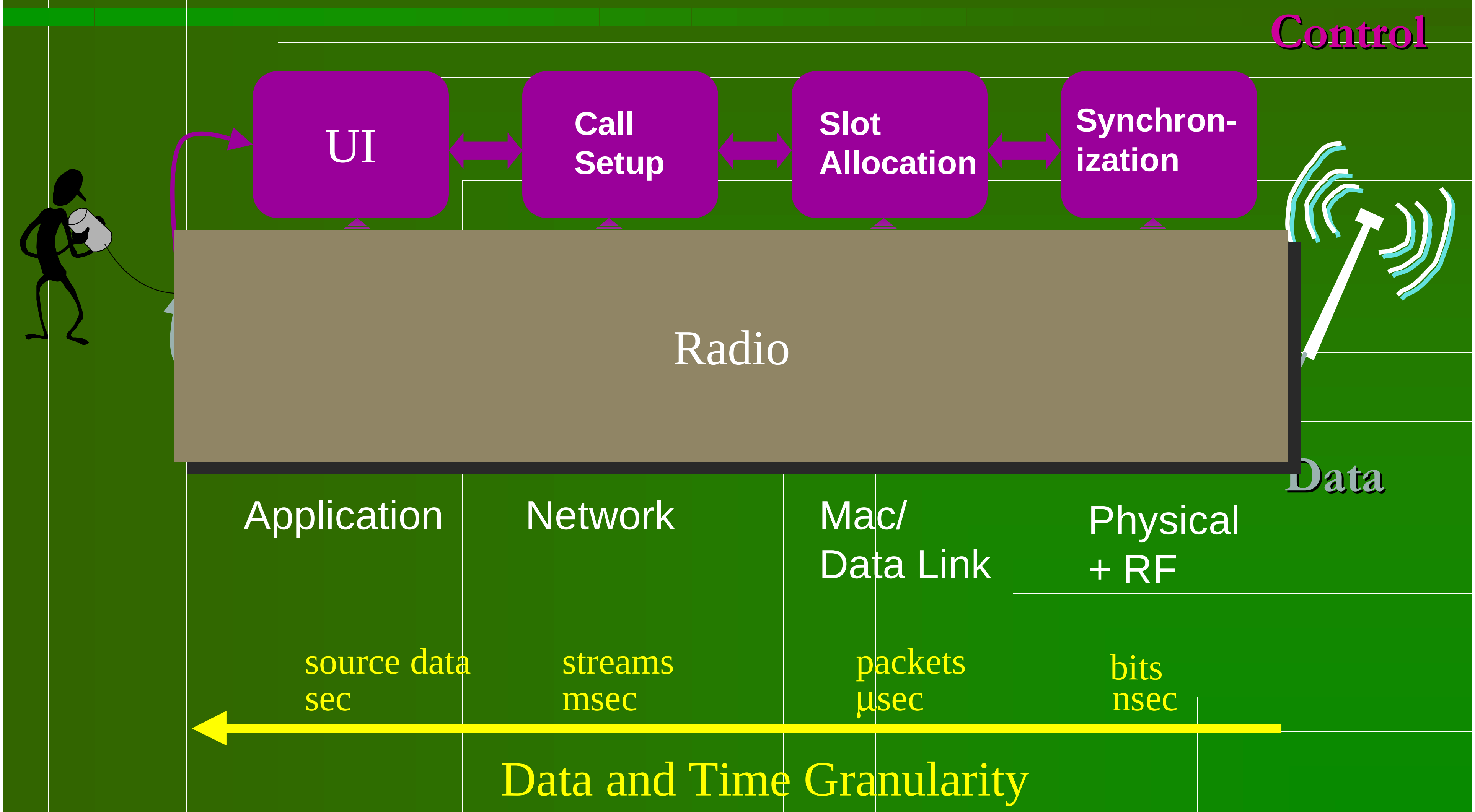
Trends in Wireless Systems

- Towards better spectrum utilization
  - **using aggressive signal and protocol processing**
    - Examples: multi-user detection, multi-antenna arrays
  - **adaptive, multi-functional networks**
    - Example: IMTS2000 / UMTS (3G)
- Towards ubiquitous wireless networking
  - Example: Bluetooth, HomeRF, FireFly

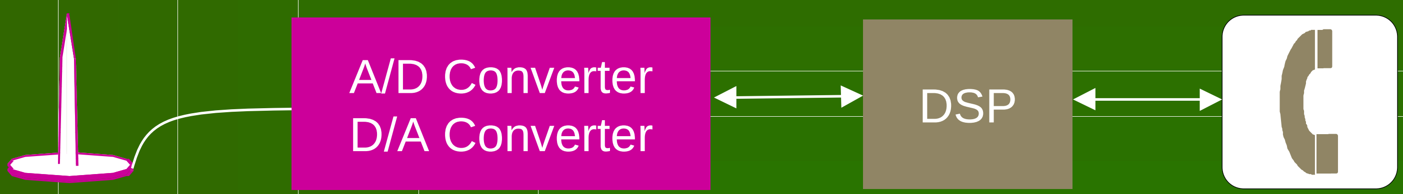
Resulting requirements

high performance, low-energy, adaptivity and flexibility

Issues in Single-Chip Radio Design

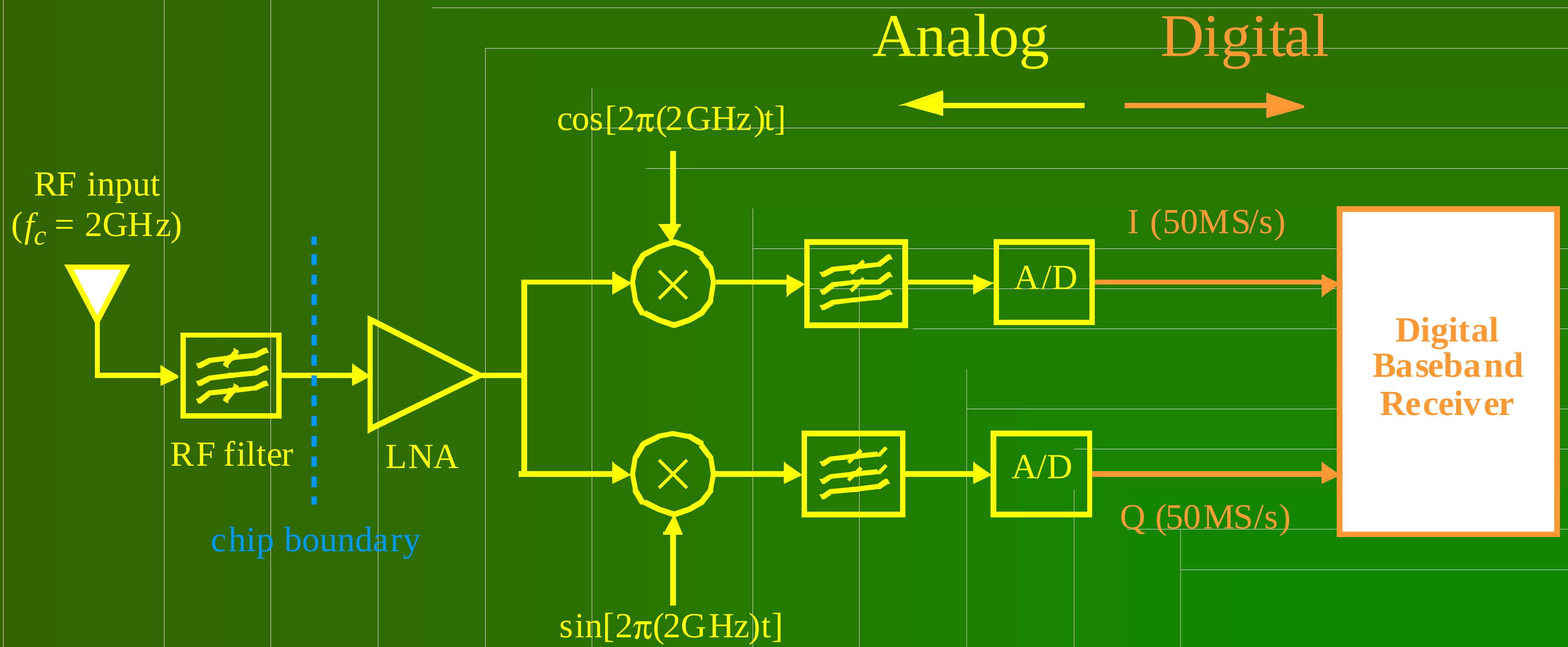


The Software Radio

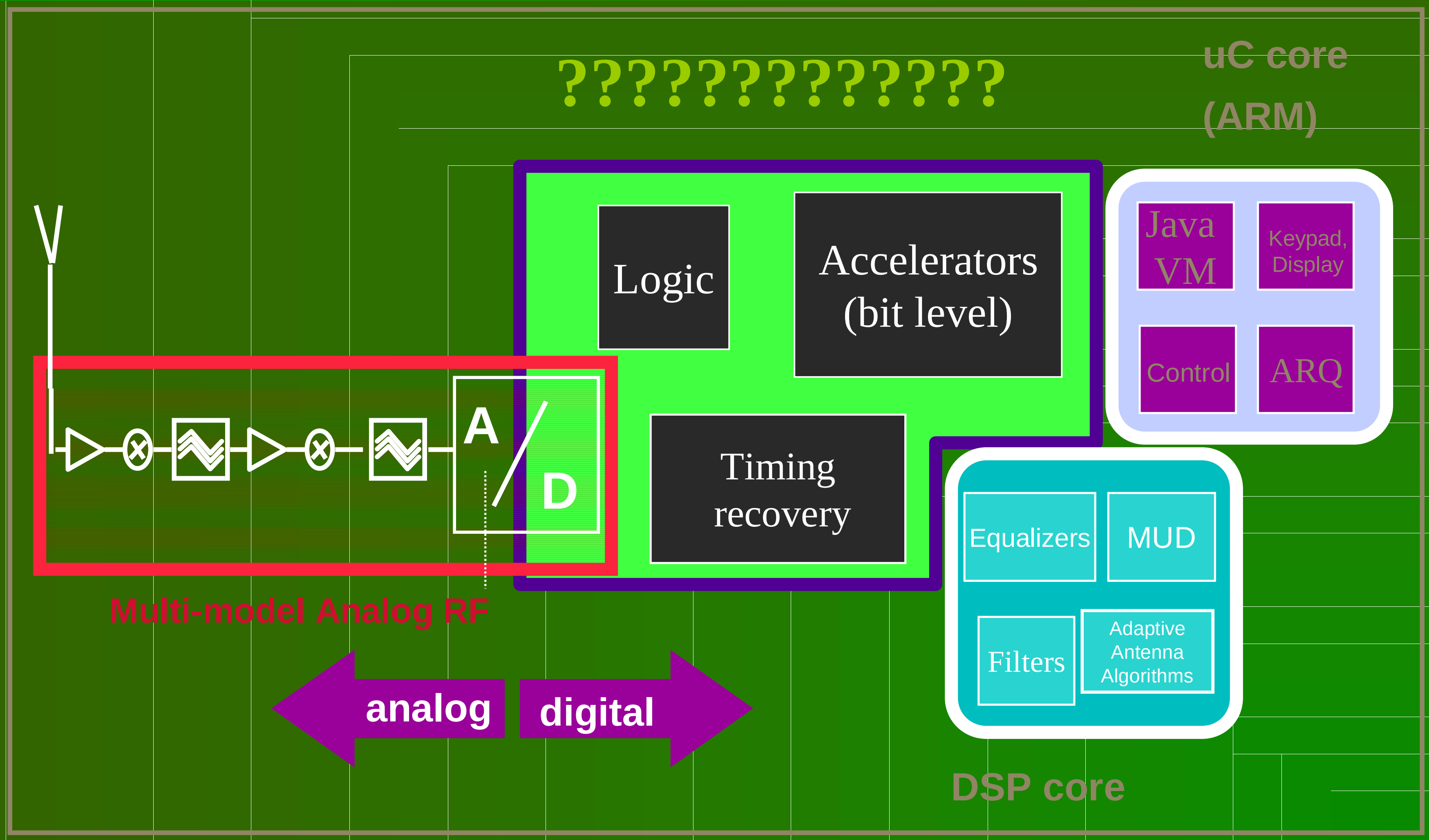


- Idea: Digitize (wideband) signal at antenna and use signal processing to extract desired signal
- Leverages of advances in technology, circuit design, and signal processing
- Software solution enables flexibility and adaptivity, but at huge price in power and cost
- 16 bit A/D converter at 2.2 GHz dissipates 1 to 10 W

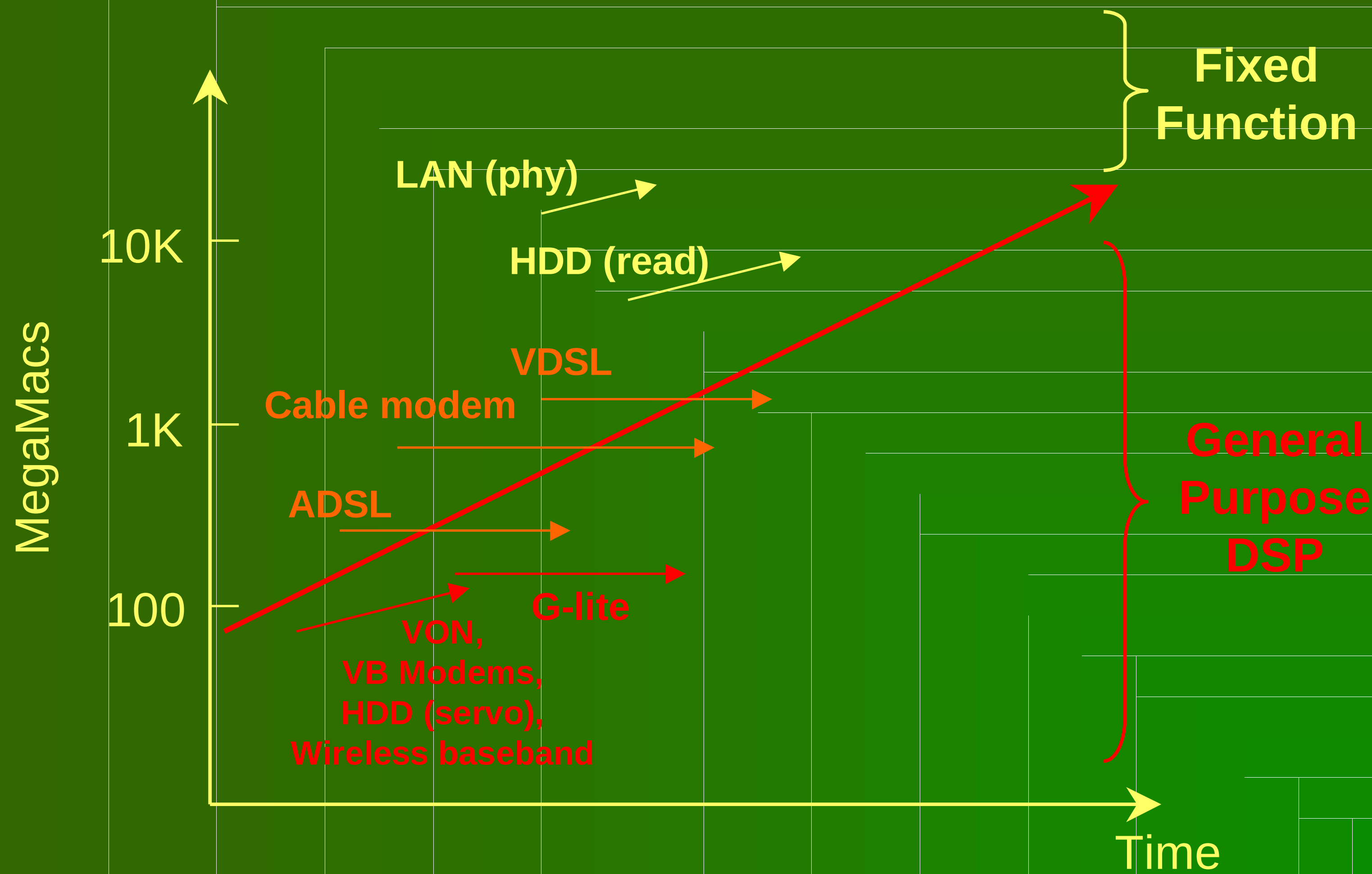
The Mostly Digital Radio



The Software-Definable Radio

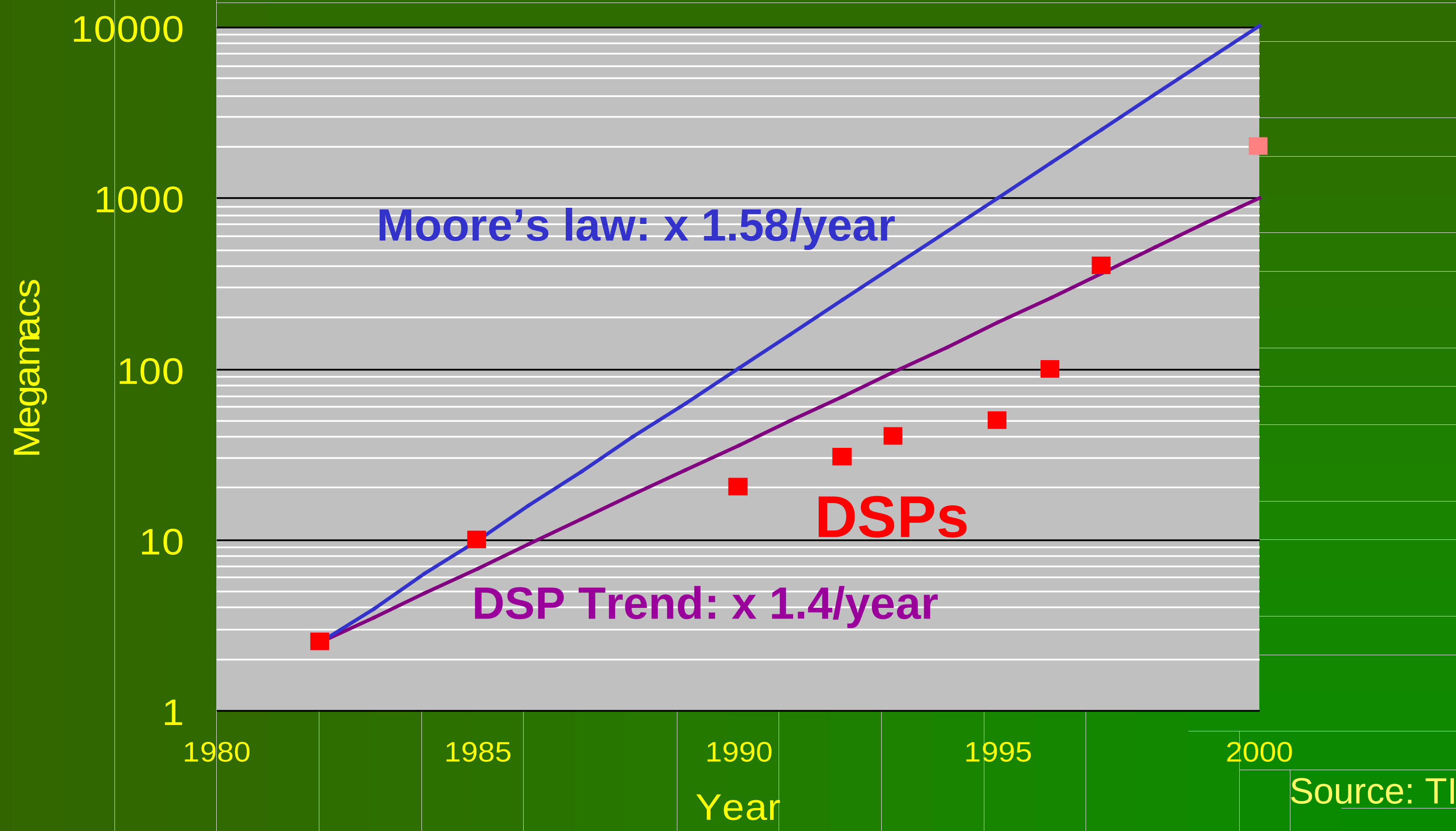


A Trend Towards GP DSPs?



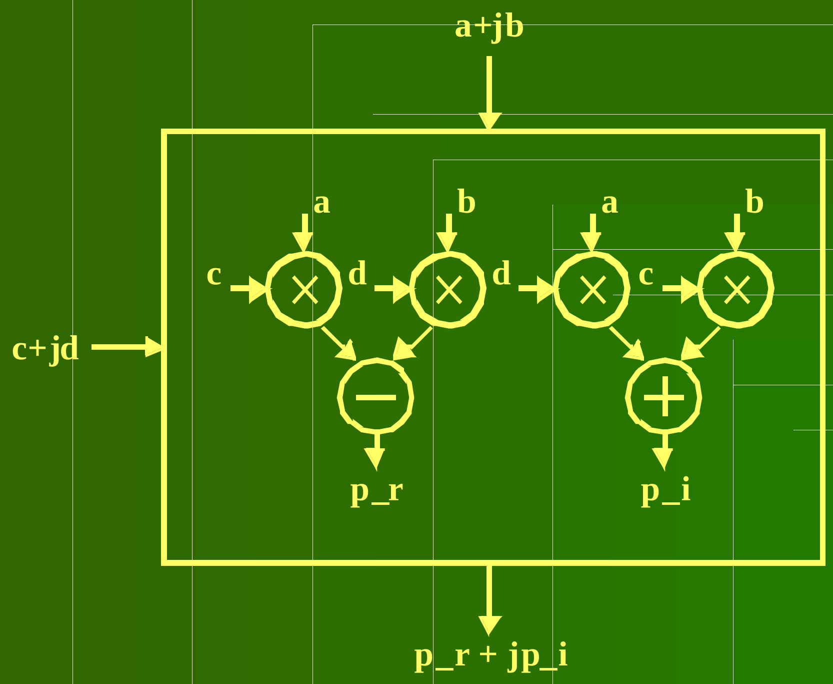
Source: TI

Single-Chip DSPs are Lagging ...

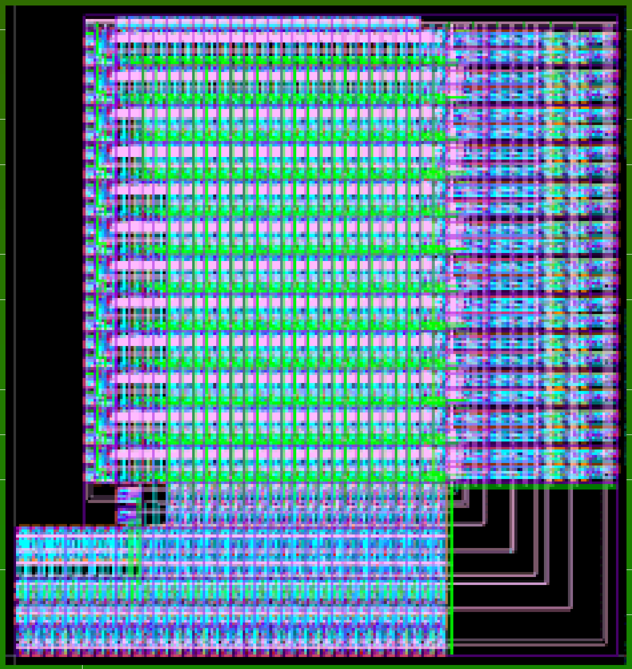


While algorithms are beating Moore's law!

And Computation Seems Almost for Free



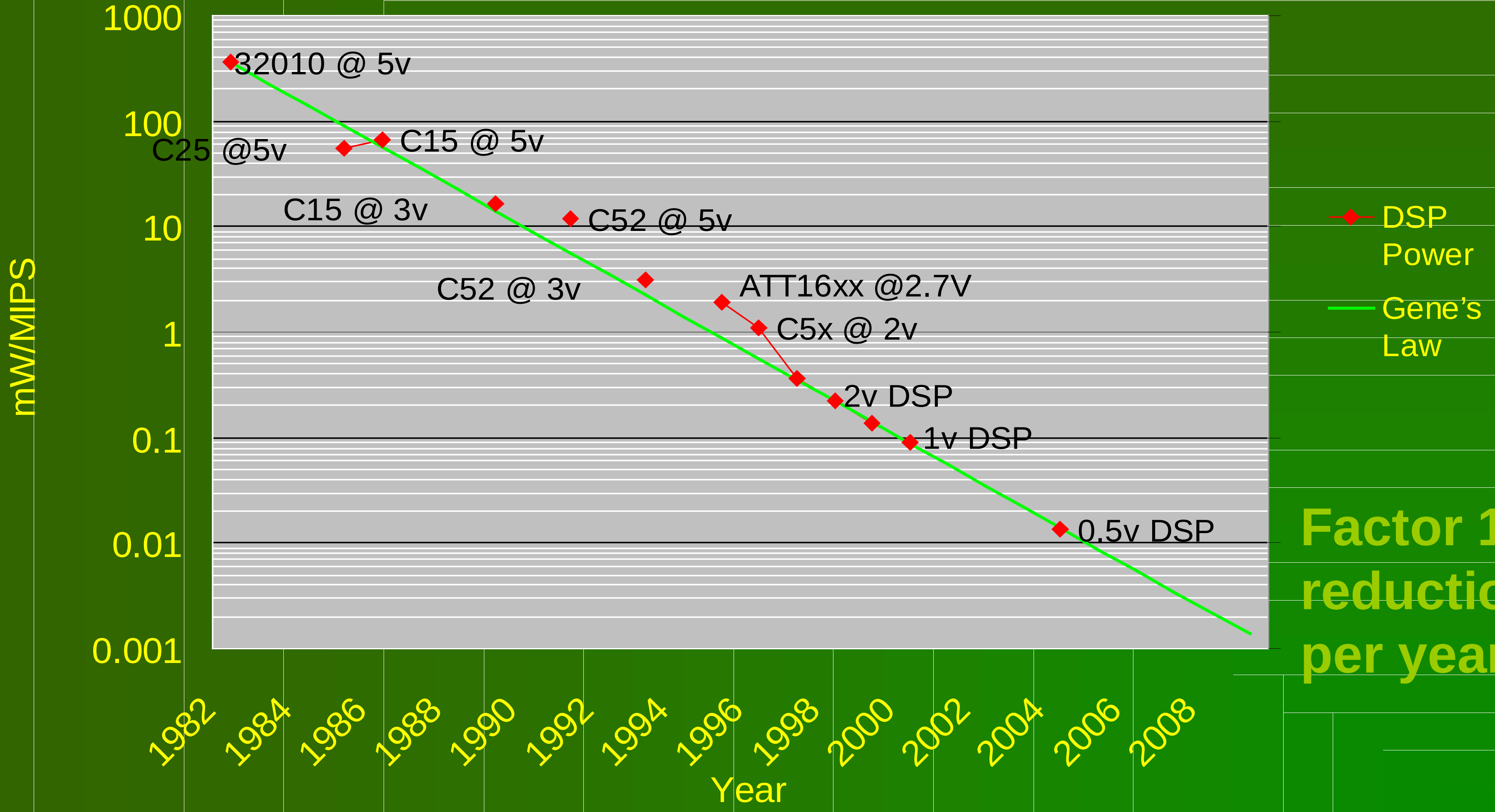
12 x 12 Complex Multiplier



0.25 µm CMOS process

- Area: 0.18 mm^2
- Perf: 25 MCMACS/sec @ 1V
- Energy: 40 MCMACs/mW

Energy Trends in DSPs

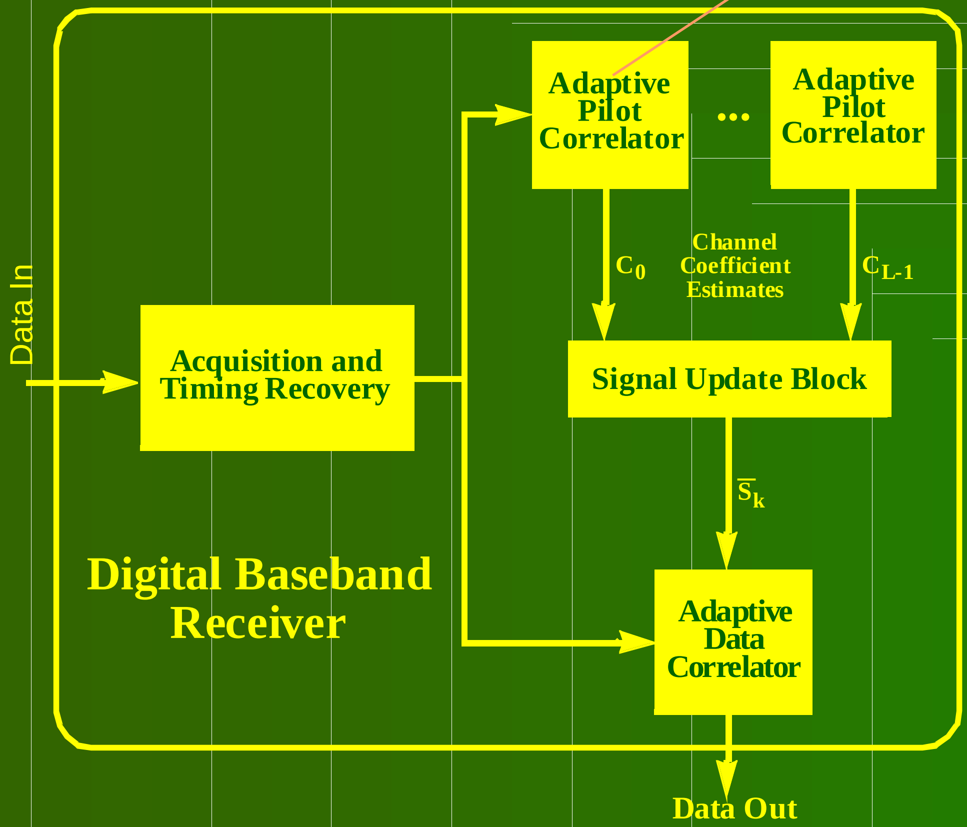


Factor 1.6 reduction per year

Source: TI

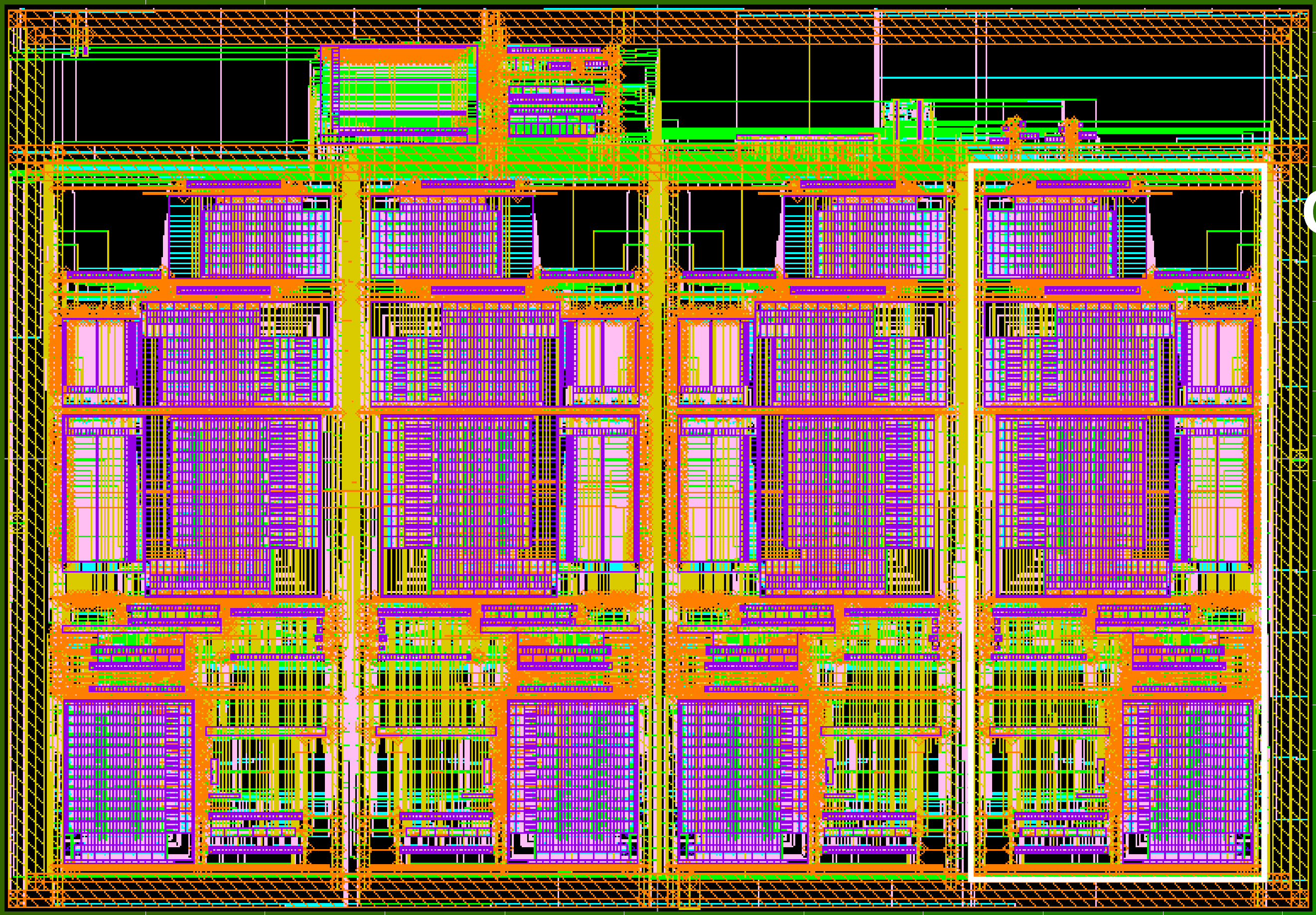
The Implementation Trade-off

300 million multiplications/sec
357 million add-sub's/sec



	DSP Implementation	Direct Mapped Implementation
Adaptive Pilot Correlator	Power Consumption: 460mW Area: 1089mm^2	Power Consumption: 3mW Area: 1.3mm^2
Digital Baseband Receiver	Power Consumption: 1500mW Area: 3600mm^2	Power Consumption: 10mW Area: 5mm^2

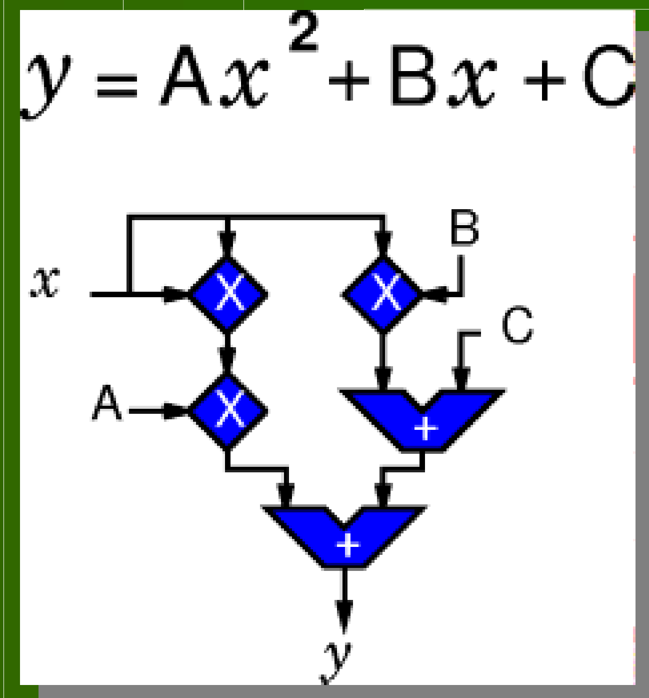
Adaptive Multi-User Detection
A Direct Mapping Approach



Power and area are dominated by MACs and multiplies
Only 36% of power of DSP-processor solution going into arithmetic

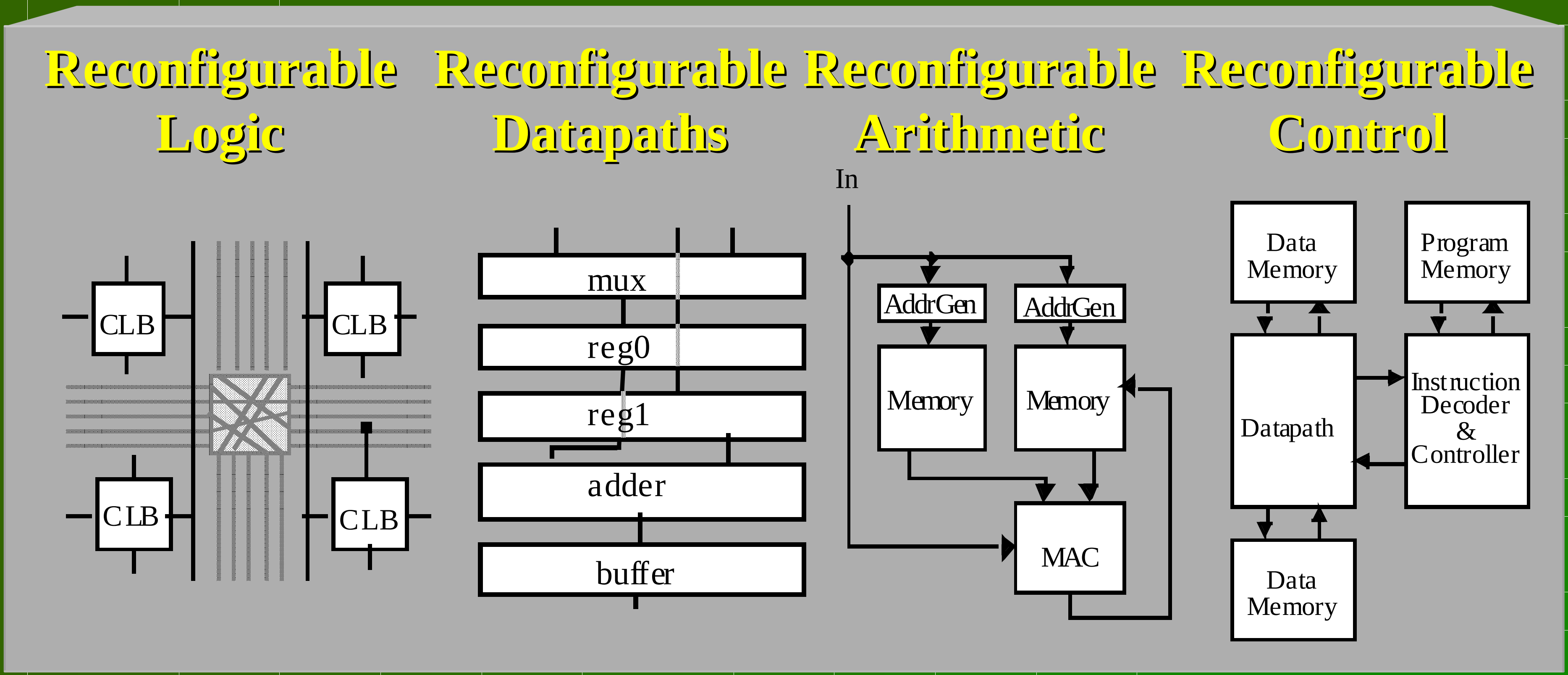
Reconfigurable Computing: Merging Efficiency and Versatility

Spatially programmed connection of processing elements.



- "Hardware" customized to specifics of problem.
  - Direct map of problem specific dataflow, control.
- Circuits "adapted" as problem requirements change.

A New Look at Architectures — Heterogeneous Reconfiguration



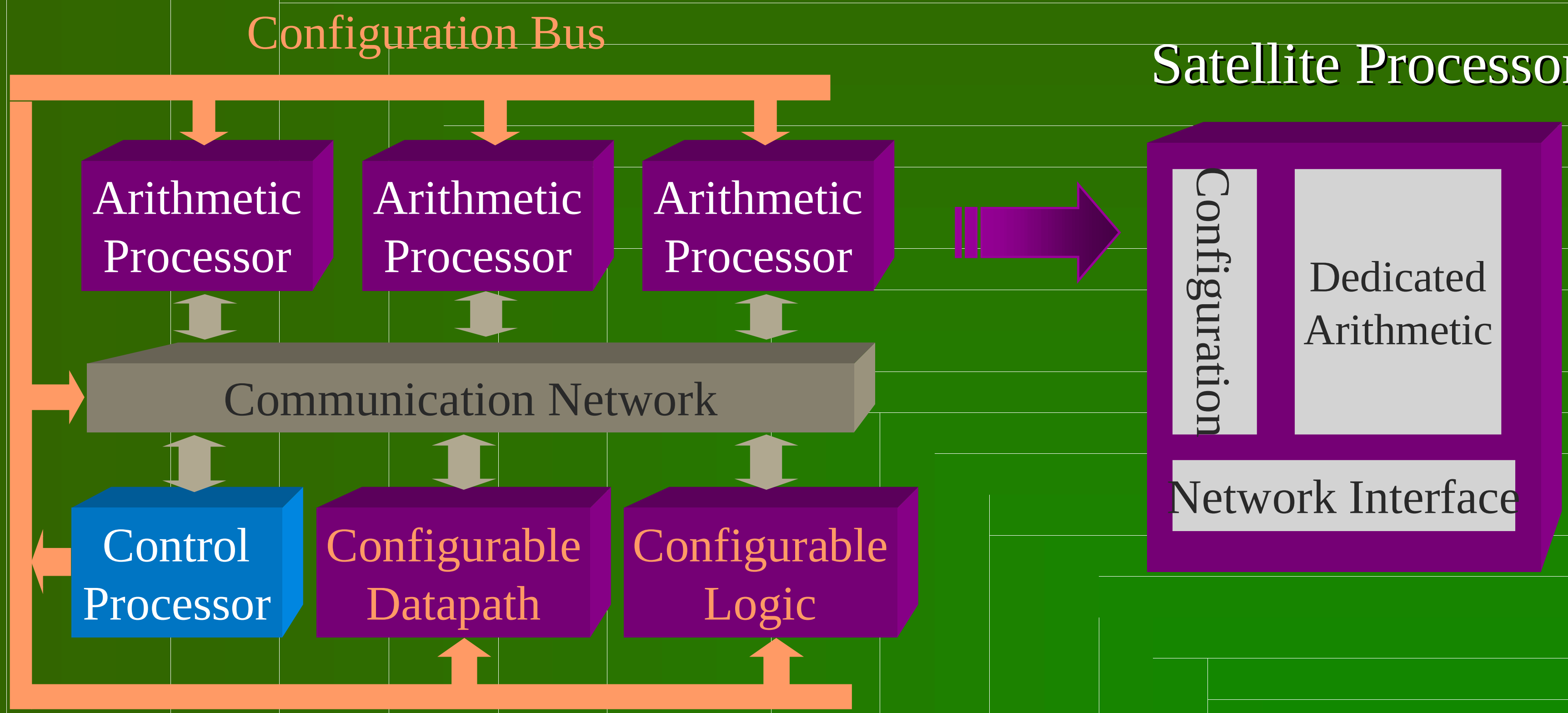
Bit-Level Operations
e.g. encoding

Dedicated data paths
e.g. Filters, AGU

Arithmetic kernels
e.g. Convolution

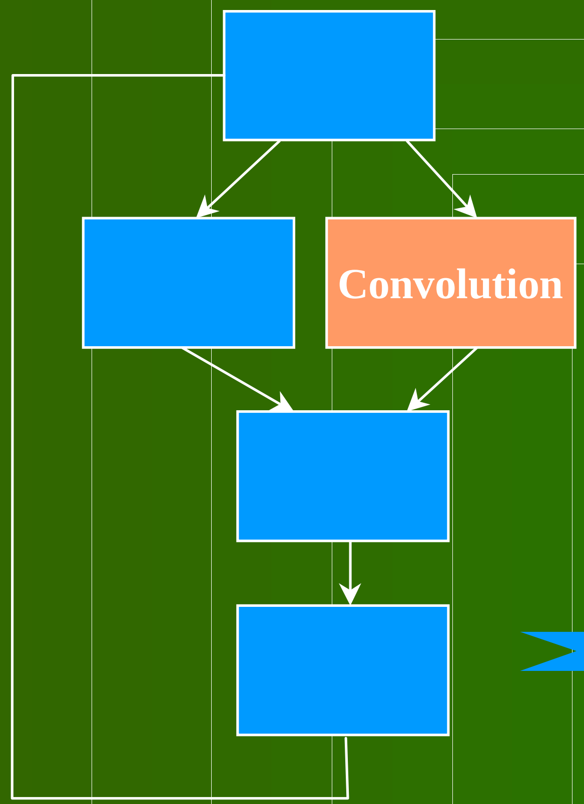
RTOS
Process management

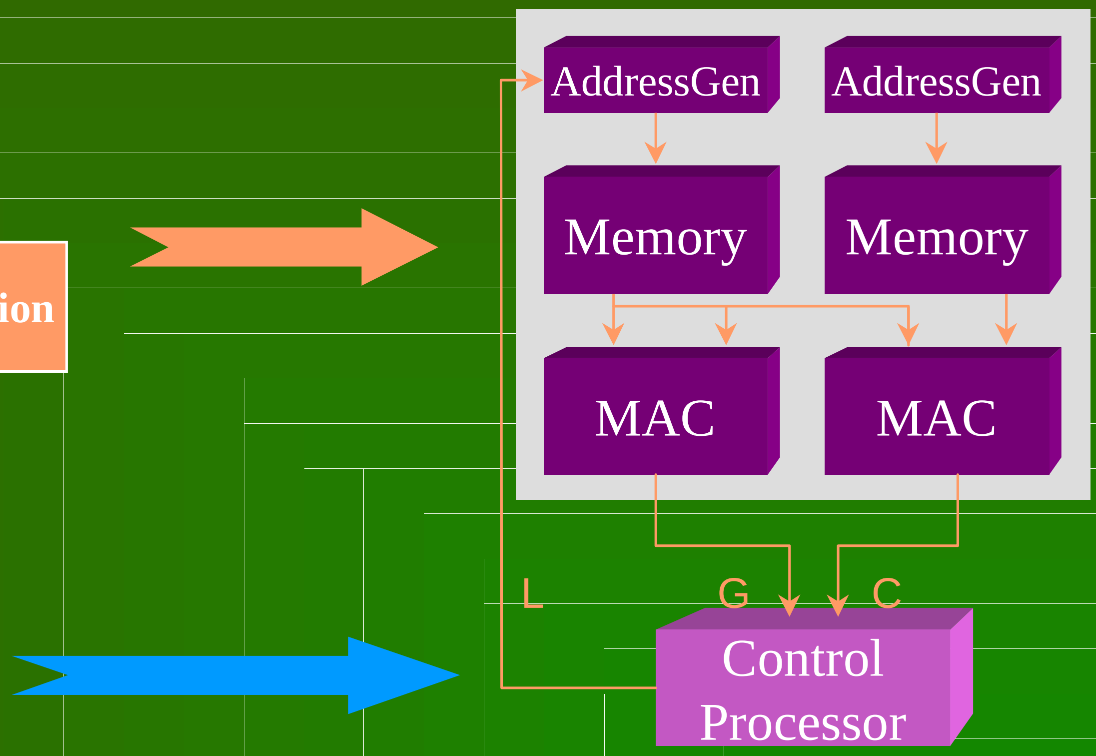
Multi-granularity Reconfigurable Architecture: The Berkeley Pleiades Architecture



- Computational kernels are “spawned” to satellite processors
- Control processor supports RTOS and reconfiguration
- Order(s) of magnitude energy-reduction over traditional programmable architectures

Matching Computation and Architecture



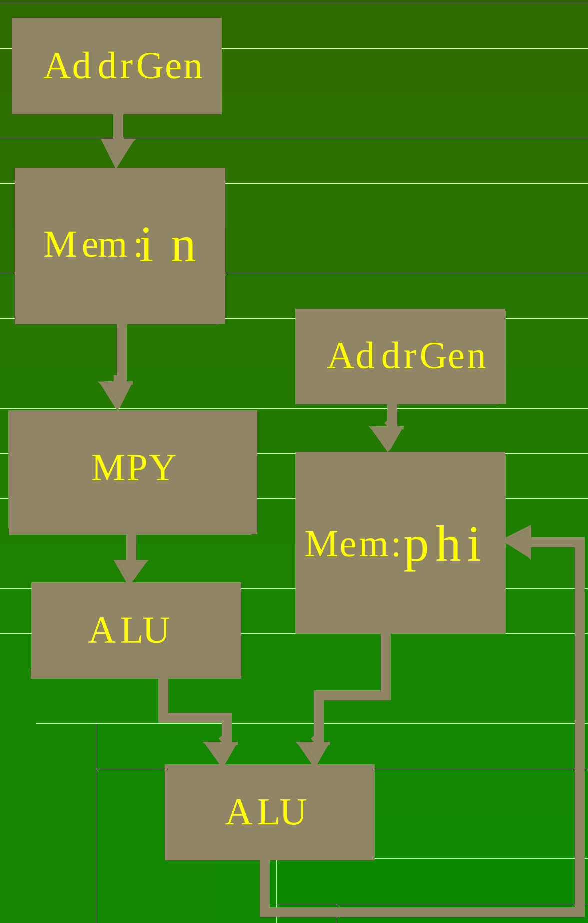


Two models of computation:
communicating processes + data-flow

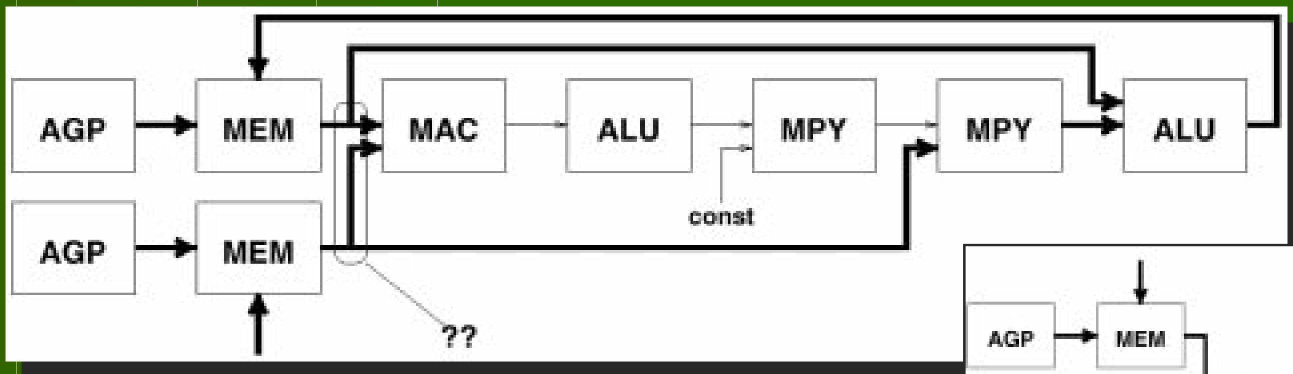
Two architectural models:
sequential control+ data-driven

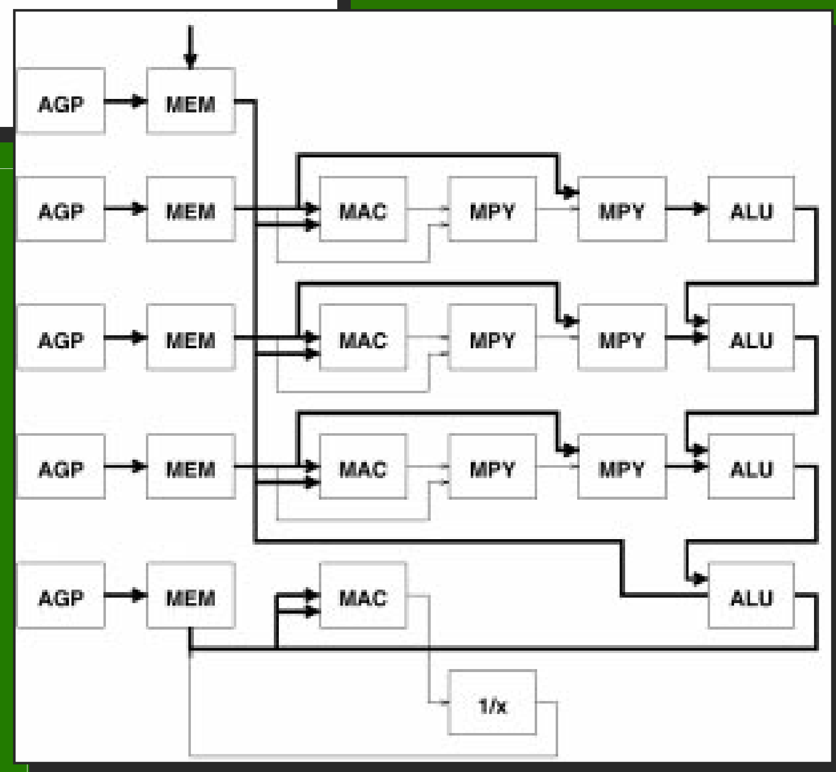
Example: Covariance Matrix Computation

```
for (i=1; i<=length; i++) {
  for (k=i; k<=length; k++) {
  phi[i][k] = phi[i-1][k-1] +
    in[NP-i]*in[NP-k] -
    in[NA-1-i]*in[NA-1-k];
  }
}
```



Reconfigurable Kernels for W-CDMA

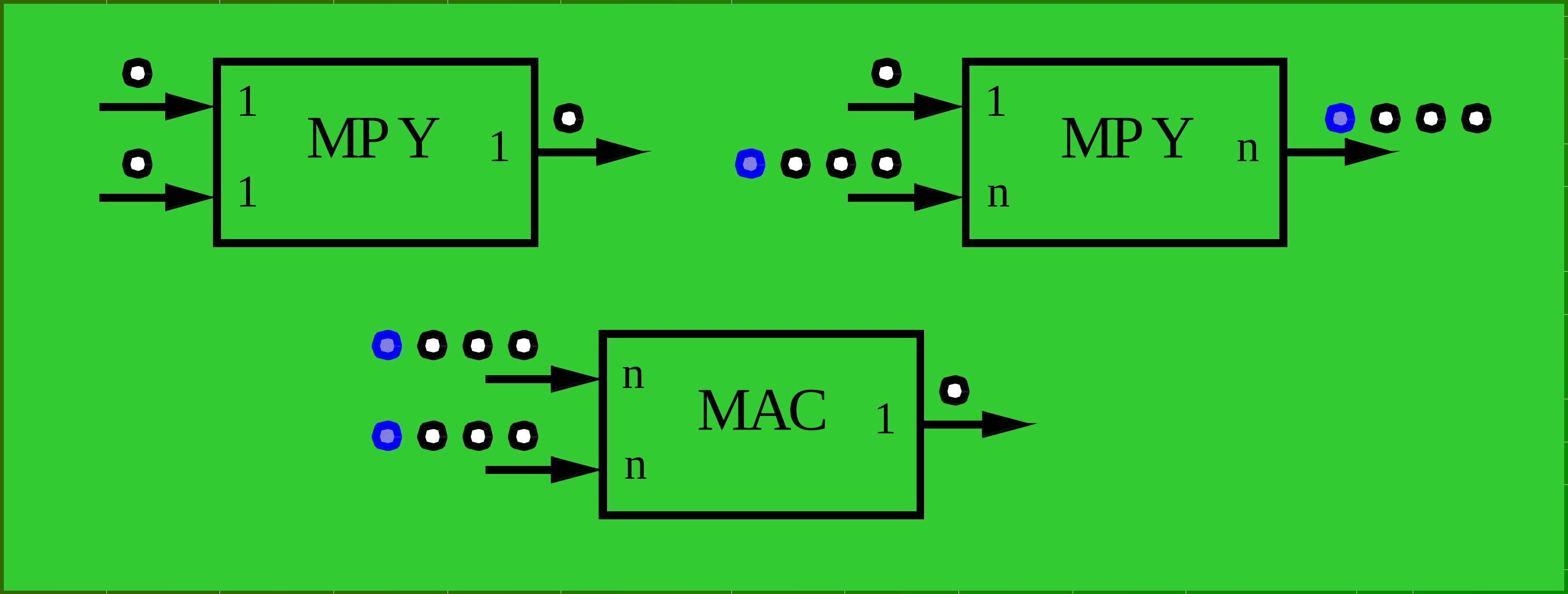




- Dominant kernel $M(M^T X)$ requires array of MACs and segmented memories
- Additional operations such as sqrt(x), 1/x, and Trellis decoding may be implemented using FPGA or cordic satellite

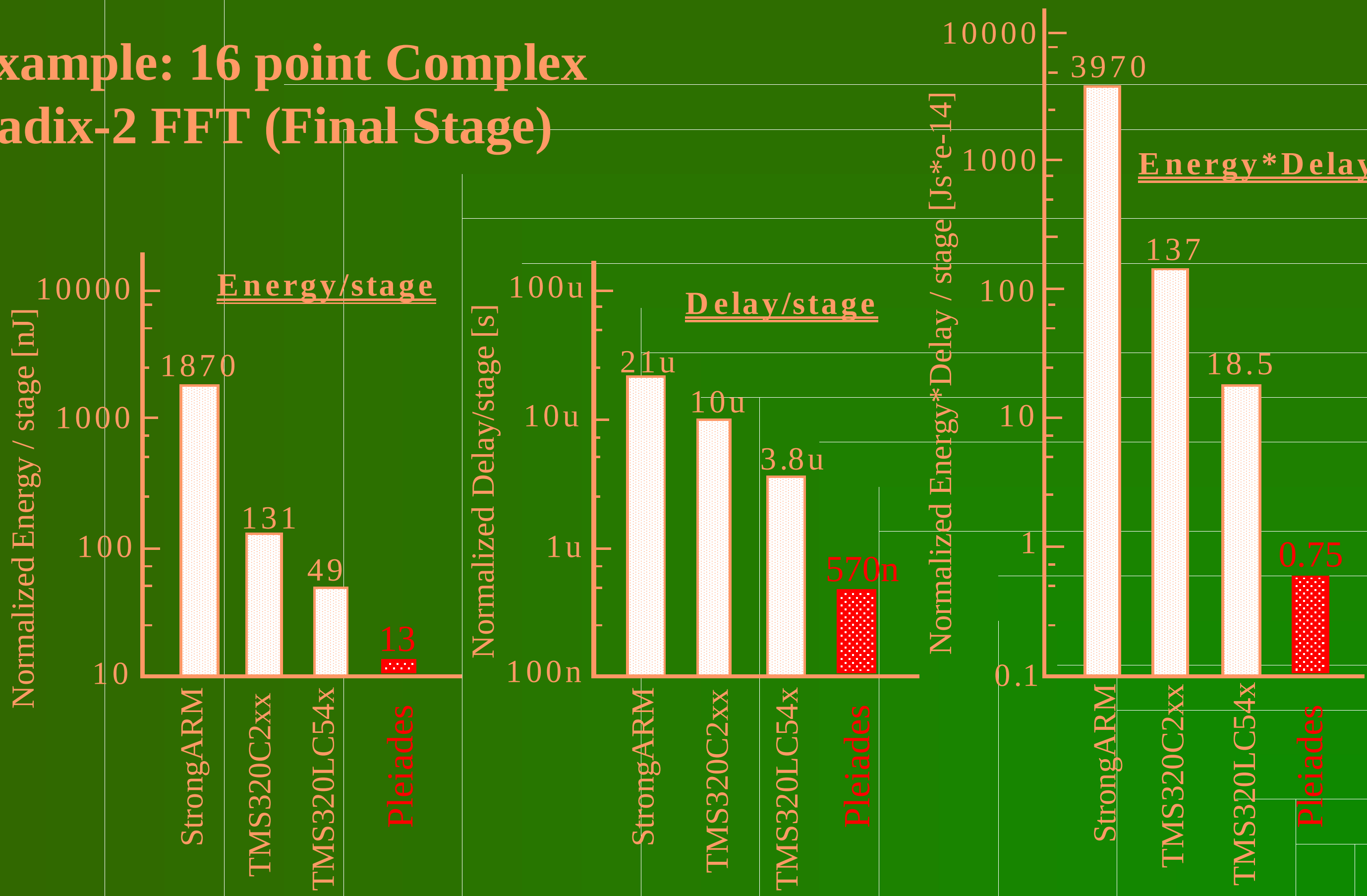
Data-driven Synchronization Based on Finite Streams

- “Smart” satellites able to handle data inputs of different types
- Support of multi-dimensional signal processing
- Introduction of data types: scalars, vectors, matrices



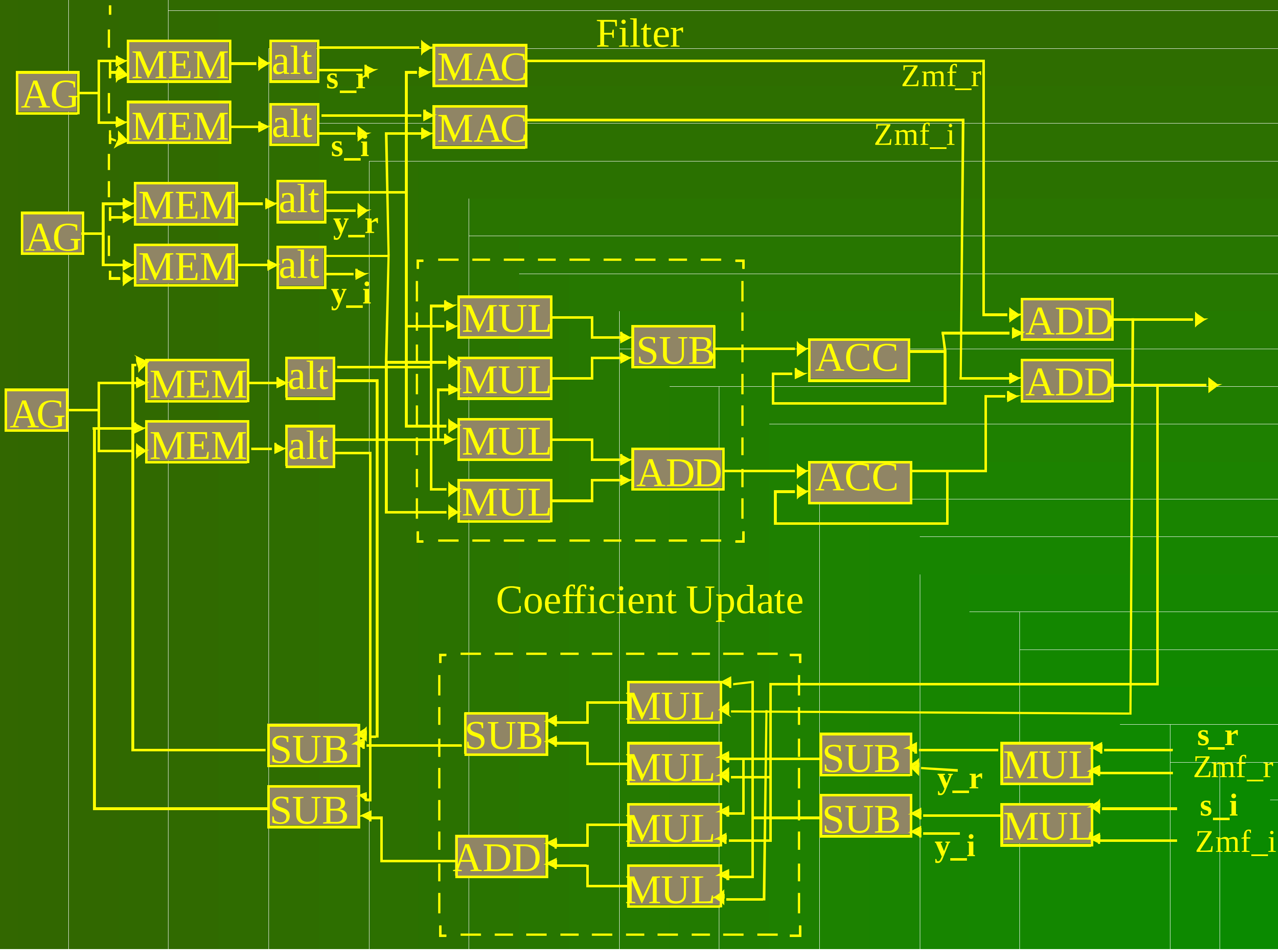
Impact of Architectural Choice

Example: 16 point Complex Radix-2 FFT (Final Stage)



Adaptive Multi-User Detector for W-CDMA

Pilot Correlator Unit Using LMS



Architecture Comparison

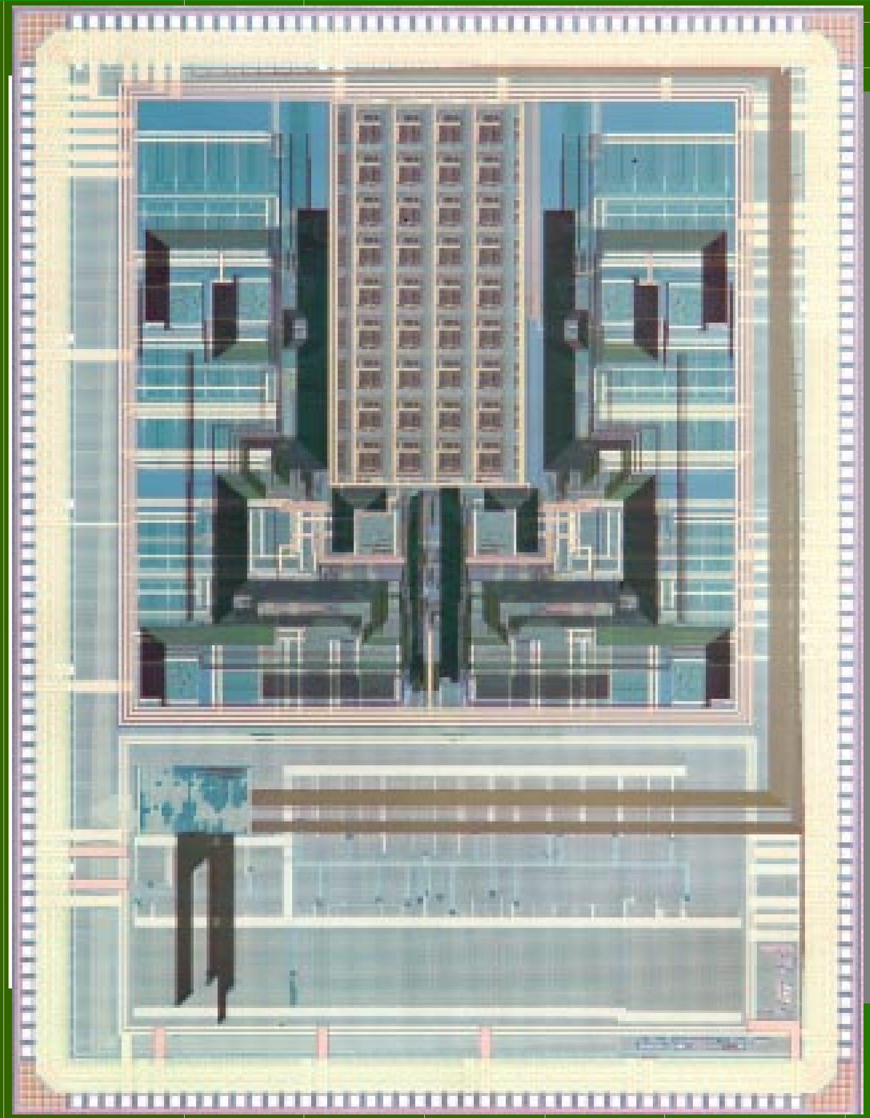
LMS Correlator at 1.67 MSymbols Data Rate
Complexity: 300 Mmult/sec and 357 Macc/sec

Type	Power	Area
TMS320C54*	460 mW	1089 mm^2
Pleiades	18.09 mW	5.448 mm^2
ASIC [Zhang]	3 mW	1.5 mm^2

16 Mmacs/mW!

Note: TMS implementation requires 36 parallel processors to meet data rate - validity questionable

Maia: Reconfigurable Baseband Processor for Wireless



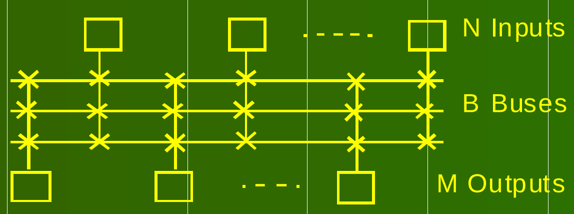
- 0.25um tech: 4.5mm x 6mm
- 1.2 Million transistors
- 40 MHz at 1V
- 1 mW VCELP voice coder
- Hardware
  - 1 ARM-8
  - 8 SRAMs & 8 AGPs
  - 2 MACs
  - 2 ALUs
  - 2 In-Ports and 2 Out-Ports
  - 14x8 FPGA

Fast Design Space Exploration Interconnect Models

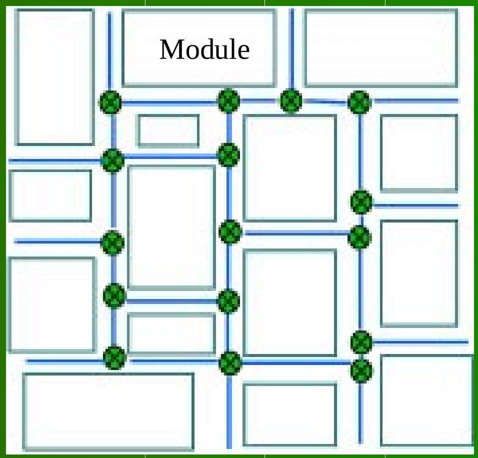
Model:
- **Interconnect energy and delay model**
- **Algorithm mapping**
- **Graph-based place and route**

		dot_ product	vector sum w/ scalar mult.	IIR
Multi-bus		50	50	138
Mesh	Best	[illegible]	[illegible]	[illegible]
	Worst	17.7	14.7	43.4
H_Mesh	Best	**4.7**	**3.8**	**18.8**
	Worst	11.1	10.2	**31.3**

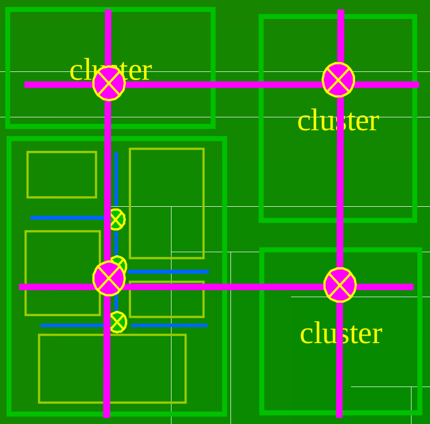
Multi-Bus



Mesh



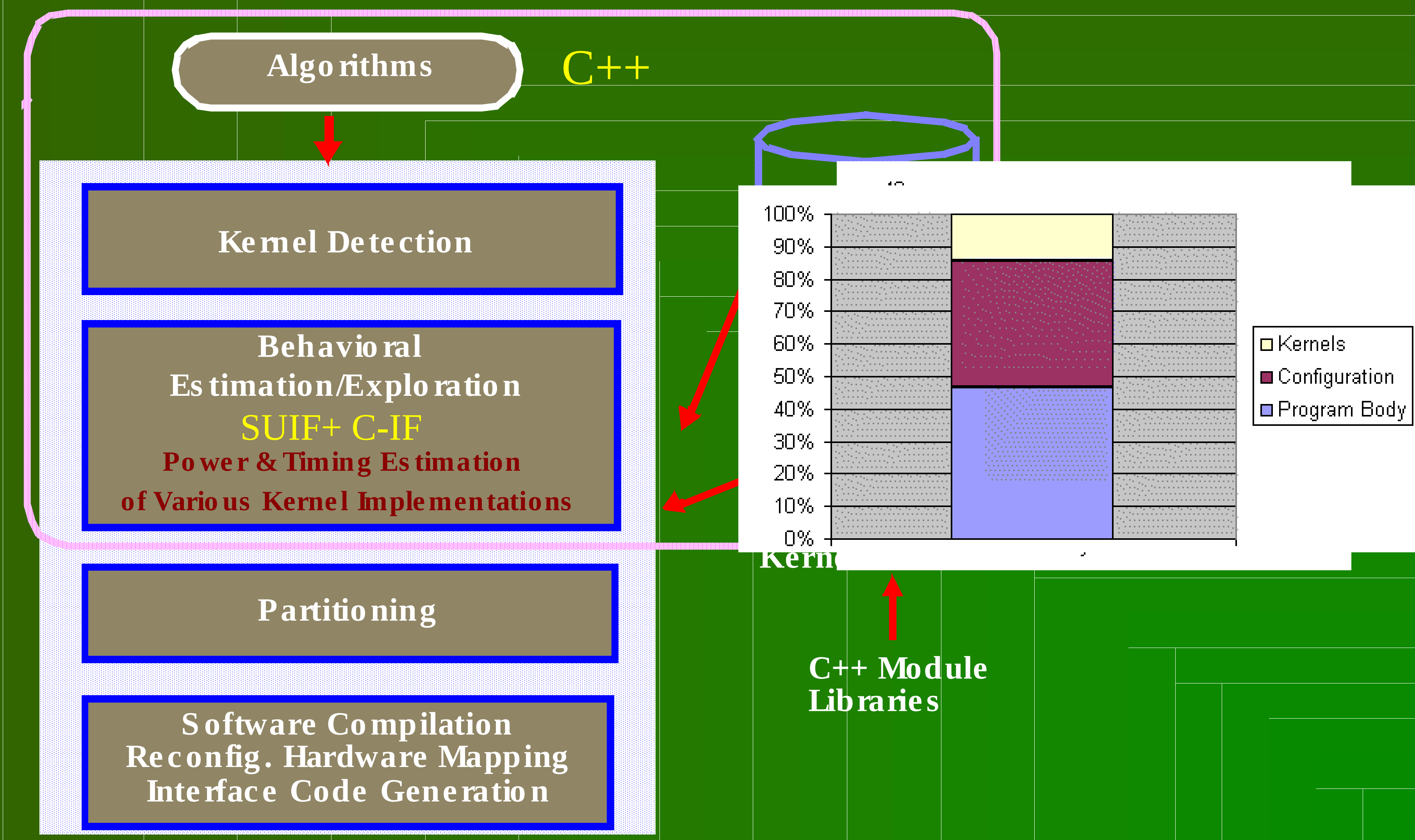
Hierarchical Mesh



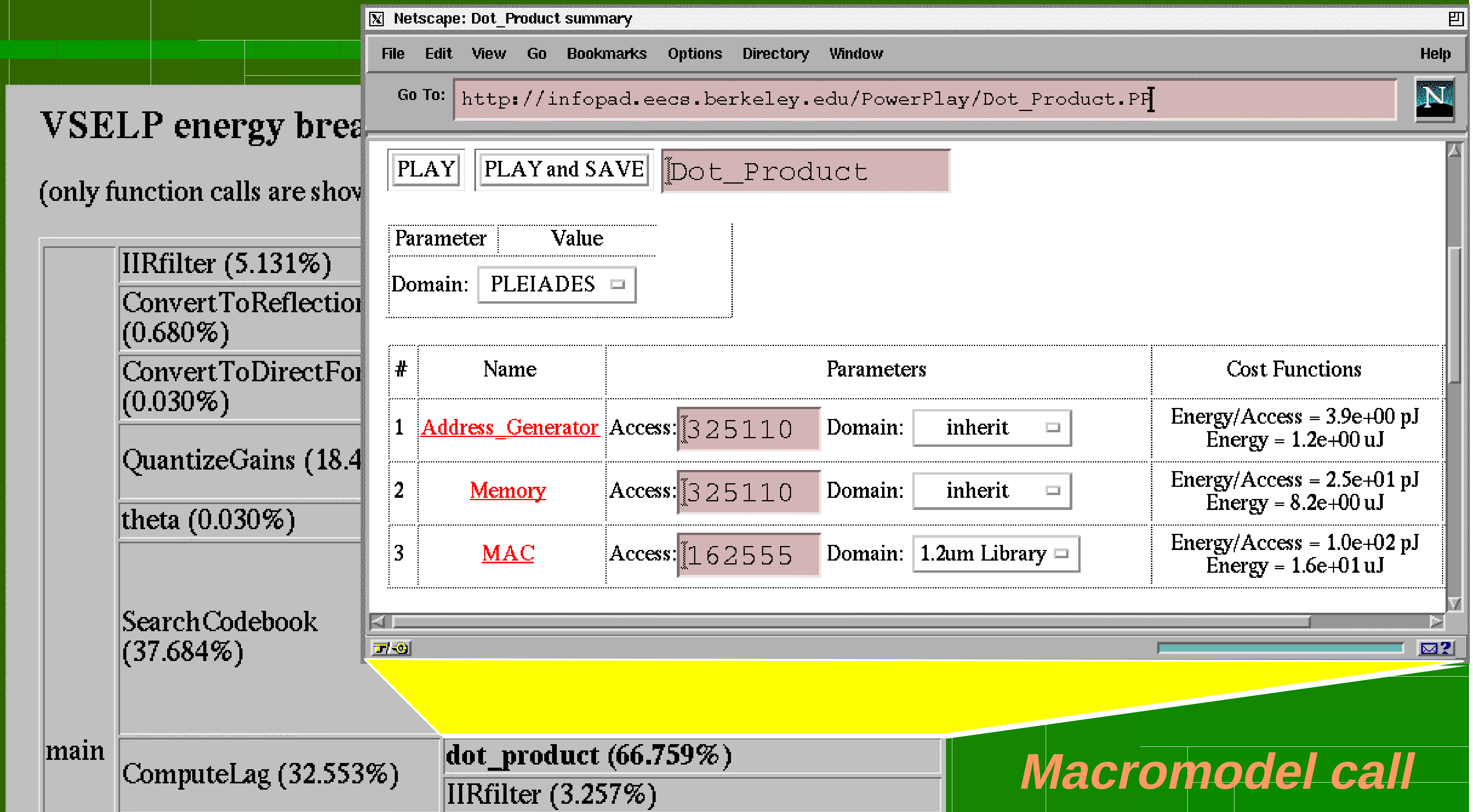
Design Methodology and Flow

- Requires architecture exploration over heterogeneous implementation fabrics
- Should support refinement and co-design of hardware and software, as well as behavior and architecture
- Should consider all important metrics, and present PDA (Power-Delay-Area) perspective

Software Methodology Flow

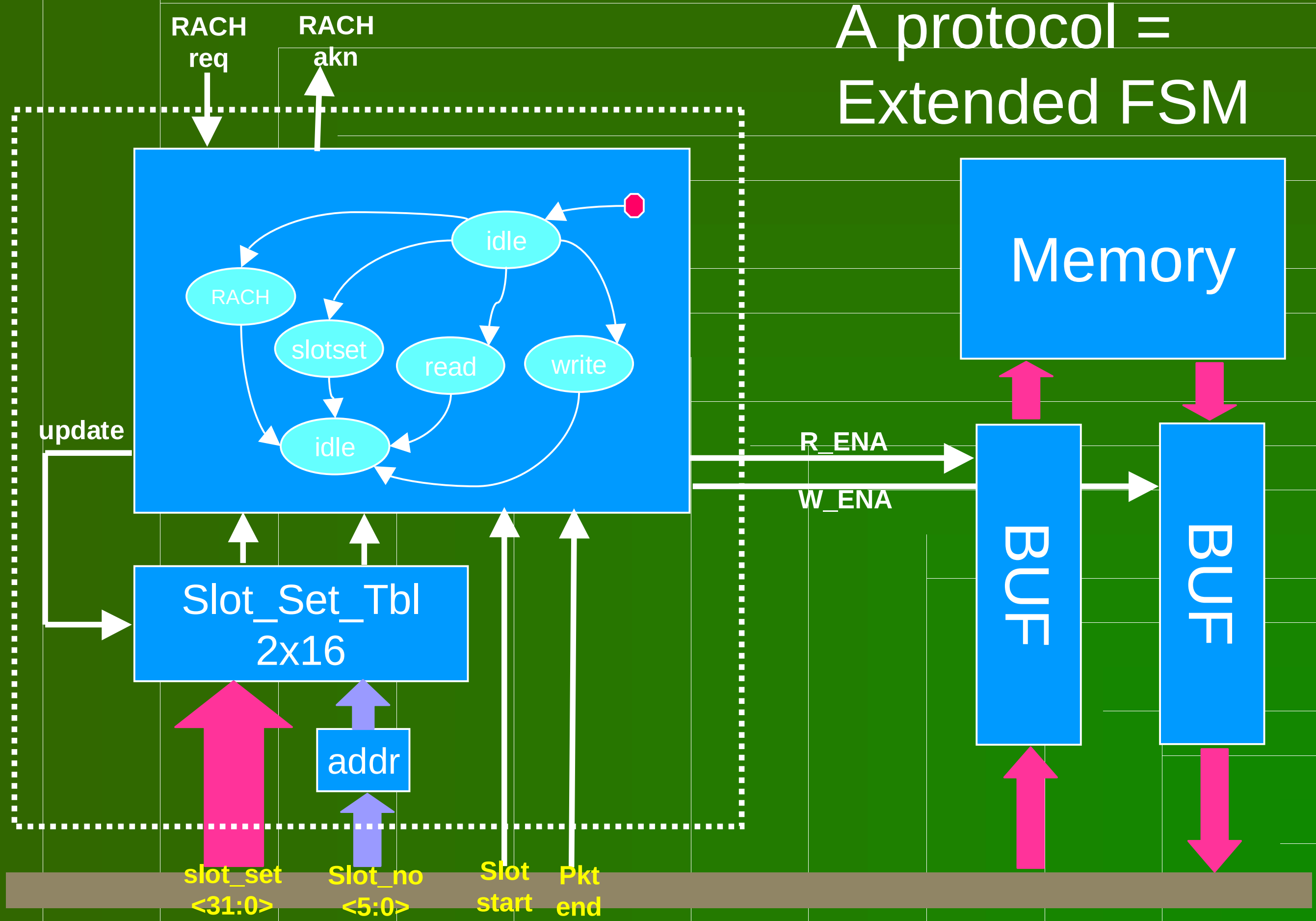


Hardware-Software Exploration



Implementation Fabrics for Protocols

A protocol = Extended FSM



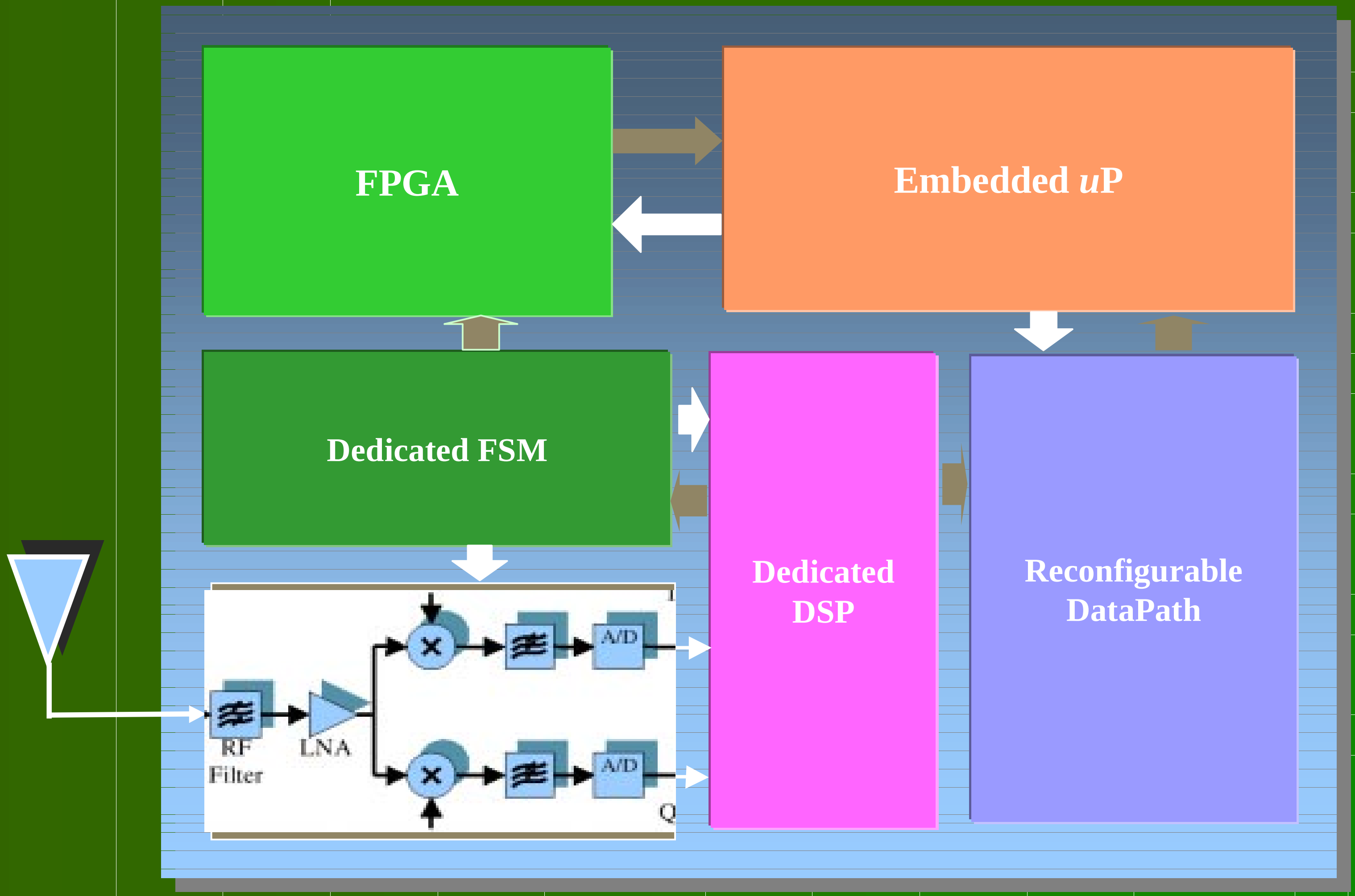
Intercom TDMA MAC

Intercom TDMA MAC Implementation alternatives

	ASIC	FPGA	ARM8
Power	0.26mW	2.1mW	*114mW*
Energy	10.2pJ/op	81.4pJ/op	*n*457pJ/op*

- ASIC: 1V, 0.25 µm CMOS process
- FPGA: 1.5 V 0.25 µm CMOS low-energy FPGA
- ARM8: 1 V 25 MHz processor; n = 13,000
- Ratio: 1 - 8 - >> 400

The Software-Defined Radio



Summary and Perspective

- Technology scaling is redefining the term “complexity”
- System-on-a-Chip fosters renaissance in processor architecture, opening the door for new models and combinations thereof:
  Component and Communication Based Design
- SOC driven by new set of metrics: how to simultaneously optimize flexibility, cost, energy, and performance?
- Reconfigurable architectures provide tantalizing combination of flexibility and efficiency
- Numerous solutions for addressing the data-intensive component of the software-defined radio — the next challenge is control