PHILIPS

The Next Level Platforms: Networks-on-Silicon

Albert van der Werf

Philips Research

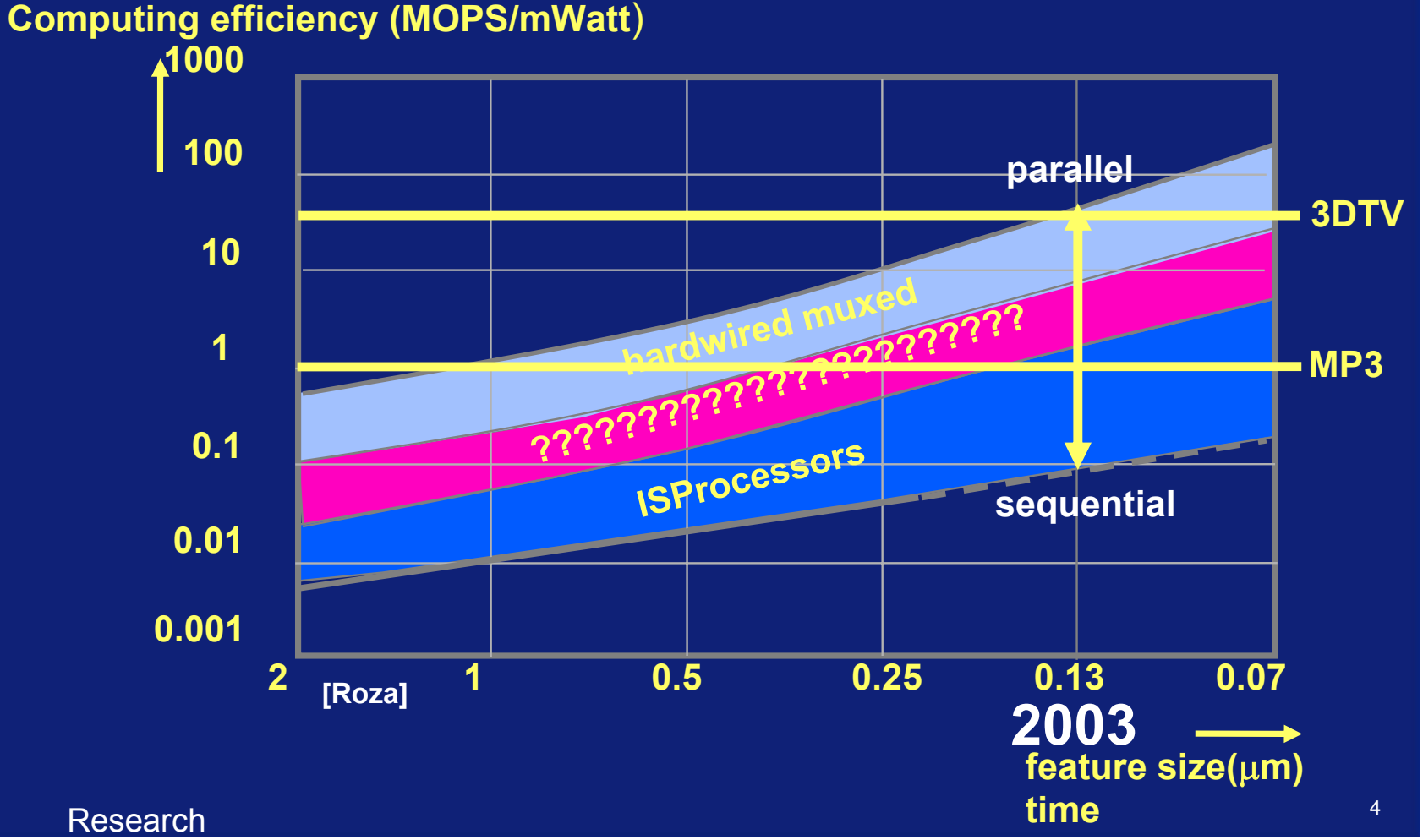
Overview

- Trends
  - Technology
  - Industry
- Networks-on-Silicon
  - Infrastructure
  - Mapping
  - Predictability
- Concluding Remarks

Trends

- Moore's Law is driving us
  - ~60% yearly growth in number of transistors
- Enabling new products with
  - *lower* cost (mm²) and *lower* power (W) dissipation and
  - *higher* flexibility (MIPS) and functional performance (GOPS)

Moore's Law: Computational Efficiency



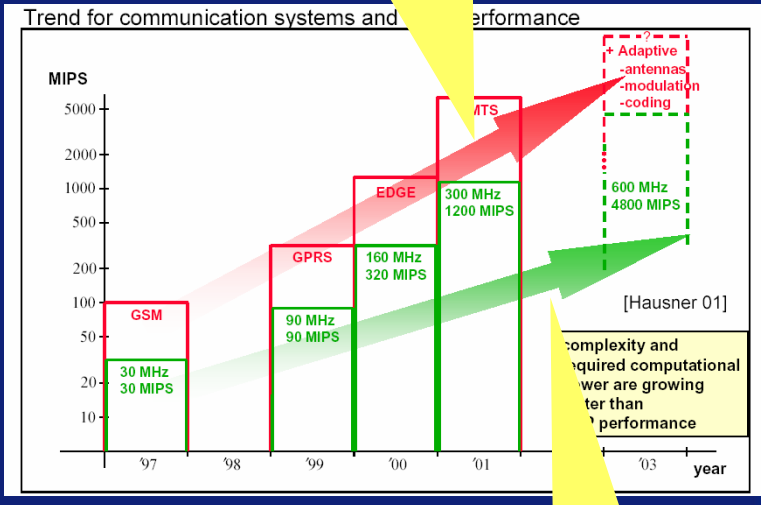
Technology Trends

- **System with one or more CPUs: Moore's law gives us both *cost down* and *feature-up*.**
  - General Purpose processors are eating more of the current application pie (e.g. audio on CPU) – with IP in SW
  - Flexibility becomes more important but also more affordable

- **New systems are demanding more compute power**

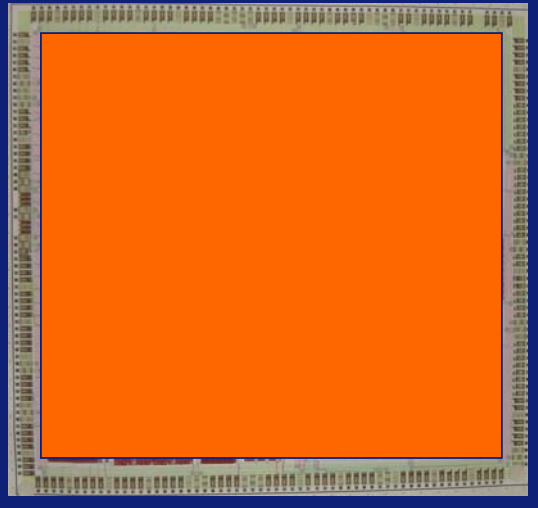
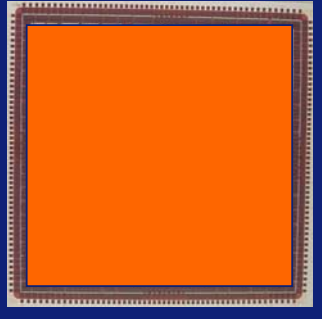
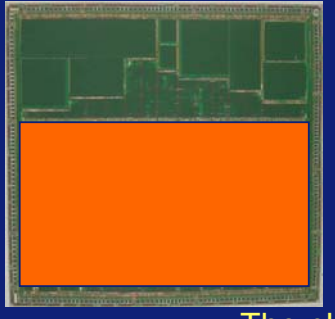
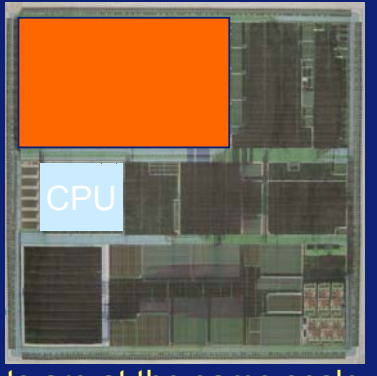
- **Challenge/opportunity: high performance at low power requires (massive) parallelism**
  - System customized compute engines
  - Multi-processor systems
  - Subsystem integration

Required by wireless standards



Offered by conventional DSPs

MPEG2 Storage: IC Roadmap

I.McIC	Empire	Empress	Chrysalis
			
0.50	0.35	0.25	0.18
1997	1998	2000	2002

The plots are at the same scale

The OEM's integration problem

- FM radio
- DVB-H

- GPS
- location-based services

- SIM card
- RFID Payment
- NFC dev pairing

- subVGA→VGA→1.3 Mpixel

Broadcast

Security & identif.

Camera

- IRDA
- USB 1.1
- Bluetooth
- multiple profiles
- WLAN
- VoIP
- UWB?

Connectivity

Software

Display

- TFT, LTPS…
- 320 x 240 pixels
- 18-bit colors
- touch screen
- 3D displays

RF Modem

Host

Audio I/O

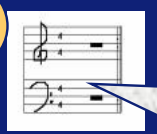
- polyphonic ring tones
- MP3
- MEMS microphones
- echo cancellation

- GSM / GPRS
- CDMA2000
- WCDMA
- UMTS
- 4G?

Flash storage

RAM

Media



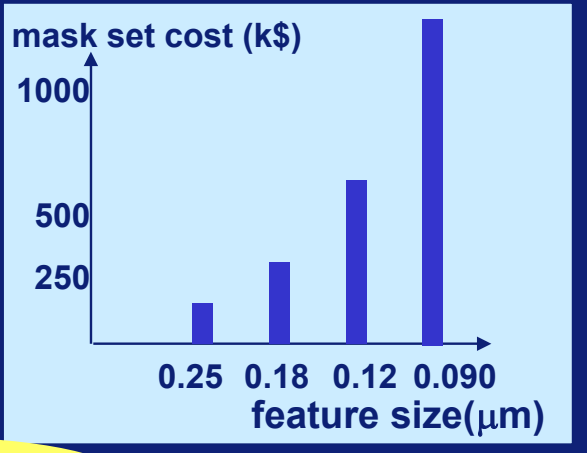
- MP3
- JPEG
- MPEG4
- Voice commands
- 3D accelerators

2002: 8MB
2004: 256 MB
2007: 1 GB
MRAM?

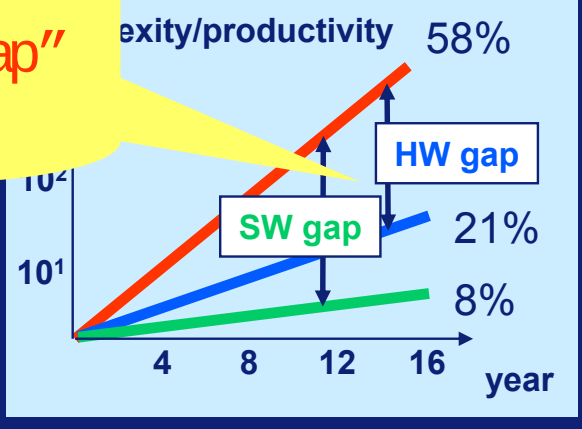
2004: 64 MB

Industry Trends (Observations)

- Mask cost is increasing above a dangerous threshold.
- Design teams are becoming very large (> 100 FTE); design
- Embedded SW content is exploding.



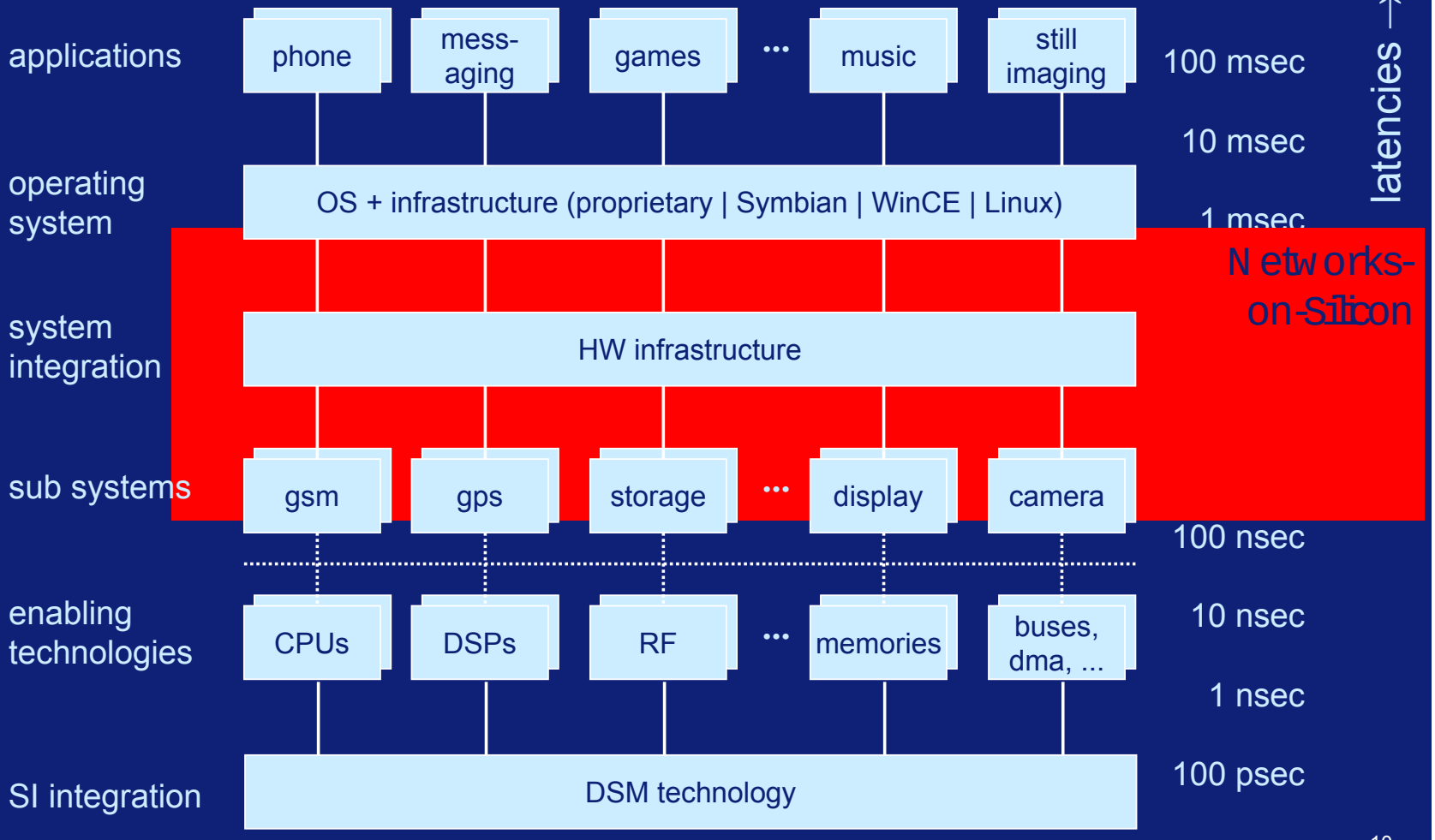
"Cost of design is the greatest threat to continuation of the semiconductor roadmap" [ITRS 2003]



Industry Trends (Approach)

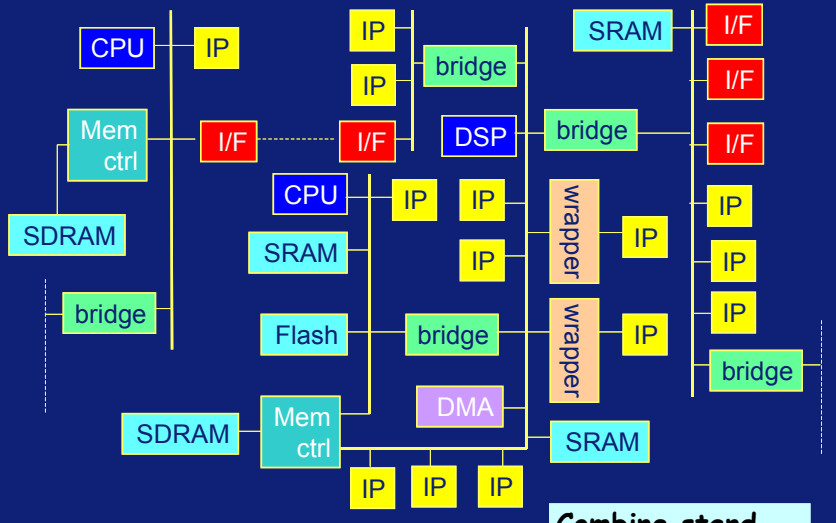
- Industry wide reuse of Intellectual Property (IP) blocks, from many IP suppliers; from captive cores to “open” cores with extensive *Ecosystems*
- Facilitated by platform choices (Nexperia Home/Mobile)
- From Application Specific ICs towards Programmable ICs for a certain Application Domain; product differentiation through SW.

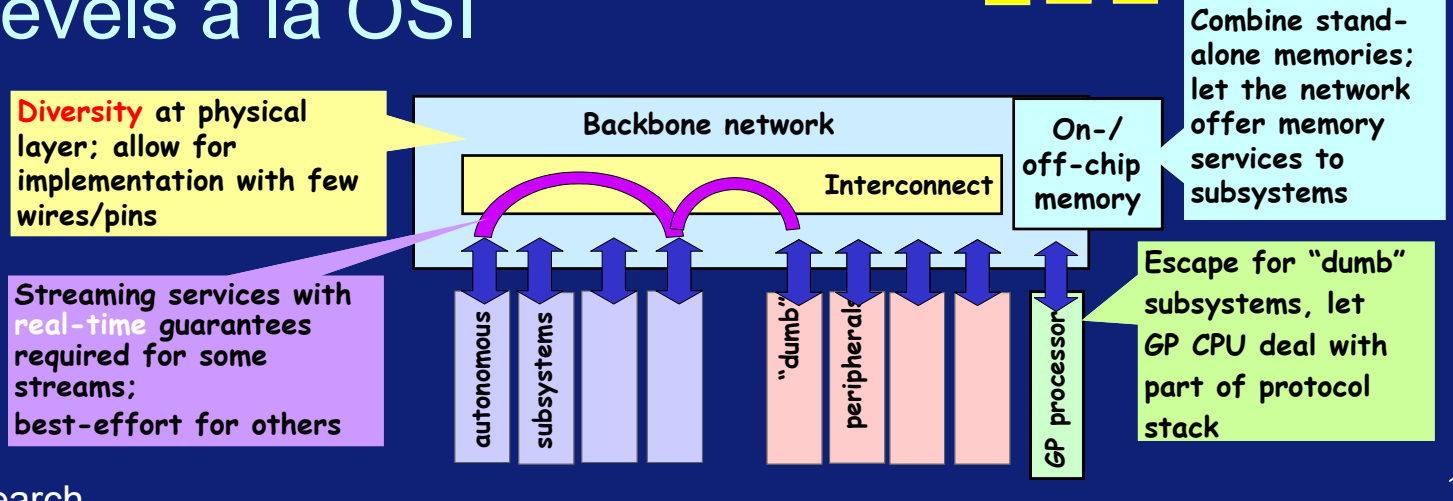
Networks-on-Silicon: Positioning



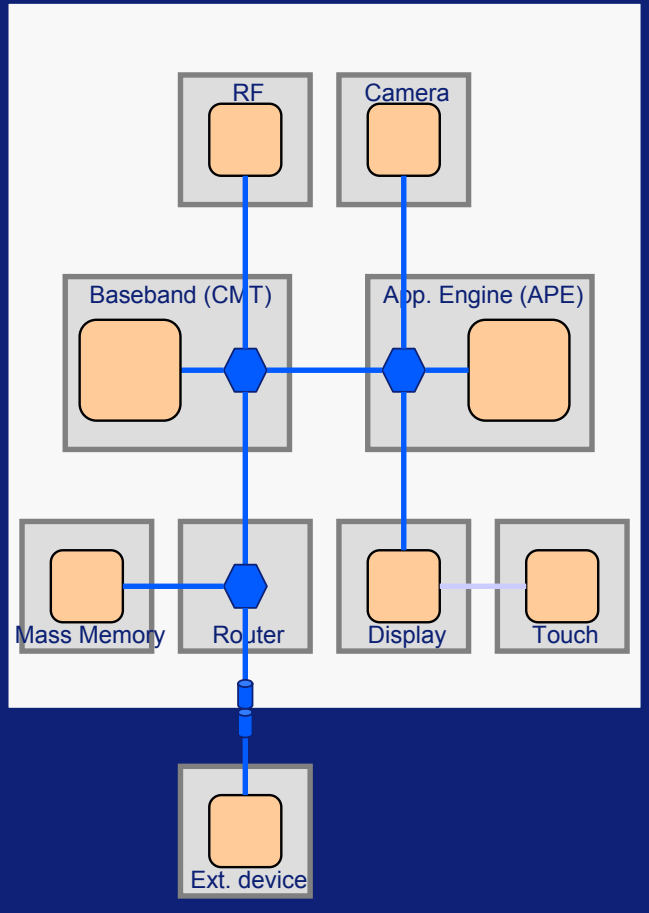
From Busses to Networks-on-Silicon

- IP blocks become subsystems
- Busses become networks offering services at various levels à la OSI





Networks-on-Silicon: Connecting ICs

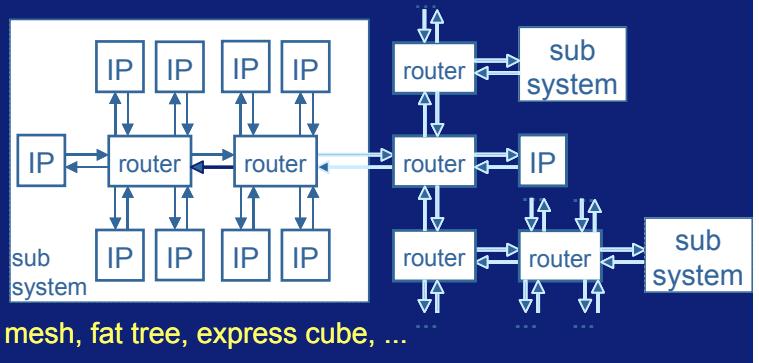


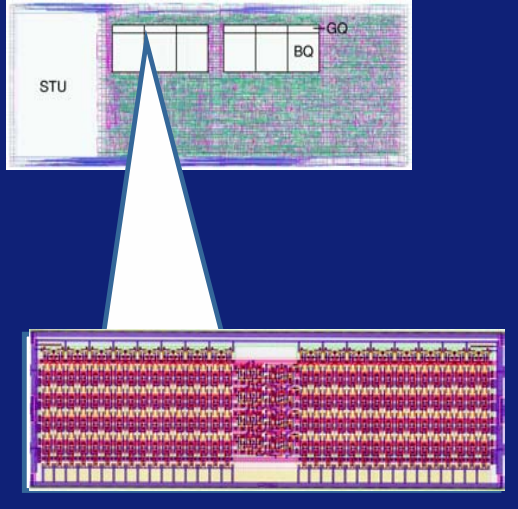
Networks-on-Silicon: Technology

- To solve the transport delay problem and to handle the complexity in future generations of SoC by defining the next generation paradigm for platforms.
  - Infrastructure and related design technology for communication.
  - To organize the communication according to a layered & standardized communication stack (OSI like).
  - To describe systems as networks and efficiently map programs on multiple processors in the network
  - To guarantee (predictable) performance and scalability with plug & play of subsystems.

Networks-on-Silicon: Infrastructure

- Main active elements in network infrastructure:
  - network interface
  - router
    - ATM-like packet-based programming model
    - 52Gb/s aggregate throughput
- QoS
  - Guaranteed Throughput
  - Best Effort
- Supporting interoperability



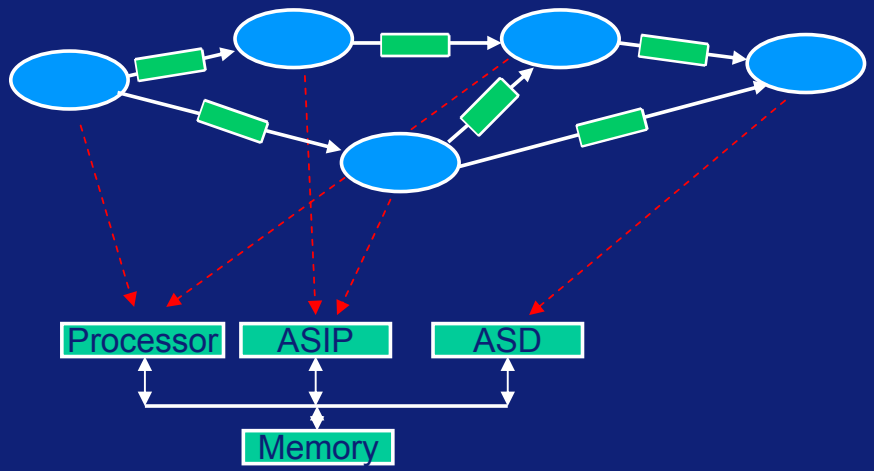


Networks-on-Silicon: Inter-Chip Link

- Standardization of physical interconnect
  - For seamless links between chips
- From SiP up to box level
  - Multiple chips in different process technologies
- Transparent to IPs / subsystems on chips
  - Physical interconnect hidden by more abstract interface
  - Abstract interface supports re-partitioning with a different physical interconnect for on-chip communication
- Variety of classes: from Mbits/s up to Gbits/s

• CMOS	Audio, Compressed Video
• Low swing differential	Standard resolution images
• Embedded clocks	High resolution images

Mapping Applications to Networks



Application
- parallel tasks
- streams

Mapping
- tasks to processors
- FIFOs to memory

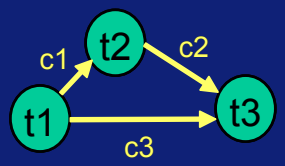
Architecture
- multi-processor
- distributed shared memory

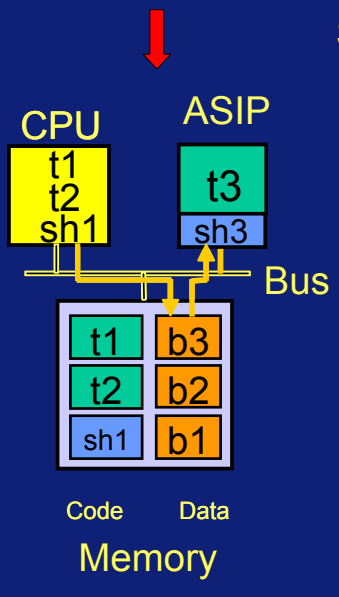
System synthesis: minimal hardware that is required to meet the timing requirements as defined in the specification.

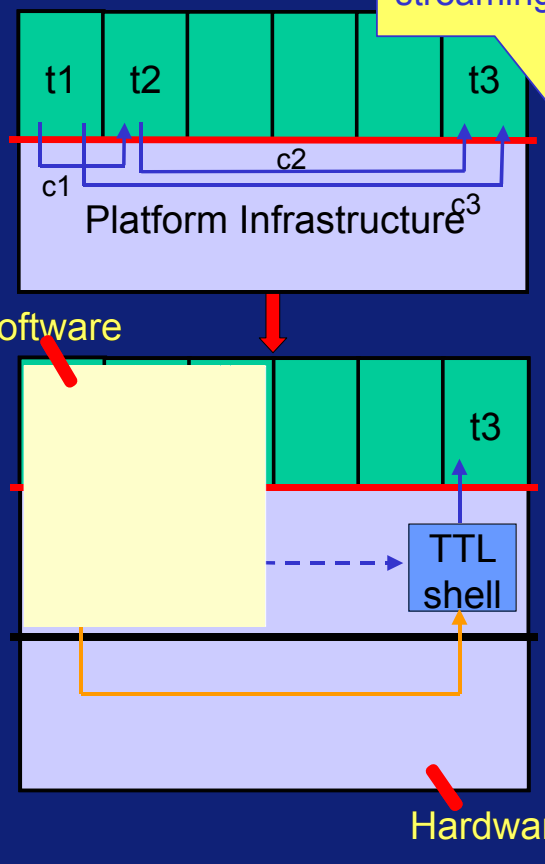
System programming: given a multiprocessor network find a mapping of the application that satisfies the timing constraints.

Networks-on-Silicon: Linking HW & SW

•Goal is to facilitate reuse of tasks through definition of interface for streaming (queues, fifos, channels)







Memory and Streaming

Support for streaming via on-chip memory

- Streaming via off-chip memory:
  - Bandwidth bottleneck
  - Power consumption
  - Pin count
- Streaming via on-chip memory
  - High sync rate is enabler (implementation challenge)
  - Small buffers (in distributed shared memory)
  - Low latencies

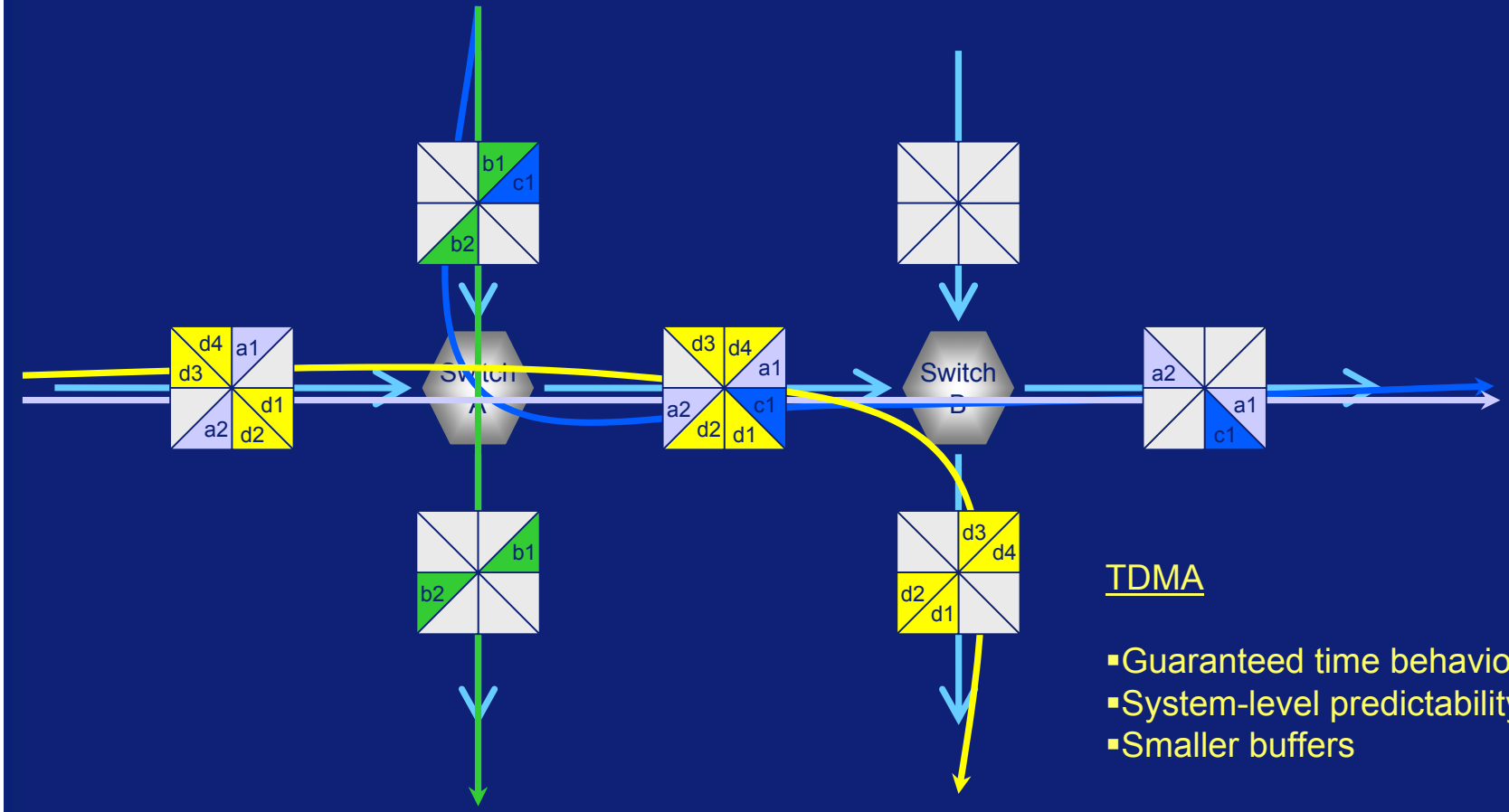
Predictable Multiprocessor Networks

- Meeting the temporal requirements is essential for many consumer systems.
  - Hard real-time:
    - don't miss a deadline (= guarantee throughput and latency)
    - graceful degradation is not supported
    - e.g. channel decoders, picture improvement, audio decoding
  - Soft real-time:
    - there is some diminished value when deadline is missed and value does not increase if result is delivered earlier
    - graceful degradation or fall back must be supported
    - objective is constant Quality of Service (QoS)
    - e.g. video decoders
  - Best-effort:
    - an earlier delivered result is appreciated
    - e.g. web browser

Related Models of Computation

- Kahn Process Networks [Kahn, 1974]
  - concurrent processes communicating through unbounded fifos
  - deterministic communication only
- Communicating Sequential Processes [Hoare, 1978]
  - concurrent processes communicating through unbuffered channels
  - non-deterministic communication through probe [Martin, 1985]
- Dataflow Process Networks [Lee and Parks, 1995]
  - special case of KPN; processes are actors plus firing rules
  - Fire & Exit: each iteration has to be one atomic action. Requires explicit state saving for data-dependent behavior
- Communicating Finite State Machines [Balarin et al.,1997]
  - broadcasting of time-stamped events
  - global notion of time difficult to implement in parallel and distributed signal processing systems

Networks-on-Silicon: QoS



TDMA

- Guaranteed time behavior
- System-level predictability
- Smaller buffers

Multi-processor network nodes

- A number of (smaller processors) communication using a protocol with cache coherence extensions
- Each processor has its own L1 cache and shares an L2 cache with interleaved memory banks
- Escaping from Pollack's rule (exploding power densities for higher performant CPUs)

Concluding Remarks

- From computation centric to communication centric architectures: Networks-on-Silicon will be at the heart of future platforms
- Digital architectures offer a wealth of implementation options. Therefore standardization is key
  - In interfaces, services, and protocols
  - In design environments (including SDK)
- Automated flow with fast performance verification is essential.
- Towards an Open Platform and Ecosystem


PHILIPS