PHILIPS

Networks on Silicon:
Blessing or Nightmare?

Paul Wielage
Philips Research Laboratories, The Netherlands

Electronic systems

- Systems on chip are everywhere

- Technology advances enable increasingly more complex designs

Central Question: *how to exploit deep-submicron technologies efficiently?*

Silicon technology roadmap

- **intrinsic capability** of ICs (transistor count / gate delay) grows with ~ 50% per year (Moore's Law)

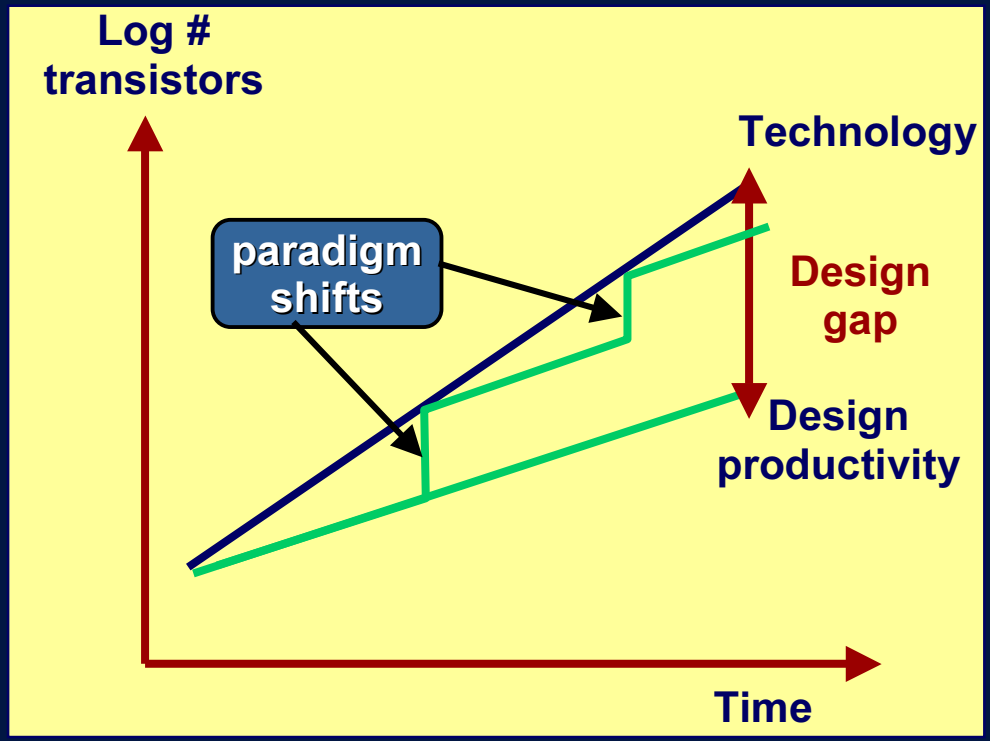
	low power SoC			high performance MPU/SoC		
	2001	2004	2010	2001	2004	2010
gate length (nm)	130	90	45	90	53	25
supply voltage	1.2	1	0.6	1.1	1	0.6
transistor count (M)	3.3	8.3	40	276	553	2212
chip size (mm^2)	100	120	144	310	310	310
clock frequency (GHz)	0.15	0.3	0.6	1.7	2.4	4.7
wiring levels	6	7	9	7	8	10
max power (W)	0.1	0.1	0.1	130	160	218

Source: ITRS 2001

- **power** limits the performance

Design Challenges

Moore's Law is a nice prophesy, but it is hard work to bring into practice



➔ Paradigms shifts in design methodology is the only escape

Design challenges

- **design productivity and design time**

  system level design paradigm shifts:

  - **component based design**
    (IP block re-use)
  - **platform based design**
    (architecture re-use)
  - **networks on silicon**
    (communication-centric view)

- **dynamic and standby power consumption**
  - low swing signaling
  - clock gating / supply switching
  - power management
  - multi-V_t transistors
  - new memory technologies

Component-based design

Design methodology:

- **IC is a composite of heterogeneous IP blocks, preferably reused**
  - e.g. processors, memories, controllers
  - or even whole sub-systems like MPEG encoders / decoders
- **composition by standard interfaces and buses**
  - e.g. virtual component interface (VCI) or AHB bus protocol
- **use wrappers to comply to chosen communication standard**
  - goal: plug and play by means of automatic wrapper generation

Weak point: *how do physical issues influence performance and functional-correctness?*

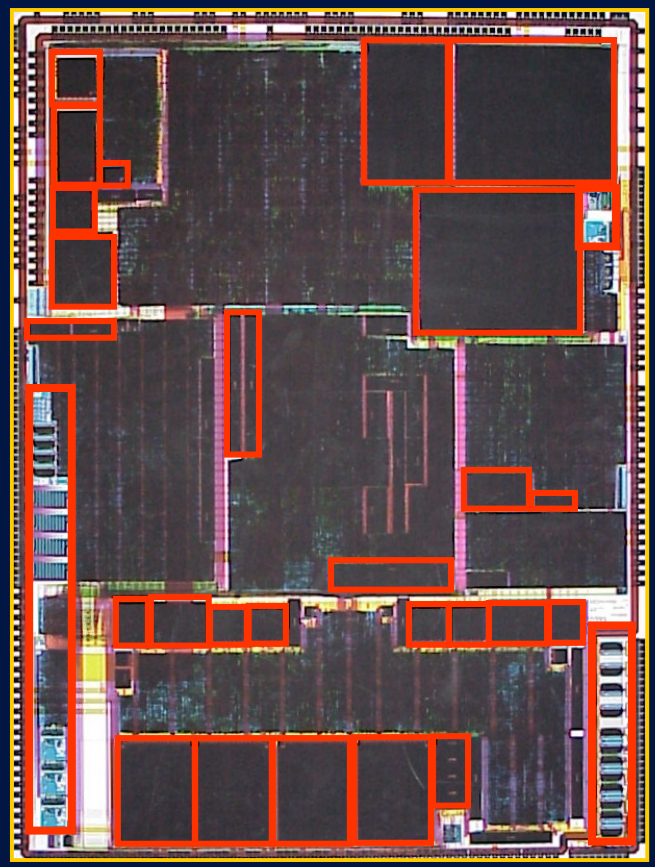
SoC design in practice

IC of heterogeneous IP blocks

- analog
- storage
- computation
- communication

Prospect

- blocks of 50K-100K gates is do-able till 2010
- however in 2010: 1000 < # blocks < 10.000
- increasingly difficult with growing # blocks
- speed and energy are crucial



One-Chip TV Nexperia™ platform

Importance of communication speed

Scaling makes transistors faster but not wires ➔ mismatch

Consequence: performance bottleneck

- faster processors need more data / instructions and more instantly
- highly concurrent processing makes hiding communication latency difficult

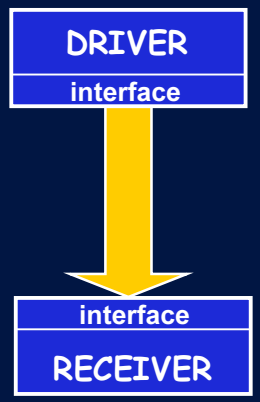
Eventually interconnect will dominate SoC performance

➔ focus shift from computation to communication required

SoC interconnect requirements

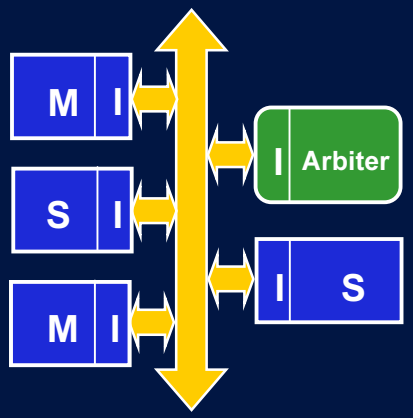
- **scalable**
  - in bandwidth and latency for any system size
- **flexible**
  - multiple applications / configurations, various bit-rates
  - connectivity between each pair of IPs
- **compositional**
  - allow to merge two sub-systems
- **deep sub-micron robust**
  - noise, cross-talk, IR drop, soft errors
  - support multiple clock-domains (e.g. GALS)
- **efficient**
  - cost
  - power

Today's communication solutions



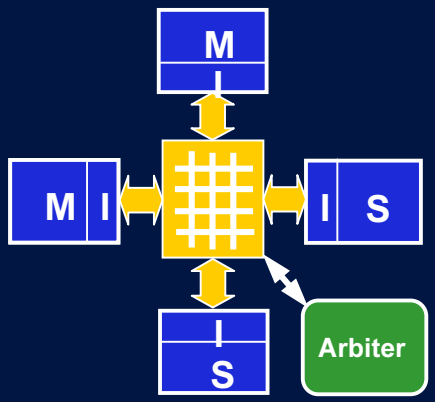
dedicated point-to-point

~~scalable~~
~~flexible~~
~~compositional~~
~~efficient~~



shared bus

~~scalable~~
~~compositional~~



cross-bar switch

~~scalable~~
~~flexible~~
~~efficient~~

In addition, today's communication solutions are not deep submicron proof

On-chip communication

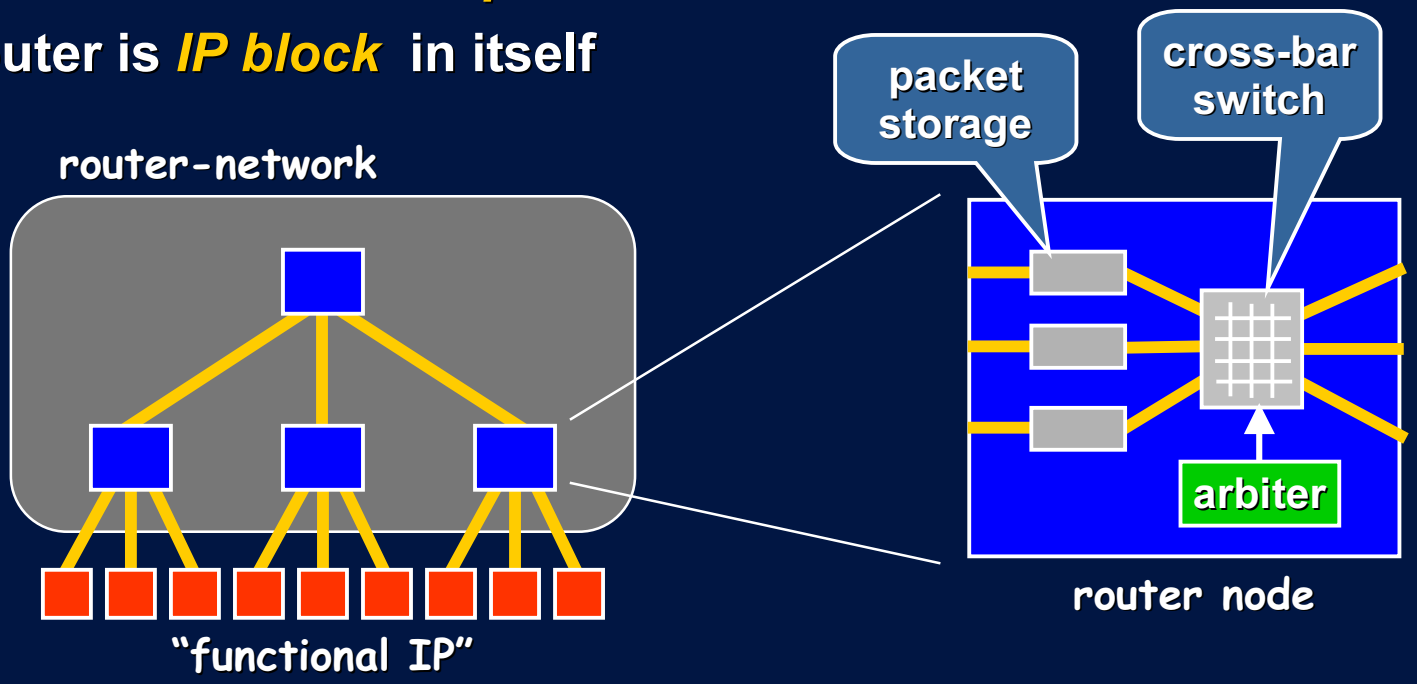
Novel approach:

- **Charles Seitz et.al., "*Let's Route Packets Instead of Wires*", 1990**
- **William J. Dally et.al., "*Route Packets, Not Wires: On-Chip Interconnection Networks*", DAC 2001**
- **Kees Goossens et.al., "*Networks on Silicon: Combining Best-Effort and Guaranteed Services*", DATE 2002**

Networks on Silicon
a paradigm shift in *on-chip communication*

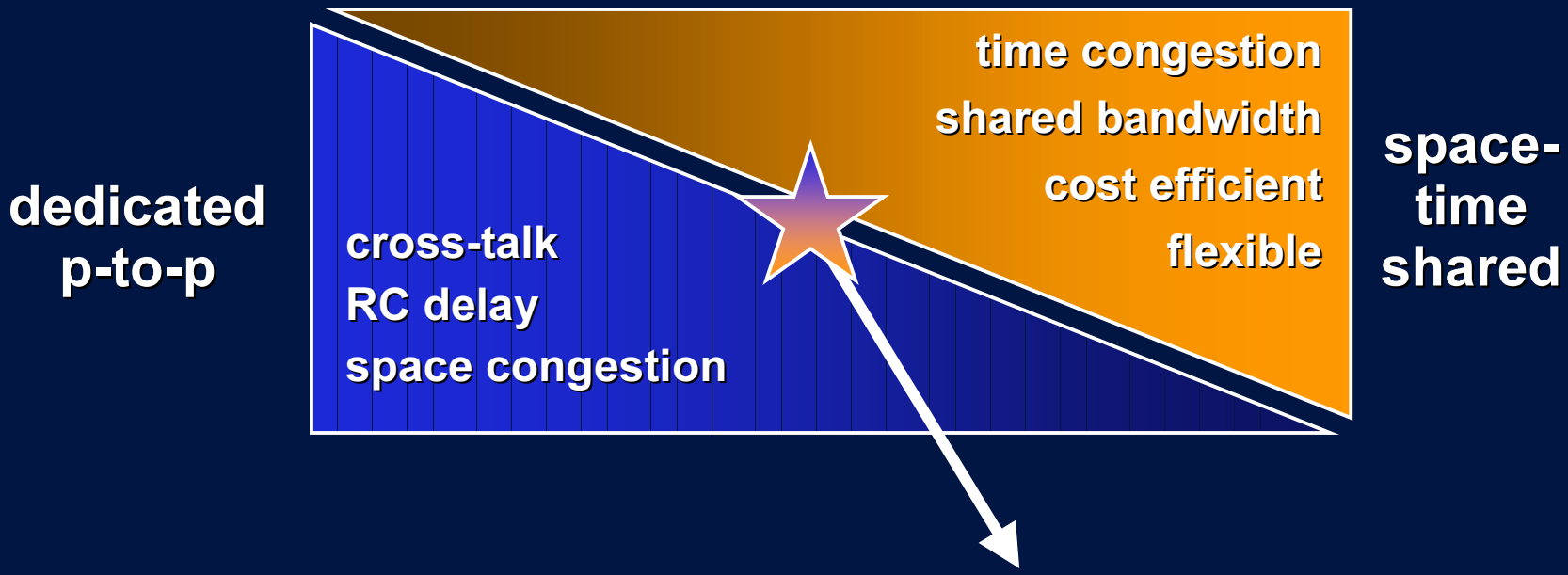
Essence of a NoS:

- all IP to IP communication via ***single network***
- network is ***multi-hop***: routers are point-to-point connected
- routers forward ***data-packets***
- router is ***IP block*** in itself



Wire usage in router-networks

The typical extremes are not favorable for chip-wide interconnect



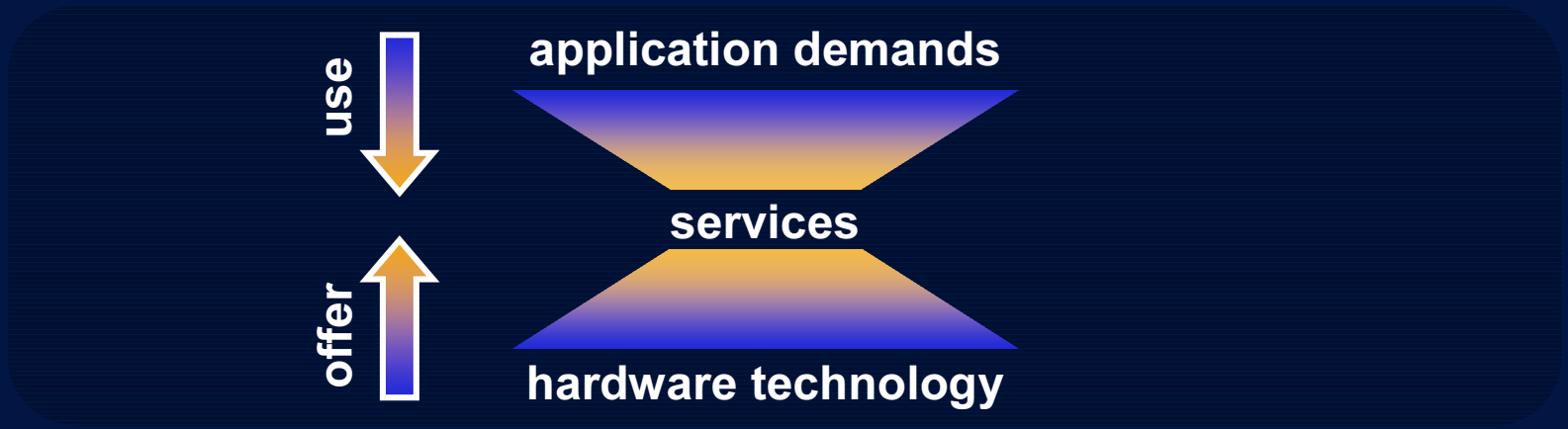
A router-network uses the proper mix: time-shared & point-to-point

- high utilization, few wires
- high frequency, pipelining & repeater insertion possible

Networks on Silicon

Abstract communication services:

- transport
  - uncorrupted, loss-less, without duplication
- performance
  - guaranteed throughput, bounded latency and jitter
  - without guarantees: best-effort
- ordering
  - in-order per transaction, connection, global, …



NoS characteristics

- ✓ **scalabe**
  - #routers, topology, traffic classes
  - size, bandwidth, latency
- ✓ **flexible**
  - every IP is reachable
  - services
- ✓ **compositional**
  - merging two networks is again a network
- ✓ **deep submicron robust**
  - routers are highly reusable: allows for DSM optimization
  - distributed implementation: no global clock required
- ✓ **efficient**
  - high wire utilization ➔ less wires needed

Issues of concern

Two issues of many:

- overhead at interface between network and functional IP
- state synchronization at system level

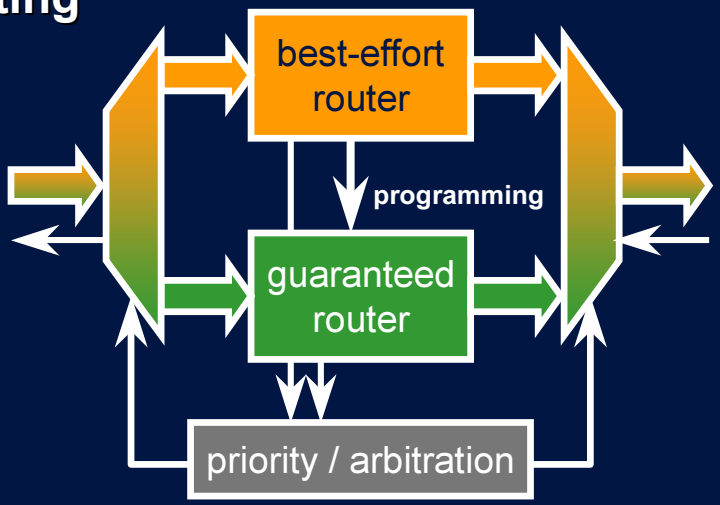
The Æthereal network on silicon

Combination of guaranteed and best-effort services

- guaranteed throughput & latency
  - circuit switching (time division multiplexed)
  - ATM-like connection set up
- best-effort for efficiency
  - virtual output queuing
  - worm-hole routing

Inherently loss-less and ordered transport

No global signals



6 port prototype router

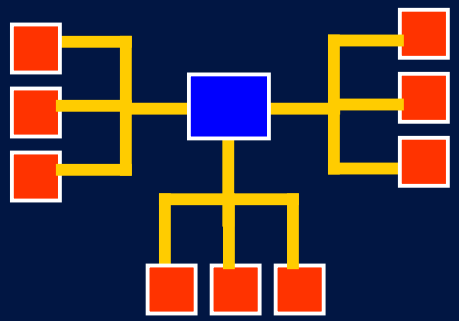
- cmos12
- 6 Kbit queuing
- 512 TDMA slots

Blessing or …

dedicated, p2p

shared bus

Hybrid:
routers + buses



Promising solution!

Summary

1. Technology offers tremendous opportunities

2. High demands from future applications

3. Communication is the problem of future SoCs

4. Networks on Silicon is the solution
   - technology wise
   - design wise