Networks on Silicon: Combining Best-Effort and Guaranteed Services

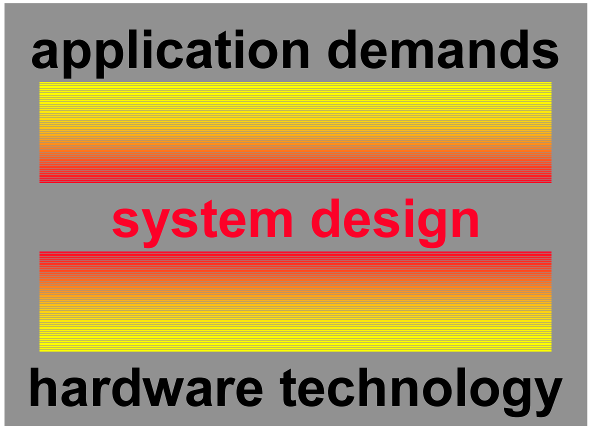
K. Goossens, J. van Meerbergen,
A. Peeters, and P. Wielage

Philips Research Laboratories

D·A·T·E 02

system design problems

- technological problems
  - global clock
  - global timing closure
  - IP partitioning based on floor plan
  - multiple busses & bridges
  - wire congestion
- increasing complexity
  - more dynamic applications
  - number of IP increases
  - diverse communication patterns
- design style problems
  - tightly-coupled components
    - clock, shared (external) memory
  - non-scalable



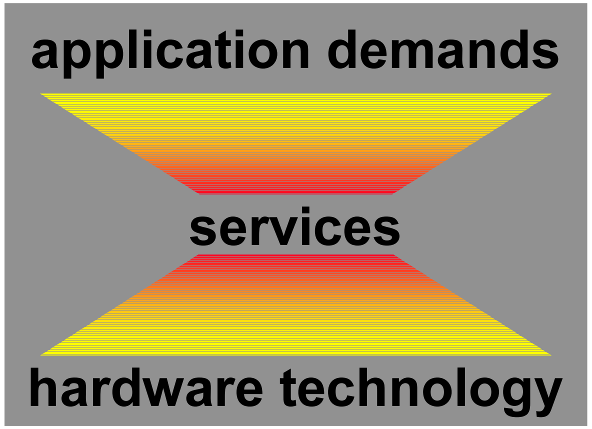
application demands

system design

hardware technology

overview

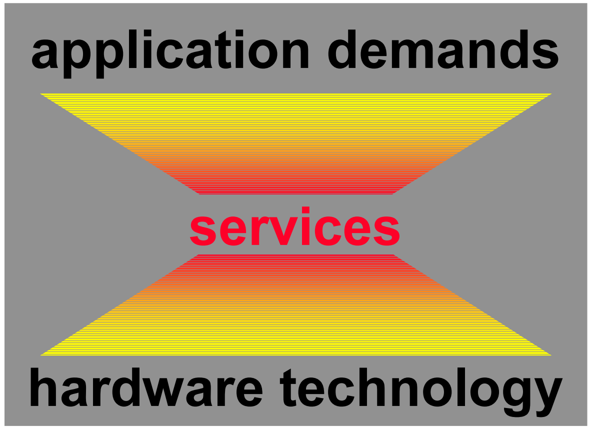
- the essence of a network on silicon
  - decouple computation and communication
  - routers are a hardware solution
  - protocol stack is a software solution
  - both solutions centre on services
- guaranteed services are essential
- the Æthereal network on silicon



1. network on silicon

decouple computation & communication

- **separate intra-IP & inter-IP communication**
- **inter-IP communication**
  - **is explicit by using services**
  - **no longer free & instantaneous**
- **no global time**
  - **GALS clocking strategy**



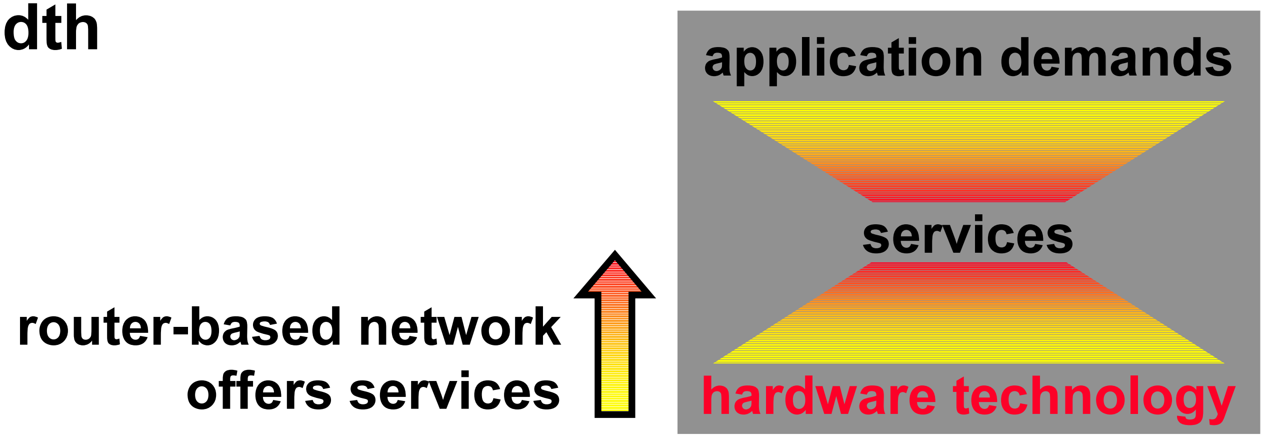
⇒ composable (plug & play)

- **synthesis, lay-out**
- **local timing closure**

Goossens, Meerbergen, Peeters, Wielage
Philips Research Laboratories

2. router-based hardware solution

- communication hardware is now re-usable IP
- offer different kinds of communication **services**
- share wires
  - less wire congestion
  - dimension for average, not worst traffic
- scalable (no central resources)
- composable
- high bandwidth



3. protocol stack software solution

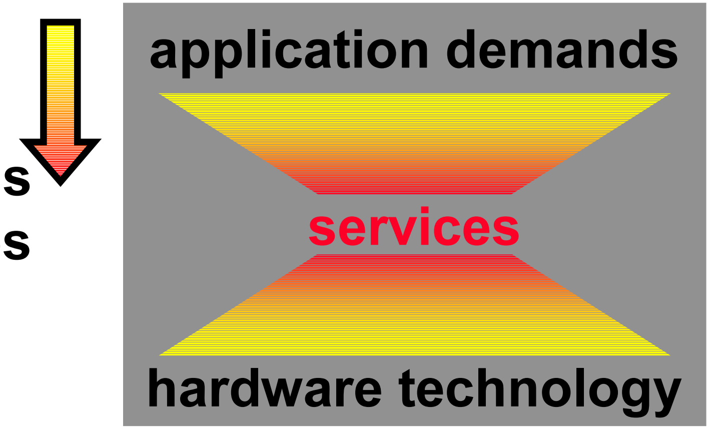
- application
- presentation
- session
- transport

- application diversity
- network independent
- peer to peer

- network
- link
- physical

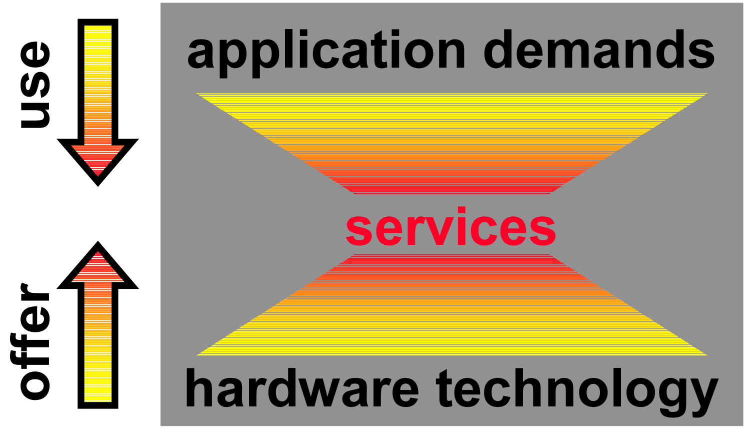
- network on silicon
- network dependent

protocol stack is
based on services



4. communication services

- uncorrupted data transmission
- loss-less transmission, duplication
  - percentage, per transaction type, ..
- data arrival order
  - per connection or transaction type, global, ..
- throughput guarantee
- latency bounds
- jitter bounds
- traffic classes/priorities
- etc.



Goossens, Meerbergen, Peeters, Wielage
Philips Research Laboratories

classes of services

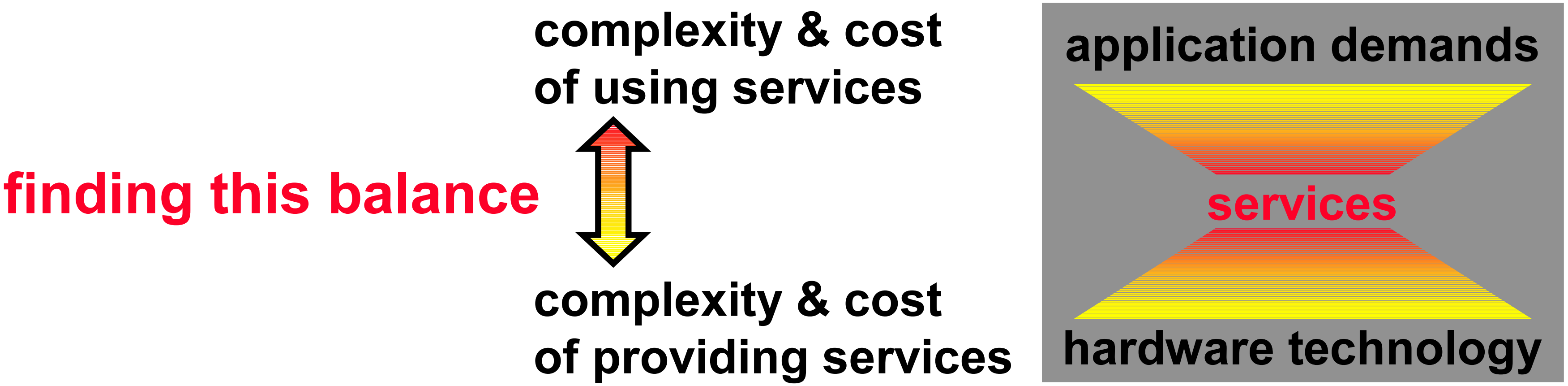
- **guaranteed**
  1. **predictable, dependable performance** **V**
  2. **shared resources must be managed** **X**
  3. **services must be requested (negotiation),**
     and are either granted or rejected **X**
- **best effort**
  - **no resource management / QoS** **V**
    - **may lead to higher performance**
  - **unpredictable performance** **X**
    - **if you know nothing of network**
      or else IP are network dependent

guaranteed services are good

1. good design practice
   - services make assumptions on partners explicit
   - service contract limits possible interactions
2. composable method
   - services & design of different IP are independent
   - no interference (cf. caches)
3. robustness
   - communication failure of IP limited to negotiation
   - no overload of communication resources
     - local IP failure, not global system failure
4. resource management
   - QoS requires observation & predictable steering of communication resources
   - independent of network architecture

services and their costs

- **guaranteed services**
  - more (static) information allows firmer guarantees
- **best-effort services**
  - shift responsibility of predictability to upper levels
  - not always possible to recover predictability
- more services means harder to offer but easier to use
  - complexity can only be shifted, never removed



complexity & cost of using services

finding this balance

complexity & cost of providing services

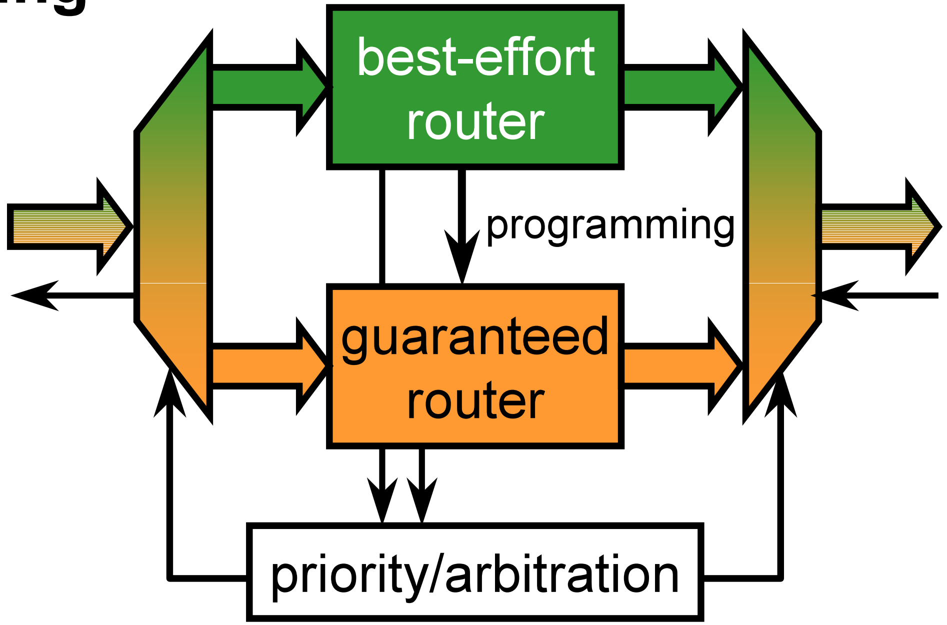
application demands

services

hardware technology

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



Goossens, Meerbergen, Peeters, Wielage
Philips Research Laboratories

conclusions

1. decouple computation & communication
2. networks on silicon
   - routers are good hardware architecture
     - structure the wiring & clocking problem
     - enable local timing closure (composable)
     - are re-usable communication IP
   - provide basis for software architecture
     - protocol stack
     - diverse applications on single architecture
   - centred around notion of services

- offering guaranteed services
  - complicates hardware architecture
  - eases system design and programming