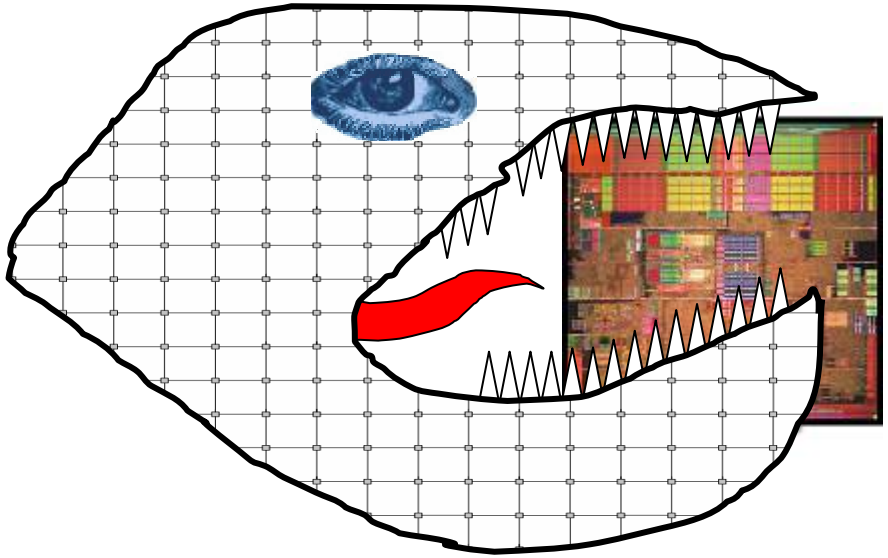
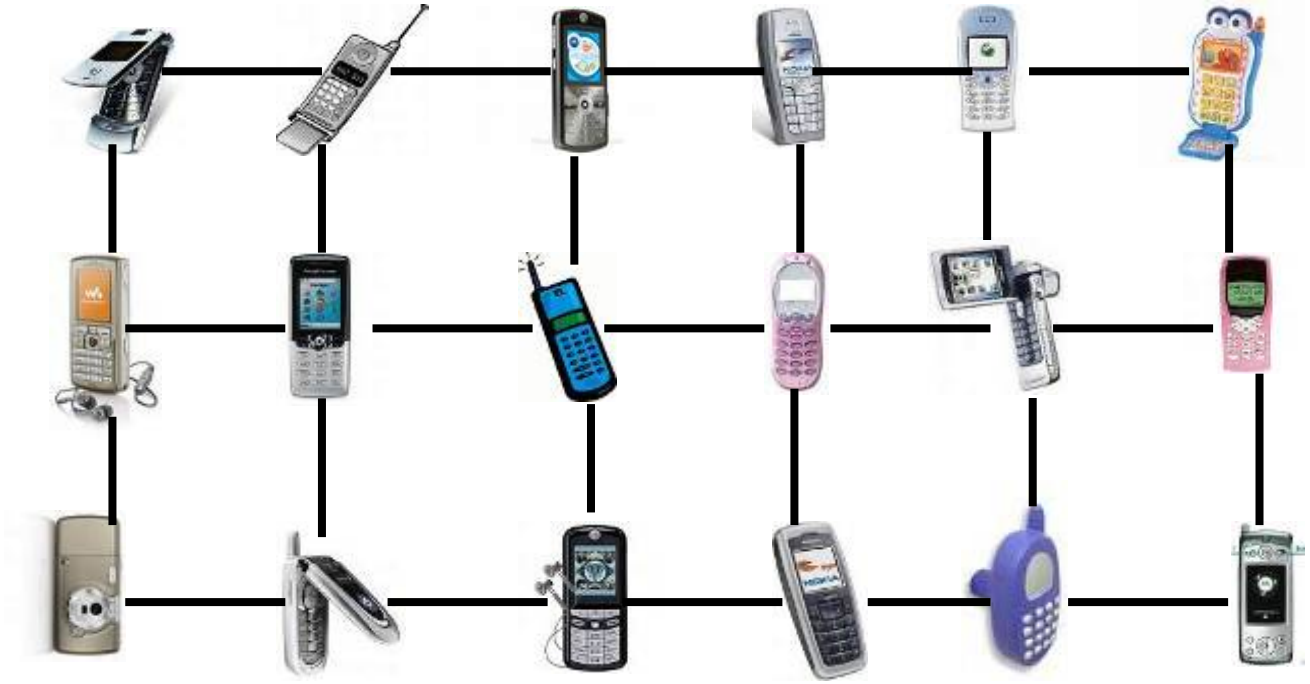
Will the *NoC* swallow the *SoC*?



Ran Ginosar
Technion—Israel Institute of Technology
Haifa

NoC = Network of Cellphones

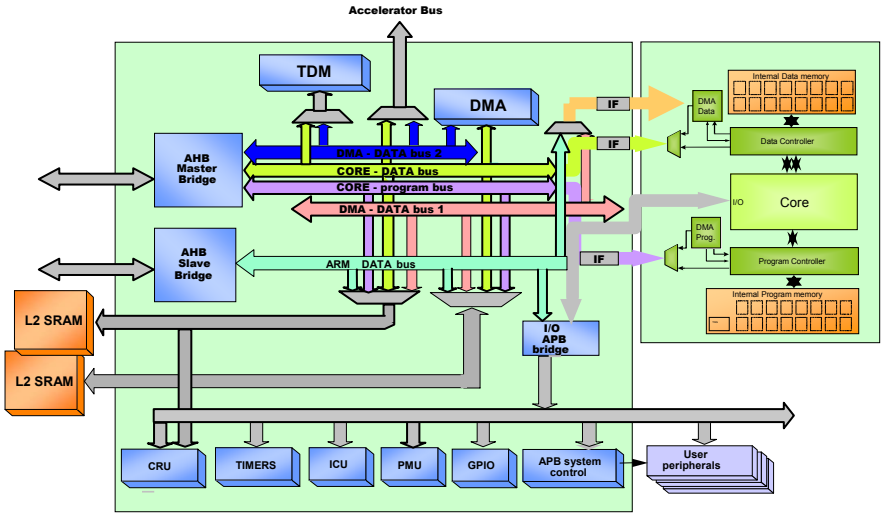


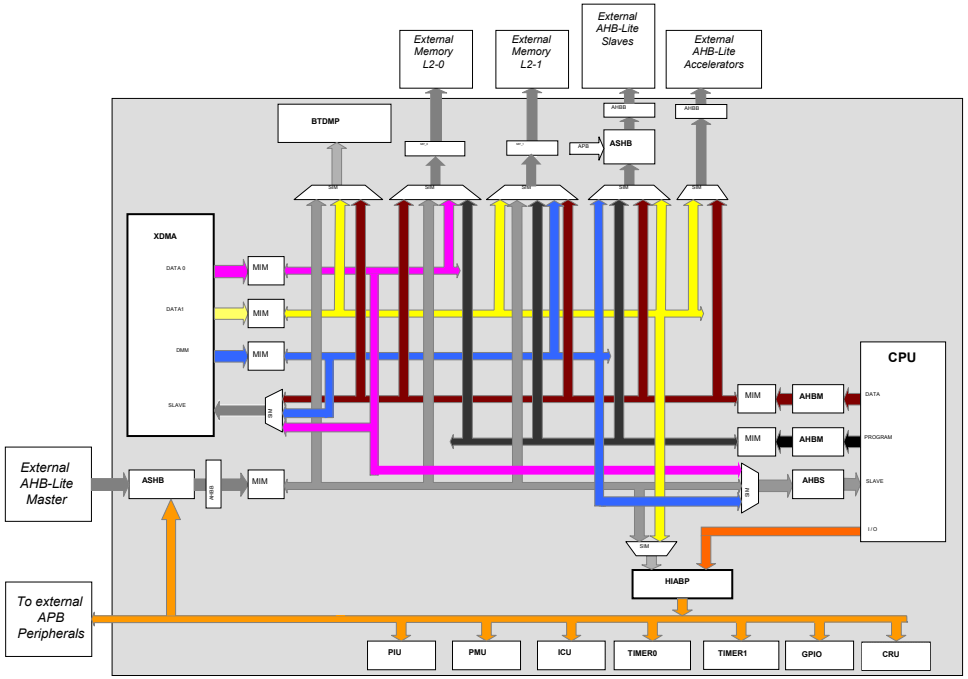
The new IEEE NoC Symposium

- May 7-9, 2007, Princeton
- Paper deadline December 1, 2006
- Scientific
- Large PC, all papers reviewed, look to accept 21-24 papers
- Excellent program!

NoCs International Symposium on Networks-on-Chips

Pre-NoC era: Spaghetti a-la Amba

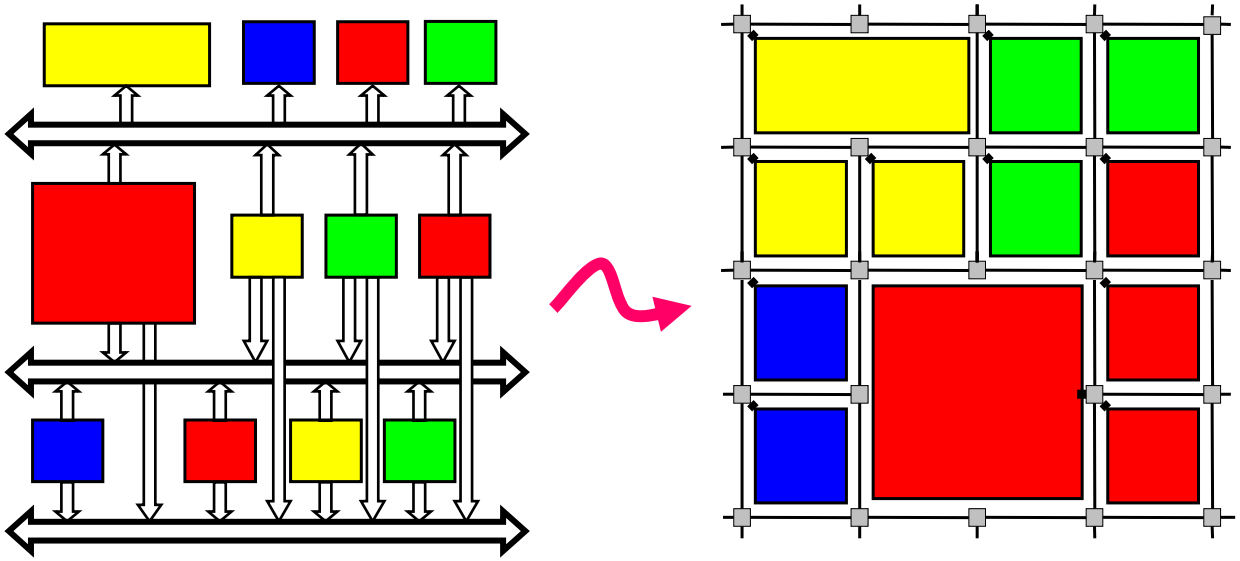




I

II

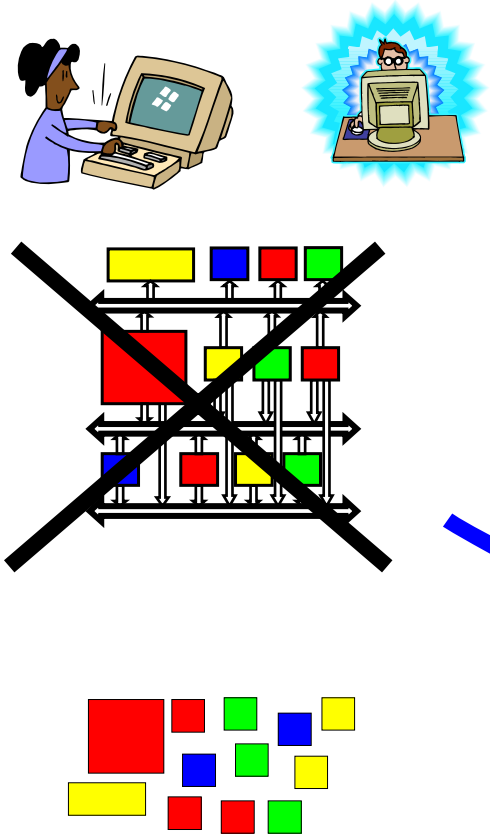
The NoC Paradigm Shift



- Architectural paradigm shift
  - Replace the spaghetti by a <u>customized</u> network
- Usage paradigm shift
  - Pack everything in packets (replace bus access semantics)
- Organizational paradigm shift
  - Confiscate communications from logic designers
  - Create a new discipline, a new back-end responsibility
    - Already done for power and clock grids

NoC as an Org Shift

Logical Design Team



Physical Design Team

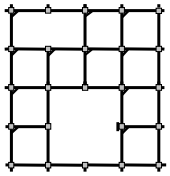
Placement

Routing

Clocking

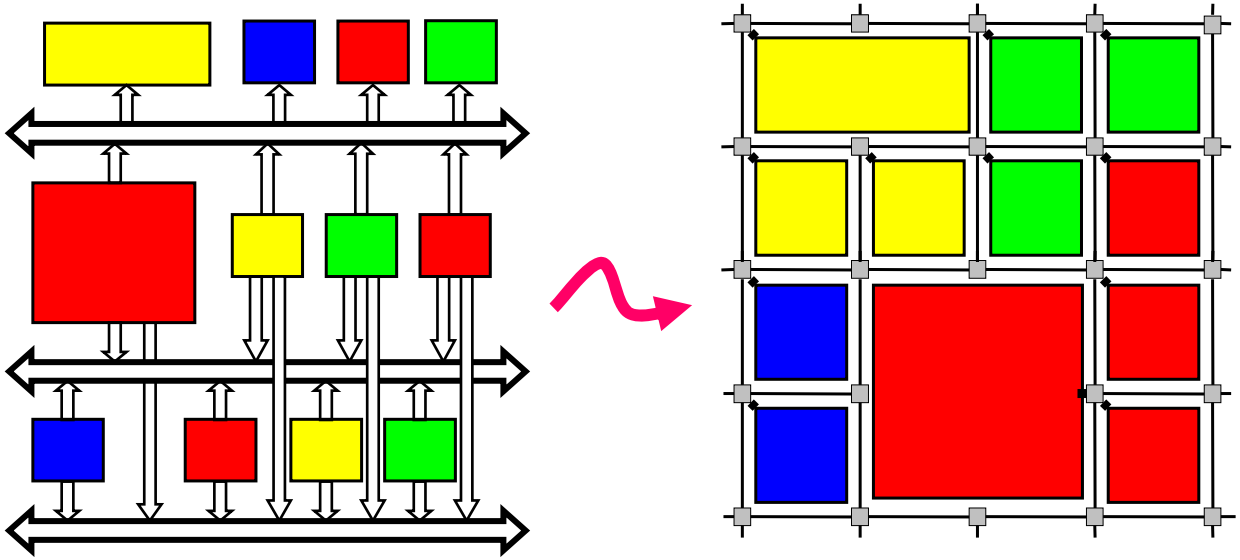
Timing

Power



NoC (EDA + Experts)

Why go there?

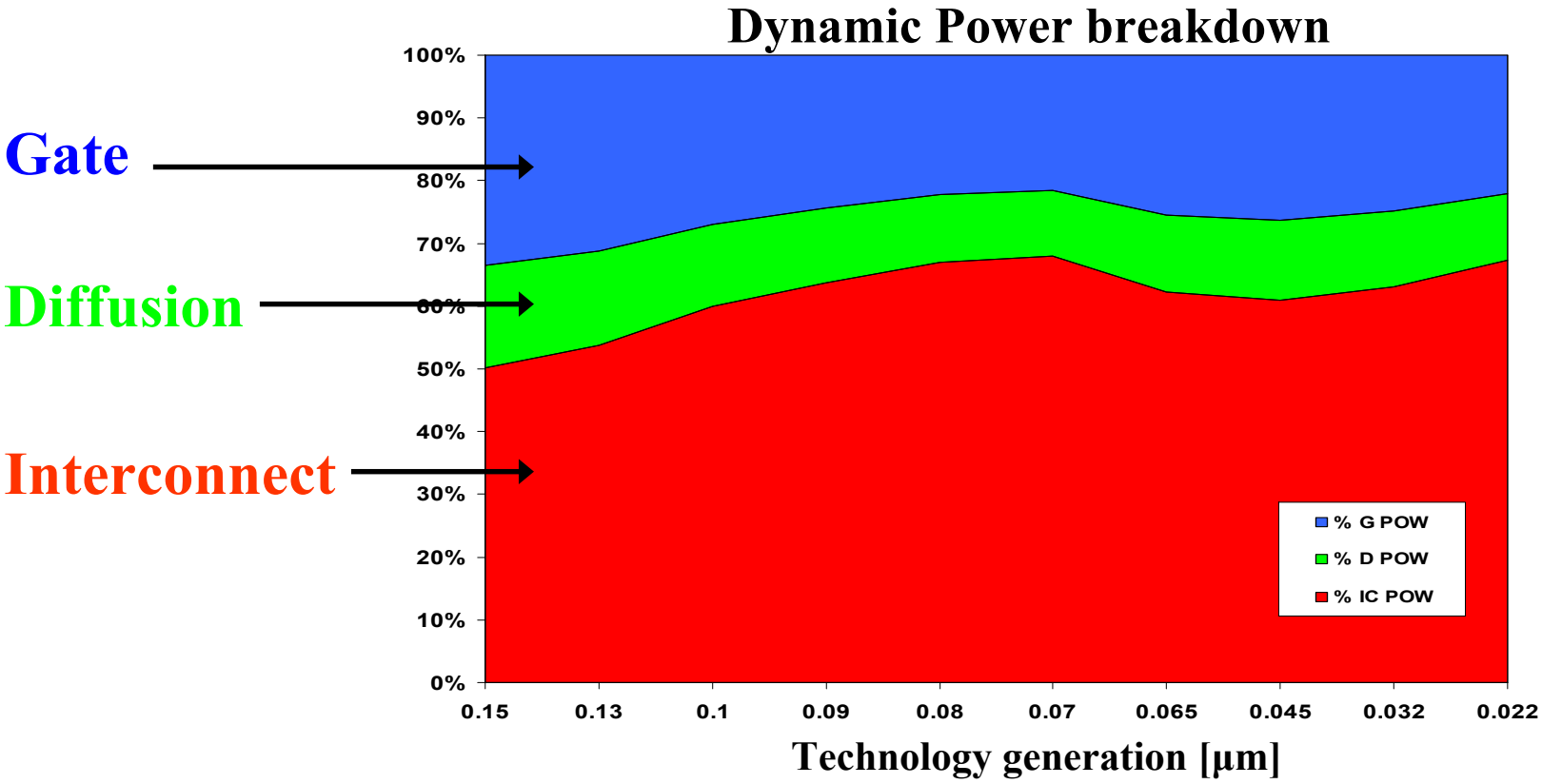


- Efficient sharing of wires
- Lower area / lower power / faster operation
- Shorter design time, lower design effort
- Scalability
- Future: Higher level services

Beware the Net !

- The networking discipline has created lots of useful solutions
- But they may not be appropriate for VLSI
- In VLSI, we must "optimize"
  - Power
  - Area
  - Power
  - Design time and complexity
  - Power
  - Achieve specific performance goals (achieve *Quality of Service*)
  - Power

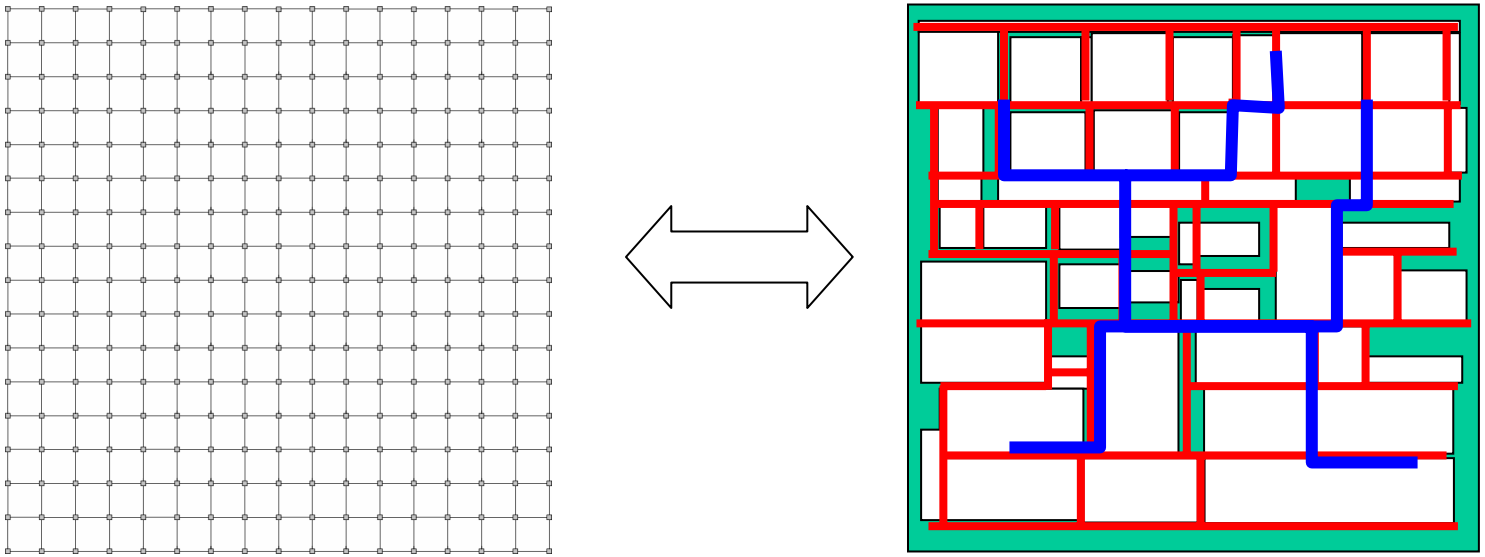
Power, power, power



Source: Magen et al., SLIP04 (analysis of Pentium M)

Beware the Net (1): Regular Topology

- ~~You can create an elegant regular topology (e.g. MESH)~~
  - But SoCs are irregular
  - Must solve issues of irregular topology and routing (QNoC)



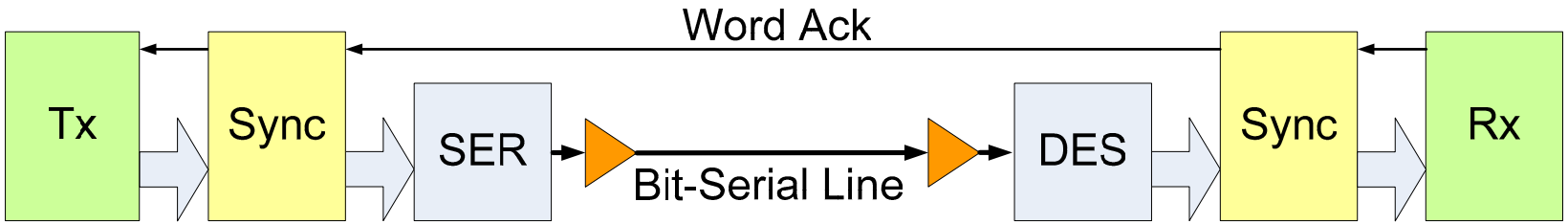
    - Combine XY with deviations
  - Also allow multi-levels (hierarchy, serial long-range) (QNoC)

The hierarchical NoC

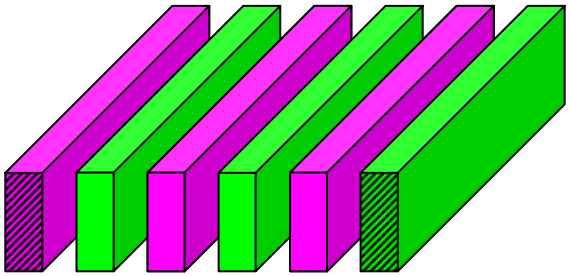
High speed, serial links

- Designed for long links (5-10 mm)
- Save area, save routing channels, save power, save pins on modules

Wires are expensive

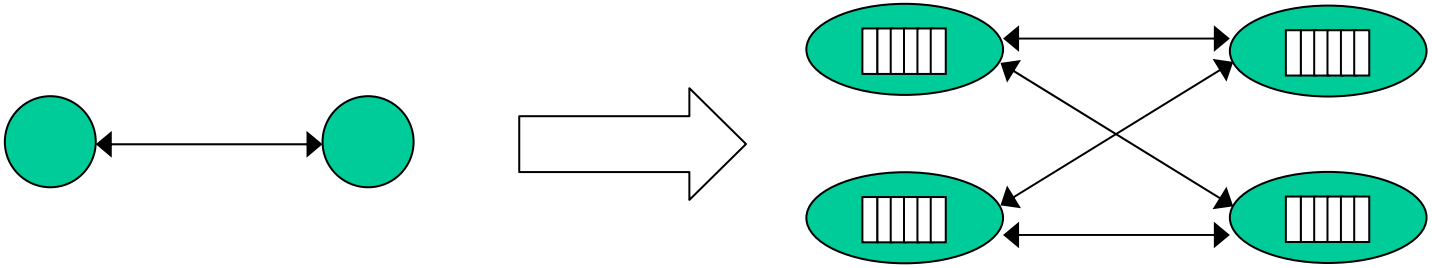


- Fastest: New bit every gate delay. E.g. 65 nm → 70 Gbps
- How?
  - Asynchronous, wave-pipelines, special wires, special circuits
- Extremely robust: 10σ variations, 24 PVT corners

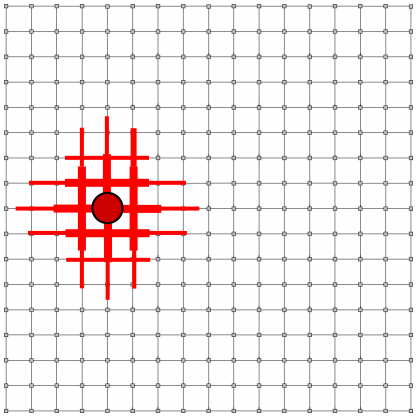


Beware the Net (2): Non-Blocking

- ~~You can create a non-blocking network~~
  - But it's expensive
    - Requires more buffers and wires



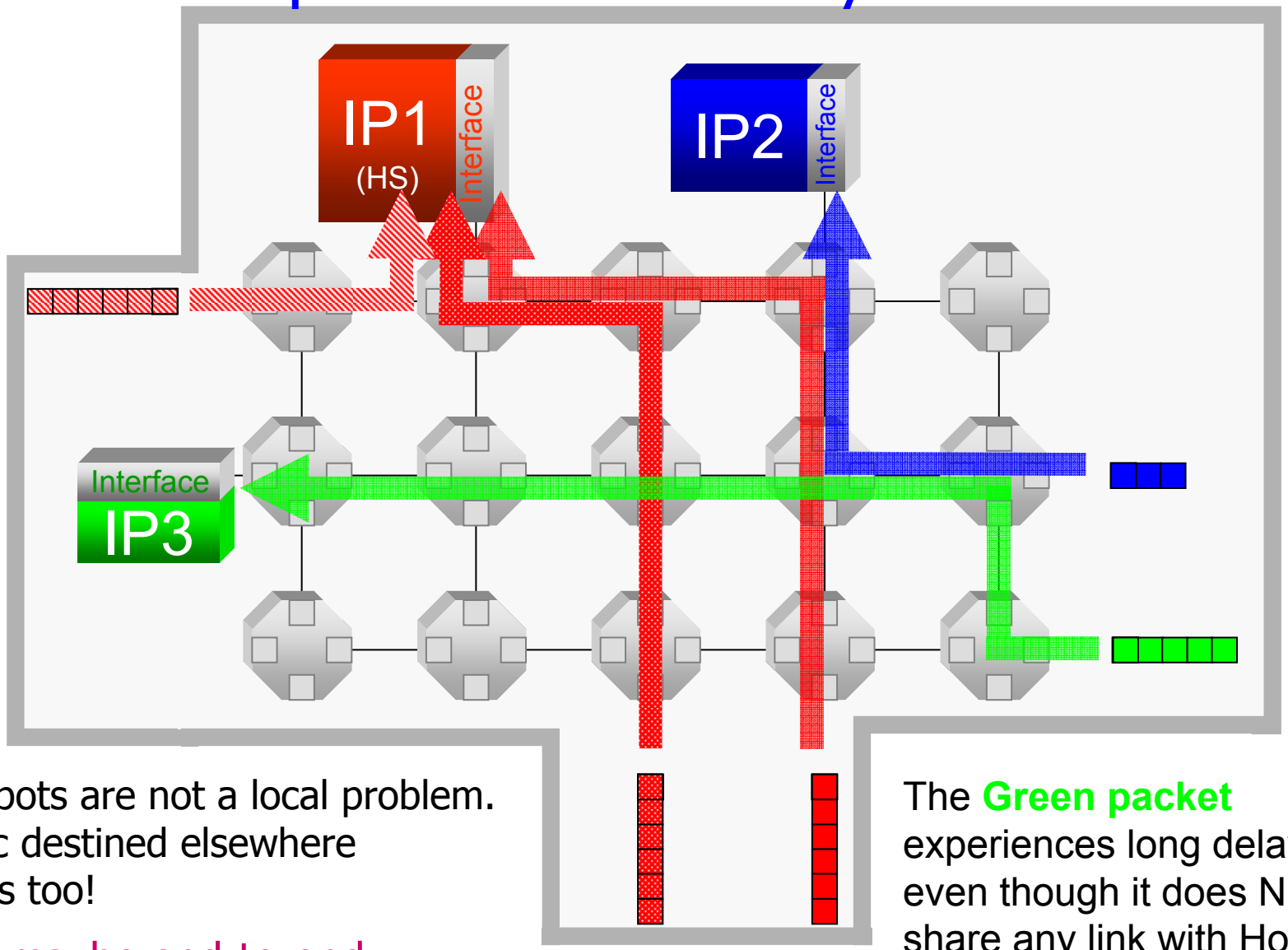
  - And it doesn't help
    - Hot spots can block networks of infinite capacity



Hot Spots in QNoC

- Hot Spot: A module (rather than a NoC element) that occasionally cannot digest all the traffic addressed to it
  - Results in temporary massive delay build-up
  - Results in blocking the net !
- This is NOT congestion on the net
  - Higher network capacity won't help
- Examples
  - Port to off-chip DRAM
  - Shared resource on chip

Hot Spots Affects the System



- Hot Spots are not a local problem. Traffic destined elsewhere suffers too!

Solution maybe end-to-end allocation (cf. STNoC)

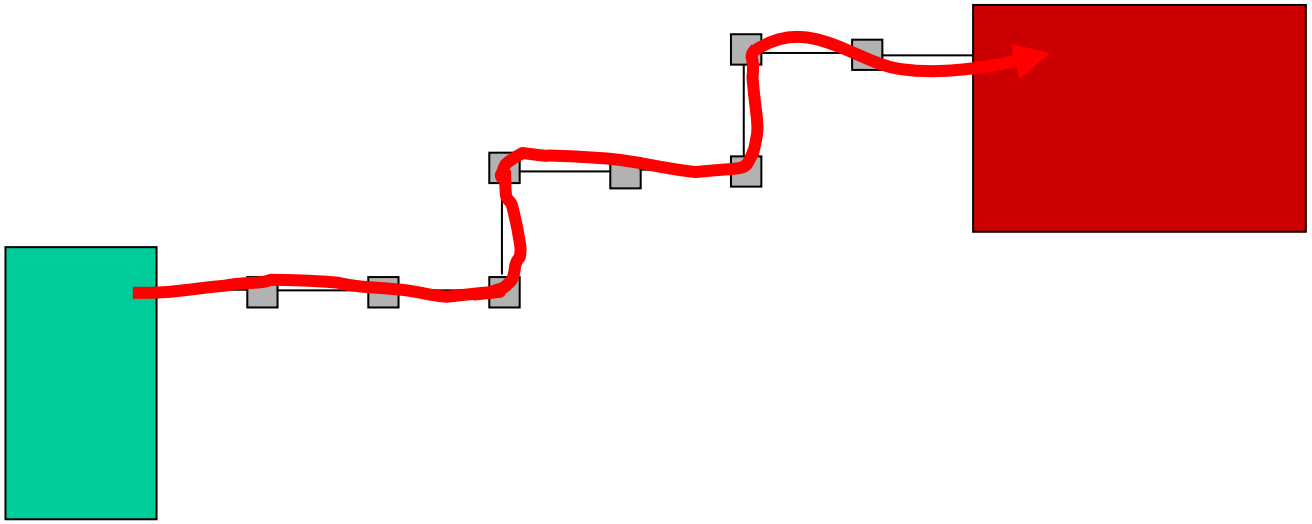
The **Green packet** experiences long delay even though it does NOT share any link with Hot Spot traffic

Beware the Net (3): Guaranteed Service ?

- You can guarantee service
  - It's nice to depend on
  - It's easy to verify
  - It's claimed to use simpler hardware
  - But it's hard to re-configure
    - Requires pre-computing multiple configurations
    - Requires distribution of configurations upon transitions
  - Best Effort is simpler ?
    - Simpler to design
    - Simpler to use
    - But more difficult to guarantee the service….
  - Anyway, both GS and BE are only means for QoS.
    - QoS is must, how to implement it is open

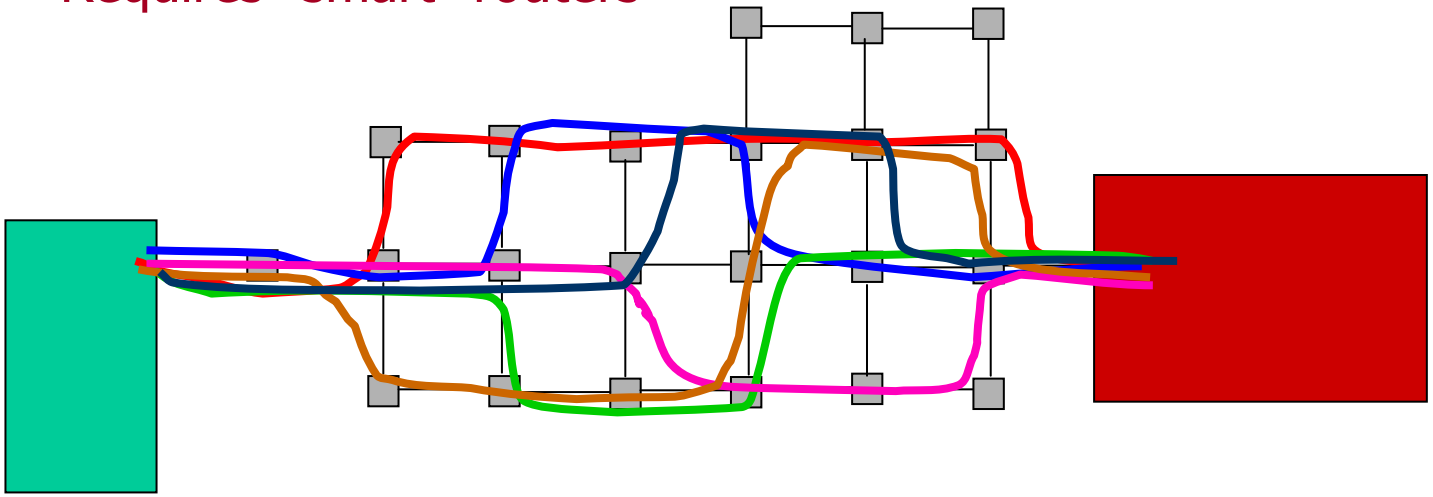
Beware the Net (4): Buffers

- ~~You can use lots of buffers in routers~~
  - But they dissipate lots of power
  - And they consume lots of area
  - And they do not improve wormhole routing

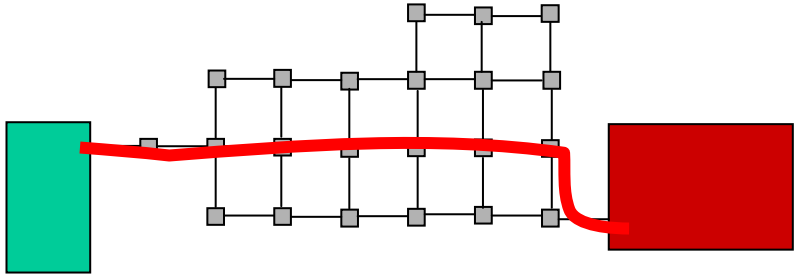


Beware the Net (5): Routing

- ~~You can create complex routing~~
  - Especially for load balancing
  - Requires "smart" routers



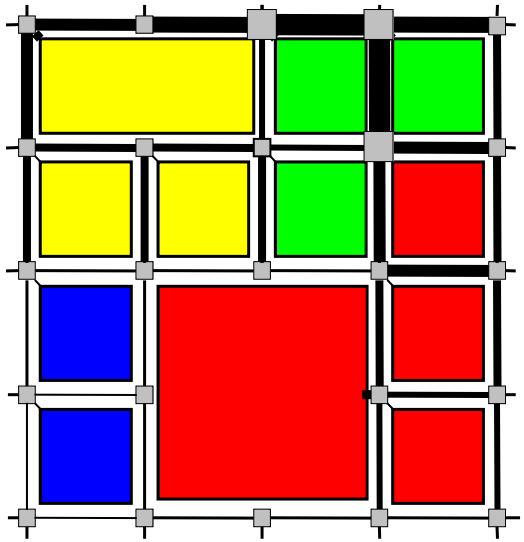
  - Fixed, simple single-path routing saves energy and area



    - Must accommodate irregular routing

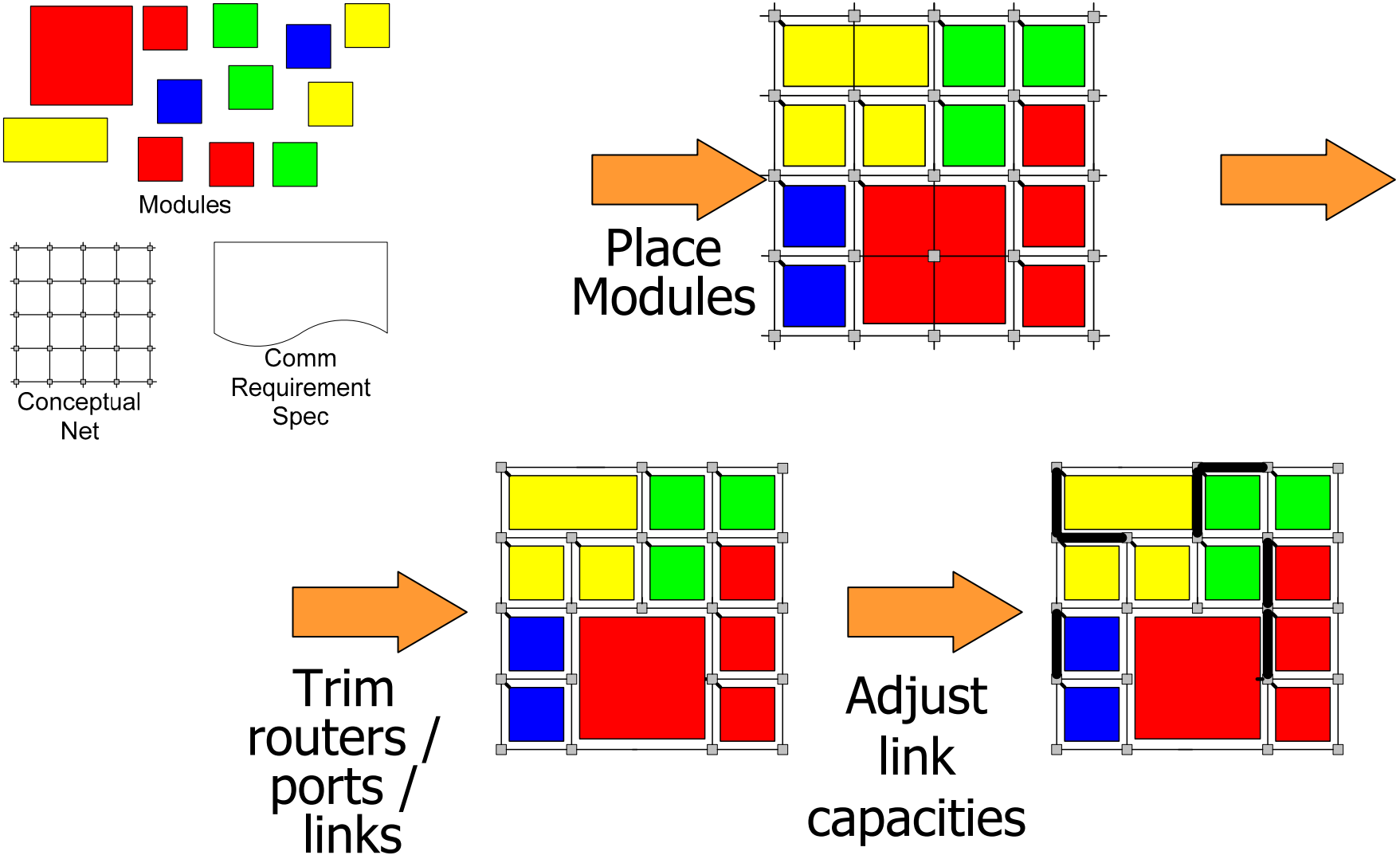
Simple Routing

- Single-path routing is best supported by irregular, custom capacity allocation (QNoC)



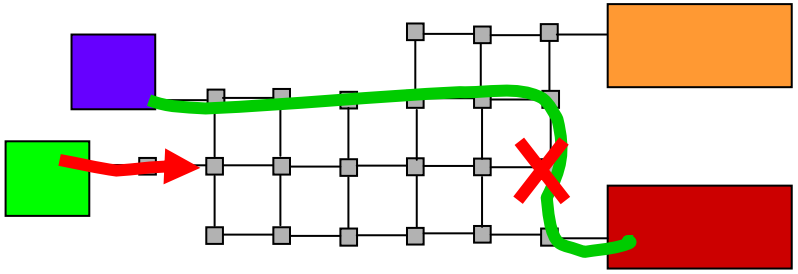
There is always enough bandwidth!

QNoC Customization

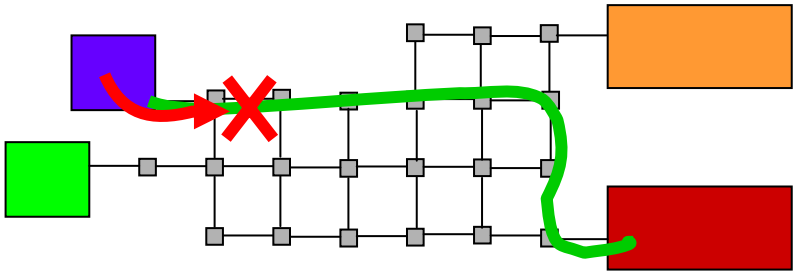


Beware the Net (6): Conflicts

- ~~You can make packets conflict with each other~~
  - E.g. fast / short msg blocked behind long DMA block



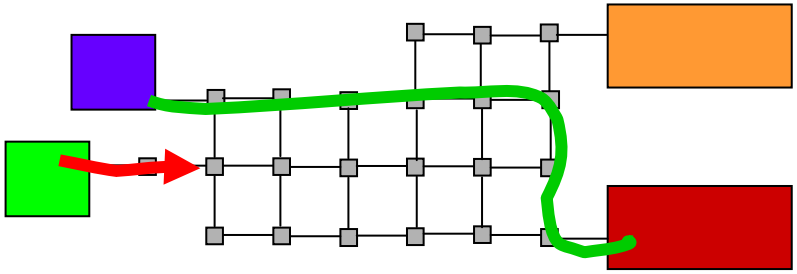
  - E.g. urgent cache fetch blocked behind long pre-fetch

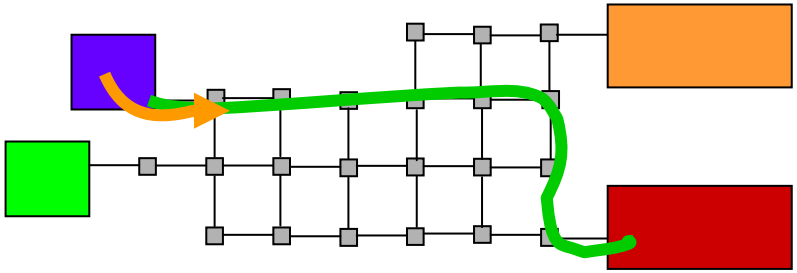


  - Better use priority levels and pre-emption

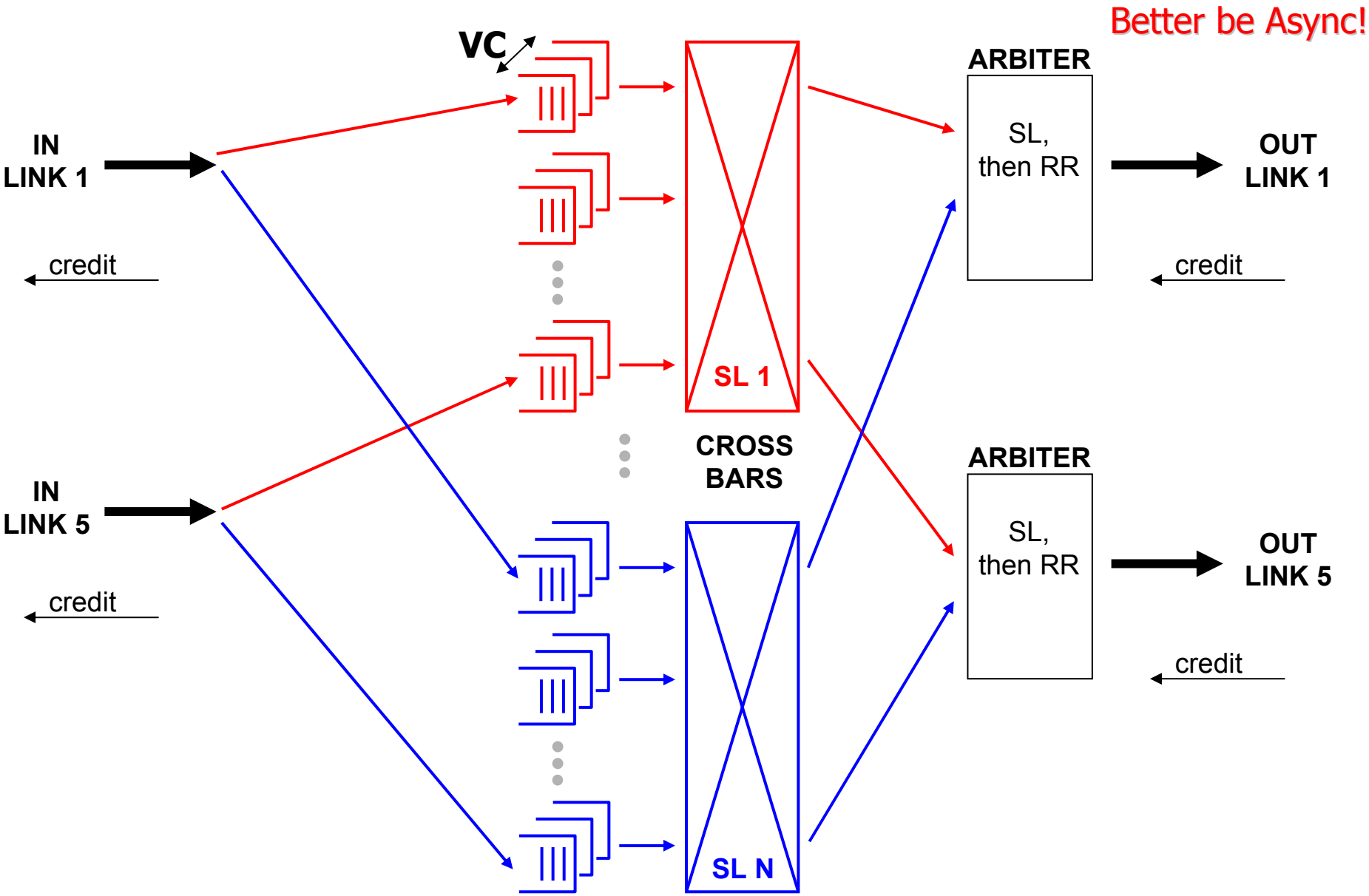
Service Levels

- Solution: Priority levels and pre-emption
  - QoS NoC→QNoC



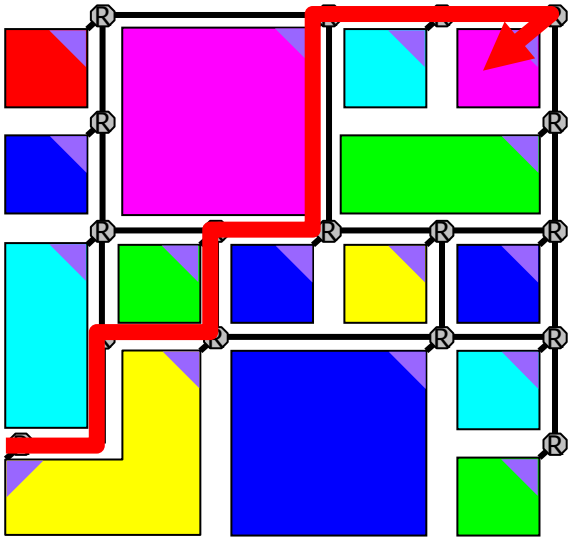


Multi-service-levels QNoC Router



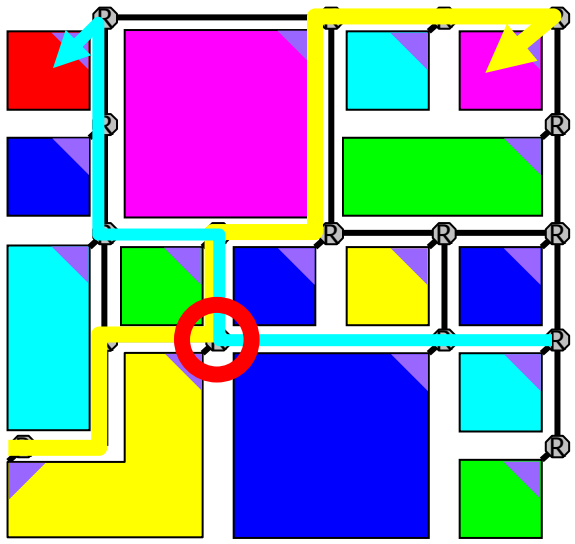
Beware the Net (7): Synchronization

- You can use synchronous communication over NoC
  - E.g. use a single, fast clock for NoC
    - But it's hard to gate the clock → high power dissipation



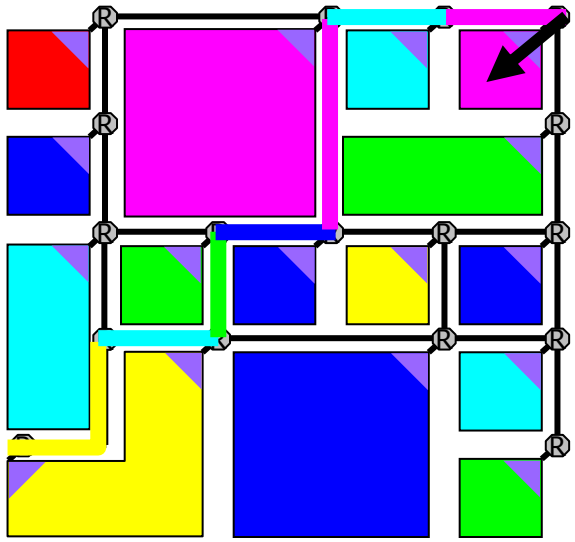
Synchronization

- You can use synchronous communication over NoC
  - E.g. use a single, fast clock for NoC
    - But it's hard to gate the clock → high power dissipation
  - E.g. use source-synchronous communication
    - But it's hard to synchronize streams from different sources



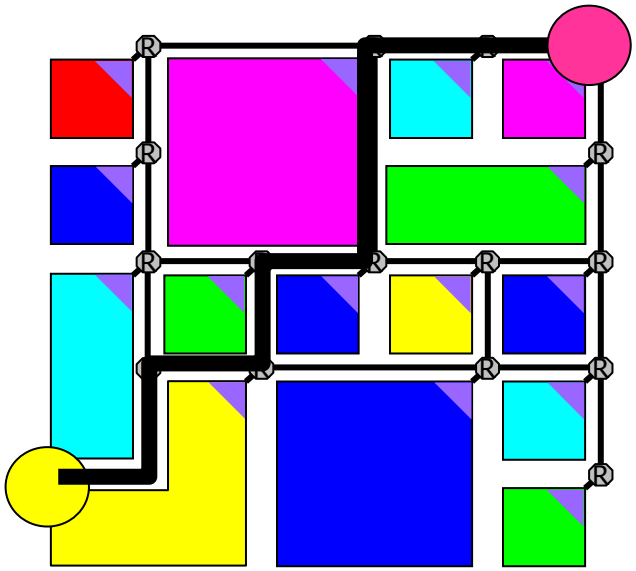
Synchronization

- ~~You can use synchronous communication over NoC~~
  - E.g. use a single, fast clock for NoC
    - But it's hard to gate the clock → high power dissipation
  - E.g. use source-synchronous communication
    - But it's hard to synchronize streams from different sources
  - E.g. use local clocks
    - But it requires many re-synchronizations along the way



Asynchronous NoC

- Synchronize only at the ends
- Asynchronous routers



Summary

- NoC is a great idea
- But it must be approached carefully, so that it will not "swallow the SoC"
- Issues:
  - Power, power, power (and area)
  - Regular topology
  - Non-blockiness
  - Service guarantee
  - Buffers
  - Routing
  - Conflicts
  - Synchronization
- Solutions
  - Customized Architecture
  - Irregular, hierarchical topology
  - Simple routers, routing, signaling, clocking
  - Support multiple service levels

The new IEEE NoC Symposium

- May 7-9, 2007, Princeton
- Paper deadline December 1, 2006
- Scientific
- Large PC, all papers reviewed
- Excellent program!

NoCs International Symposium on Networks-on-Chips