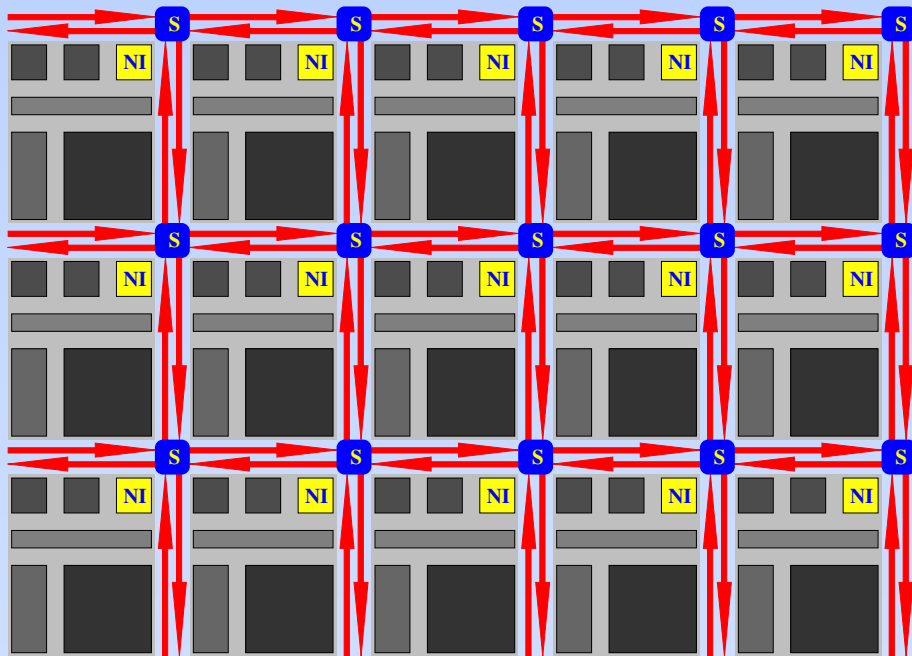


The Nostrum Network on Chip



Axel Jantsch

Royal Institute of Technology, Stockholm

November 2005

Overview

Topology and Structure

Protocol Stack

The Network Layer and the Switch

Data Protection

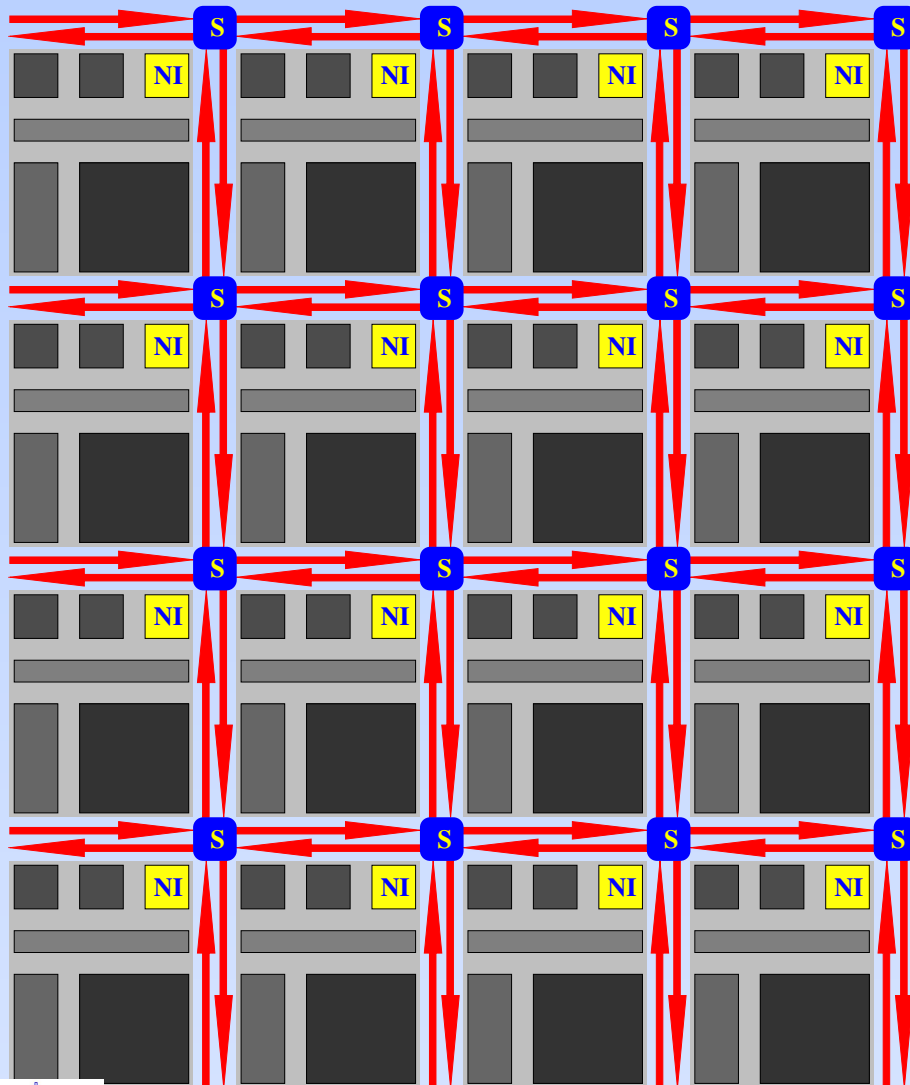
Clocking

Dynamic Voltage Scaling

Network Simulator



Nostrum Topology: Mesh



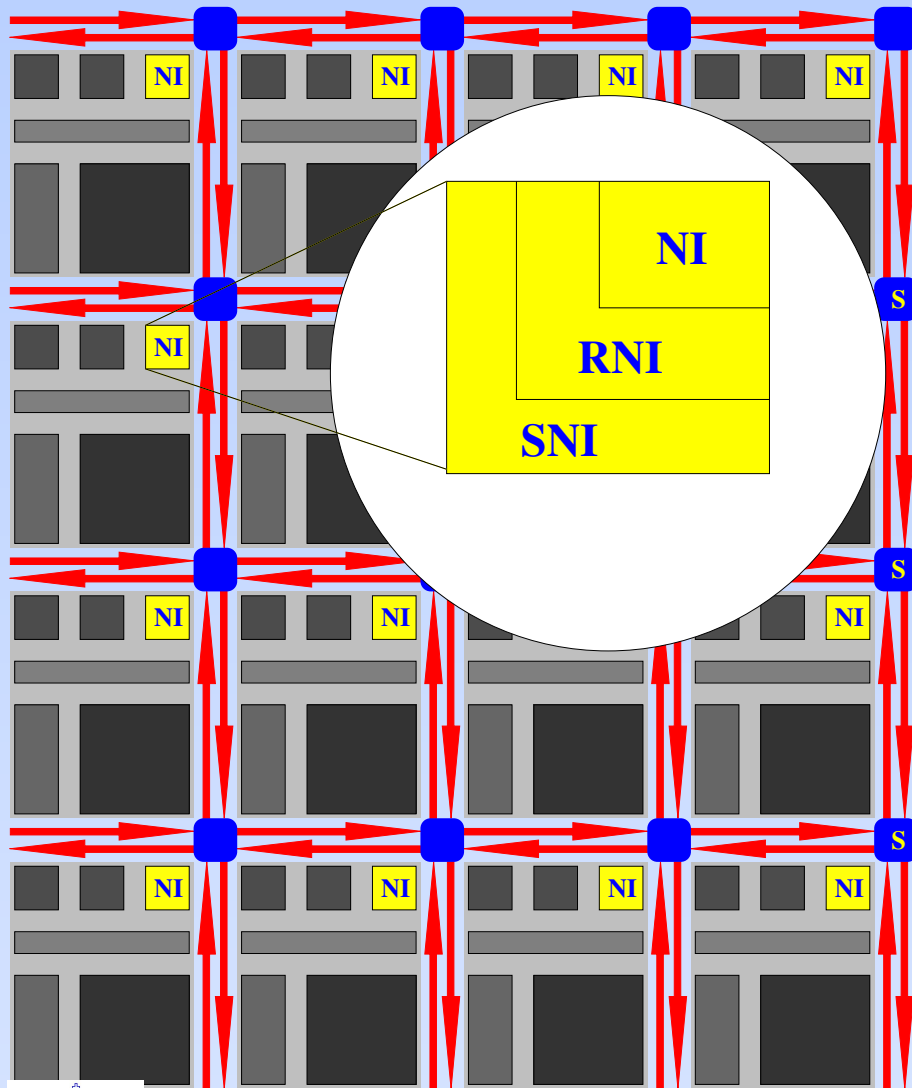
Characteristics:

- Resource-to-switch ratio: 1
- A switch is connected to 4 switches and 1 resource
- A resource is connected to 1 switch
- Average distance: $2/3n$
- Bisection bandwidth: $2n$

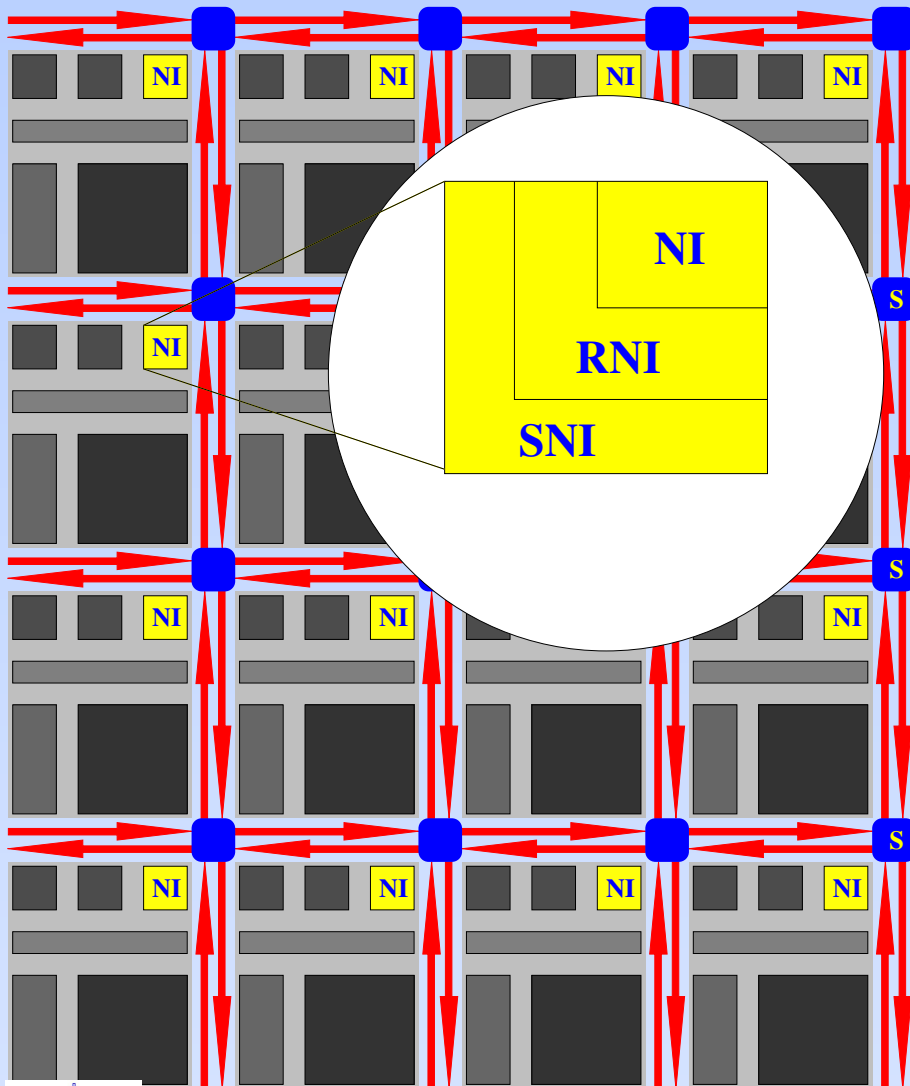
Motivation:

- Regularity of layout; predictable electrical properties
- Expected locality of traffic

The Node in a Mesh



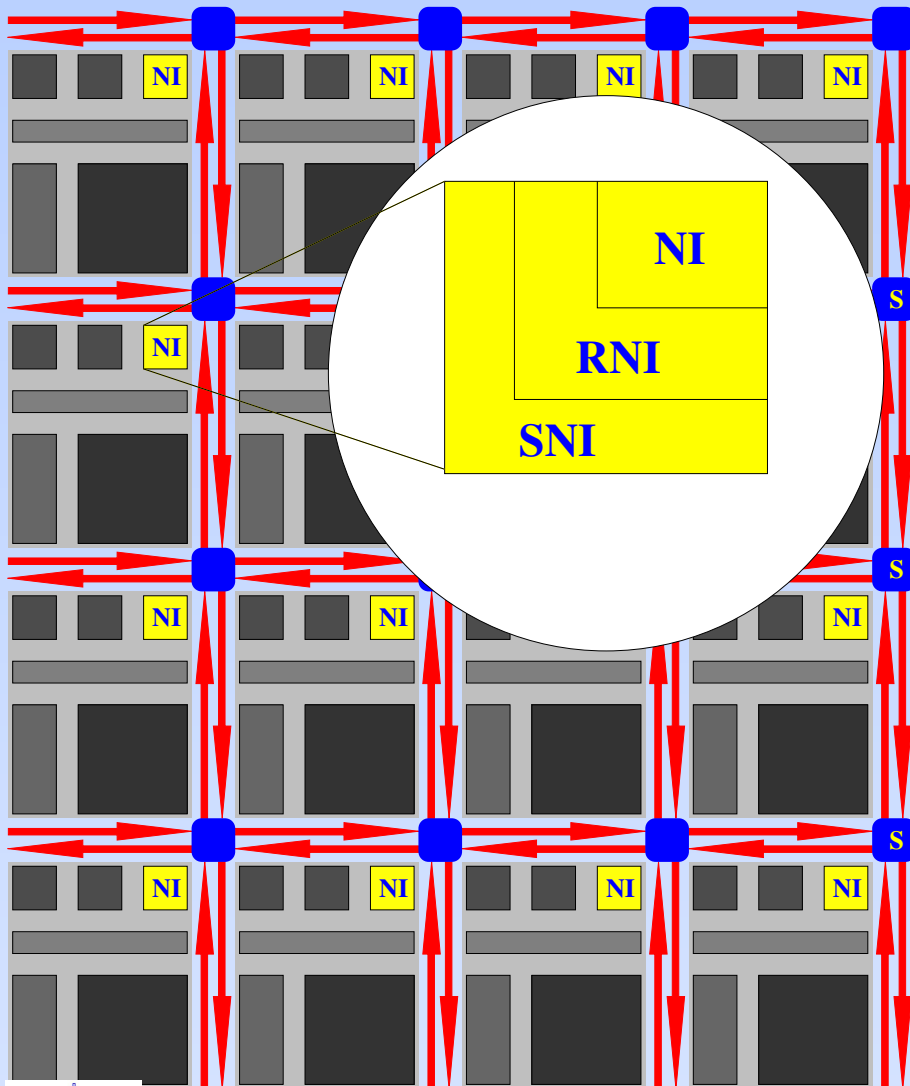
The Node in a Mesh



NI: Network Interface:

- Compulsory
- Hardware
- Implements the network layer protocol

The Node in a Mesh



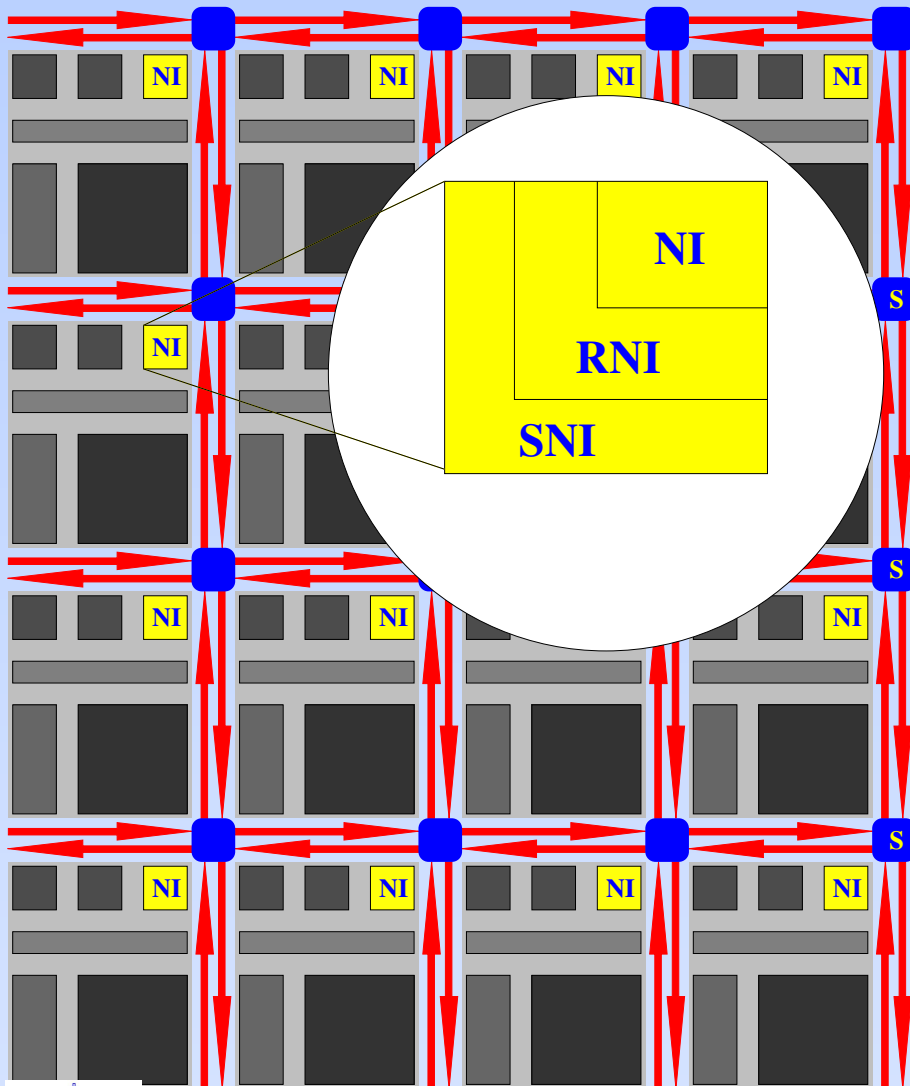
NI: Network Interface:

- Compulsory
- Hardware
- Implements the network layer protocol

RNI: Resource Network Interface:

- Optional
- Hardware and/or Software
- Implements transport layer
- Provides resource specific interfaces

The Node in a Mesh



NI: Network Interface:

- Compulsory
- Hardware
- Implements the network layer protocol

RNI: Resource Network Interface:

- Optional
- Hardware and/or Software
- Implements transport layer
- Provides resource specific interfaces

SLI: Session Layer Interface:

- Optional
- Hardware and/or software
- Implements the session layer protocol

Overview

Topology and Structure

Protocol Stack

The Network Layer and the Switch

Data Protection

Clocking

Dynamic Voltage Scaling

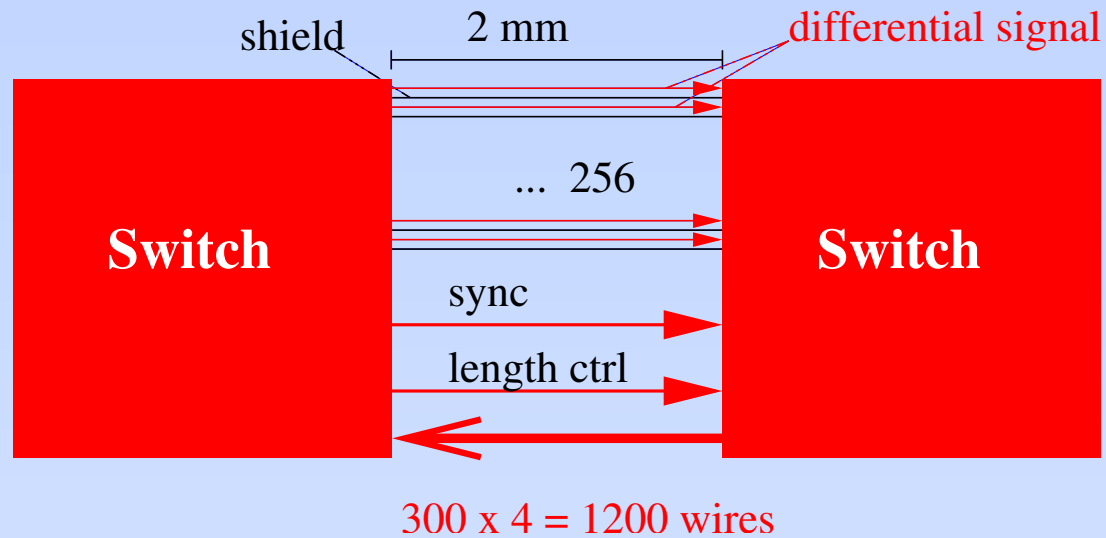
Network Simulator



Physical Layer

Parameters:

- Physical distance
- Number of lines
- Activity control
- Buffers and pipelining



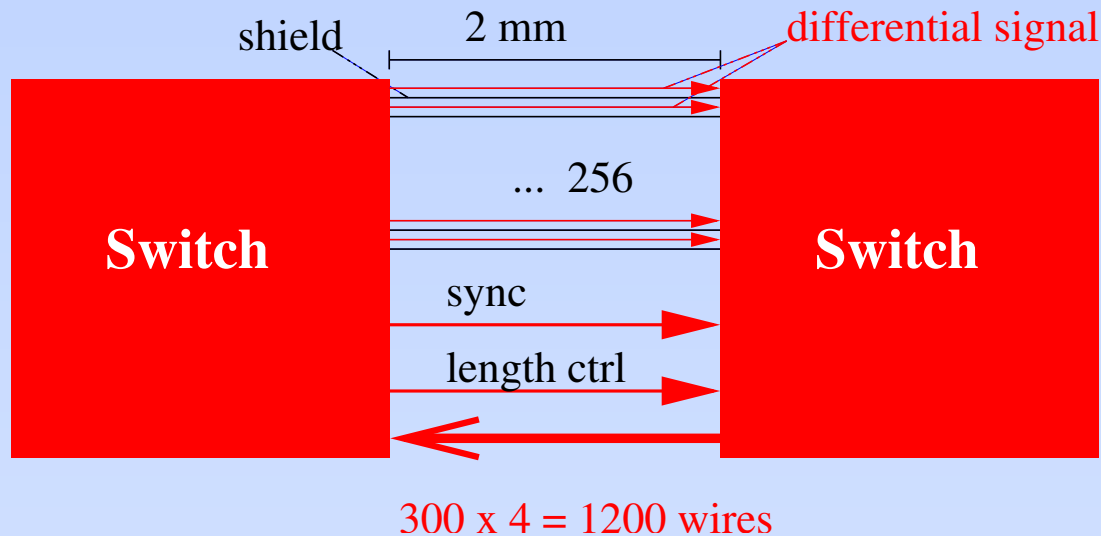
Physical Layer

Parameters:

- Physical distance
- Number of lines
- Activity control
- Buffers and pipelining

Nostrum status:

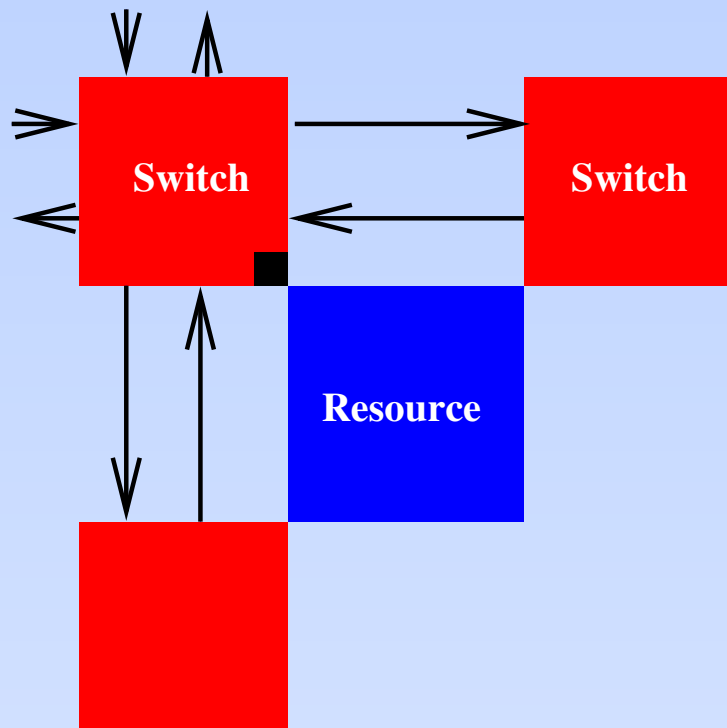
- 128 data lines in each direction
- No pipelining
- On/off control for power saving



Data Link Layer

Parameters:

- Line frequency versus switch frequency
- Buffering
- Error correction
- Power optimization encoding



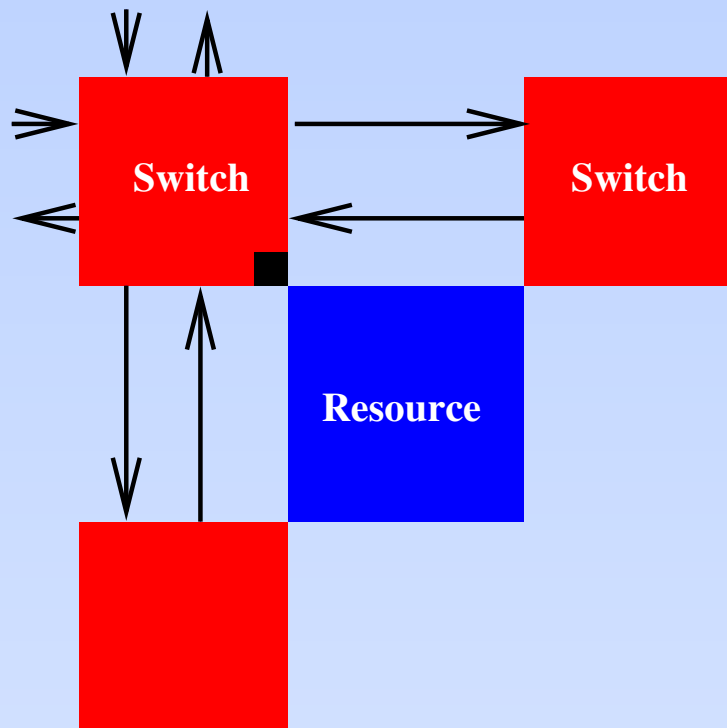
Data Link Layer

Parameters:

- Line frequency versus switch frequency
- Buffering
- Error correction
- Power optimization encoding

Nostrum status:

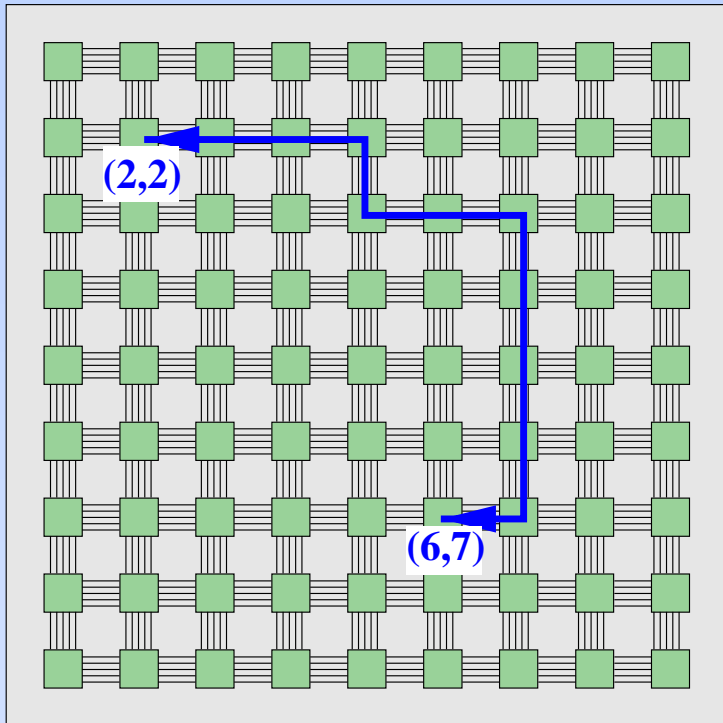
- Physical packet = data link packet
- Physical clock = data link clock
- Single packet input buffer
- Error correction
- On/off activity control



Network Layer

Parameters:

- Link layer cell size vs. network layer packet size
- Network address scheme
- Routing algorithm
- Switching policy
- Priority classes
- Error correction



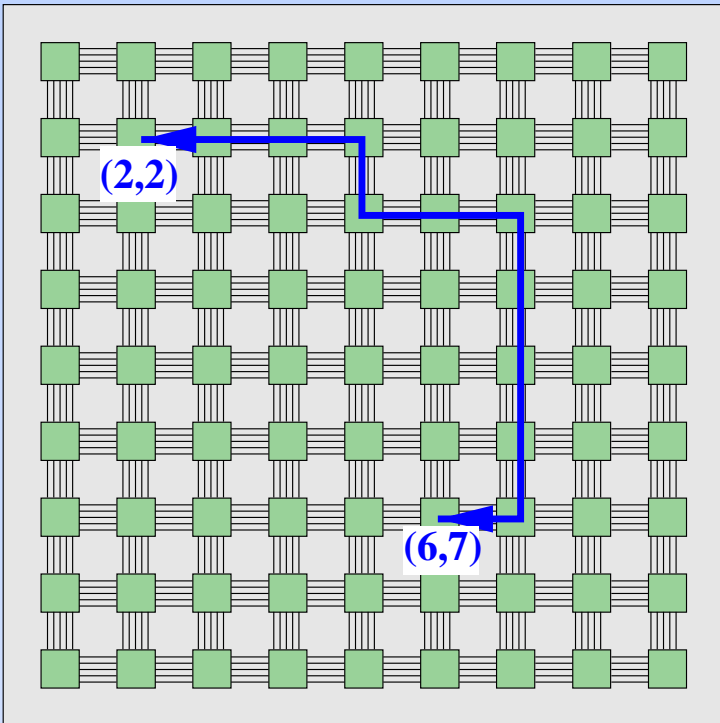
Network Layer

Parameters:

- Link layer cell size vs. network layer packet size
- Network address scheme
- Routing algorithm
- Switching policy
- Priority classes
- Error correction

Nostrum status:

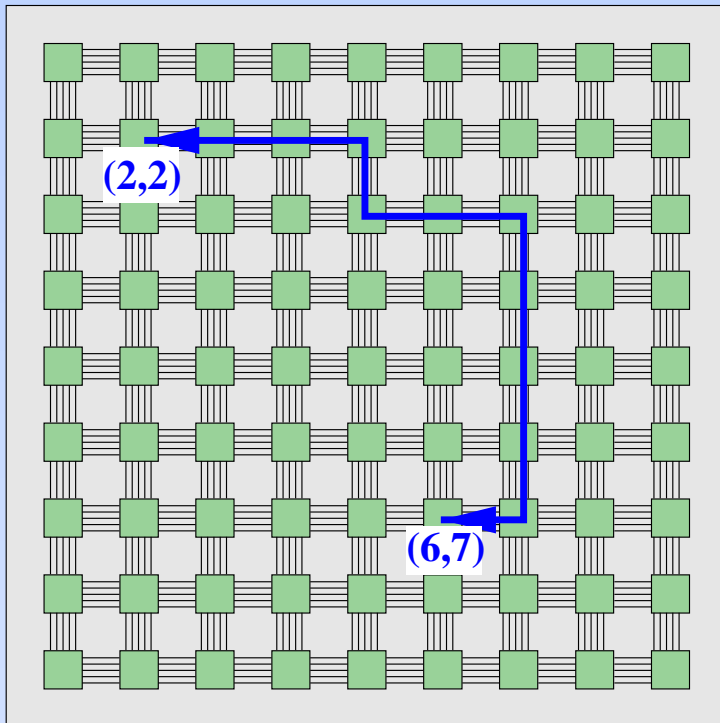
- Link layer packet = network layer packet
- Relative x-y addresses
- Deflective routing with no buffers and no routing tables
- Virtual circuits with guaranteed bandwidth and delays
- No data protection



Transport Layer

Parameters:

- Network layer packet size vs. transport layer message size
- Offering and establishing connections, e.g. sockets
- Negotiating quality of service
- Error correction



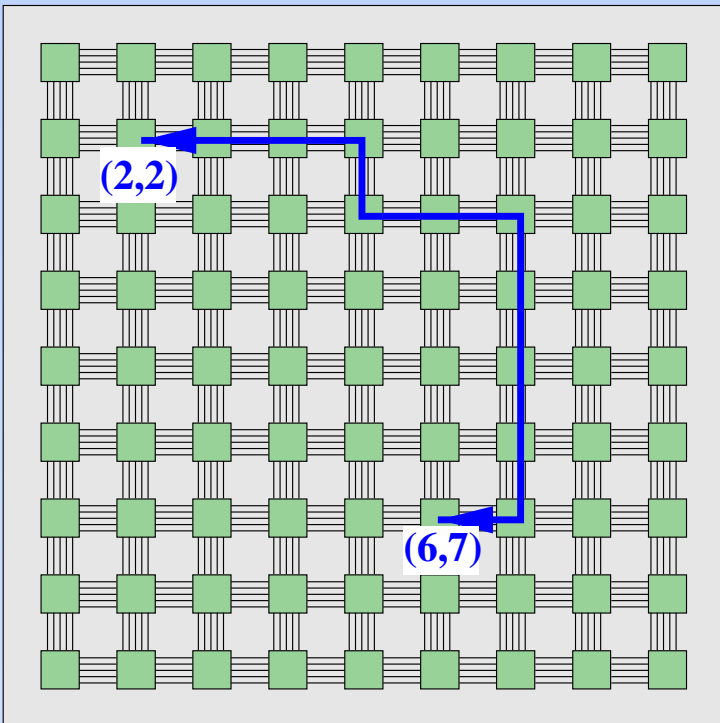
Transport Layer

Parameters:

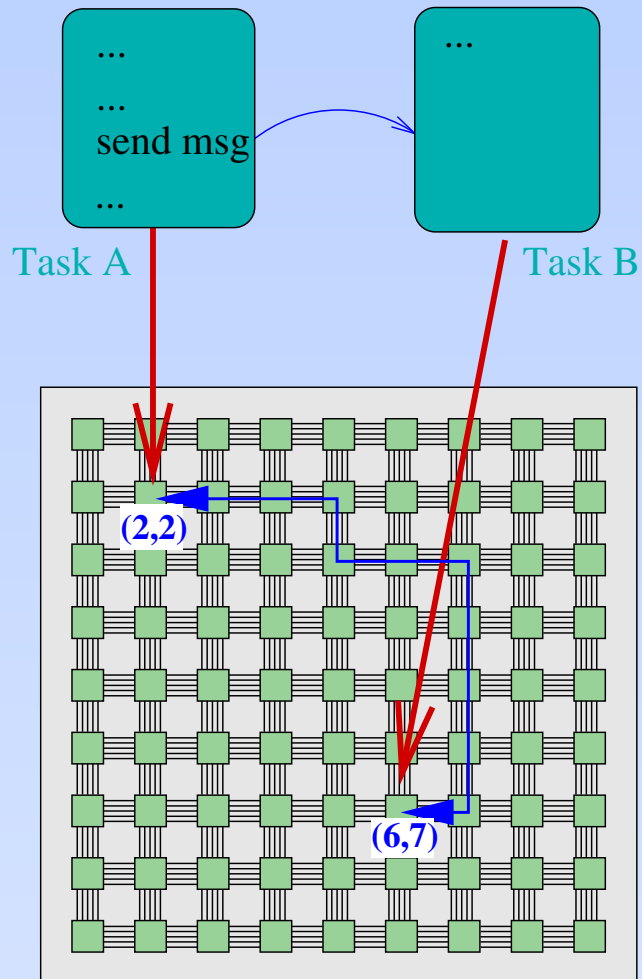
- Network layer packet size vs. transport layer message size
- Offering and establishing connections, e.g. sockets
- Negotiating quality of service
- Error correction

Nostrum status:

- Transport layer message can be of unlimited size
- Static and semi-static establishment of connections
- Static bandwidth assignment and simple admission protocol
- Two levels of data protection



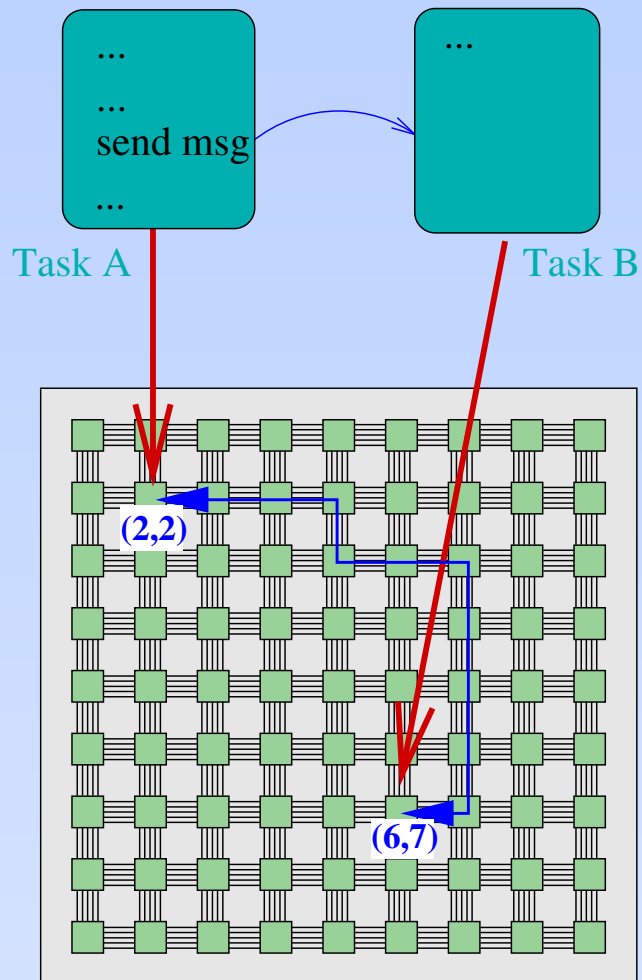
Session Layer



Parameters:

- Task level communication primitives
- Message passing
- Shared memory based communication
- Synchronization
- Error correction

Session Layer



Parameters:

- Task level communication primitives
- Message passing
- Shared memory based communication
- Synchronization
- Error correction

Nostrum status:

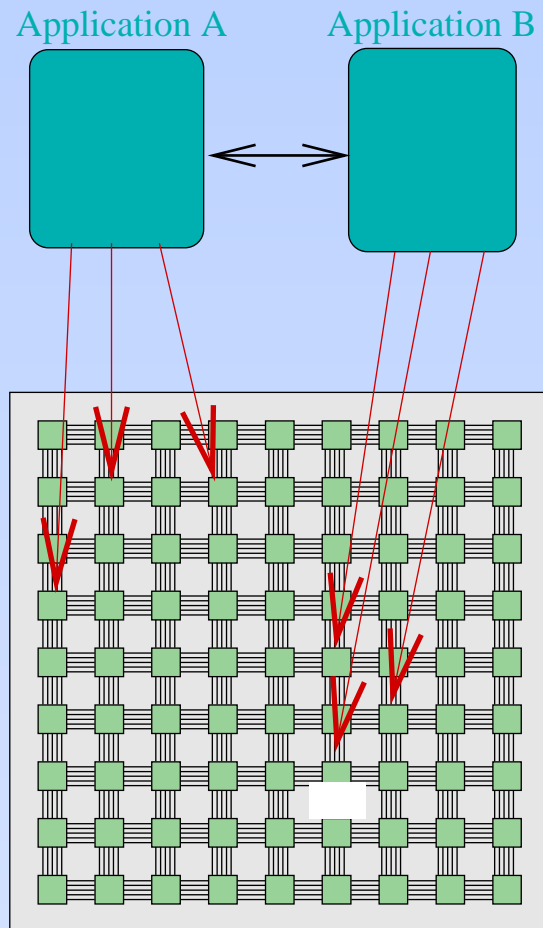
- Set of communication primitives defined
- Both message passing and shared memory
- User controlled synchronization
- No data protection

Session Layer Communication

- Message passing communication:
 - ★ open/listen/accept/bind primitives to open a channel
 - ★ send/receive to communicate
 - ★ close to tear down the channel
 - ★ blocking/non-blocking send/receive
- Shared memory communication:
 - ★ allocation
 - ★ read/write
 - ★ free
 - ★ interruptible/non-interruptible
- VHDL,C and SystemC libraries under development



Application Layers



Application specific communication services;
E.g. the NoC operating system could use:

- Task/resource database access protocol
- Task migration protocol
- Power management services

Overview

Topology and Structure

Protocol Stack

The Network Layer and the Switch

Data Protection

Clocking

Dynamic Voltage Scaling

Network Simulator



The Network Layer

- Packet switched best effort service
 - ★ Packets are guaranteed to arrive
 - ★ Packet payload may be protected (4 levels of protection)
 - ★ Load dependable delay in the network
 - ★ Load dependable delay at the network access point
 - ★ Admission policy for best effort traffic:
 - * Network load should be below 60%
 - * Load is measured locally in switch and based on neighboring stress values

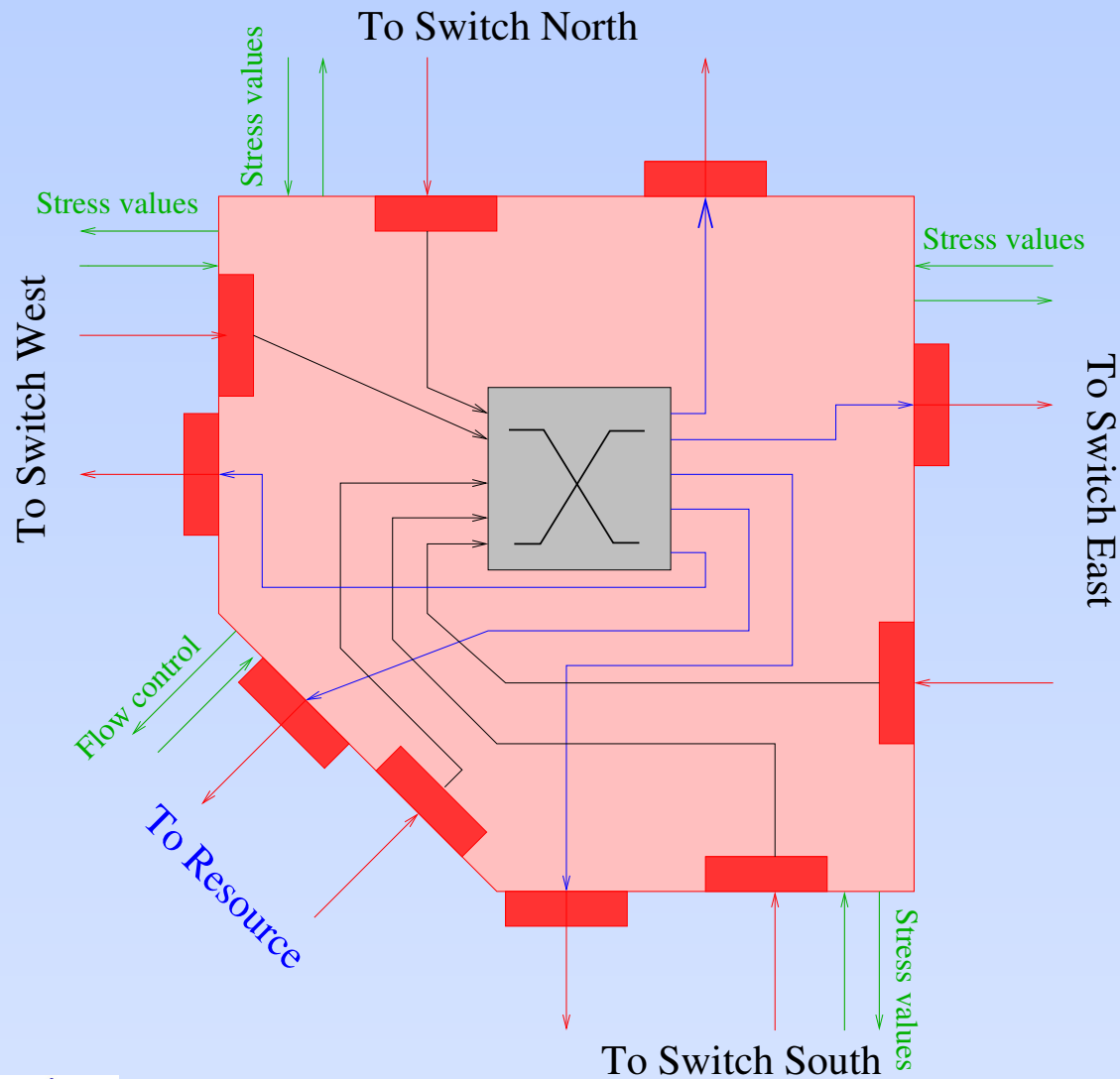


The Network Layer

- Packet switched best effort service
 - ★ Packets are guaranteed to arrive
 - ★ Packet payload may be protected (4 levels of protection)
 - ★ Load dependable delay in the network
 - ★ Load dependable delay at the network access point
 - ★ Admission policy for best effort traffic:
 - * Network load should be below 60%
 - * Load is measured locally in switch and based on neighboring stress values
- Virtual circuit service
 - ★ Guaranteed bandwidth
 - ★ Guaranteed maximum delay
 - ★ Multicast circuits
 - ★ Static and semi-static virtual circuits
 - ★ Based on packet switching service

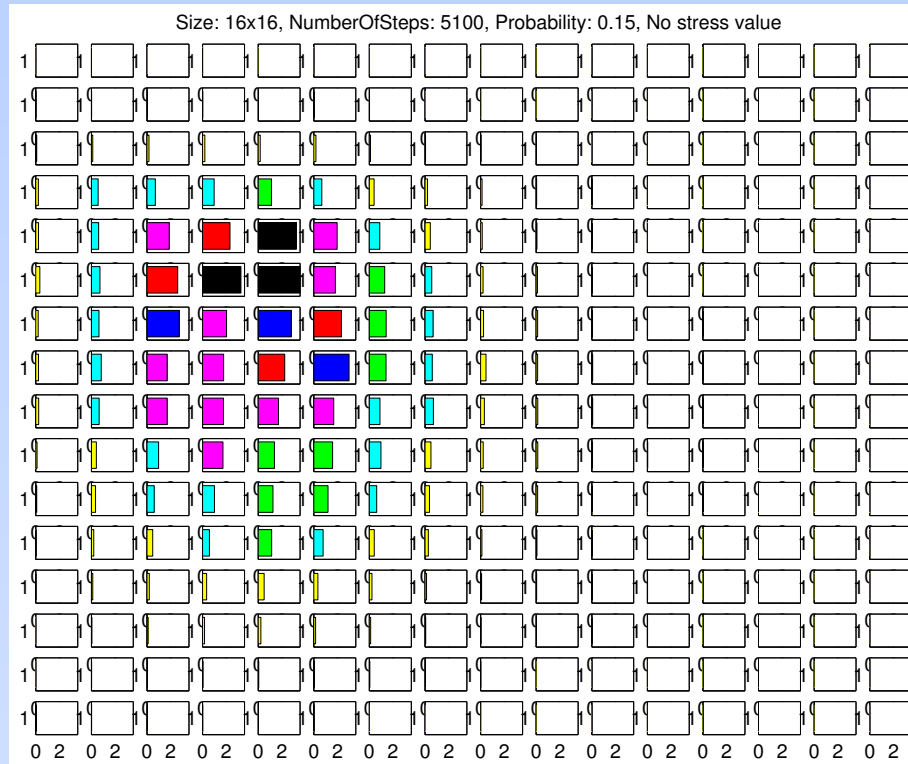


The Bufferless Switch



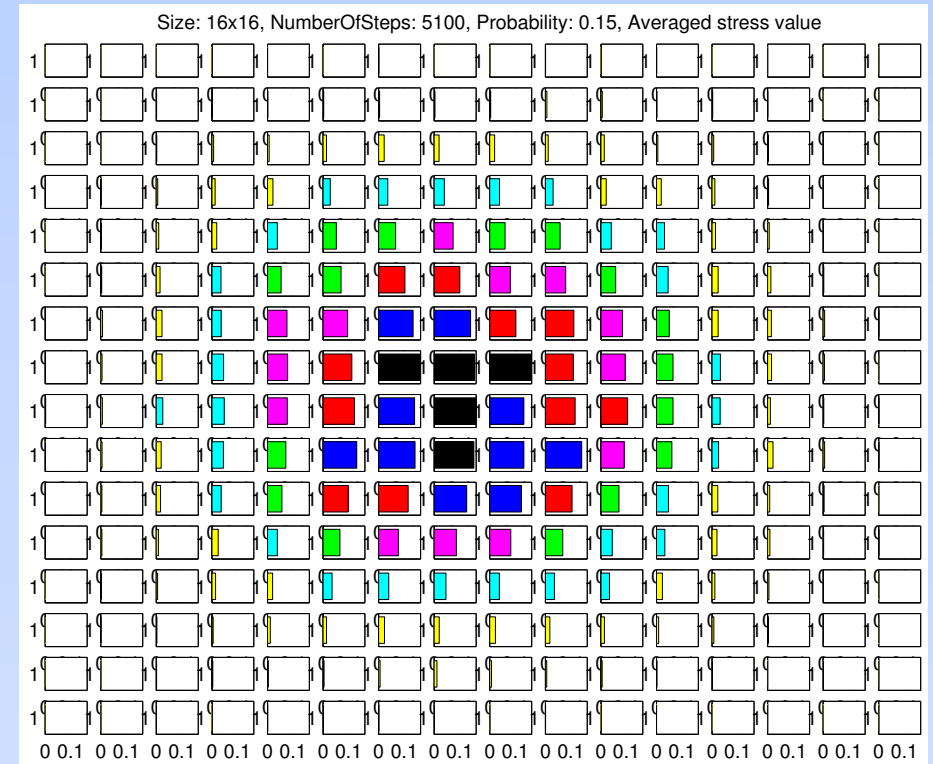
- + No buffers
- + No routing table
- + Small area
- + Short delay
- + Low power consumption
- Non-shortest path
- Header overhead due to destination address

Stress Value Effect on Buffer Sizes and Delays



No stress value control

Largest average buffer size: 3.2 (black)

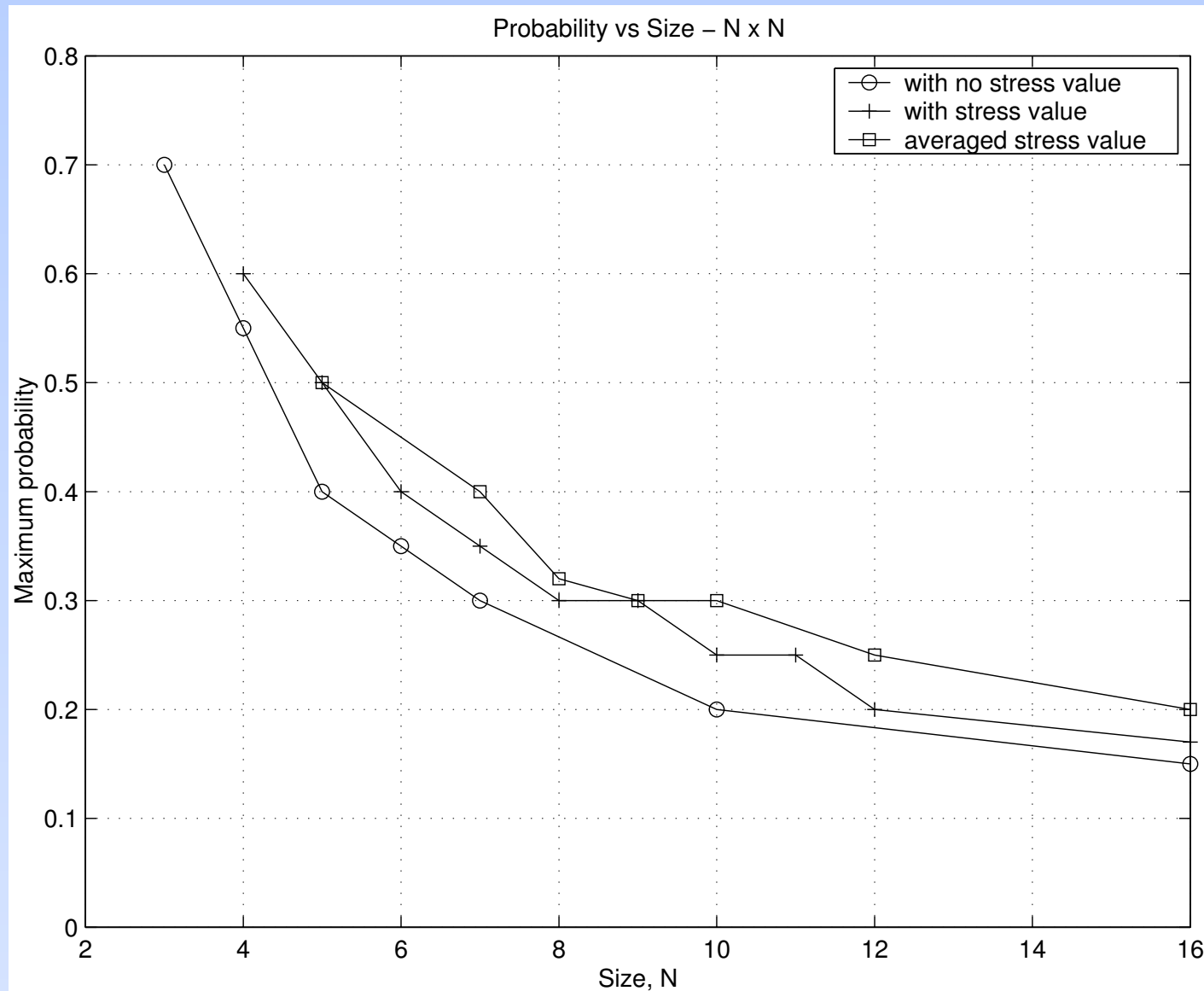


Averaged stress value control

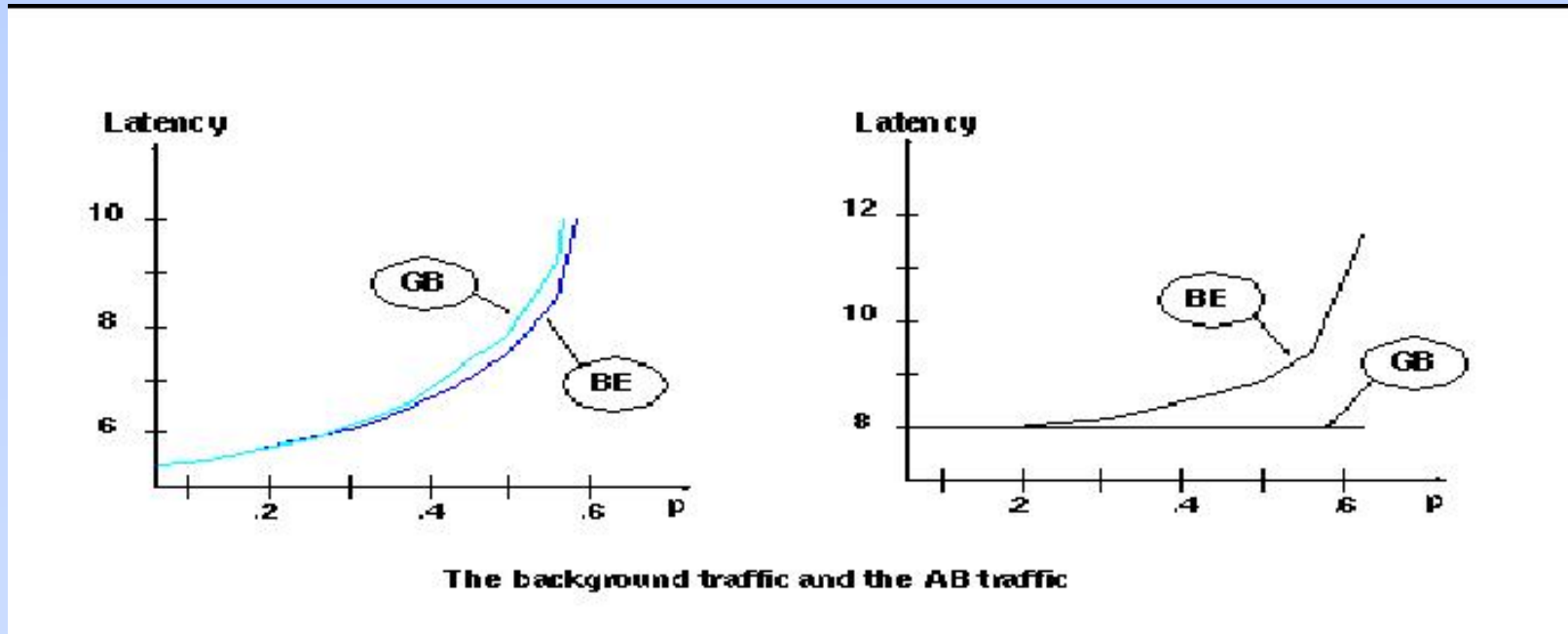
Largest average buffer size: 0.1 (black)



Stress Value Effect on Maximum Load



Best Effort and Guaranteed Bandwidth Traffic



Implementation of Static Virtual Circuits

- Bandwidth allocation and circuit setup at design time
- Implementation alternatives:
 - ★ Channel containers have higher priority
 - ★ Look-up tables in switches
- Semi-static circuits:
 - ★ Active circuits: Circulating containers
 - ★ Inactive circuits: Containers removed
 - ★ Activation of circuits subject to traffic load dependent delay
 - ★ NI can increase stress value to activate virtual circuits



Overview

Topology and Structure

Protocol Stack

The Network Layer and the Switch

Data Protection

Clocking

Dynamic Voltage Scaling

Network Simulator



Data Protection

- Two level protection: Link layer and transport layer
- Data link layer protection:
 - ★ SEC-DED header protection (16/26 bits)
 - ★ Four levels of payload protection:
 - * Maximum bandwidth - no protection (102/102 bits)
 - * Guaranteed integrity - DED protection (90/102 bits)
 - * Minimum latency - SEC protection (90/102 bits)
 - * High reliability - SEC-DED protection (81/102 bits)
 - ★ Parity based codes used (Hamming or Hsiao codes) to allow for low logic depth implementations
- Transport layer:
 - ★ Normal mode: Send-and-Forget (SaF) service
 - ★ Reliability mode: Acknowledgement-and-Retransmit (AaR) service
 - * window size N , $1 \leq N \leq 64$
 - * $2N$ packets are buffered in sender and receiver
 - * End-to-end flow control mechanism
- in total 8 modes available



Error Protection for Low Power

Scenario I:

- 8×8 network
- 80 bits payload
- 15 bits header

Scenario II: Link layer error protection

- Block code with DED/SEC capability
- 20 payload bits and 5 protection bits per block;
- 80 payload bits
- 15 header bits
- 30 protecting bits
- 125 total bits

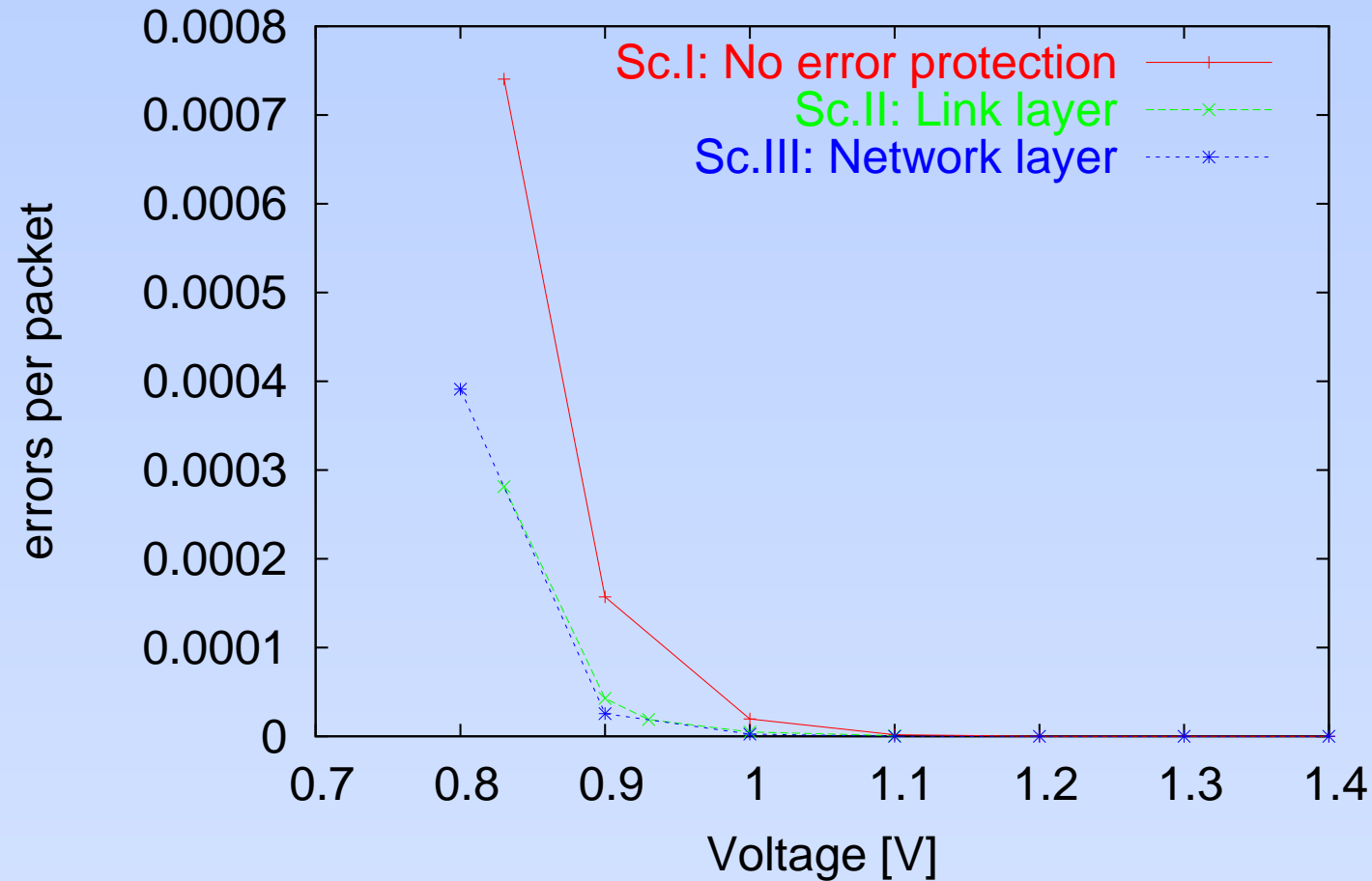
Scenario III: End-to-end protection

- Header is protected at the link layer as in Scenario II
- Payload is protected by a block code with SEC/DED capability
- 80 payload bits
- 15 header bits
- 24 protecting bits
- 119 total bits



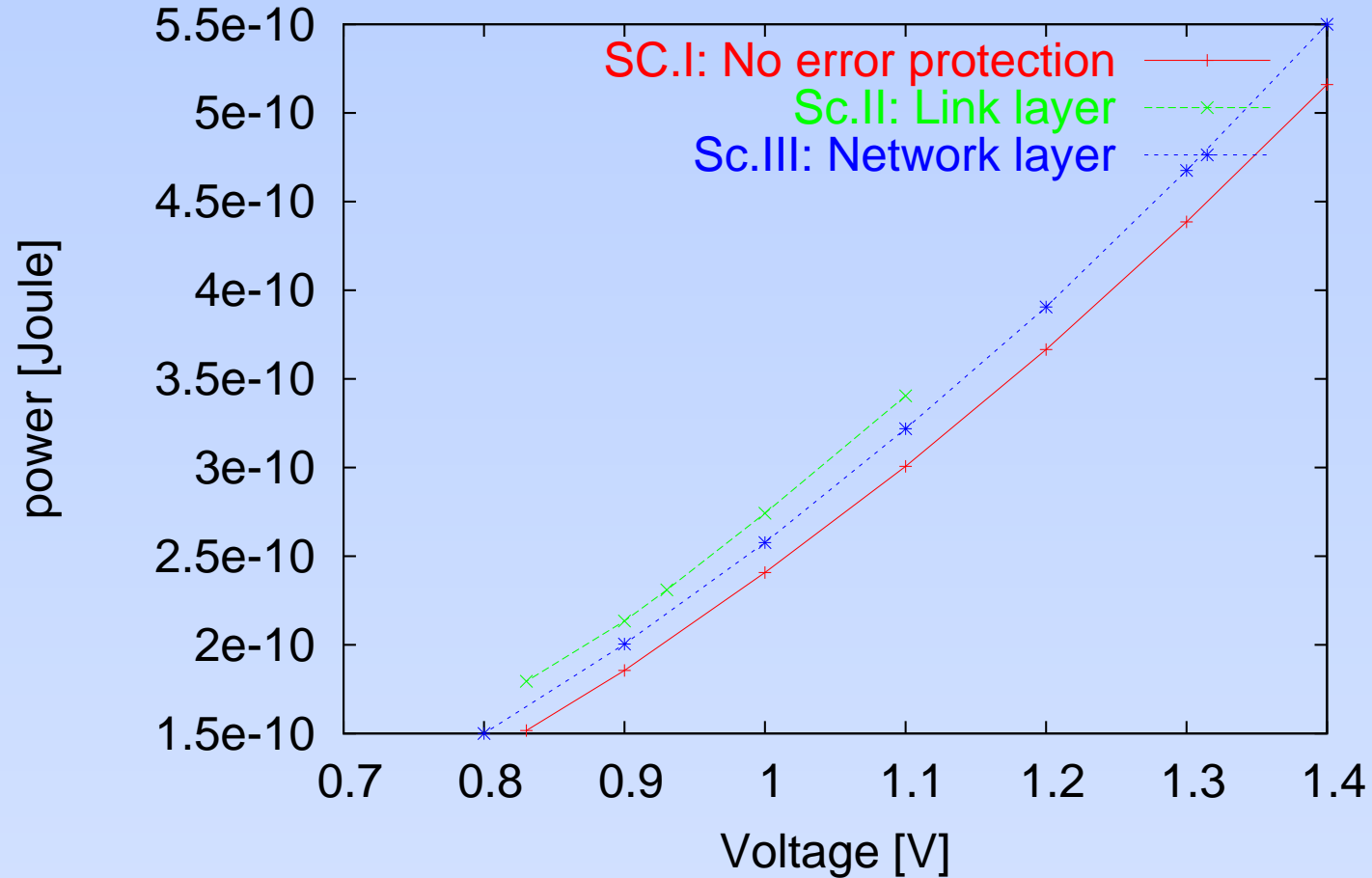
Errors per Packet

Errors per packet depending on the voltage

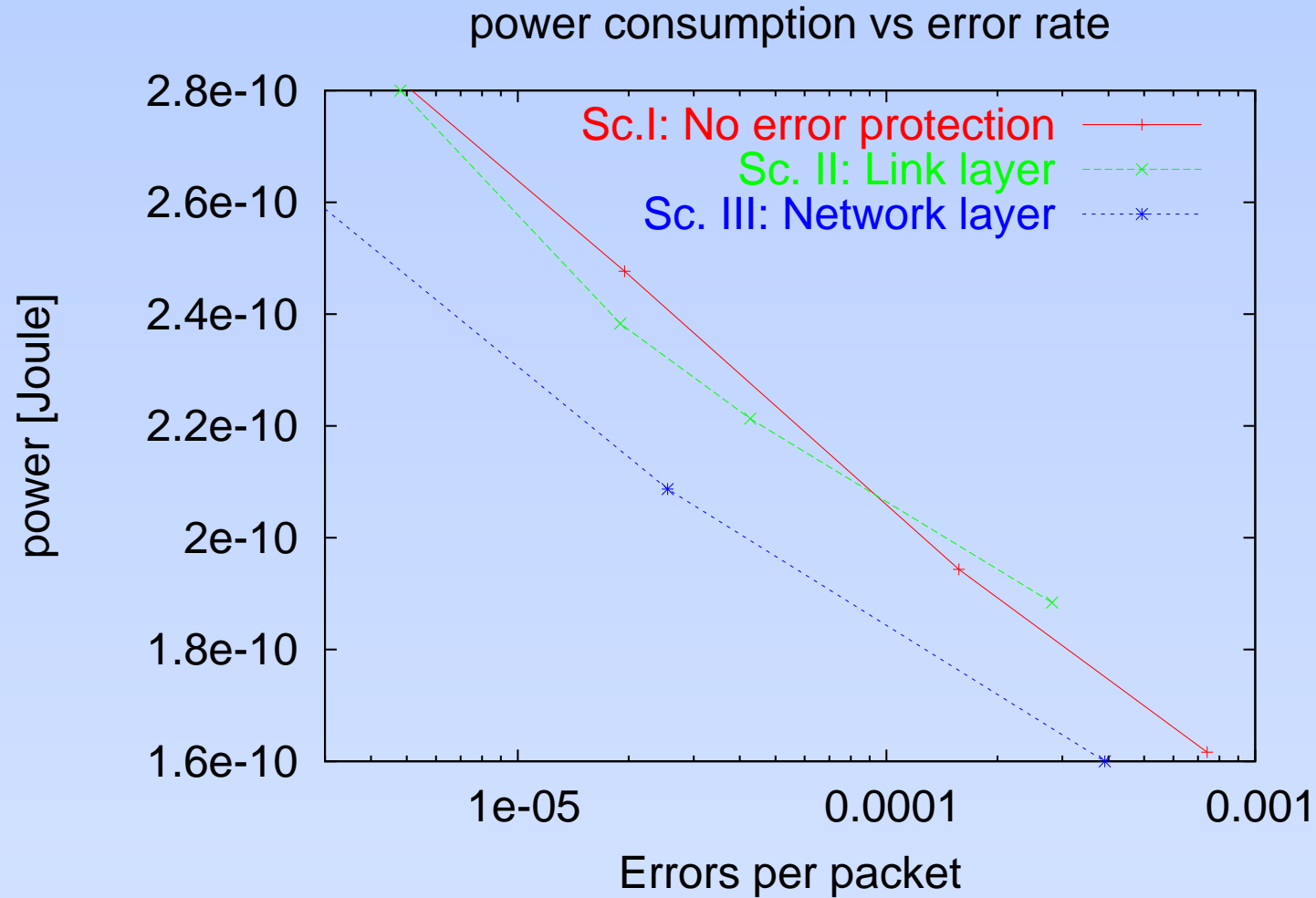


Power Consumption per Useful Bit

Power consumption per useful bit depending on the voltage



Power Consumption vs. Error Rate



Low Power encoding - Conclusion

- Low power bus encoding is of limited value and probably increases the overall power consumption.
- Link-level error protection to allow for lower voltage does not give significant improvements.
- End-to-end data protection decreases power consumption for 8×8 networks, with slowly increasing gain for larger networks.



Overview

Topology and Structure

Protocol Stack

The Network Layer and the Switch

Data Protection

Clocking

Dynamic Voltage Scaling

Network Simulator



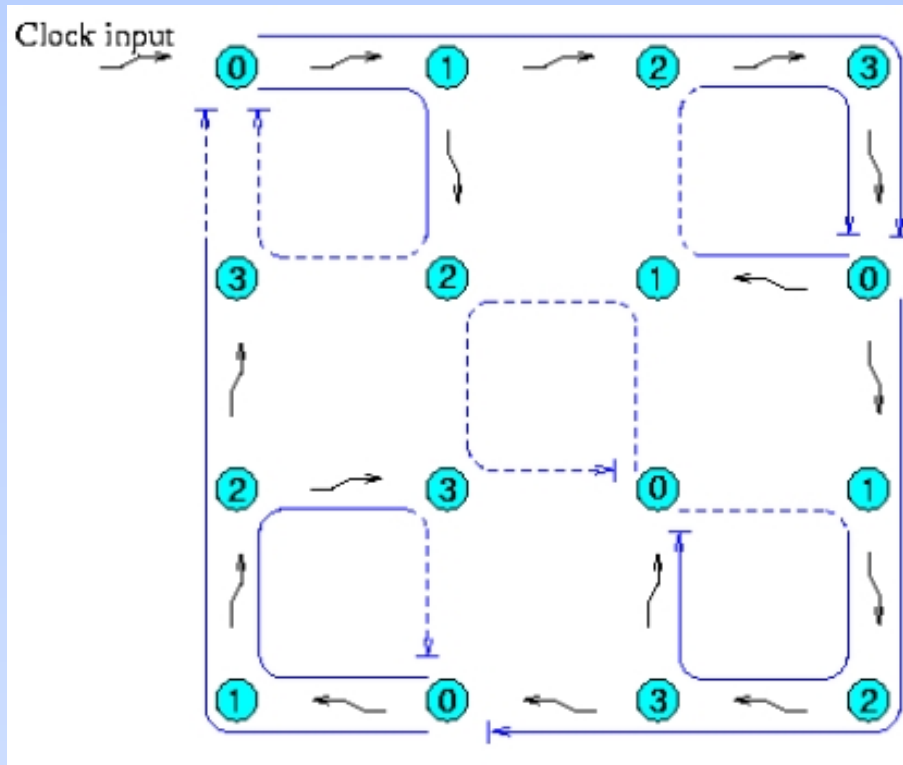
Globally Pseudosynchronous - Locally Synchronous Clocking

Every switch uses same frequency; phase difference is constant and known.

- Latency reduce with 29% at low load; 40% at high load
- Can handle 10% higher load
- More skew tolerant
- Clock skew and jitter is depending only on local constraints
- No global clock distribution with associated power gains
- Reduced peak power with 50% at best
- Jitter reduced significantly

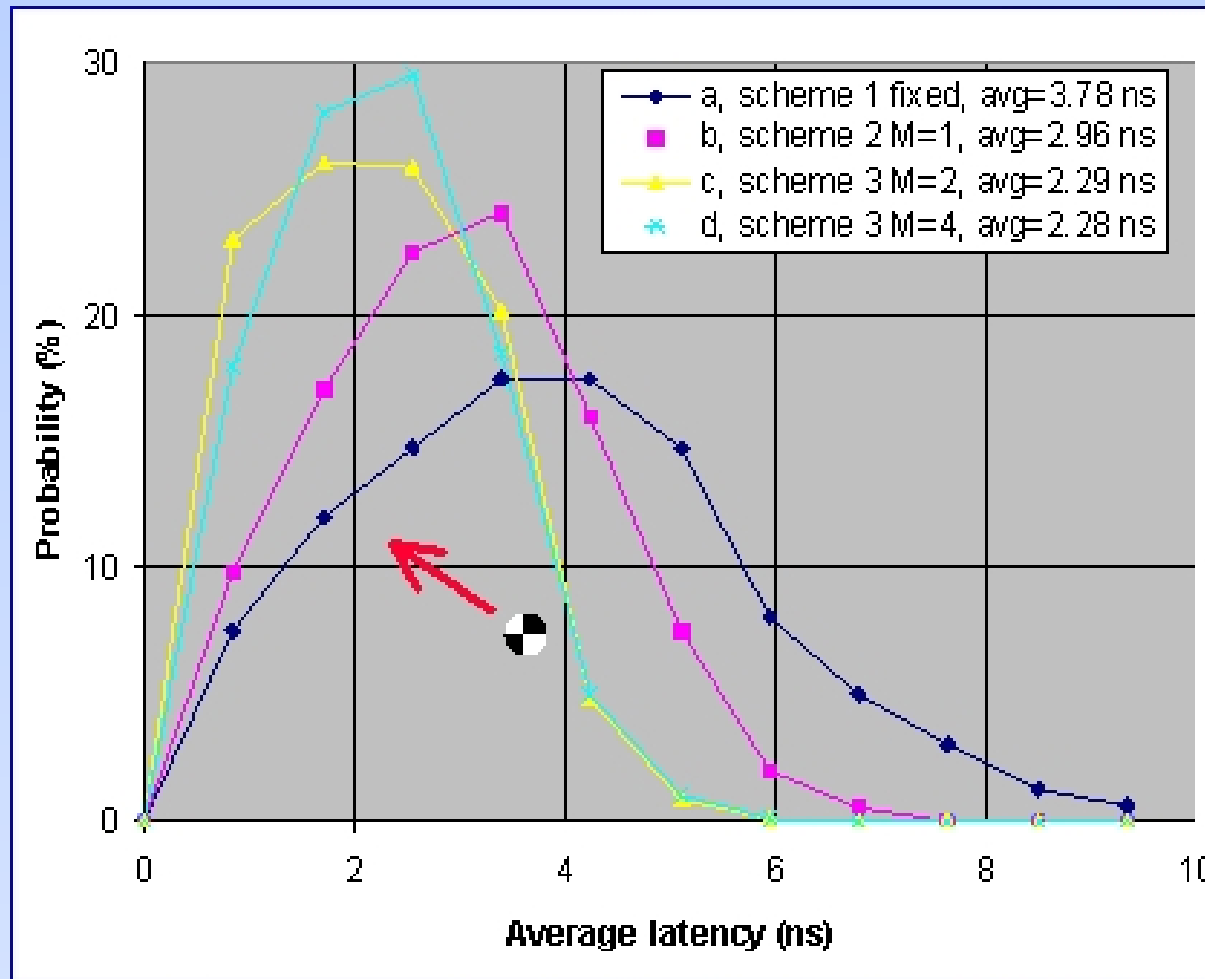


Globally Pseudosynchronous Clocking - cont'd



- Downstream data create low latency paths (Data Motorways)
 - ★ Guaranteed data motorways
 - ★ Phase related data motorways
- Periphery roundtrip:
 - ★ 14 cycles downstream
 - ★ 21 cycles upstream
 - ★ 24 cycles synchronous

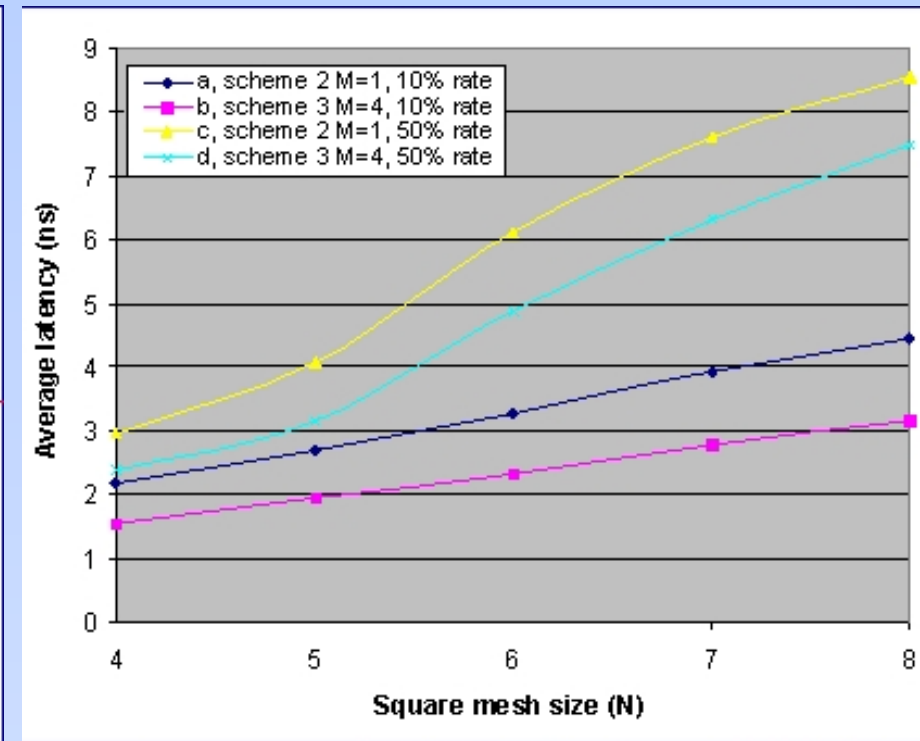
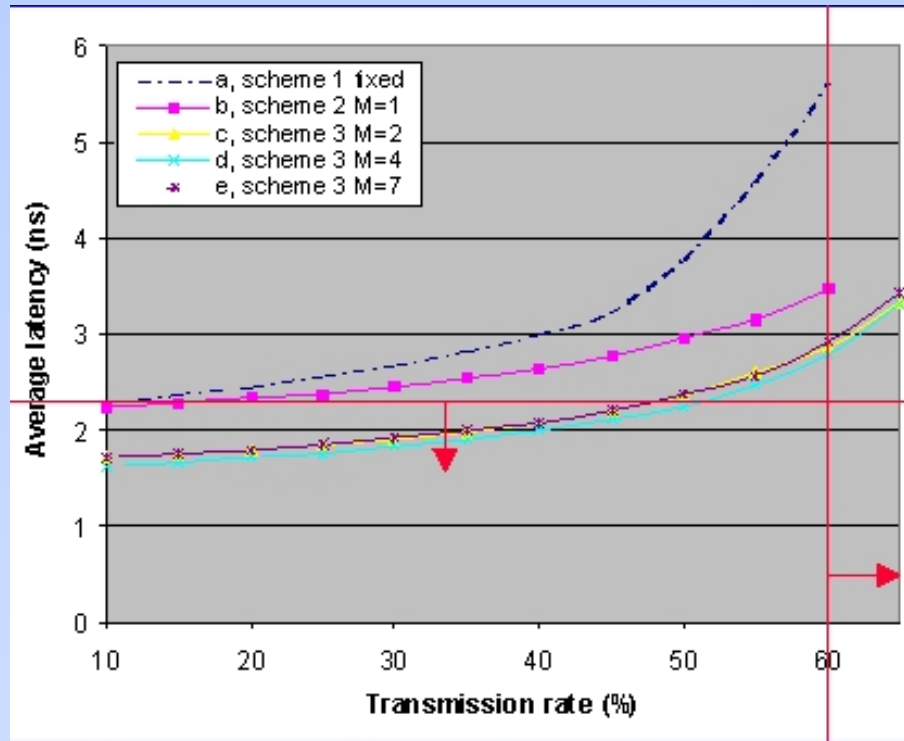
Globally Pseudosynchronous Clocking - cont'd



4x4 network, 50% emission rate

- (a) Synchronous clocking, uniform routing
- (b) Synchronous clocking, centrifugal routing
- (c) pseudosynchronous clocking, 2 phases
- (d) pseudosynchronous clocking, 4 phases

Globally Pseudosynchronous Clocking - cont'd



Overview

Topology and Structure

Protocol Stack

The Network Layer and the Switch

Data Protection

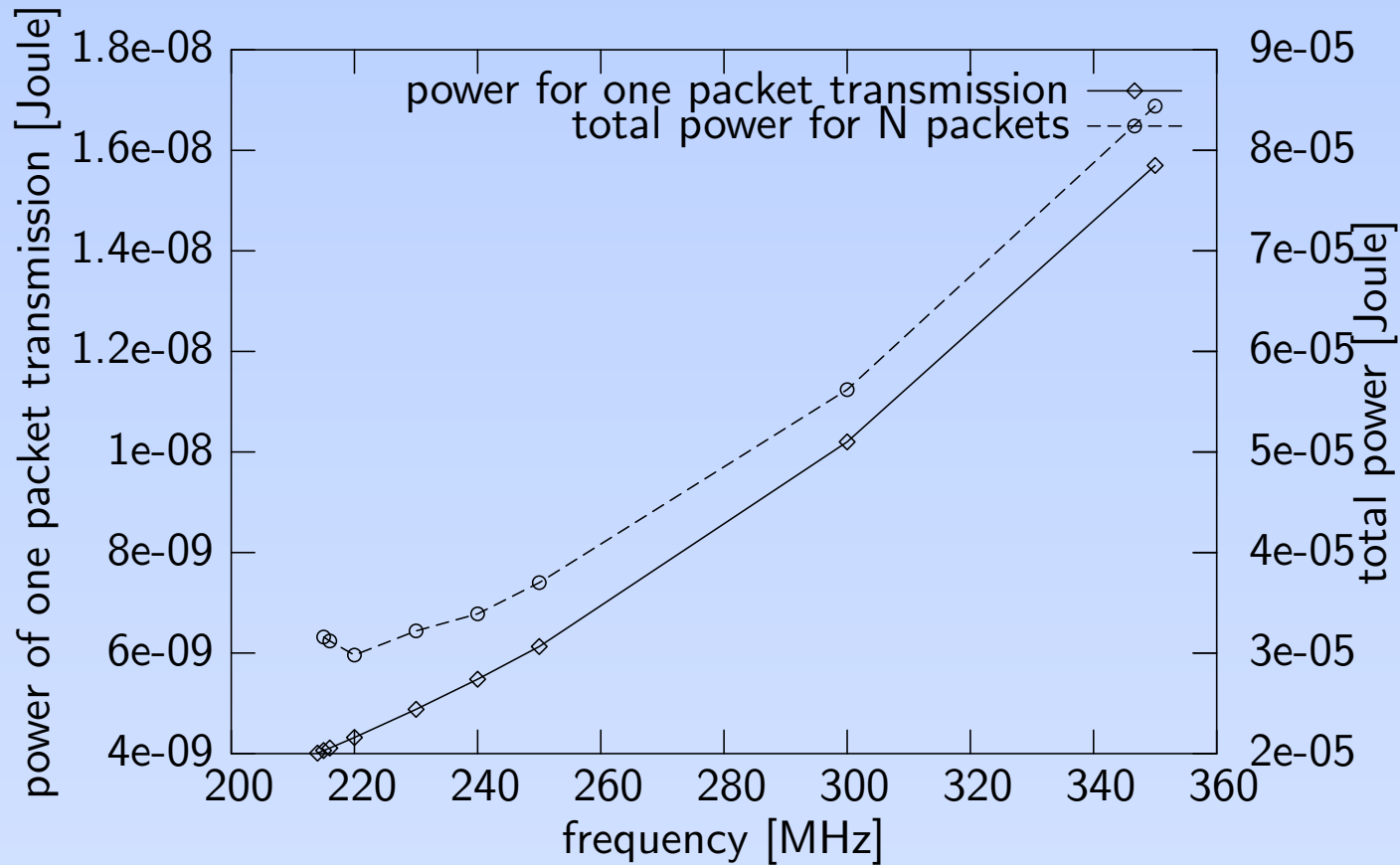
Clocking

Dynamic Voltage Scaling

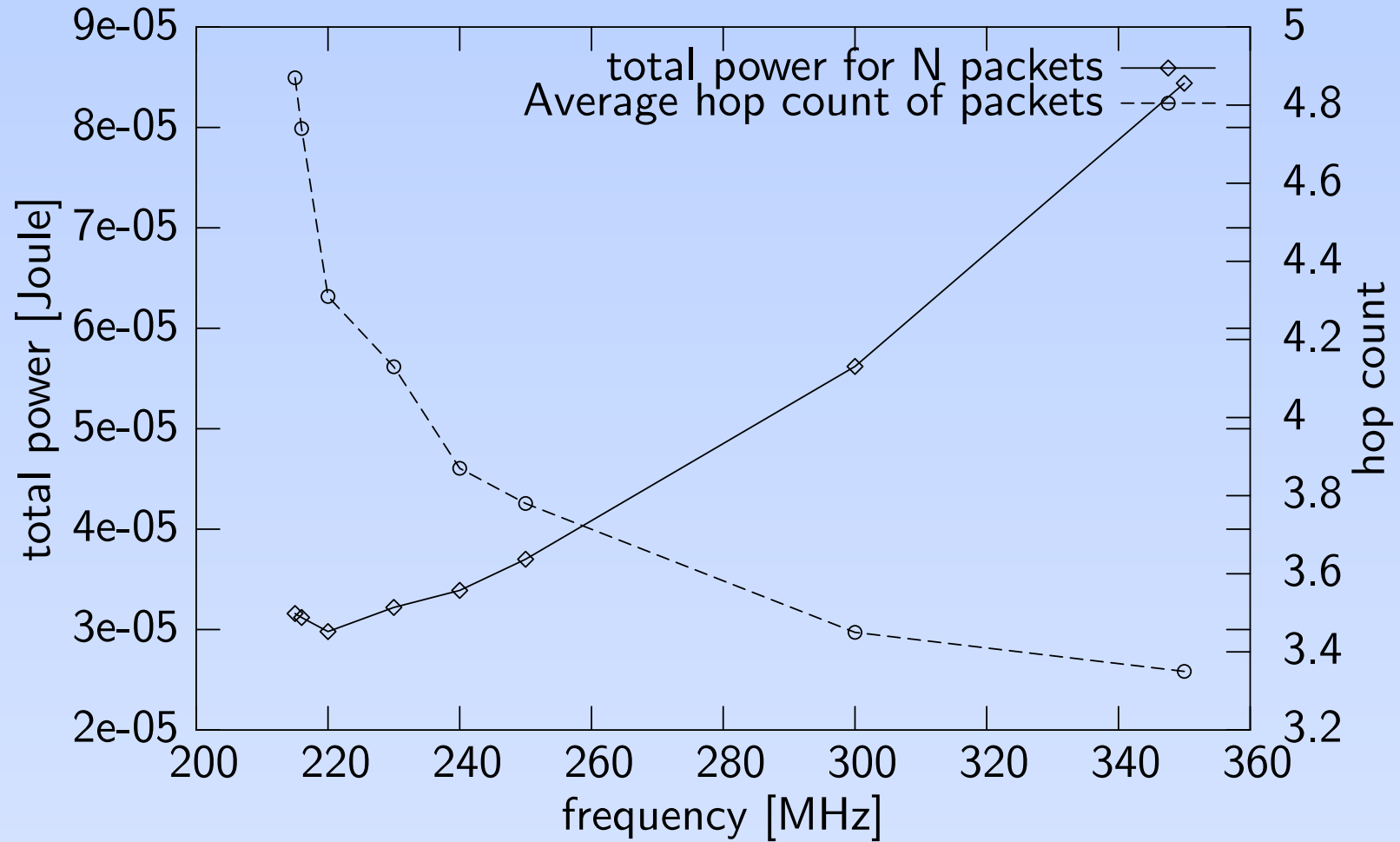
Network Simulator



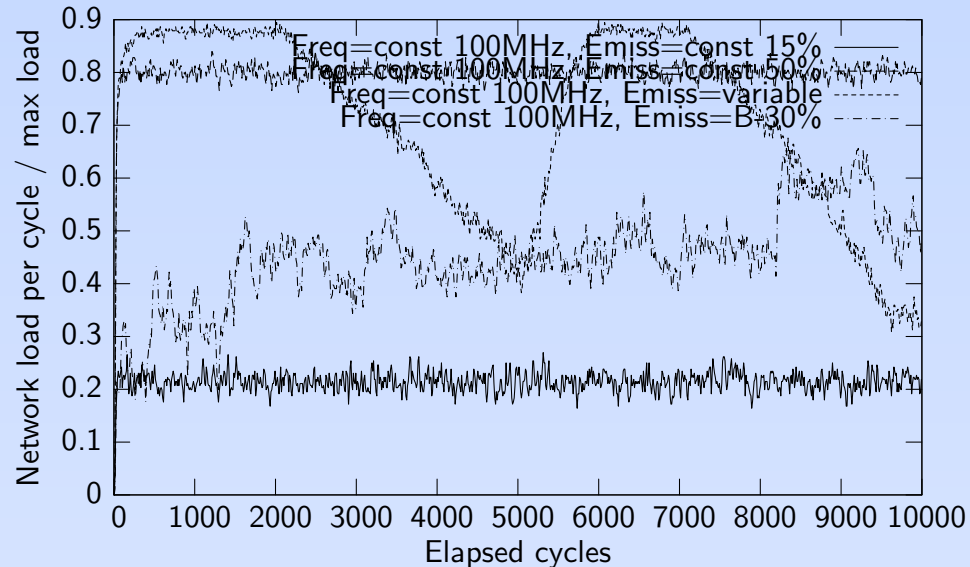
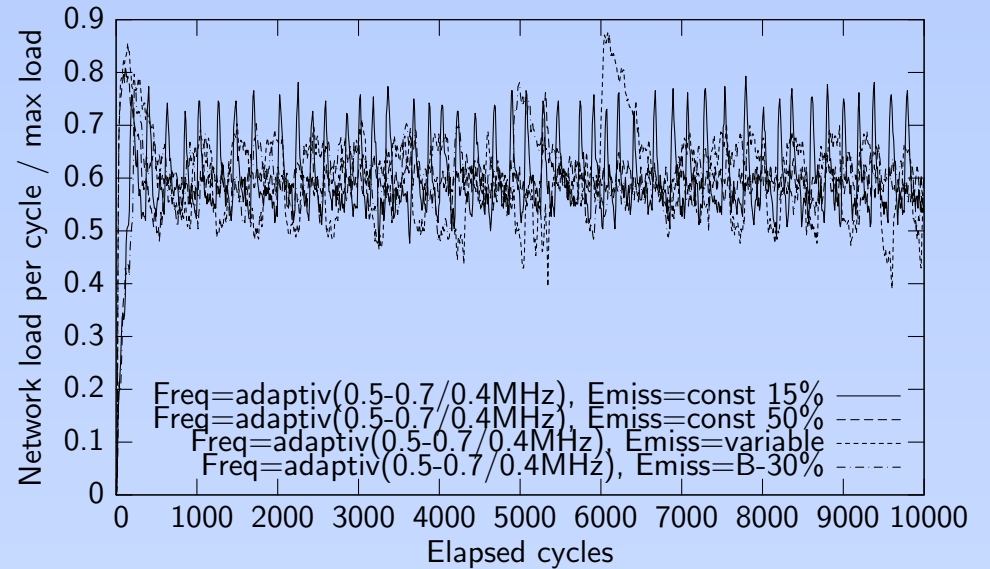
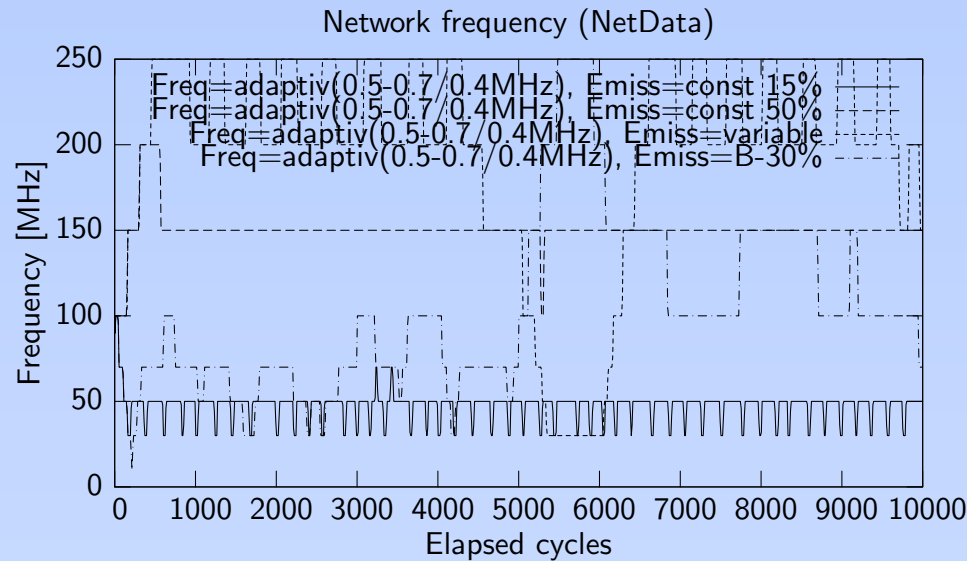
Potential of Dynamic Voltage Scaling - Power



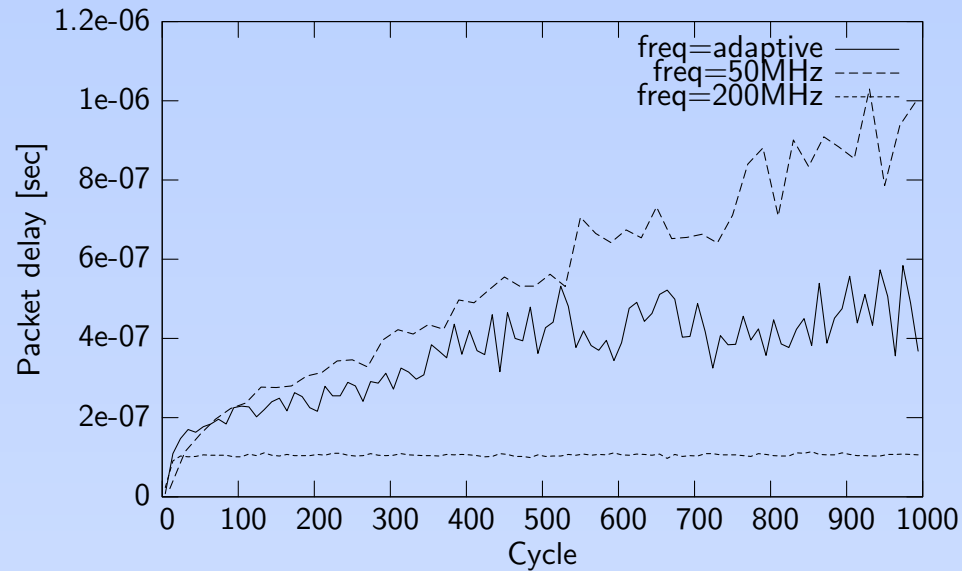
Potential of Dynamic Voltage Scaling - Hops



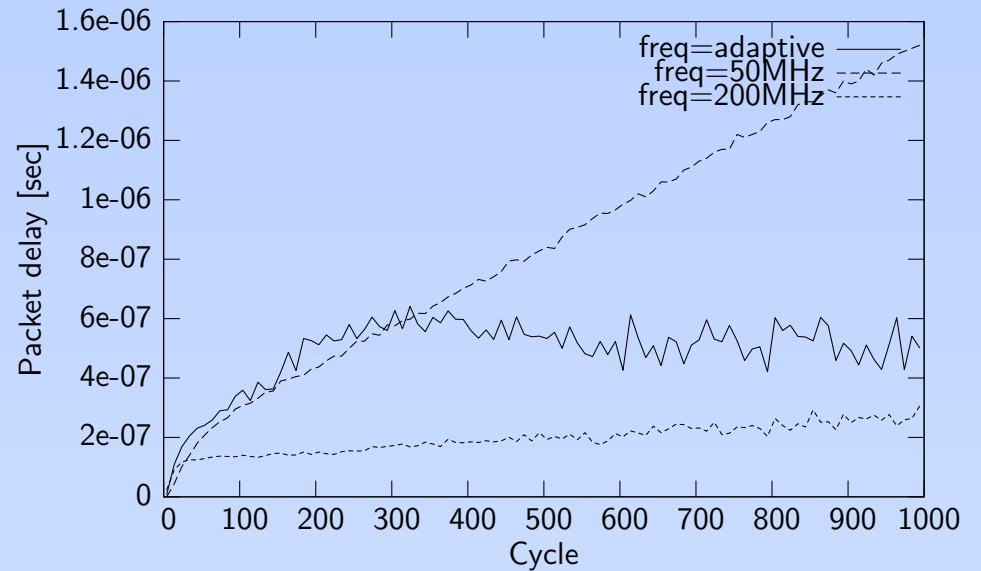
Potential of Dynamic Voltage Scaling - Load variations



Potential of Dynamic Voltage Scaling - Delay

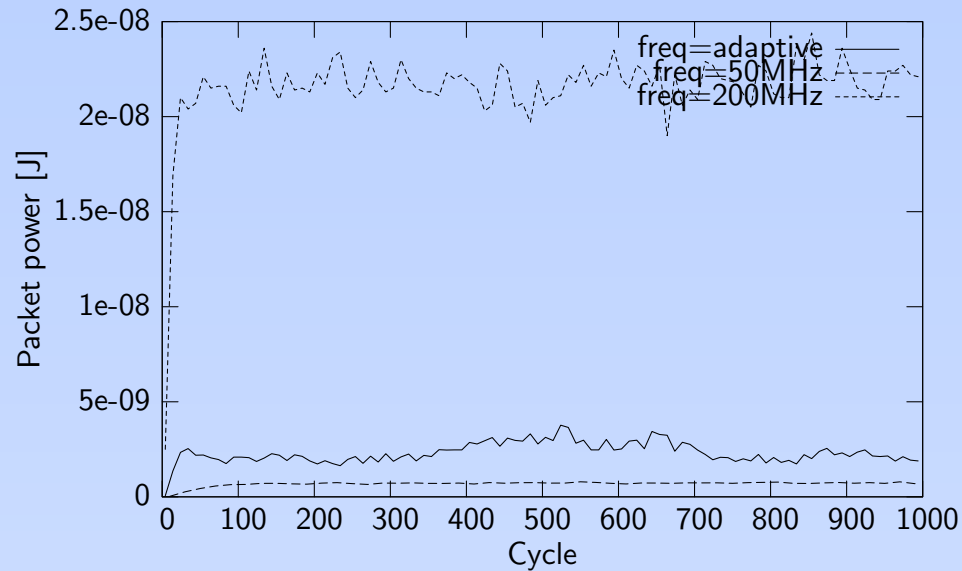


Emission probability = 30%

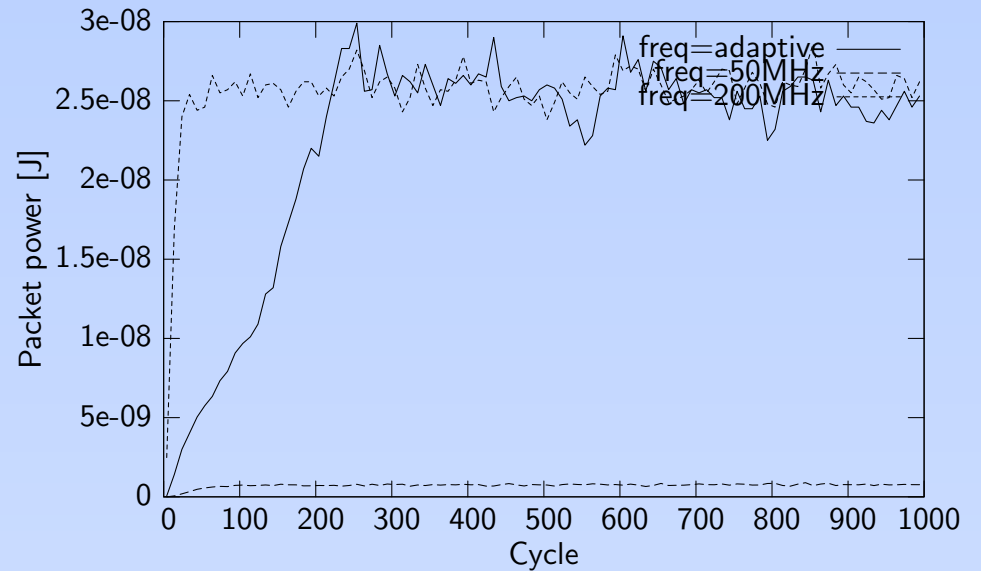


Emission probability = 70%

Potential of Dynamic Voltage Scaling - Power



Emission probability = 30%



Emission probability = 70%

Overview

Topology and Structure

Protocol Stack

The Network Layer and the Switch

Data Protection

Clocking

Dynamic Voltage Scaling

Network Simulator



Nostrum Simulation Environment

- Nostrum NoC Simulation Environment (NNSE)
- Based on SystemC simulator
- Configuration parameters:
 - ★ Size
 - ★ Topology (Mesh, Torus)
 - ★ Switching and routing (deflective, wormhole)
 - ★ Traffic pattern (temporal, spatial, random, locality, per channel, ...)
 - ★ Analysis plots (delay, load, power, ...)
- Useful to analyse the zillion trade-offs in NoC design.



NNSE Network Configuration

Project Network Traffic Simulation Help

Project: testproject.p

Configured Network:
testproject.xml

Configured Traffic:
traffic-1.xml
traffic-2.xml

Evaluation Results:

Network and Traffic configuration files

Network Configuration

Network topology

Number of nodes on X [2, 8]: Choose structure:

Number of nodes on Y [2, 8]: Choose connection:

Link bandwidth (data bits):

Deflection routing **Wormhole routing**

Deflection routing

Routing algorithm:

Deflection policy:

Wormhole routing

Number of VCs per PC [2, 4]:

Number of buffers per VC [2, 8]:

Routing algorithm:

Menu status:



NNSE Traffic Configuration

Project Network Traffic Simulation Help

Project: testproject.p

Configured Network:
testproject.xml

Configured Traffic:
traffic-1.xml
traffic-2.xml

Evaluation Results:

Network and Traffic configuration files

Traffic Configuration

Spatial specification

Distribution: Locality

Distribution specification

Array of source nodes: All

Array of destination nodes: All

Array of locality factor: 1

Temporal specification

Distribution: Normal

Inter-arrival Time Specification

Mean Interarrival: [3, 5, 7, 9]

Standard Deviation: [0, 0, 0, 0]

Menu status:



Summary of Nostrum Status

- Nostrum defines a 2 D mesh topology;
- Protocol stack for link layer, network layer and session layer;
- Packet switched and virtual circuit communication services;
- Buffer-less, loss-less switch with no routing tables;
- 2 level data protection scheme;
- Session layer communication primitives;
- Flexible NoC Simulator;

Further information: www.imit.kth.se/info/FOFU/Nostrum/

