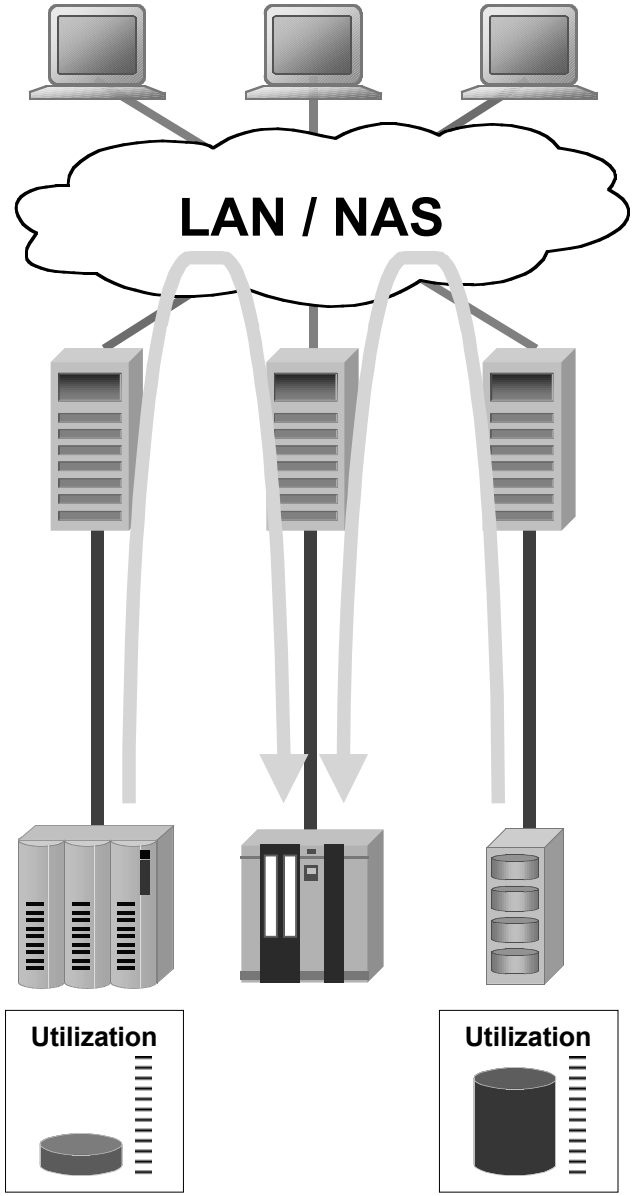
NISHAN SYSTEMS

Tyrant: A High Performance Storage over IP Switch Engine

Stuart Oberman,
Rodney Mullendore, Kamran Malik,
Anil Mehta, Keith Schakel,
Michael Ogrinc, Dane Mrazek

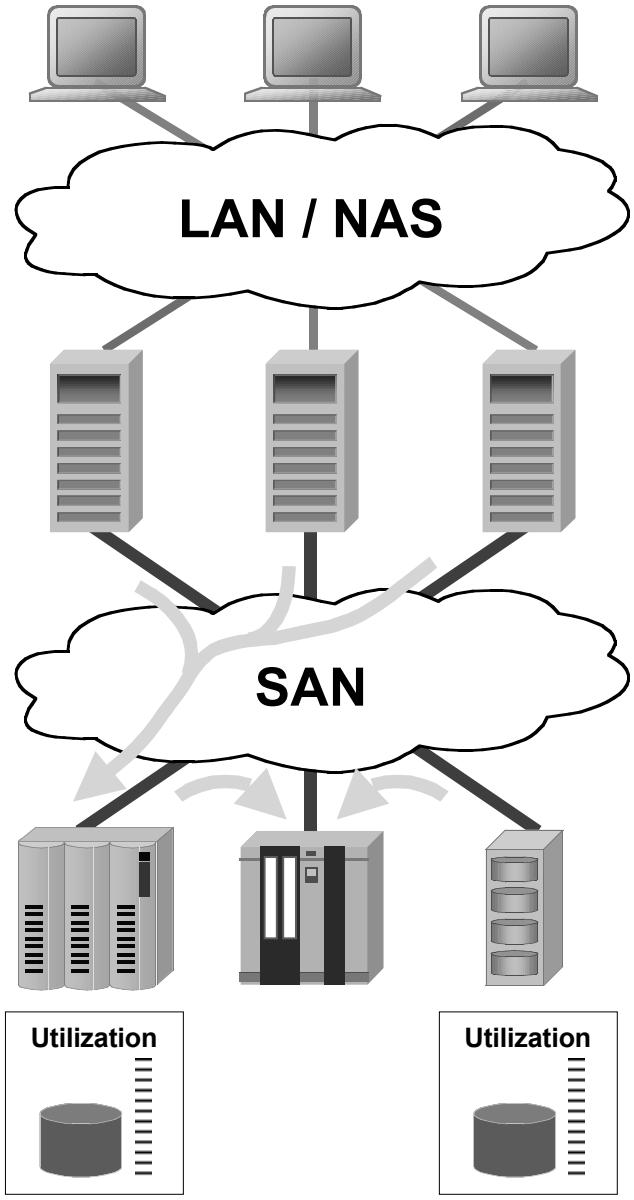
Background: Direct-Attached Storage



Direct-Attached Storage

- Expansion beyond server's internal drive capacity
- Tape backup
- Often used with Network-Attached Storage (NAS)
- High performance SCSI or Fibre Channel connections

Storage Area Networks – SANs

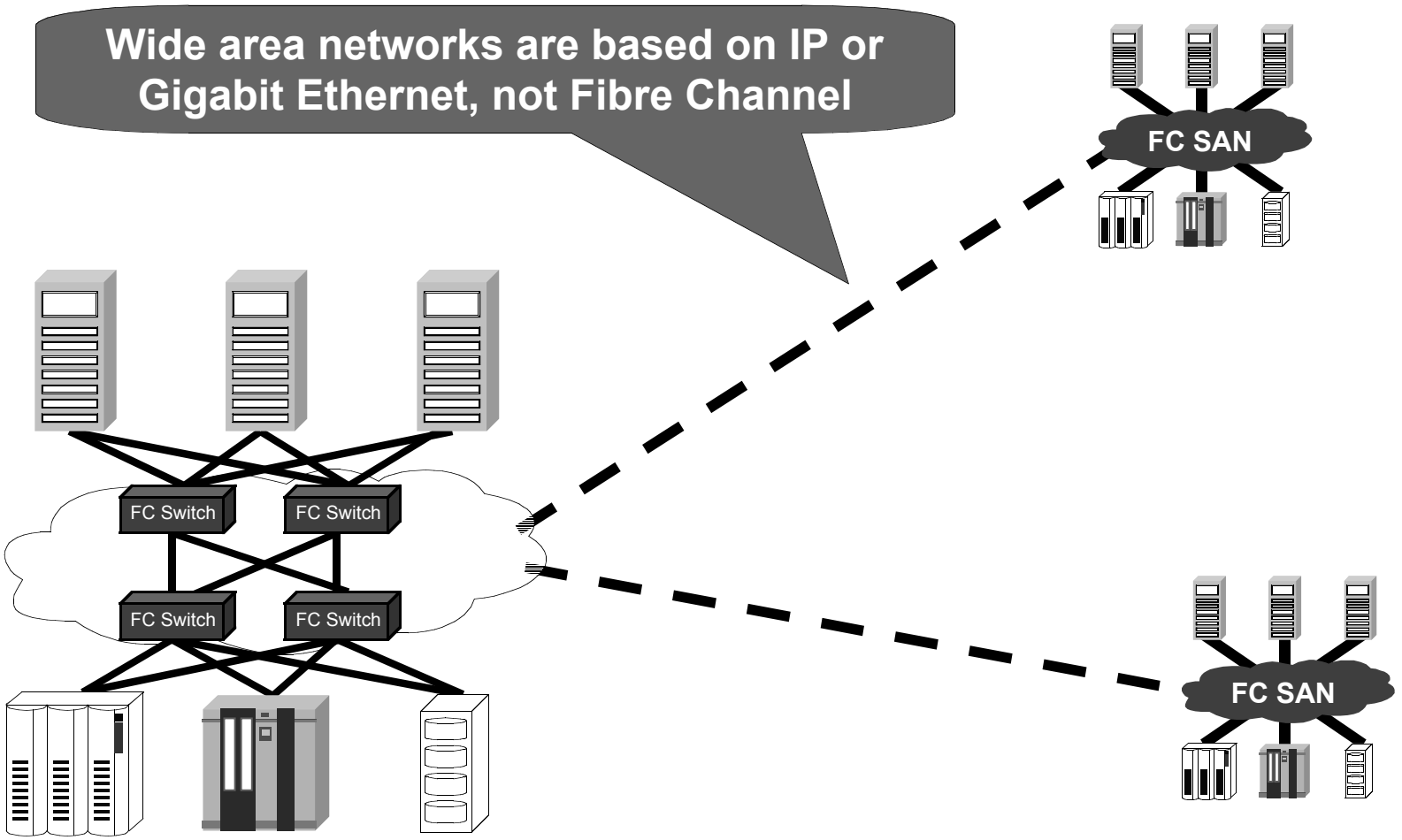


Storage Area Networks

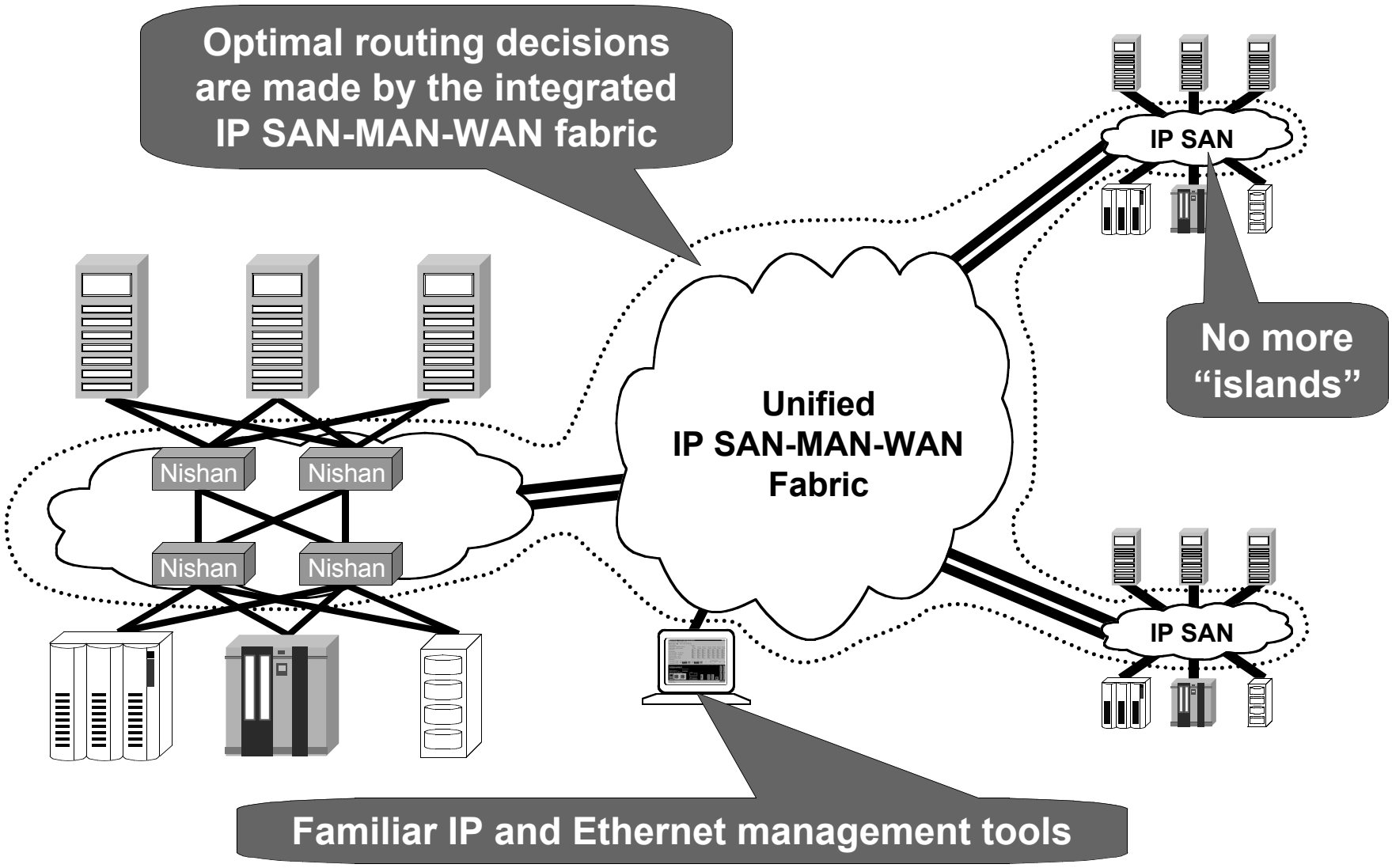
- Pooling of external storage devices for better utilization and availability
- LAN-free backup
- Serverless backup
- Non-disruptive expansion and maintenance
- Leverage existing staff to manage three or four times more storage
- SAN ROI estimates* range from 65-297 percent

**Source: CSFB, June 2001*

No Native Fibre Channel Extension



Integrated IP SAN, MAN, and WAN Fabric



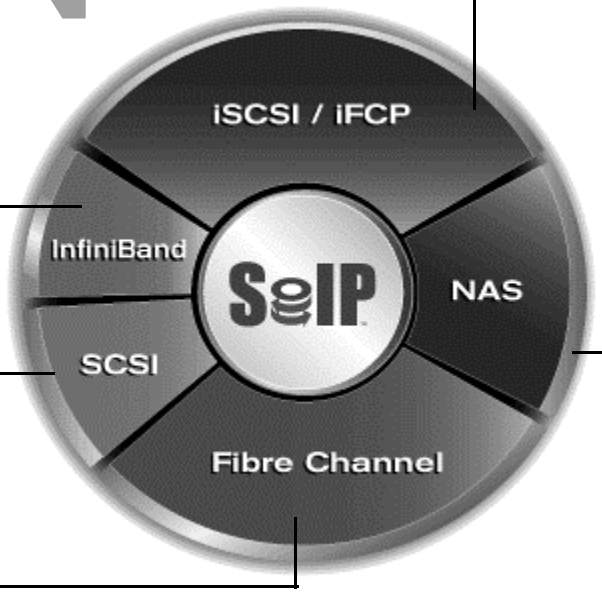
NISHAN SYSTEMS

Nishan Systems' Framework for IP Storage

OPEN

- SAN, LAN, MAN, WAN
- IP Routers and Gigabit Ethernet Switches
- 10 Gigabit Ethernet

- Future Interfaces



- SCSI Disk, RAID, Tape
- SCSI Host Bus Adapters

- Network Attached Storage

- FC Disk, RAID, Tape
- FC Host Bus Adapters
- FC SANs

Standards-Based

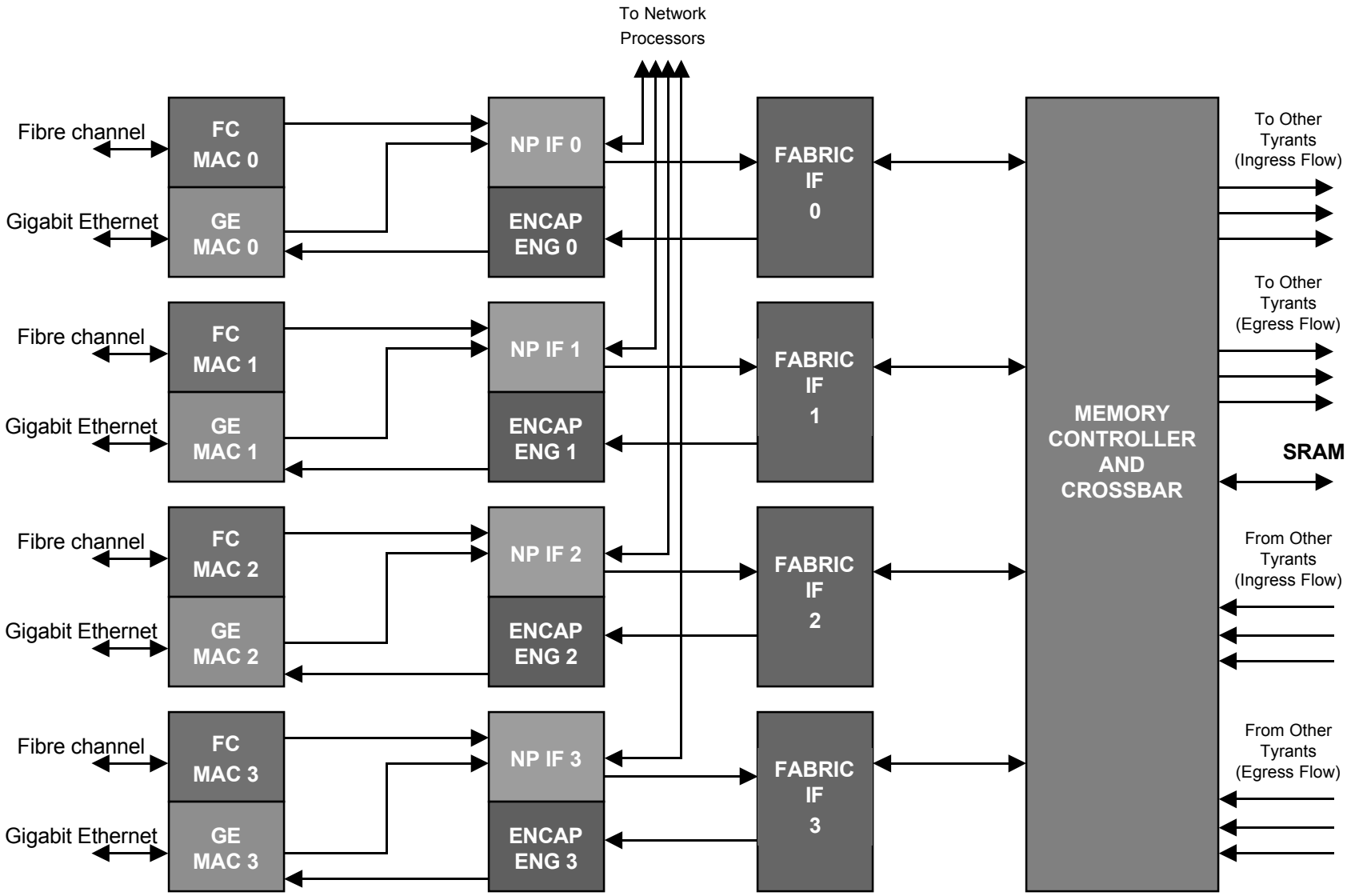
Tyrant: An Introduction

- Tyrant: a switch engine used to build many sizes of IP Storage Area Network switches
- Two categories of switches:
  - Low density: <=32 ports
  - High density: > 32 ports
- Tyrant integrates hardware to build both categories supporting
  - Fibre Channel switch ports, 1Gbps and 2Gbps
  - Fibre Channel arbitrated loop ports, 1Gbps and 2Gbps
  - Gigabit Ethernet
  - Switching between all port types

Tyrant: Major Features

- Four 1Gbps Ethernet MACs
- Four 1Gbps / 2Gbps Fibre Channel MACs
  - 1Gbps or 2Gbps
  - FC-AL and FC-SW
- Four Gigabit network processor interfaces
- Four flexible packet schedulers
- Four programmable encapsulation engines
- Distributed shared-memory controller supporting 2MB SRAM per chip
- Chip-to-chip fabric interface

Tyrant Block Diagram



Cut-thru

- Each Tyrant has side-band cut-thru request / grant interface
- Allows for each ingress port to request cut-thru to any egress port
- If cut-thru is granted, packets are switched in distributed fashion, with no writes to SRAM
- First-byte-in to MAC to first-byte-out of MAC latency:
  - Less than 2us
  - Any port to any port
  - FC packets using no external network processing

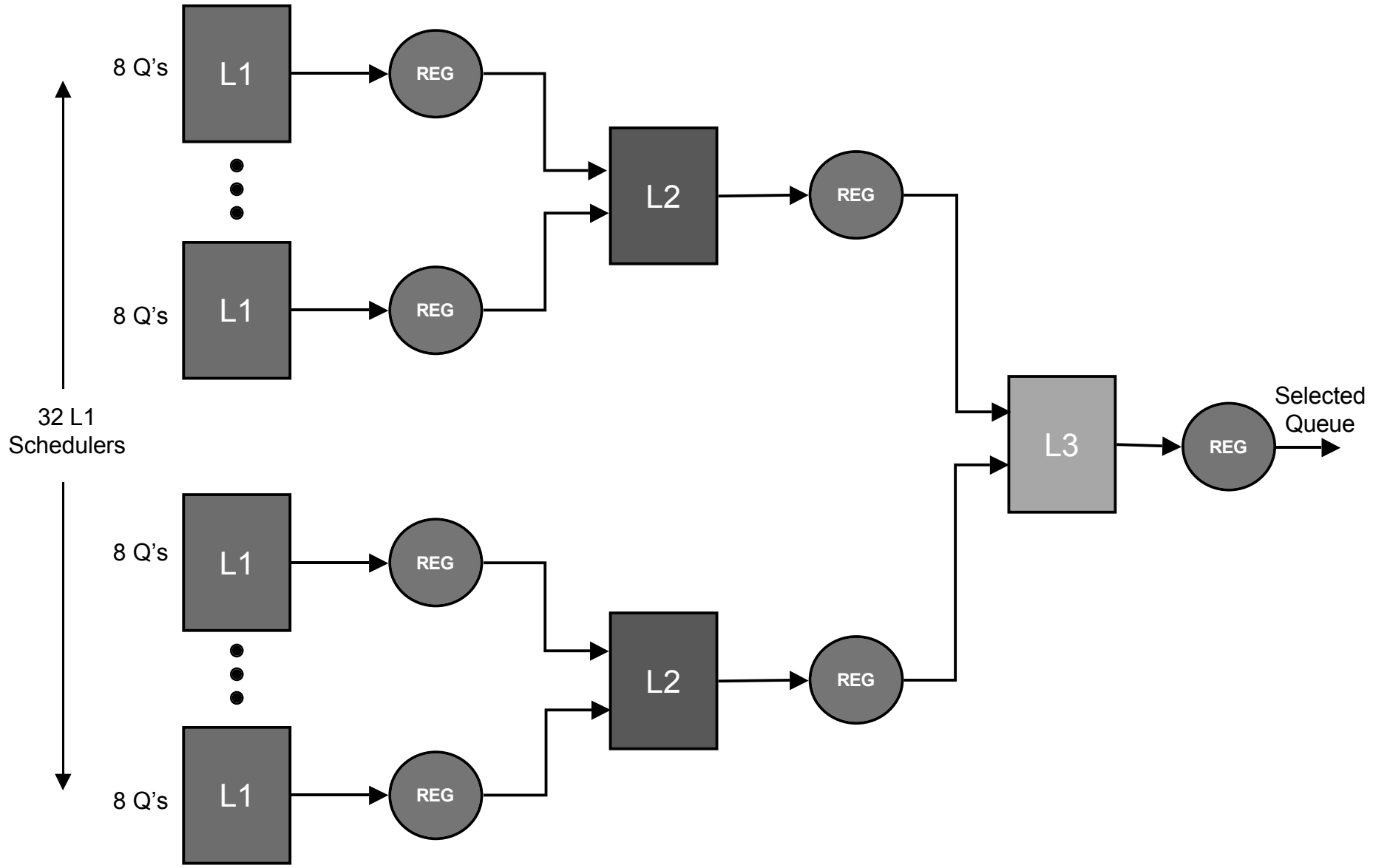
Early Forwarding

- Reduced switch latency possible even when packet stored to SRAM
- Entire store-and-forward not required
- Store only minimum amount of packet:
  - Programmable stored amount based upon link speed difference of ingress and egress ports (GE, FC, or 2G FC)
  - Ensures no underrun of egress port
- 4-5us latency possible

Packet Scheduler

- True output queues at each egress port
  - Requires fast and wide shared memory
  - Guarantees maximum switching throughput
  - Single trip to memory
- 256 unique output queues per port
- Scheduler features:
  - Three level hierarchy
  - Each level may separately use strict priority or weighted-fair-queuing (WFQ) / weighted-round-robin (WRR)
  - Virtual time based scheduling => very low jitter
  - Minimum and maximum bandwidth guarantees

Scheduler Organization



Full-Duplex FC-AL

- FC arbitrated loop supports 126 devices
- Loop operates either half-duplex or full-duplex
- In full-duplex configuration, target device may “open” switch port to send it data
- If switch port can reorganize its output data while receiving target device’s data, bandwidth is maximized

Optimized Full-Duplex FC-AL Performance

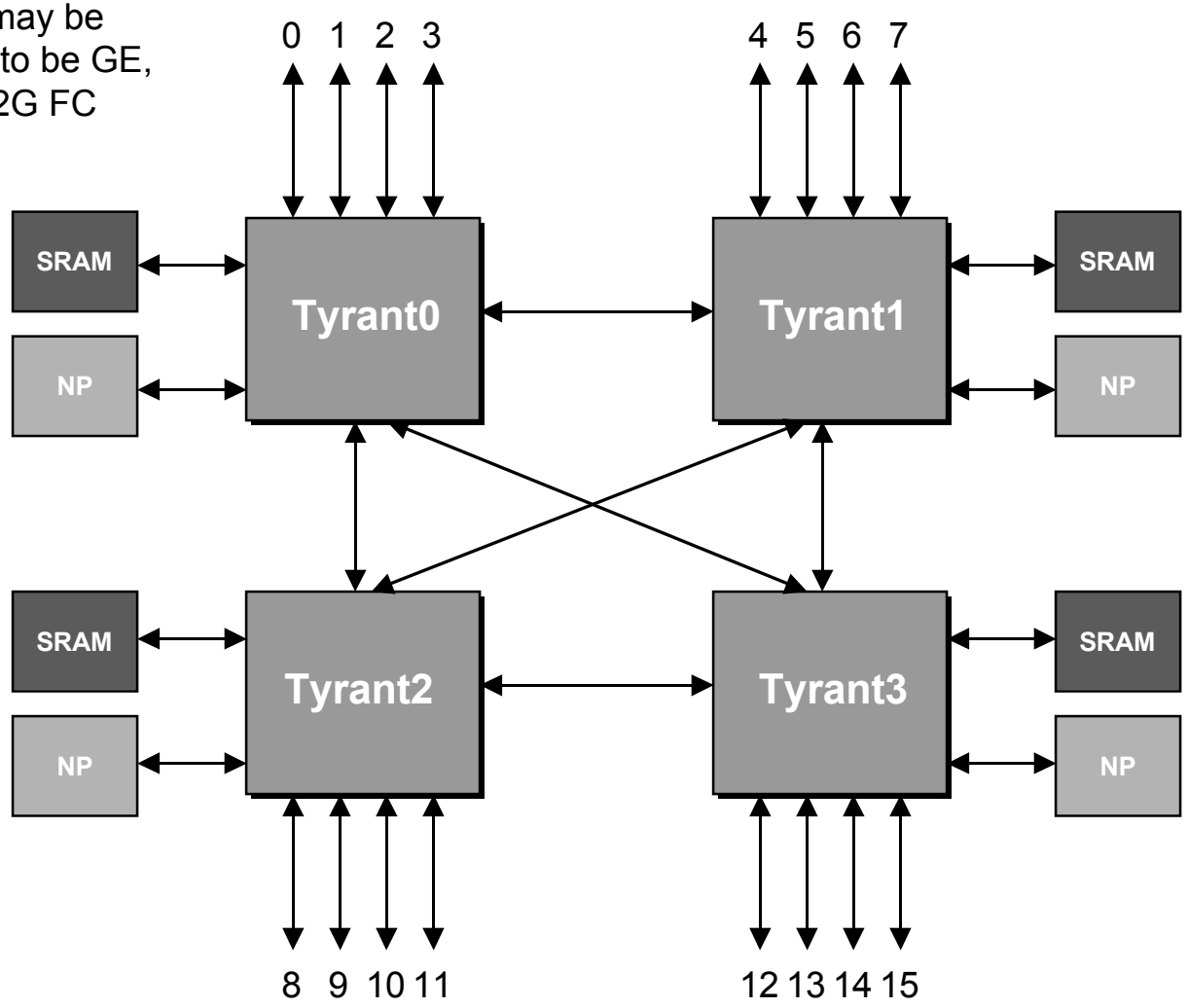
- Bin packets destined for different target devices using one output queue per loop device
- Separate packet scheduler interface allows FC MAC to directly select particular queue for transmission
- Switch port opened by loop device:
  - MAC “pulls” packet from correct queue
  - No head-of-line blocking

Programmable Encapsulation Engine

- Programmable encapsulation engine per egress port used to implement IP storage protocols
  - Allows flexibility for evolving IP storage standards
- Network processor on ingress FC port adds Encapsulation Instructions to header of packet
- Encapsulation engine at egress IP port executes microcode to form encapsulated IP storage packets
- Encapsulation engine performs:
  - IP address lookups
  - Checksum and CRC computations
  - GE and IP header creation

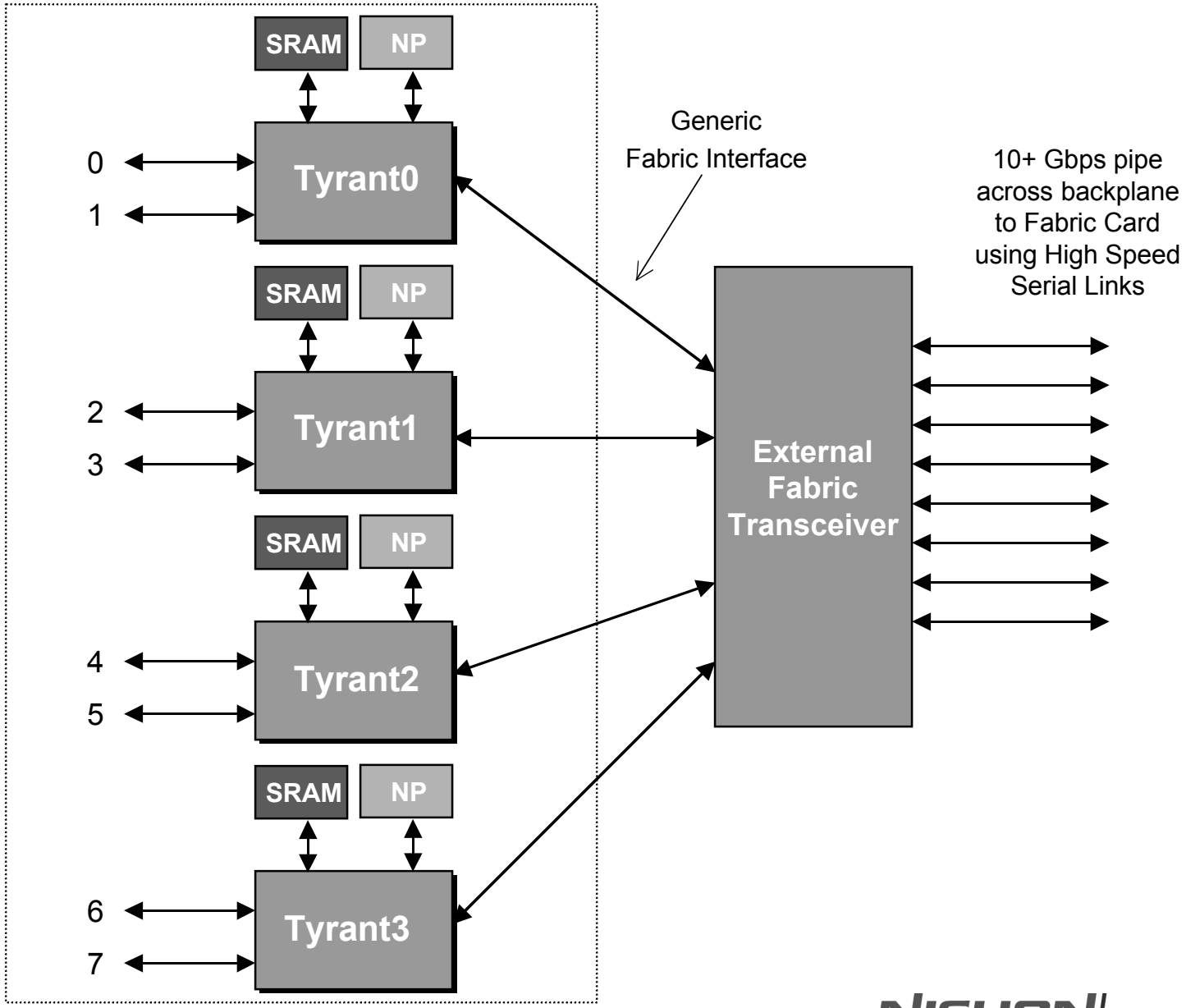
Application Example 1: 16-port Switch

Each port may be configured to be GE, 1G FC, or 2G FC



Application Example 2: Chassis Line Card

Each port may be configured to be GE, 1G FC, or 2G FC

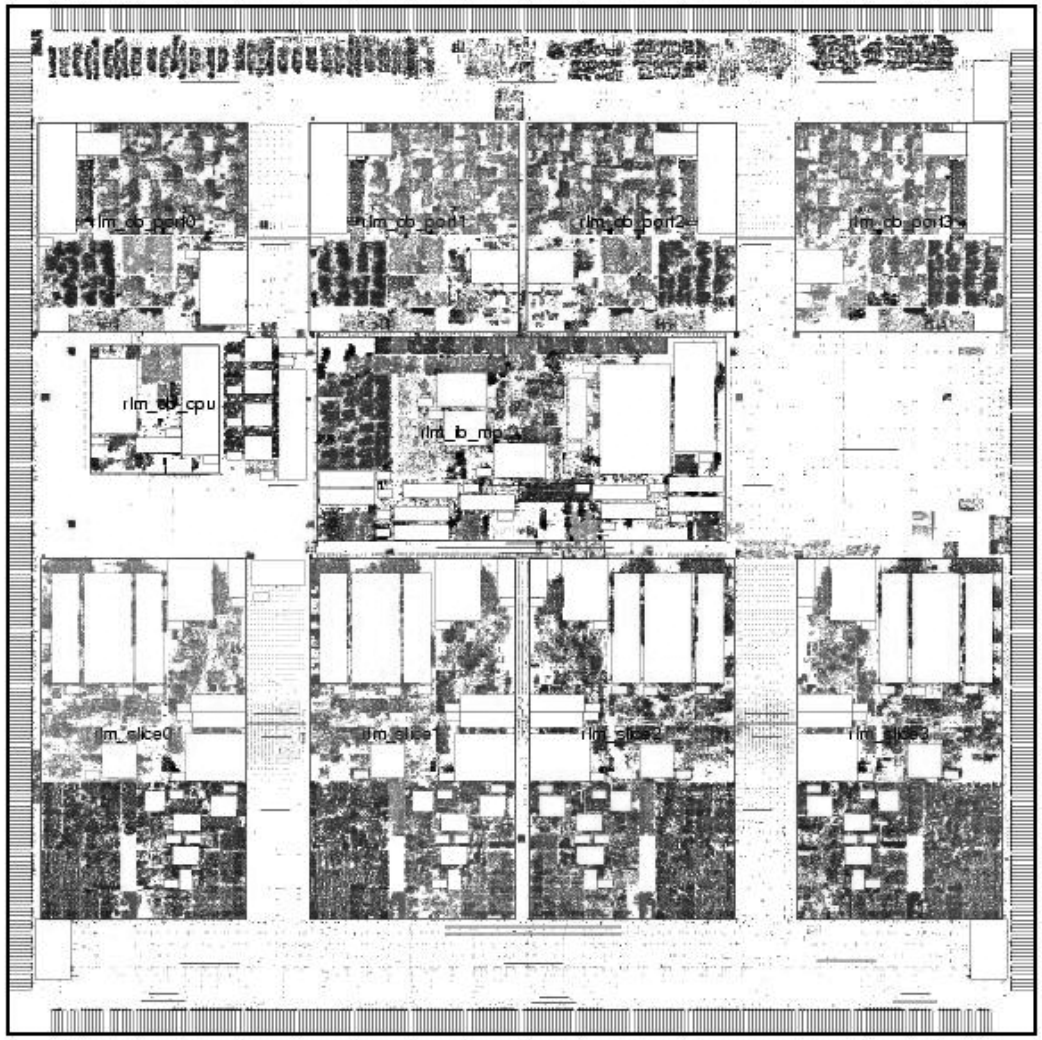


NISHAN SYSTEMS

Physical Details

- IBM 0.18um ASIC process: SA-27E
- 25 million transistors
- 15mm x 15mm die size
- 928 signal pins
- Nominal power consumption: 16W
- Three major clock domains
  - 125 MHz
  - 104 MHz
  - 53.125 MHz
- Implementation status: internal testing

Chip Layout



ty_rlm_top

NISHAN SYSTEMS

Tyrant Photo