Arquitectura CrayOrigin S^{2}MP

- Arquitectura S2MP: consideraciones previas
- Arquitectura convencional SMP
- Otras especificaciones
- Memoria compartida escalable + coherencia de caches
- Arqutectuta S2MP: idea básica
- Alternativas de interconexión
- Alternativas de interconexión: Bandwidht y Latencia
- S2MP Builing Block
- S2MP: switch technology
- I/O subsystem
- El módulo
- Modulos se convierten en sistemas
- Esquema arquitectónico de los sistemas
- Optimización con express links
- Analisis de la eficiencia

Arq. S²MP: consideraciones previas

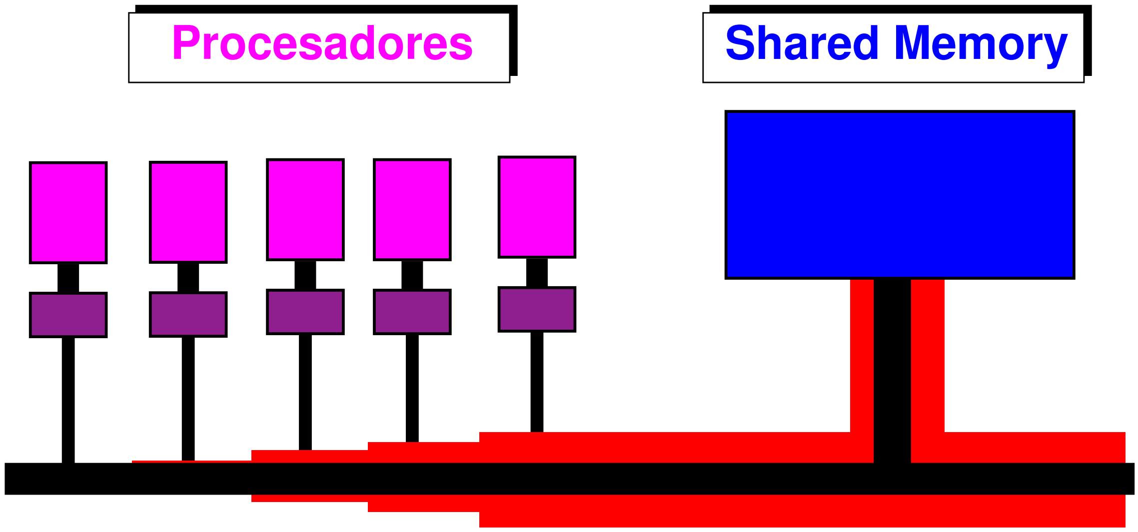
3

- **SMP**
- **MPP**
- **SIMD**

- PowerPath–2:   256 bit * 43 MHz = 1.3 GB/seg [1.2GB/s]
- 400 Mflop * 3var/flop * 8 Bytes/var ~= 1 GB/seg –> 200 MB/seg.
- SIMD...

Arquitectura convencional SMP

4



La arquitectura bus tiene problemas de escalabilidad porque:

– a partir de un cierto N el tráfico colapsa el bus (bandwidth)

– la cache coherency empieza a ser un problema

La latencia es superior en los sistema multiprocesador

Otras especificaciones

5

1. **Modelo de programación consistente y razonable .**
   Las aplicaciones existentes deben seguir funcionando.
   No se puede pedir perfectas aplicaciones en un mundo imperfecto.
   Algo de tuning puede, pero no un rediseño total para que funcione.

2. **Poder empezar desde algo muy pequeño.**
   Que las configuraciones pequeñas sean también competitivas.

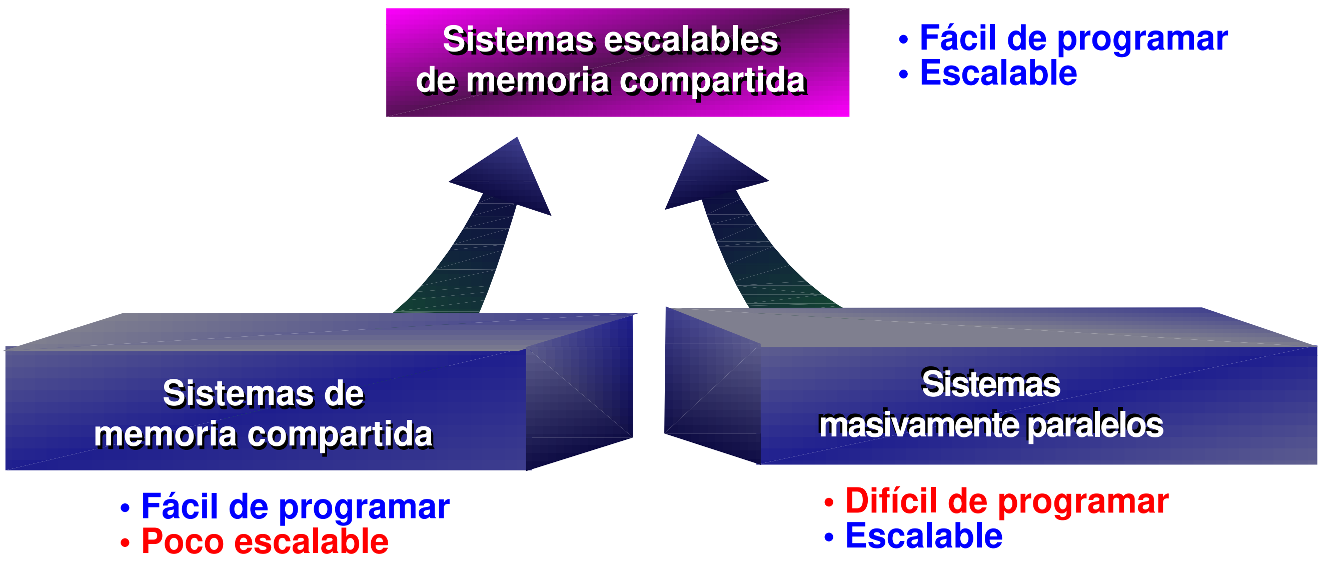
3. **Poder ampliar incrementalmente.**
   CPU, Memory, I/O independientes; poder hacer pequeñas ampliaciones.
   Para llegar a la máx. configuración sólo sea preciso añadir, no cambiar.

4. **Poder crecer hasta donde sea requerido.**
   "Escalabilidad" significa no tener que decir nunca "lo siento".

5. **Tiene que funcionar !**
   Estable, fácilmente instalable y con alto nivel de servicio.

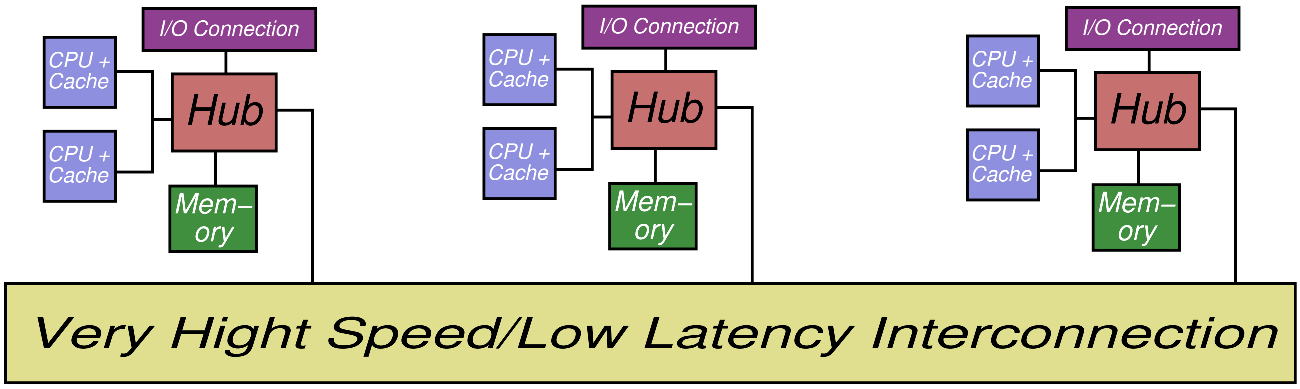
Memoria compartida escalable + coherencia de caches

6



Arquitectura S2MP: idea básica

7



- **PowerPath–2:**

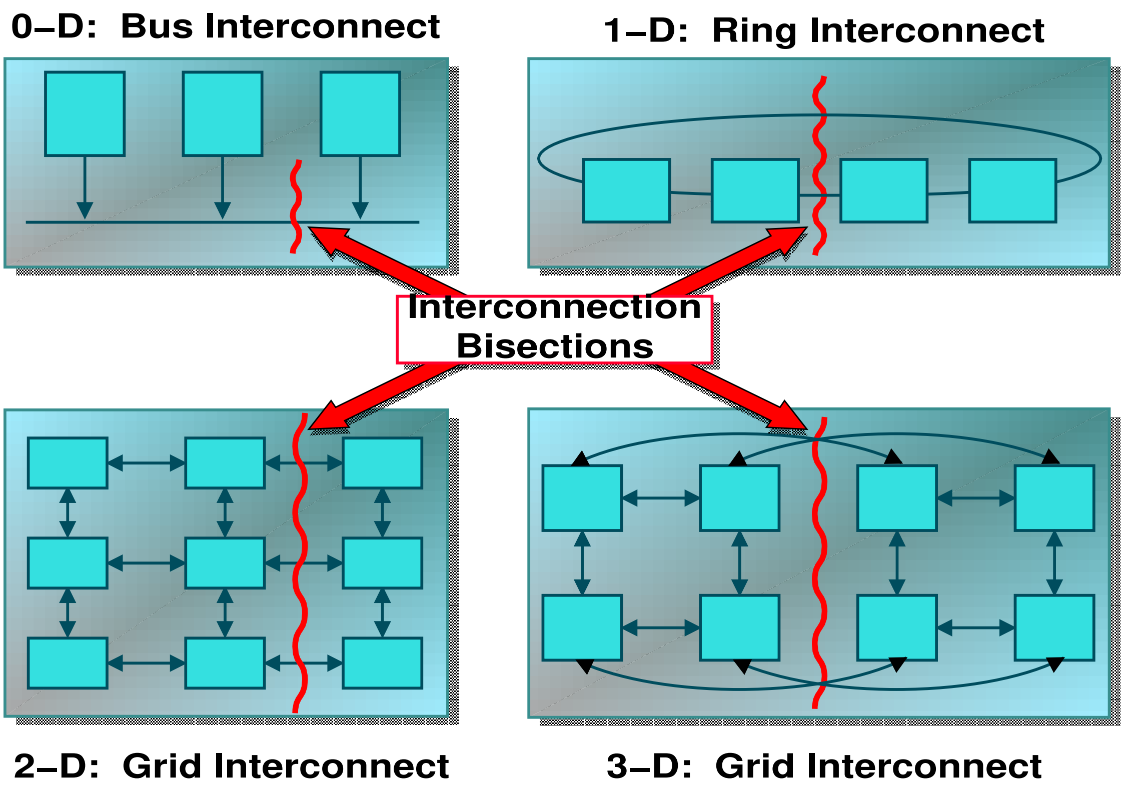
  BANDWIDTH: 256 bit * 43 MHz = 1.3 GB/seg
  LATENCIA: 1.100 ns

- **Xtalk:**

  BANDWIDTH: 16 bit * 400 MHz = 800 MB/seg (bidirec.)
  LATENCIA: 350 + 120 + 150*N ns

Alternativas de Interconexión

8



Alternativas de Interconexión – Bandwidth y Latencia

Interc. network	Dimension	Maximum sys. bandwidth	Bisection bandwidth	Average latency	Wire count per data bit
Bus	0	C	C	1	1
Ring	1	c*P	2c	P	P
2D mesh	2	c*P	$c\sqrt{P}$	$\sqrt{P}$	4P
3D mesh	3	c*P	$c\sqrt[2/3]{P}$	$\sqrt[1/3]{P}$	6P

Note: C>>c

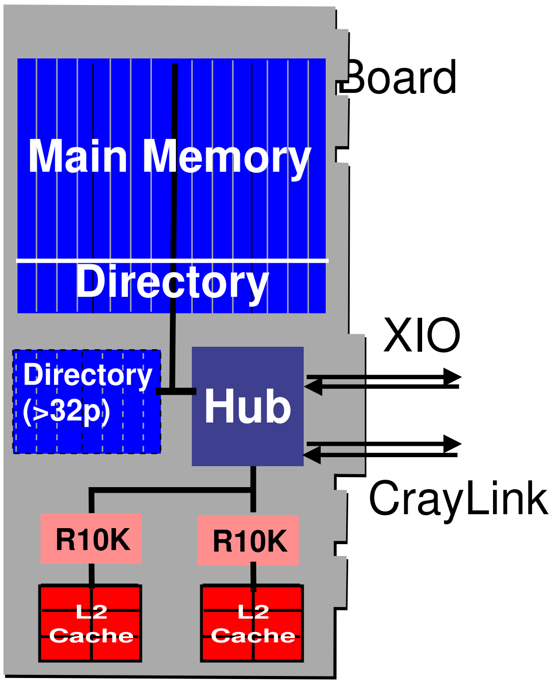
S2MP Building Block

Node Board: System on a Board

- 2 X R10000 Processors
- 64 MB to 4 GB memory

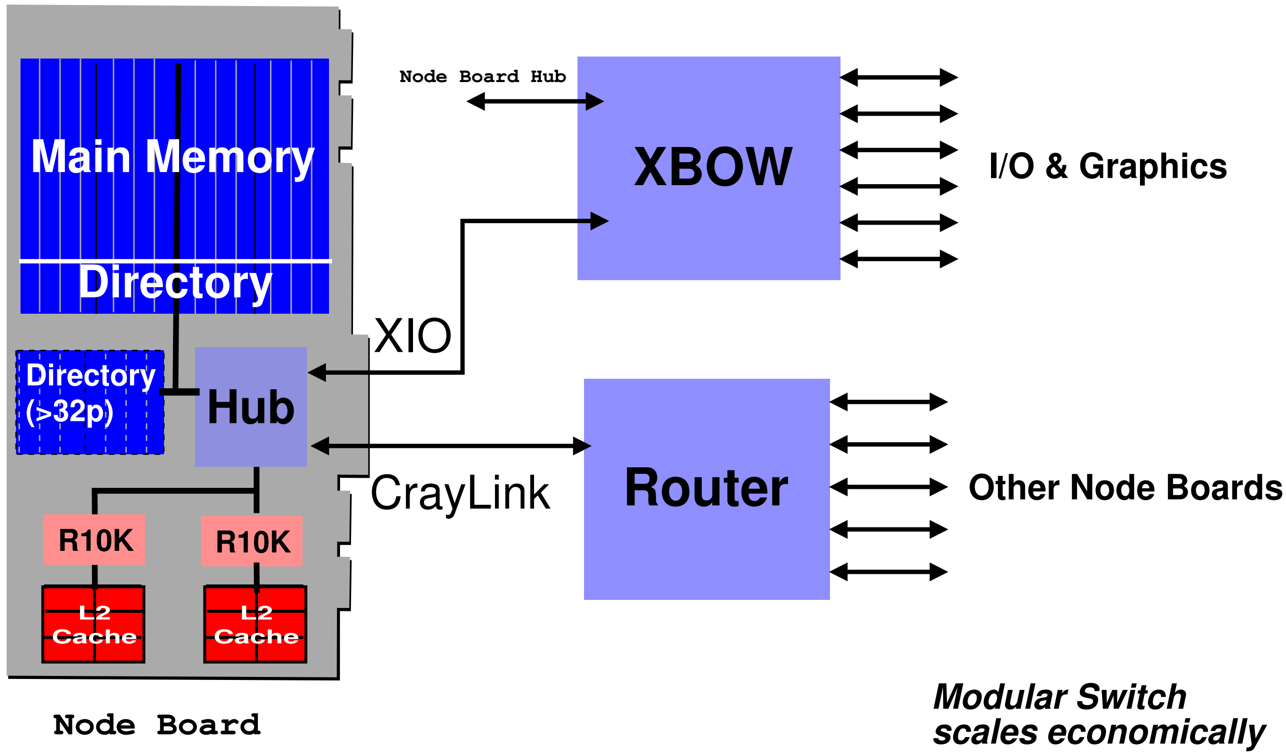
Hub B/W sustained [peak]

- 640 [800] MB/s – CPUs
- 700 [800] MB/s – memory
- 1.28 [1.6] GB/s XIO link
- 1.28 [1.6] GB/s CrayLink



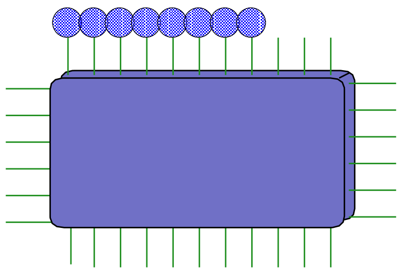
1 1

S2MP Switch Technology:

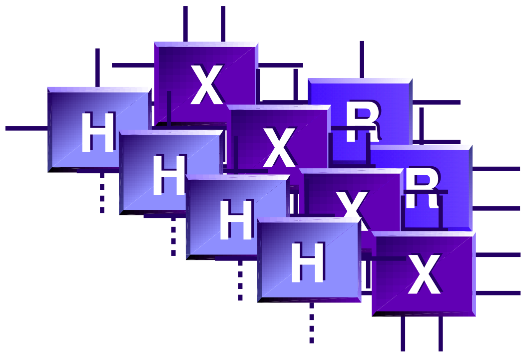


1 2

Arquitectura S2MP: un nuevo modelo de inversión



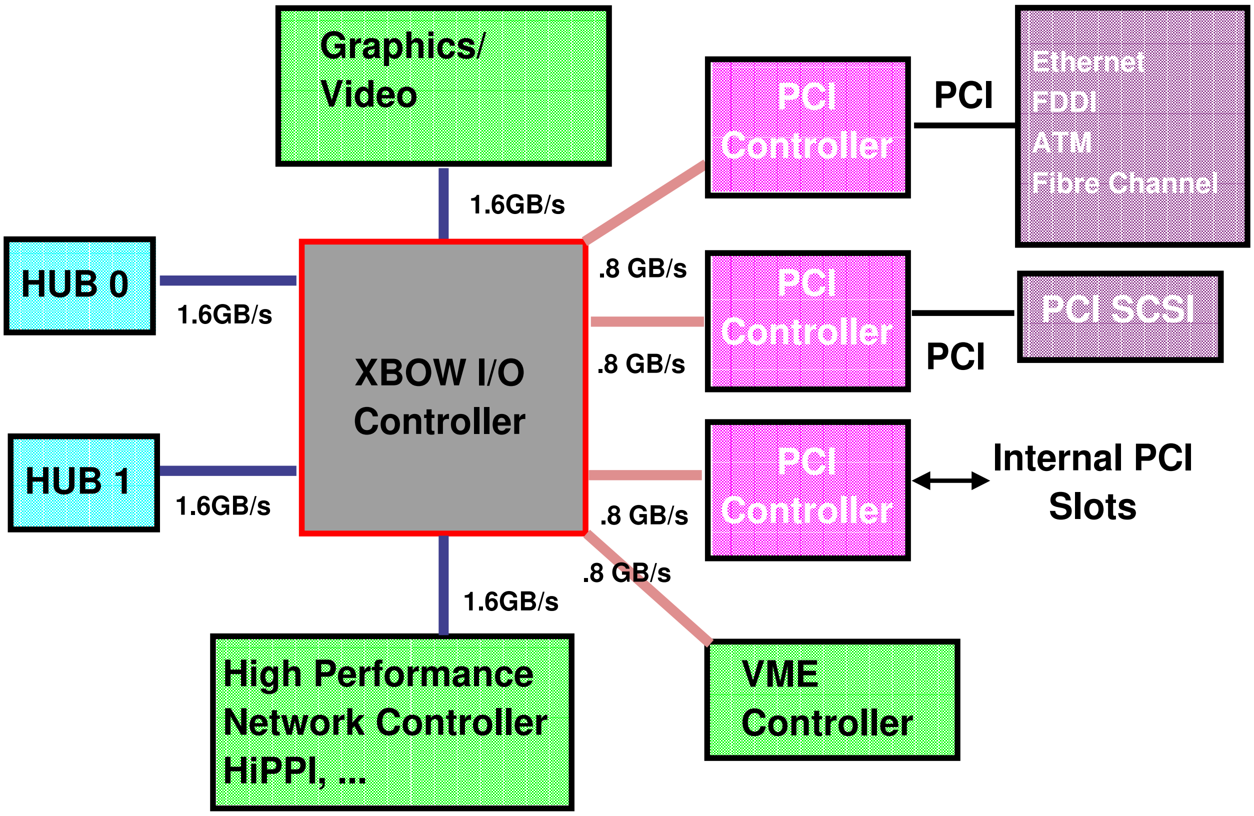
Los cross–bars convencionales, monolíticos son rentables sólo cuando se configuran en su máxima capacidad



El cross–bar modular CrayLink escala de forma económica

I/O Subsystem

13

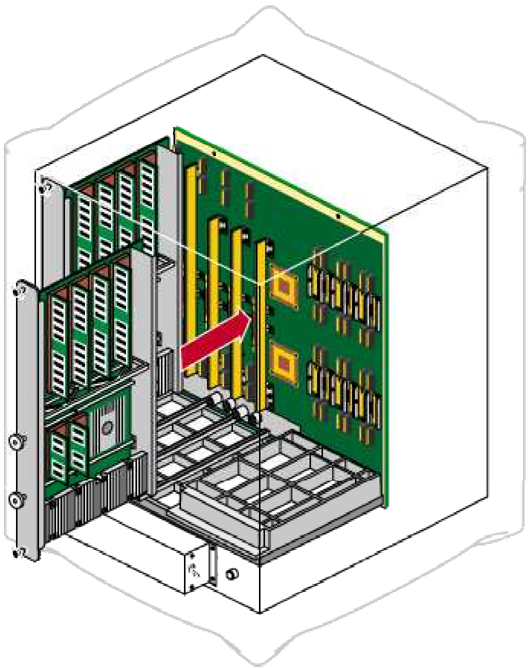


El módulo S2MP

14

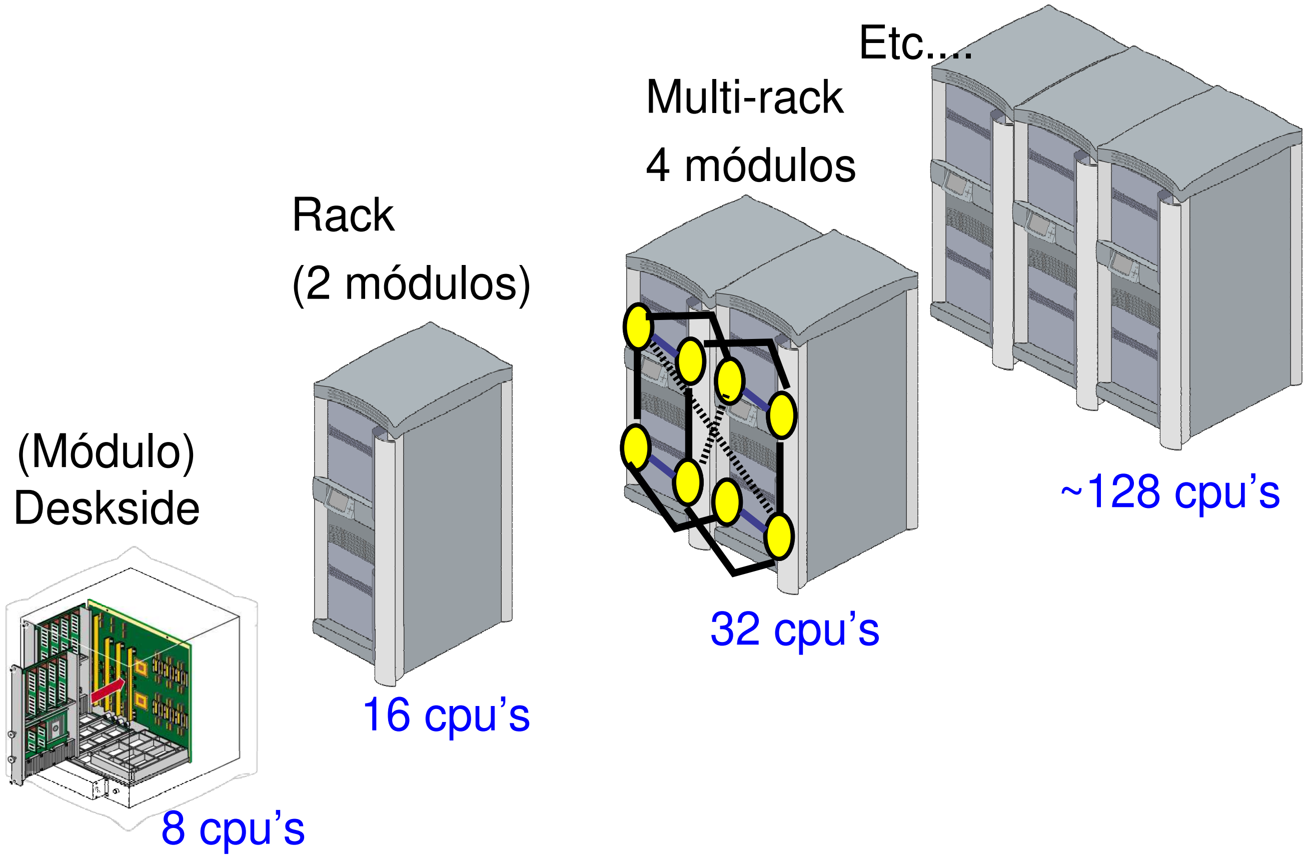
Module Features:

- Up to 8 R10000 CPUs (1–4 Nodes)
- Up to 16 GB physical memory
- Up to 12 XIO slots
- 2 XBOW Switches
- 2 Router Switches
- 64 bit internal PCI Bus (optional)
- Up to 2.56 [3.2] GB/sec system bandwidth
- Up to 5.1 [6.4] GB/sec I/O bandwidth



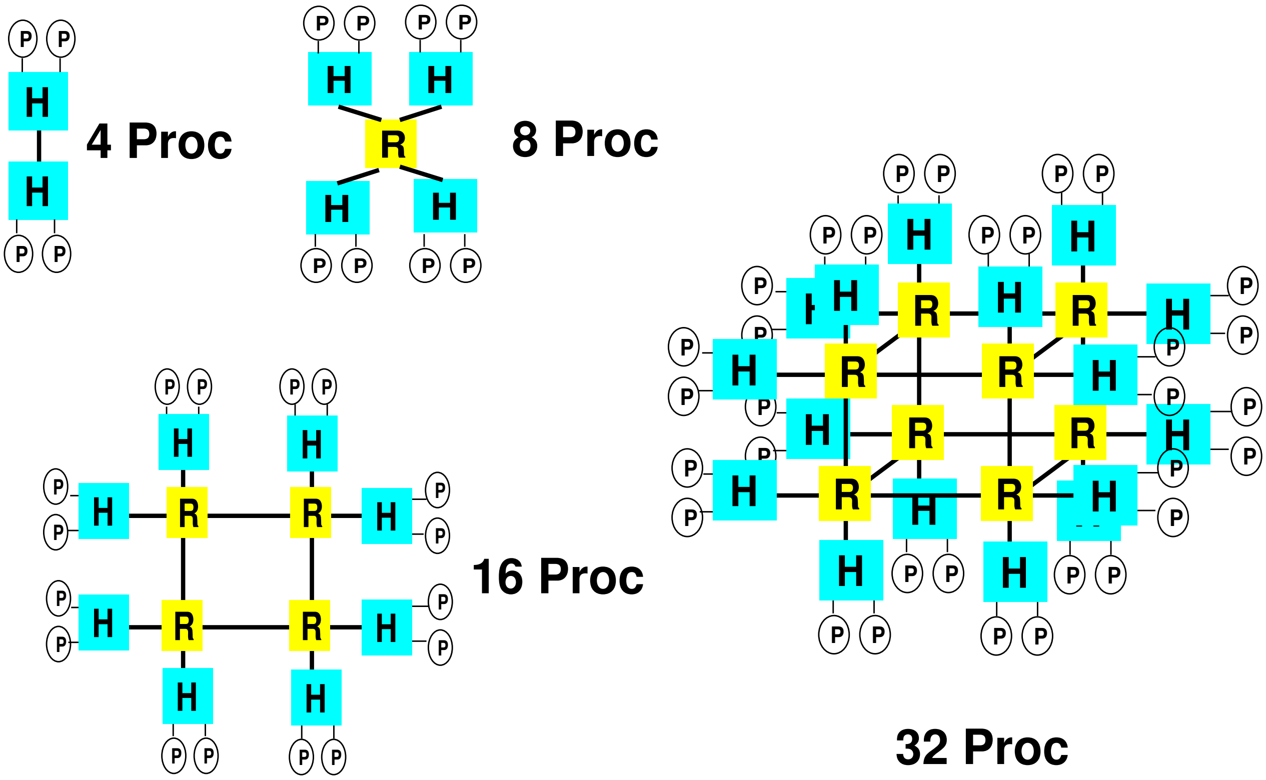
Módulos se convierten en sistemas

1 5

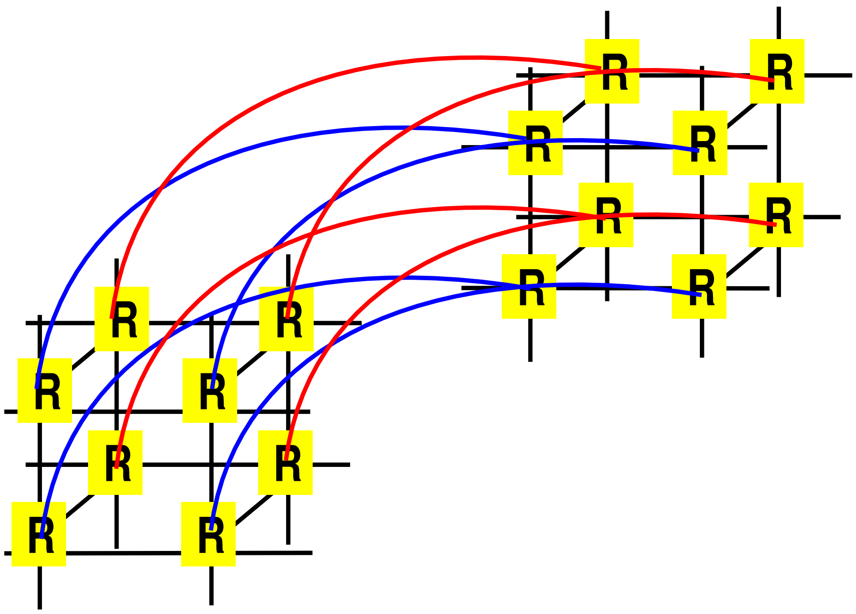


Esquema arq. de los sistemas

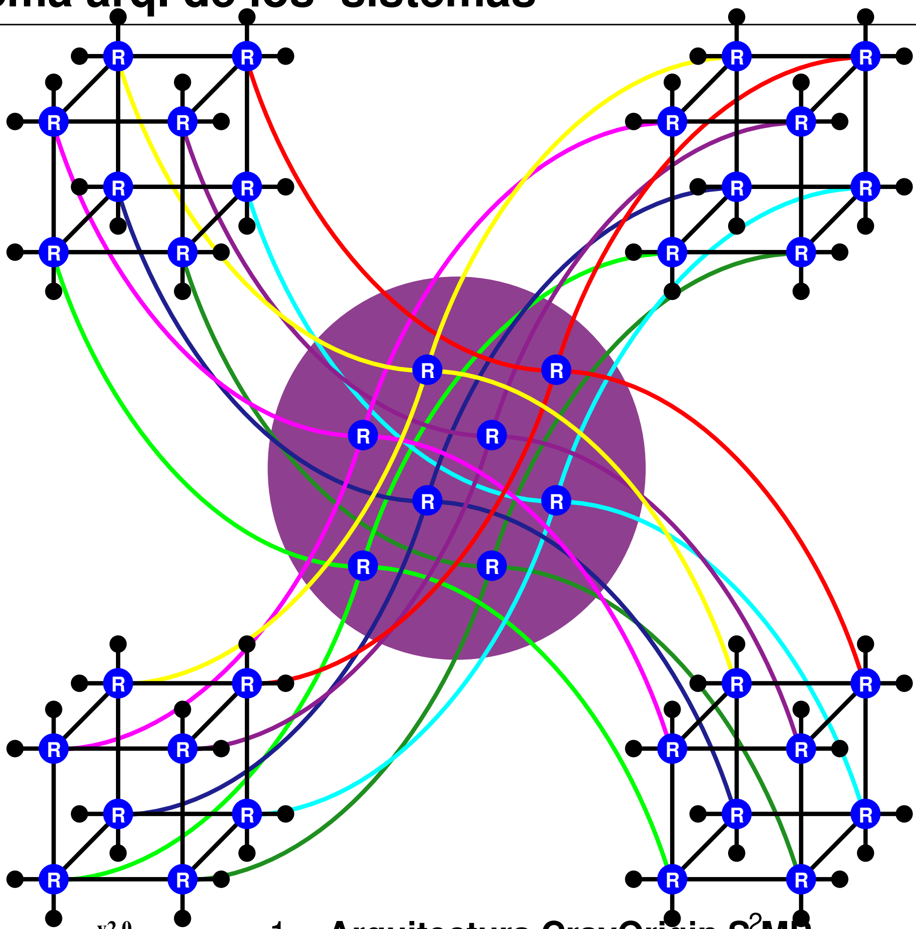
1 6



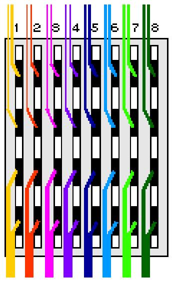
Esquema arq. de los sistemas



Esquema arq. de los sistemas



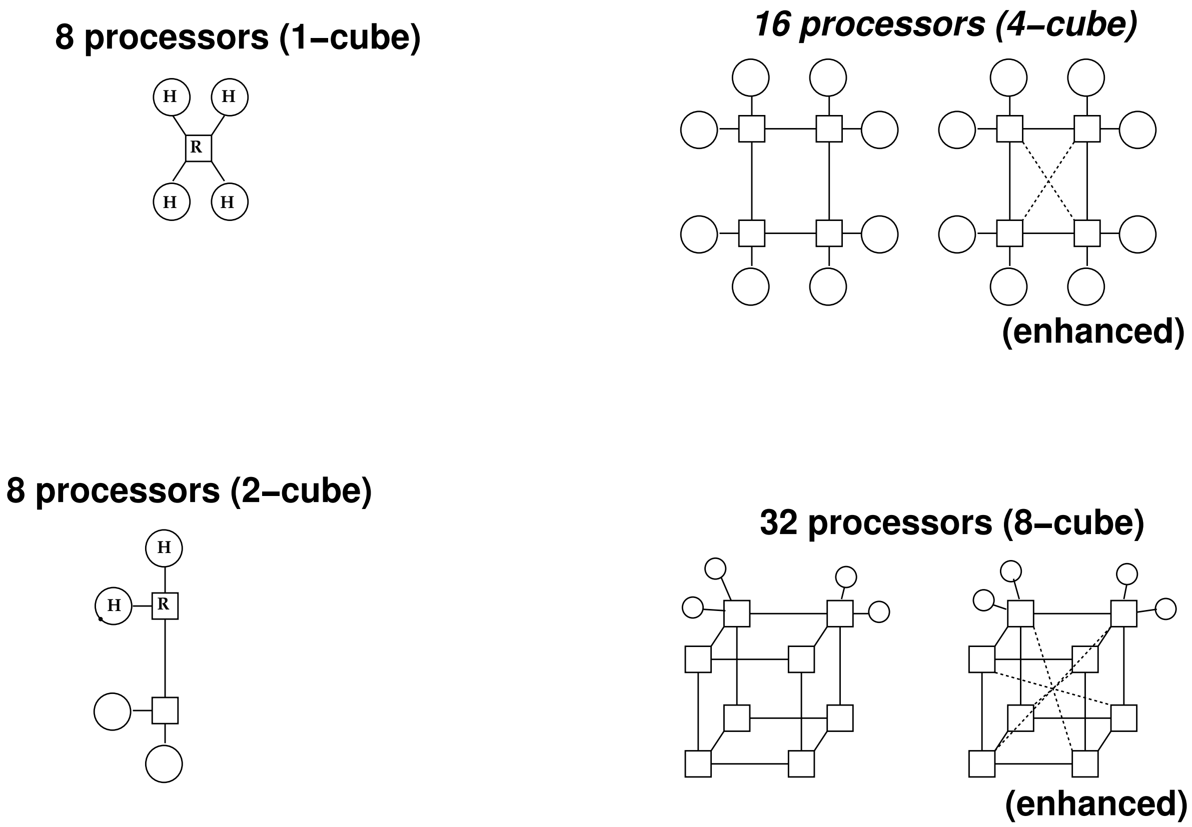
Cray
Router



>64 processors

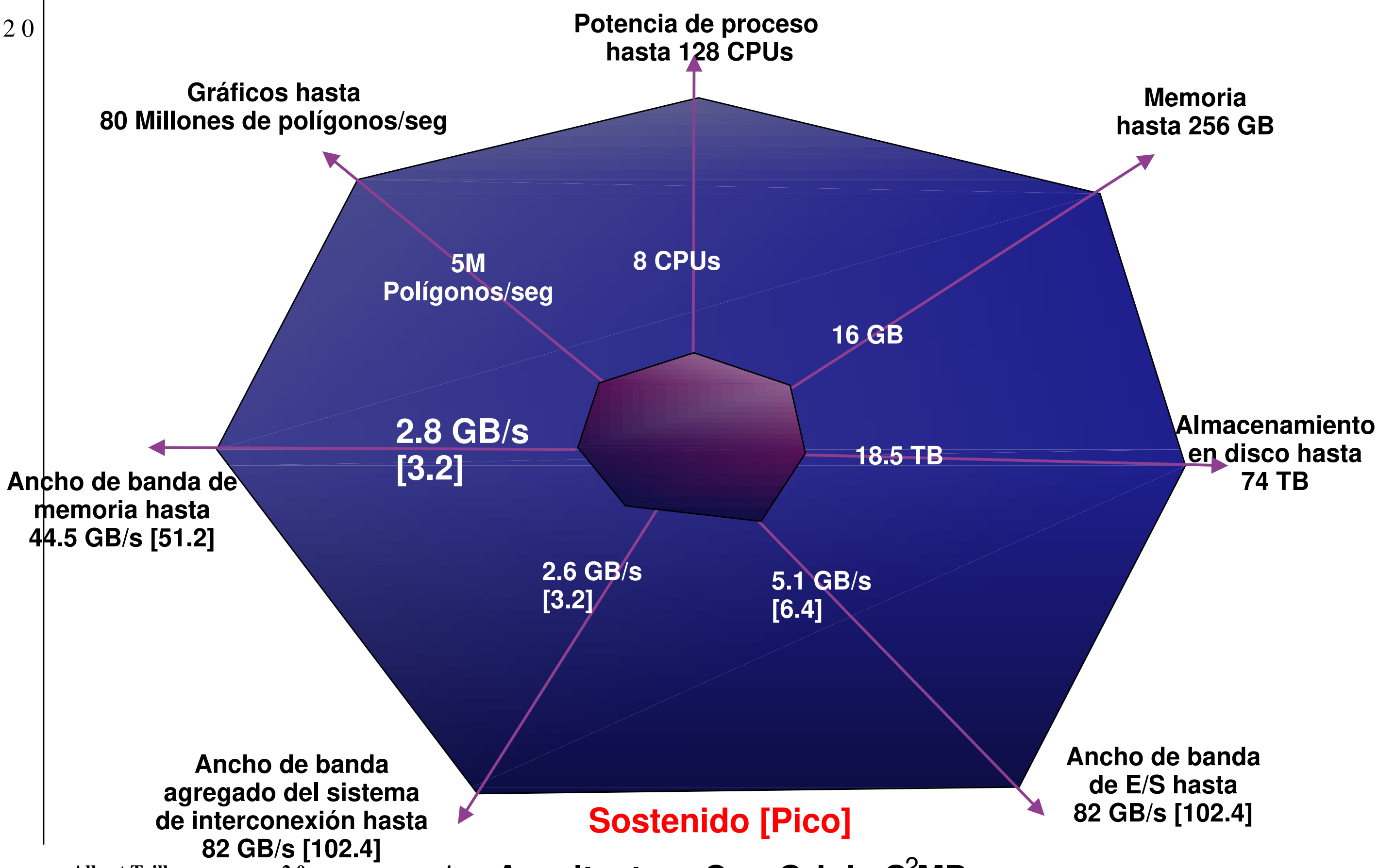
optimización con Express–Links

19



Crecimiento en todas las dimensiones

20

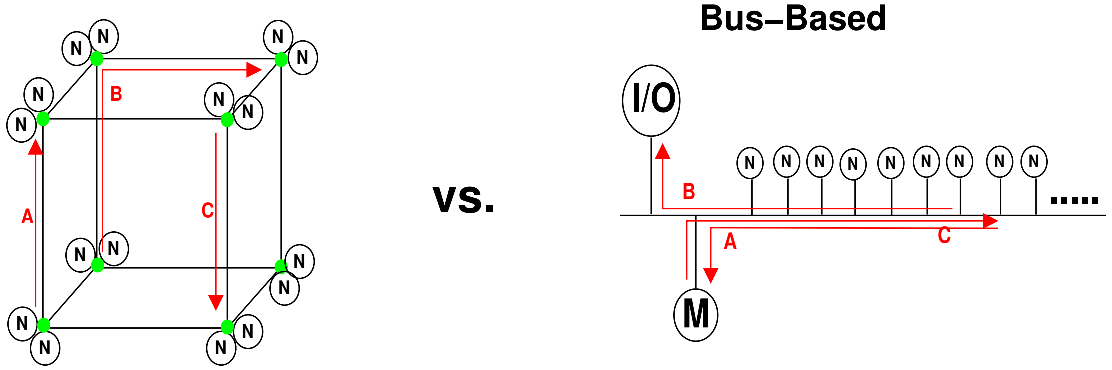


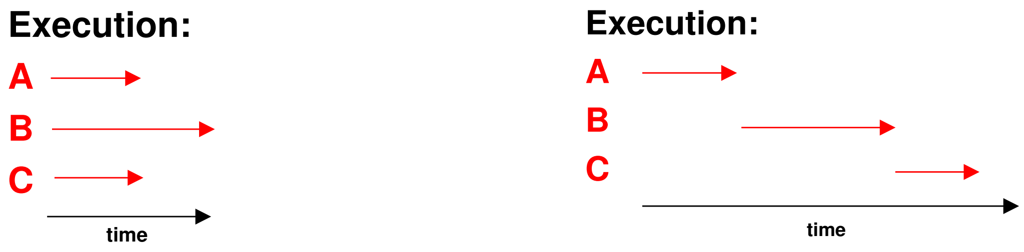
Eficiencia de la ejecución en paralelo

21

S2MP 3D Interconnect

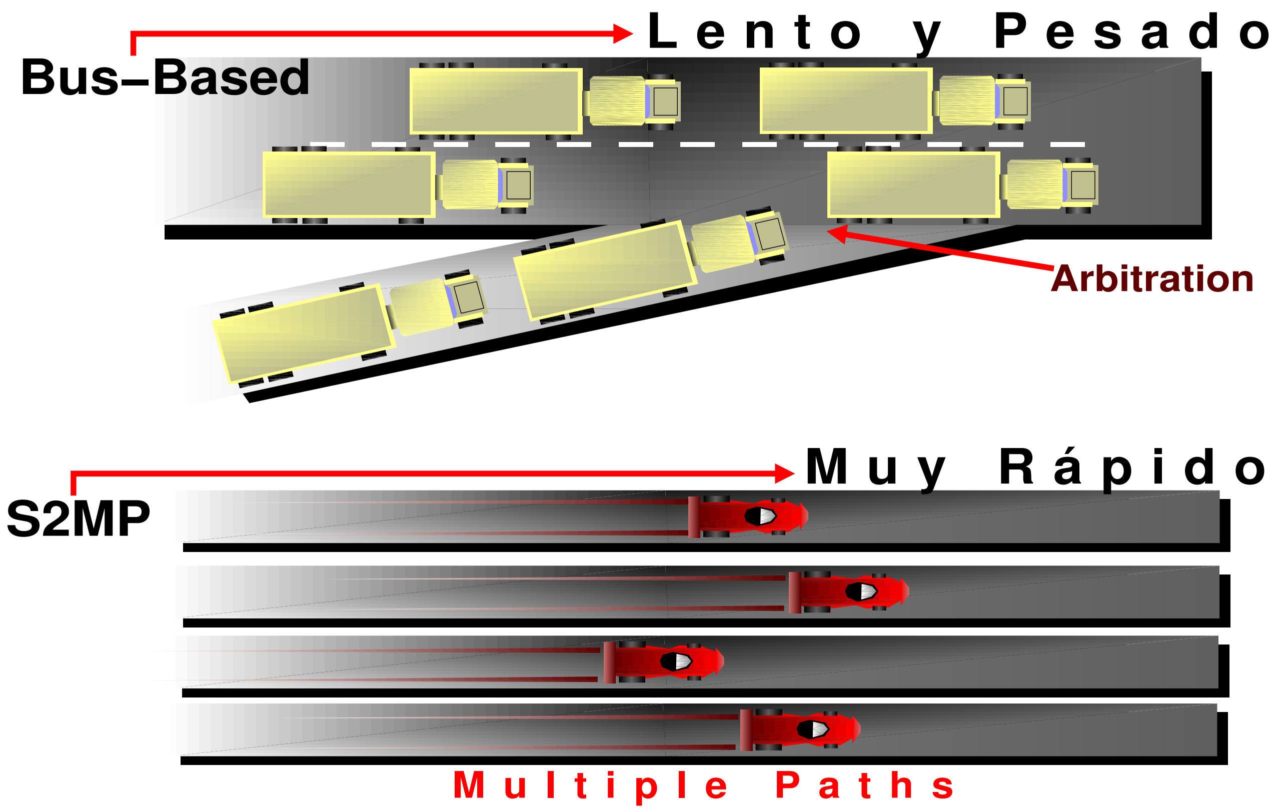
Bus–Based





Arquitectura S2MP: concepto

22

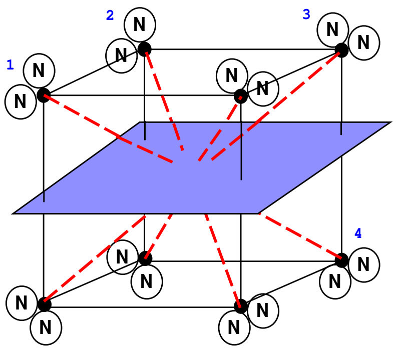


Incrementar linealmente el System Bandwidth

2 3

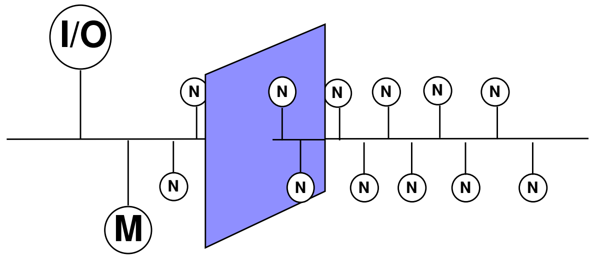
S2MP 3D Interconnect

(with ExpressLinks)



vs.

Bus–Based



Interconnect Bandwidth: 1.28 GB/sec sustained [1.6 GB/sec peak]

Number of Craylinks Bisected: 8

Bisection Bandwidth: 8 X 1.28 = 10.2 GB/sec [12.8 GB/sec]

Number of Processors: 32

Bandwidth/Processor: 10.2/32 = 320 MB/sec [400MB/sec]

Bus Bandwidth: 2.5GB/sec

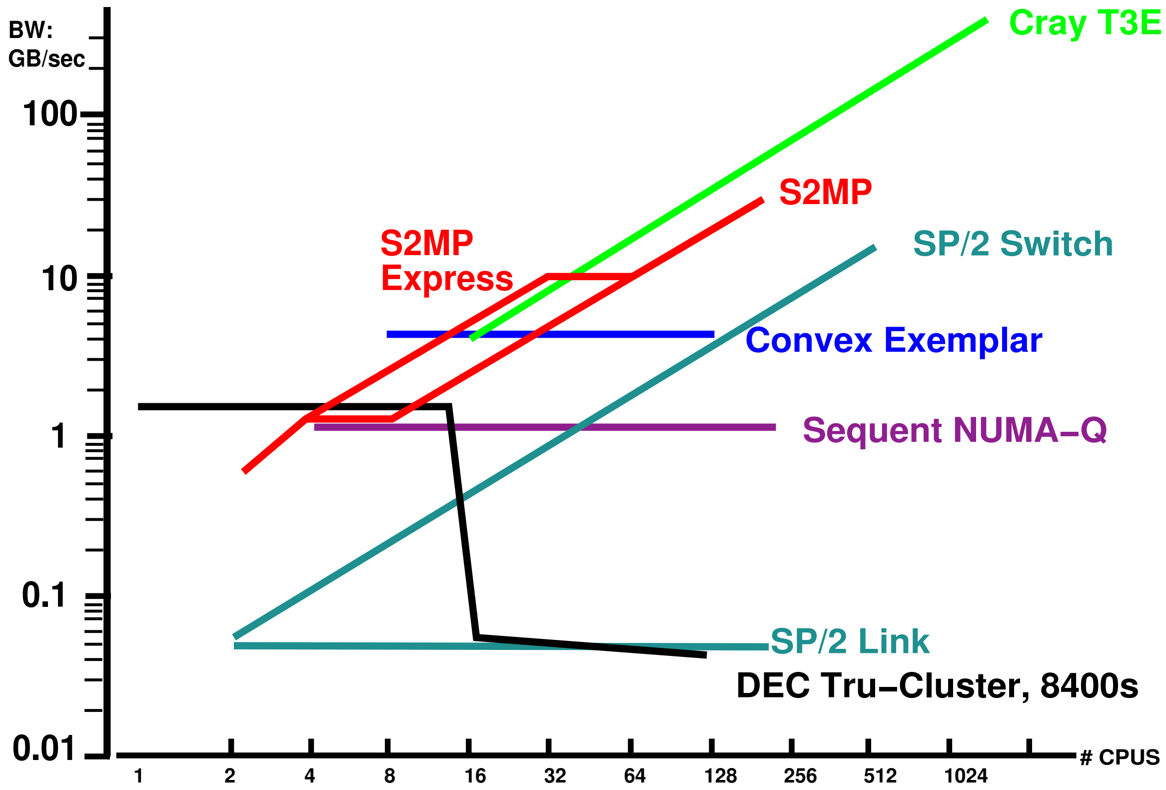
Bisection Bandwidth: 2.5GB/sec

Number of Processors: 32

Bandwidth/Processor: 2.5/32 = 78MB/sec

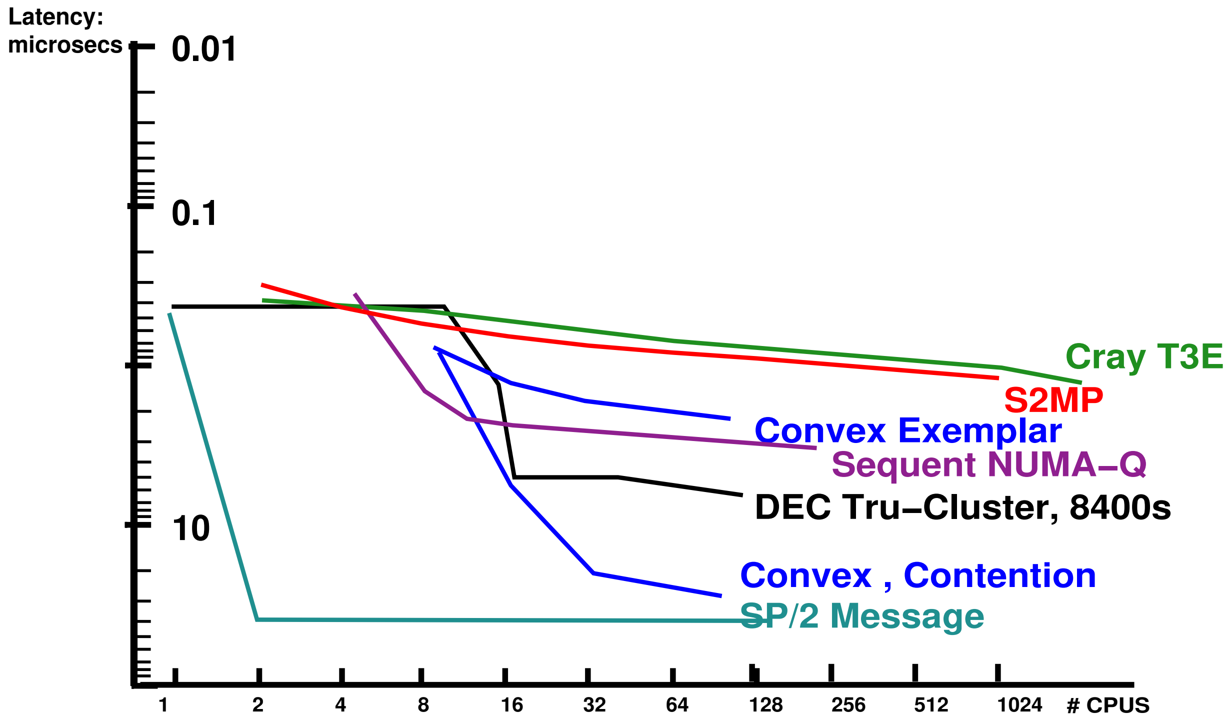
Bisection Bandwidths

2 4



Latencias Medias de Lectura a Memoria

25



BB – Lat – I/O

26