Evolução das Redes de Interconexão

Marino Hilário Catarino
Programa de Ciências da Computação
IME-USP

MAC5742 - Computação Paralela e Distribuída

12 de junho de 2015

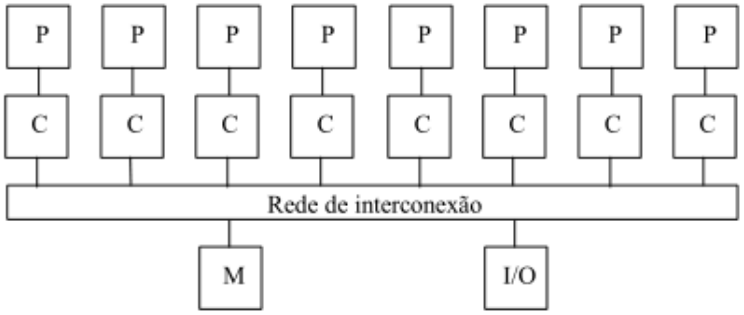
Sumário

- Introdução
- Caraterísticas
- Tipos de redes
- • Redes estáticas
- • Redes dinâmicas
- Características de Topologias
- Topologias de redes estáticas
- Topologias de redes dinâmicas
- Top 500
- O futuro

Introdução

- Redes de interconexão podem ser usadas: para conexão interna entre processadores, módulos de memória, e I/O; ou para formar uma rede distribuída de nós em um sistema multicomputador.

- Podem ser classificadas em estáticas ou dinâmicas, para sistemas multicomputador e multiprocessador respectivamente.



Arquitetura de um multiprocessador do tipo SMP (UMA)

Características

Escalabilidade:

- Aumento de tamanho da rede;

Desempenho:

- Distância
- Desempenho:
  - Latência
  - Taxa de transferência (quantidade)
  - Uni ou bidirecional.

Características

Custo:
Desempenho/número de ligações;

Confiabilidade:
Caminhos redundantes entre os componentes;

Funcionalidade:
Demais serviços como:
armazenamento temporário e ordenação.

Tipos de redes

Redes estáticas (ponto-a-ponto)

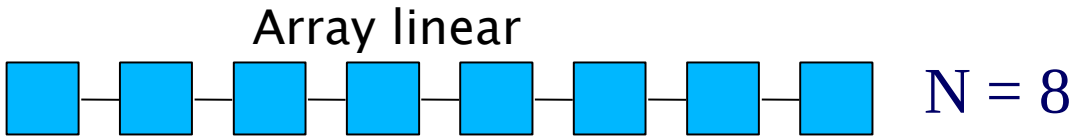
- Ligações fixas (diretas e dedicadas) entre os componentes;
- Usadas em multicomputadores;
- Topologia determina as características da rede;
- Grau do nós: o número de ligações diretas de cada componente;
- Diâmetro da rede é a maior distância em número de ligações entre dois componentes quaisquer.

Tipos de redes

Redes dinâmicas

- Sem topologia fixa
- Ligações estabelecidas conforme necessário
- Usadas em multiprocessadores e multicomputadores atuais

Características de topologias



- **Ligações (L)** entre os N nós/componentes

Exemplo: L = N-1 =7

- **Grau (g) do nó**: número de ligações diretas de cada componente

Exemplo: g = máximo 2

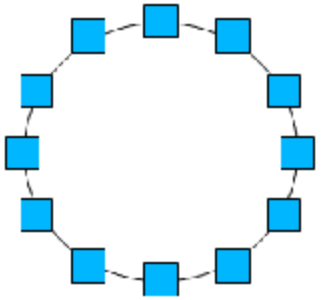
- **Diâmetro (D)**: A maior distância (em número de ligações) entre dois componentes quaisquer é chamada de diâmetro da rede.

Exemplo: D = N – 1 = 7

Anel

Grau máximo dos nós	Diâmetro	Número total de ligações	Simetria
2	N/2	N	Sim

Exemplo:
N=12
D=6

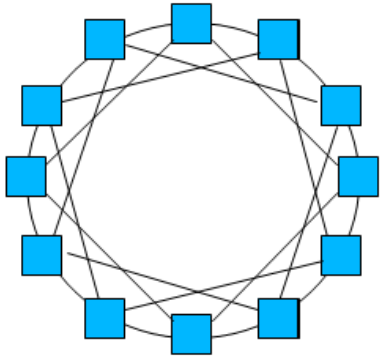


Anel cordal

- Menos tráfego no anel central
- Caminho alternativos

Grau máximo dos nós	Diâmetro	Número total de ligações	Simetria
>2	<N/2	>N	Sim

Exemplo:
N=12
D=3
g=4
L=24

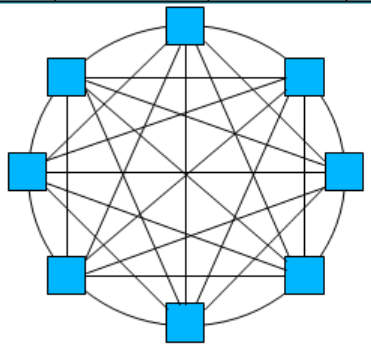


Totalmente conectada

- Alto custo
- Grau de nó = número de nós - 1
- Diâmetro 1 (ideal)

Grau máximo dos nós	Diâmetro	Número total de ligações	Simetria
N-1	1	N(N-1)/2	Sim

Exemplo:
N=8
D=1
L=28

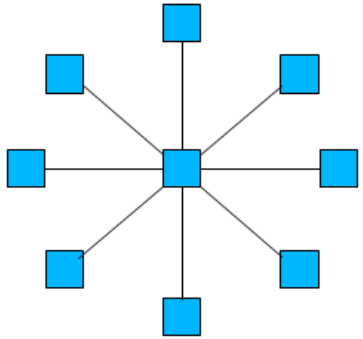


Estrela

- Tráfego intenso no nó central
- Problemas no nó central bloqueiam a rede

Grau máximo dos nós	Diâmetro	Número total de ligações	Simetria
N-1	2	N-1	Não

Exemplo:
N=9
g=8
L=8



Árvore binária - ideal para a execução de algoritmos do tipo divisão e conquista

- Diâmetro cresce de forma linear com a altura h
- Grau de nó máximo 3
- Sem caminhos alternativos
- Nó raiz é um gargalo

Grau máximo dos nós	Diâmetro	Número total de ligações	Simetria
3	2h	N-1	Não

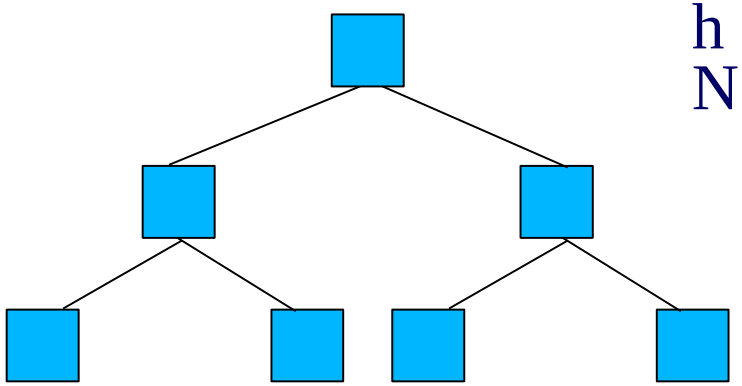
Exemplo:

N=7

h=2

D=4

L=6



h = altura da árvore

$N = 2^{(h+1)} - 1$

Malha

- Caminhos alternativos aumentam confiabilidade e diminuem risco de gargalos;

- Adequada para problemas em que uma estrutura de dados bidimensional precisa ser processada de forma particionada (matrizes, imagens, etc.)

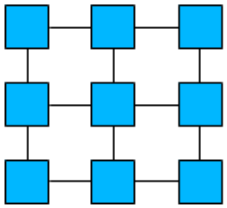
Malha bidimensional

- Grau de nó máximo 4
- Facilidade de incremento de elementos

Grau máximo dos nós	Diâmetro	Número total de ligações	Simetria
4	2(r-1)	2N-2r	Não

N = r * r

Exemplo:
r=3
N=9
D=4
L=12



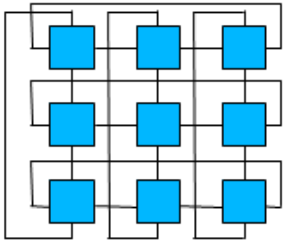
Torus

- Grau de nó 4
- Diâmetro reduzido em relação ao número de nós

Grau máximo dos nós	Diâmetro	Número total de ligações	Simetria
4	2[r/2]	2N	Sim

N = r * r

Exemplo:
r=3
N=9
D=2
L=18



Hipercubo

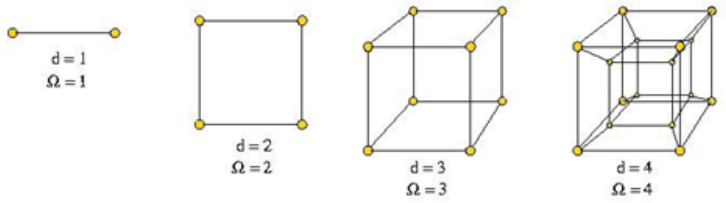
- Pequeno diâmetro é o principal atrativo desta topologia
- Conveniente para aplicações com pouca localidade de dados
- Escalabilidade: número de nós deve ser restrita em potências de 2
- Diâmetro cresce logaritmicamente
- Grau de nó = dimensão do cubo

Hipercubo

d = dimensão

N = 2^d do cubo

Grau máximo dos nós	Diâmetro	Número total de ligações	Simetria
d	d	dN/2	Sim



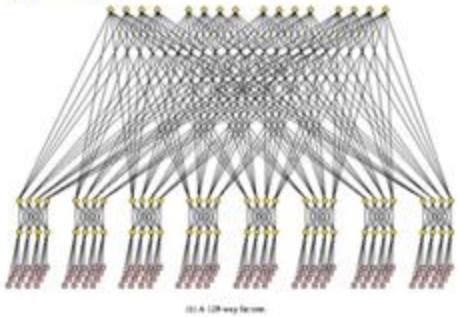
(a) Hypercubes, dimension 1-4.

Fonte: van der Steen, 2005 (distributed memory MIMD)

Fat tree

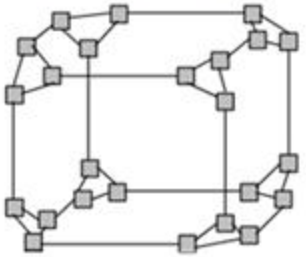
- Mais ligações que árvore binária (um filho pode ter vários pais)
- Múltiplos caminhos e capacidade de transmissão no canal aumenta das folhas para a raiz. É uma solução para o maior problema em uma árvore binária comum, onde o gargalo do sistema é justamente na raiz, onde o transito de mensagens passa a ser maior.

Fonte: van der Steen, 2006
(distributed memory MIMD)



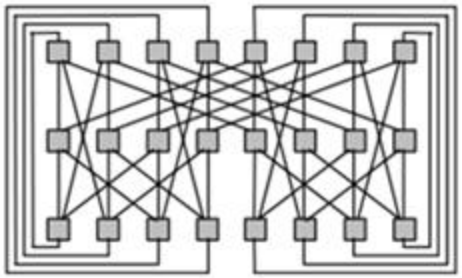
Cubo CCC (Cube Connected Cycles)

- Hipercubo em que cada nó é um anel
- Hipercubo de dimensão d = anel com d nós
- Diâmetro cresce logaritmicamente
- Grau de nó 3 para qualquer diâmetro



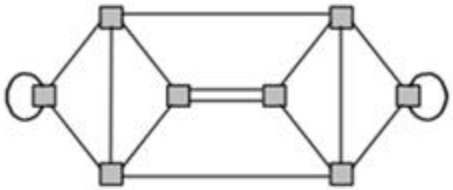
Butterfly

- Grau de nó 4
- Diâmetro menor que um cubo CCC
- Diâmetro cresce logaritmicamente
- O exemplo é de dimensão 3



Grafo de DeBrujn

- Grau de nó 4
- Grafo de dimensão d = 2 x d nós
- Diâmetro cresce logaritmicamente
- O exemplo é de um grafo de dimensão 3

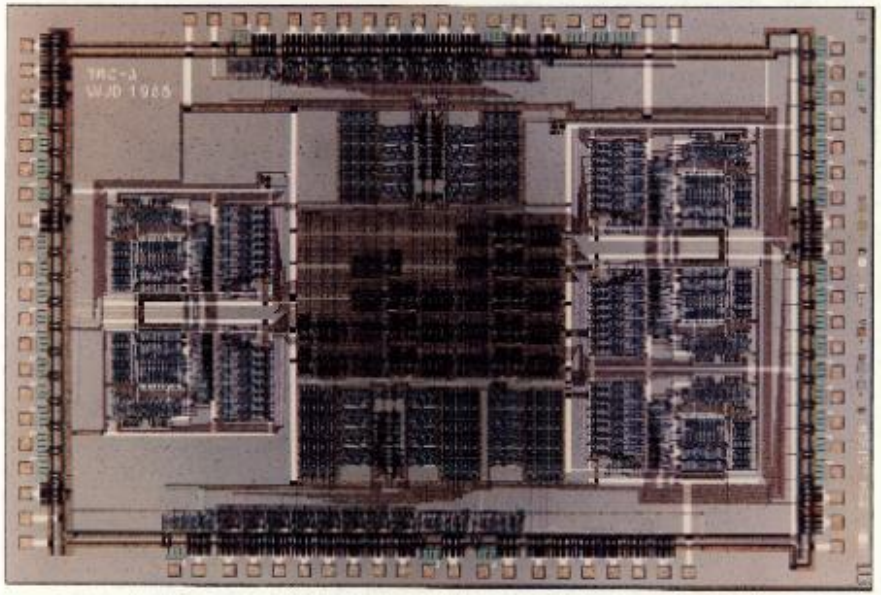


História

Caltech – 1983

Topologia Hipercubo

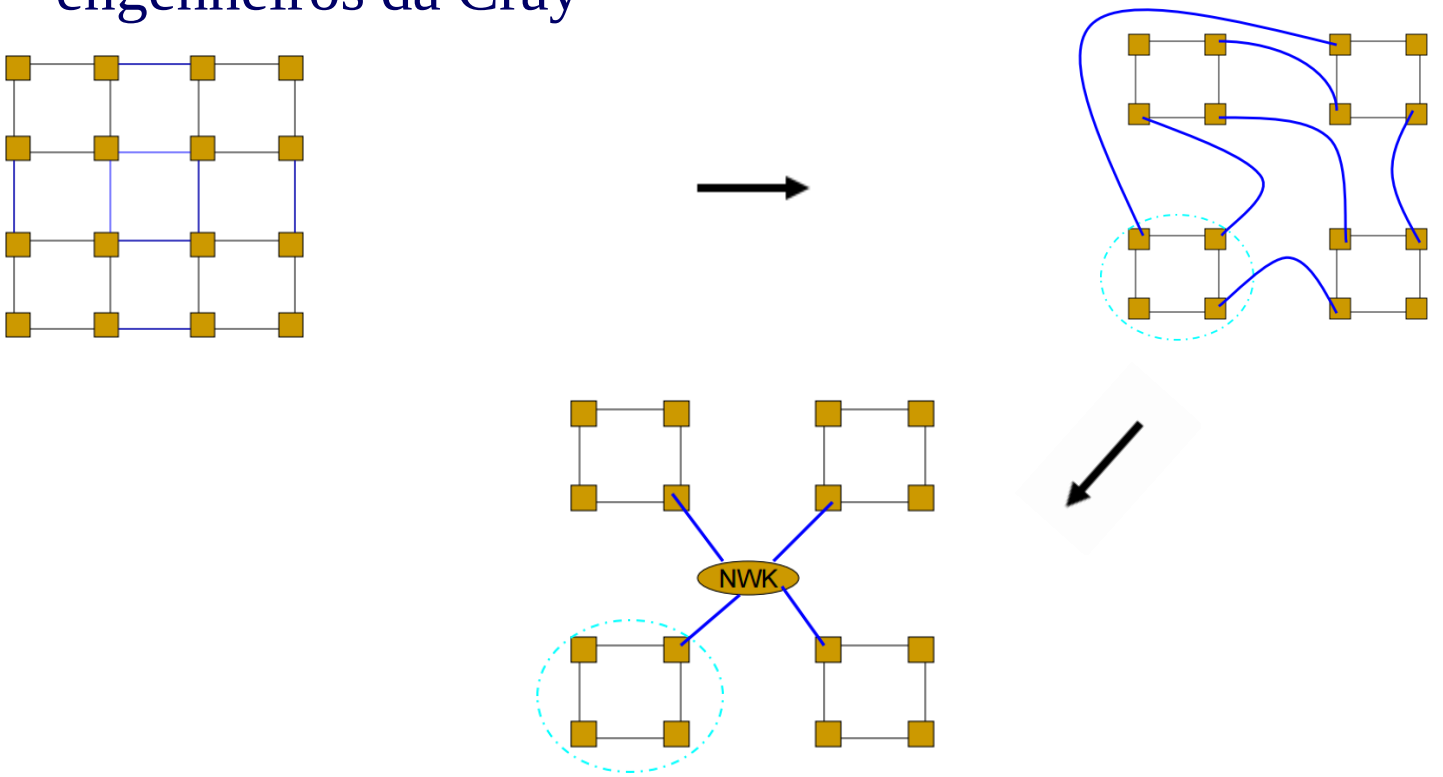
Caltech - 1985



Topologia Torus

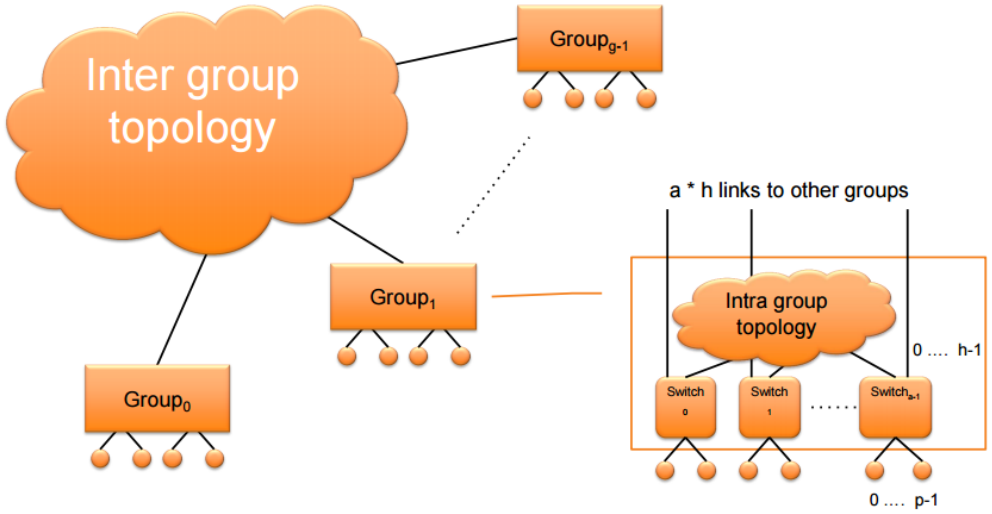
Evolução para o Dragon Fly - 2008

- Malha 2D - Universidade de Stanford, juntamente com engenheiros da Cray



Dragon Fly

- Vários grupos estão ligados em conjunto com ao menos uma ligação direta, utilizando todas as ligações.
- A topologia dentro de cada grupo pode ser qualquer topologia. Recomenda-se a butterfly.
- Foco na redução do número de ligações de longa duração e diâmetro da rede.



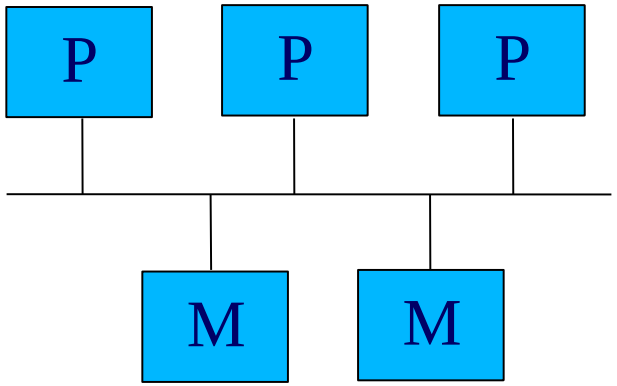
Topologias de redes dinâmicas

- Não há topologia fixa;
- Rede se adapta dinamicamente para permitir transferência de dados;
- Usadas para conectar M-P em multiprocessadores e P-P em multicomputadores modernos;
- Redes bloqueantes/não-bloqueantes: estabelecimento de uma conexão impede/não impede que outras se estabeleçam.

Barramento

- Baixo custo
- Canal compartilhado por todas as possíveis conexões
- Baixa confiabilidade
- Altamente bloqueante
- Escalabilidade comprometida

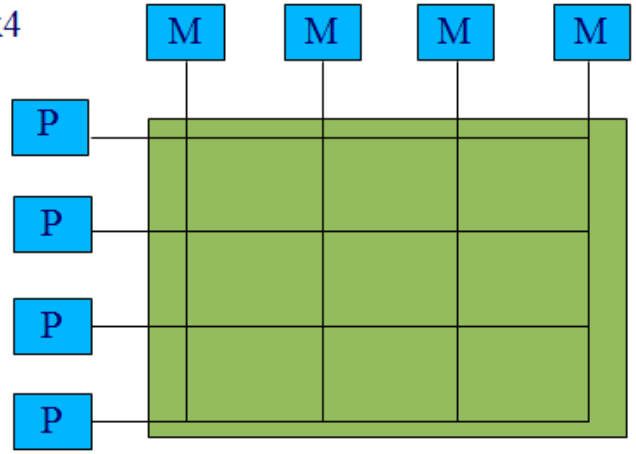
barramento
pode ser único
ou múltiplo



Matriz de chaveamento

- Permite chaveamento entre 2 componentes quaisquer;
- Não é bloqueante (comunicação simultânea entre diferentes pares P-M ou P-P);
- Boa escalabilidade (componentes acrescentados aos pares);
- Alto custo (utilização inviável para grande número de componentes).

Ex.: matriz 4x4

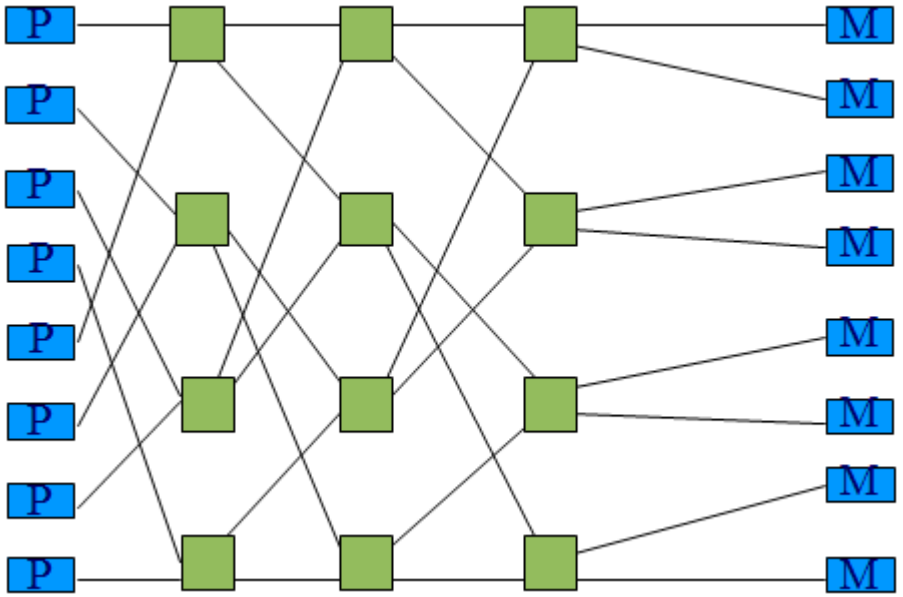


Redes multinível

- Redes hierárquicas com de várias matrizes de chaveamento;
- Associação de matrizes de chaveamento diminui redundância (menor confiabilidade);
- Redes multinível podem ser bloqueantes.

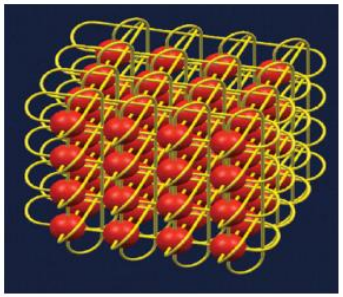
Exemplo: Rede Omega

- 16 × 16 comutador (2×2) com 4 estágios
- Diferentes combinações podem ser obtidas para a interconexão entre as entradas e saídas.



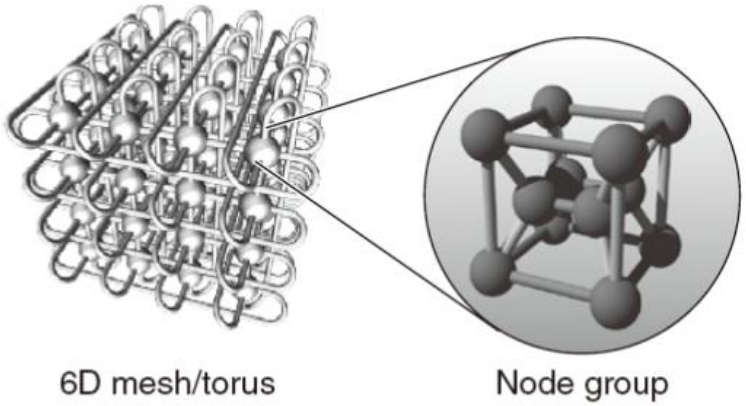
Top 500

- IBM Blue Gene/L e Blue Gene/P, Cray XT3: 3D torus



- IBM Blue Gene/Q: 5D torus

- Fujitsu K: 6 D torus chamado Tofu

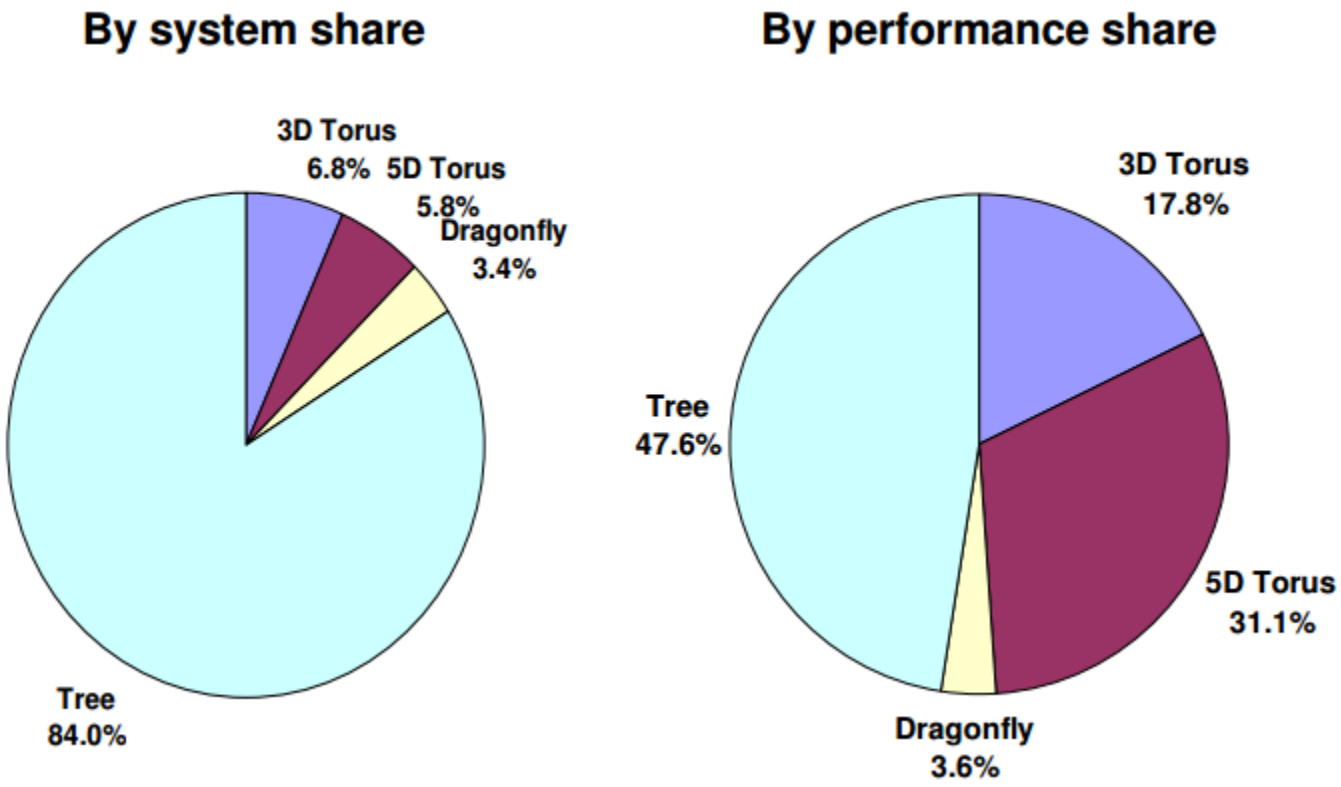


Fujitsu - K Computer

O k Computer foi concebido por meio de uma parceria da Fujitsu e a RIKEN. Este supercomputador em 2011 ocupava a segunda posição da lista Top500.

Esse sistema utiliza uma conexão direta e foi projetado para suportar mais de 80 mil nós. Com este tipo de conexão, juntamente com algoritmos especializados, é possível alocar um grupo de k nós a uma aplicação ou usuário específico.

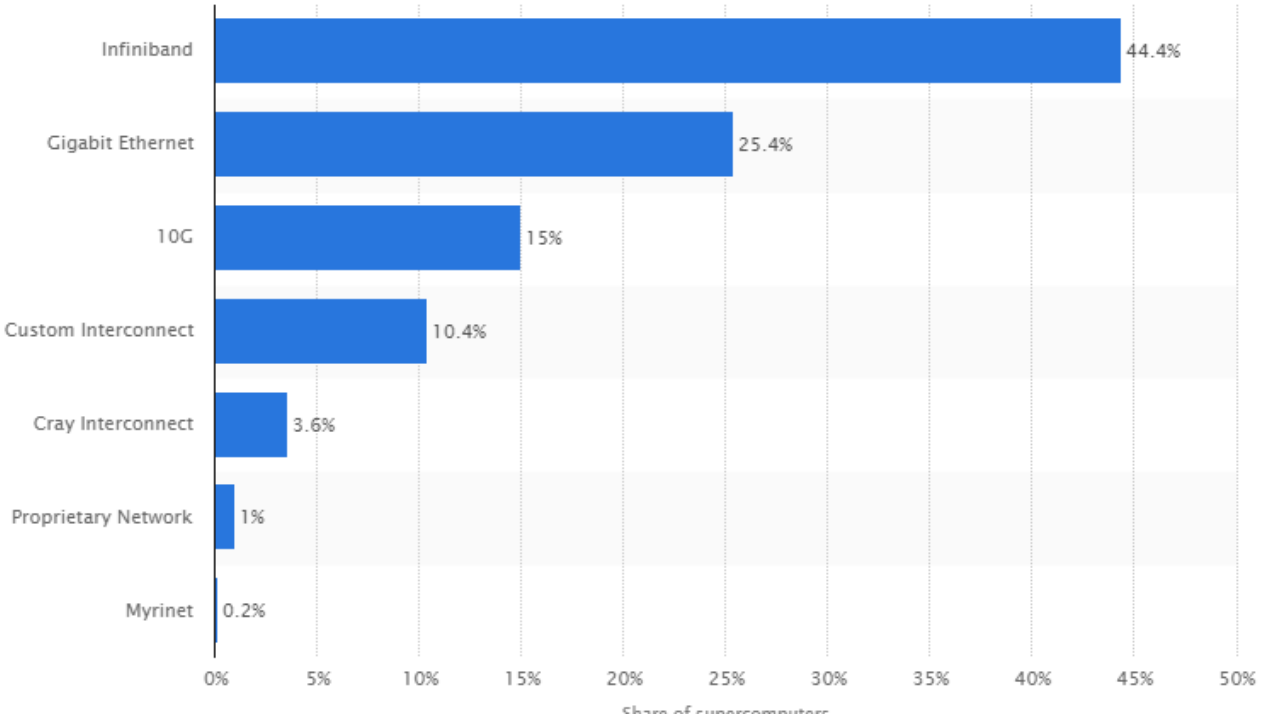
Topologias de Interconexão - 2012



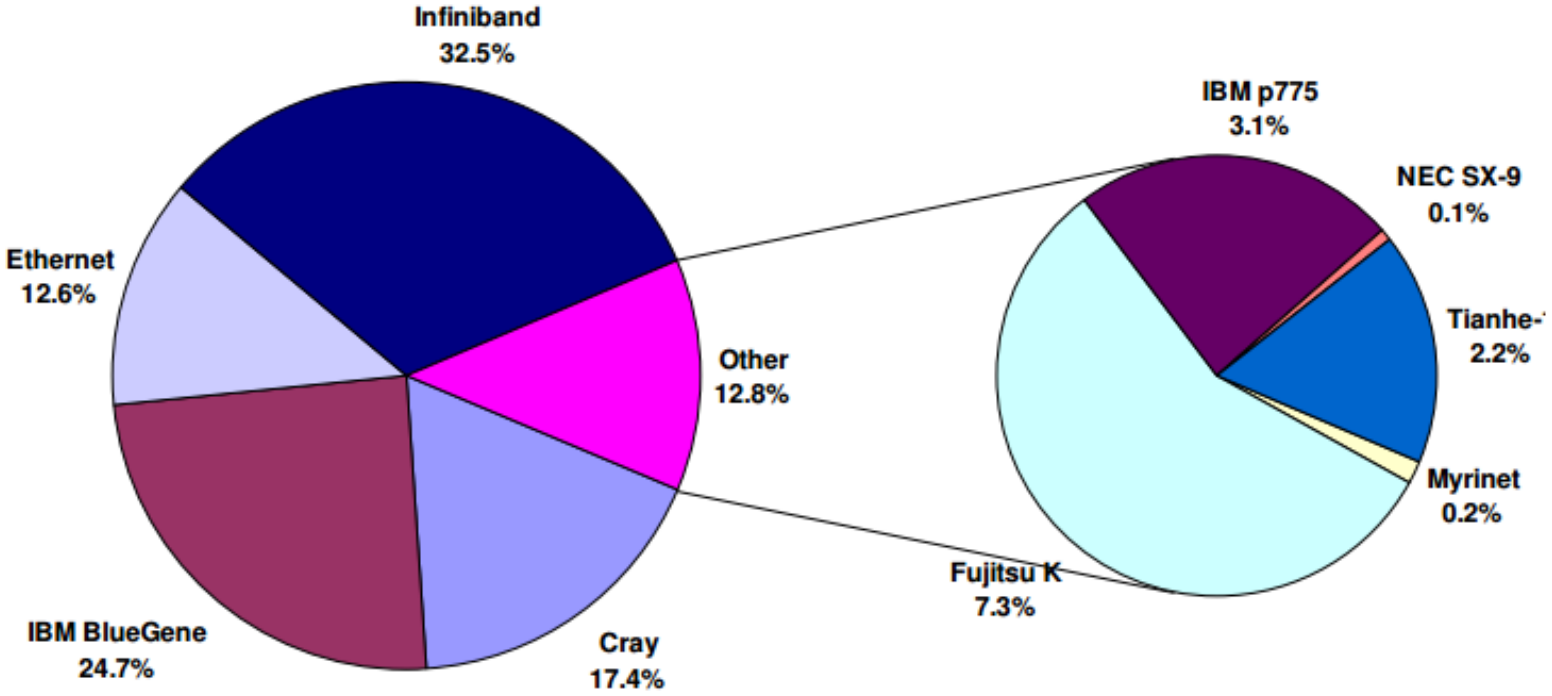
Dispositivos - 2014

System share of interconnect families used in the most powerful 500 supercomputers worldwide as of June 2014

This statistic shows the system share of interconnect families used in the 500 most powerful supercomputers around the world as of June 2014. As of this date Infiniband was the interconnect family used in 44.4 percent of the leading supercomputers.



Dispositivos – 2012
Interconexões por desempenho das ações



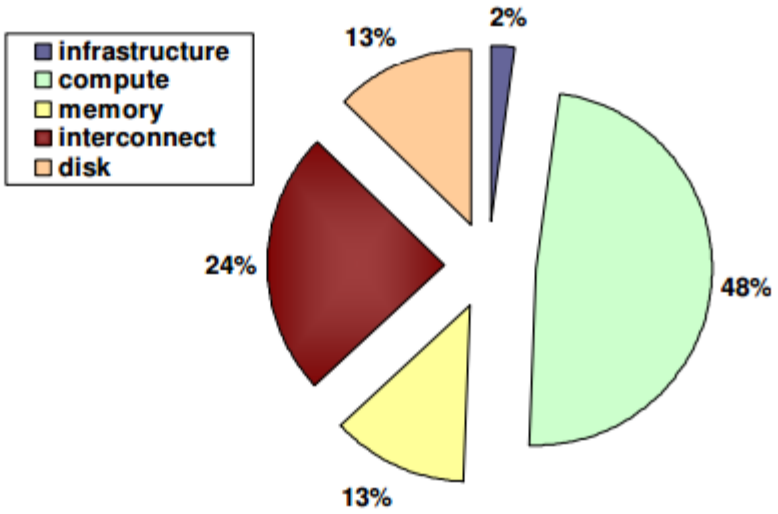
Dispositivo: InfiniBand

Barramento serial que oferece 2.5 Gigabits (312 MB/s) por segundo por par de cabos;

Comunicação é bidirecional totalizando um barramento total de 625 MB/s.

Também é possível aumentar a largura do barramento usando mais cabos, A especificação original fala em links com até 12 pares, que permitiria links de até 3.75 GB/s em cada sentido.

O futuro: Interconexão se torna um fator de custo



	Fat Tree	Torus 1	Torus 2	Torus 3
Compute	63.3%	72.5%	76.5%	79.7%
Adapters + cable	10.4%	11.9%	12.6%	13.1%
Switches + cables	26.3%	15.6%	10.9%	7.2%
Total	100%	100%	100%	100%

Referência bibliográfica

- Hwang, K.; Xu, Z. Scalable Parallel Computing: technology, architecture, programming. McGraw-Hill, 1998. (Capítulo 6)
- de Rose, C. A. F. Fundamentos de Processamento de Alto Desempenho. Curso Permanente. Escola Regional de Alto Desempenho, 2006.
- A 39ª lista Top500. Os computadores mais rápidos do mundo. LEONARDO GARCIA TAMPELINI
- Interconnection Network Architectures for High-Performance Computing, Cyriel Minkenberg IBM Research ─ Zurich, 2013.
- Nedialkov, N., Interconnection Networks, CS/SE 2015
- Dally, W. J. From Hypercubes to Dragonflies, a short history of interconnect, Stanford University, 2008.
- Minkenberg, C.,Interconnection Network Architectures for High-Performance Computing, IBM, 2013.
- Silva, J. M. O., Estudo e construção de um Ambiente de Alto Desempenho utilizando Cluster Computacional, UFPI, 2009.

Obrigado!

Dúvidas?