

## ANÁLISE COMPARATIVA DAS REDES DE INTERCONEXÃO CROSSBAR E DELTA

João Angelo Martini\*

**RESUMO.** Este artigo apresenta uma análise comparativa entre as redes de interconexão *crossbar* e delta em termos de custo e de performance para sistemas multiprocessadores. Apresenta também um algoritmo para o elemento de chaveamento da rede delta. Uma descrição do padrão de interconexão da rede delta também é realizada.

**Palavras-chave:** rede crossbar, rede delta, rede de interconexão.

## COMPARATIVE ANALYSIS OF CROSSBAR AND DELTA INTERCONNECTION NETWORKS

**ABSTRACT.** This article presents a comparative analysis between crossbar and delta interconnection networks with respect to their cost and performance for multiprocessor systems. It also presents an algorithm for the delta network switching element. A description of the delta network interconnection pattern is also carried out.

**Key words:** crossbar network, delta network, interconnection network.

## INTRODUÇÃO

As redes de interconexão (RI) desempenham um papel fundamental em projetos de sistemas de computação paralela. São responsáveis pelo tráfego da informação que circula pelo sistema paralelo. A Figura 1 mostra um sistema multiprocessador genérico e dá noção do papel crucial de uma RI num sistema paralelo.

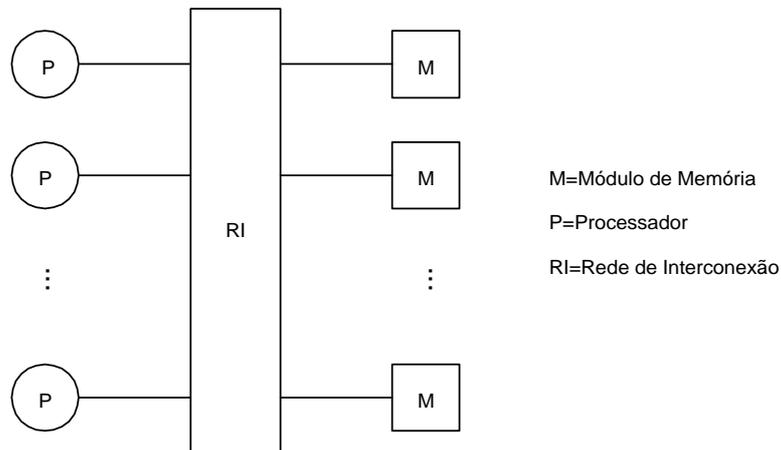
---

\* Departamento de Informática, Universidade Estadual de Maringá, Av. Colombo, 5790, Câmpus Universitário, 87020-900, Maringá-Paraná, Brasil.

Correspondência para João Angelo Martini.

Data de recebimento: 10/11/97.

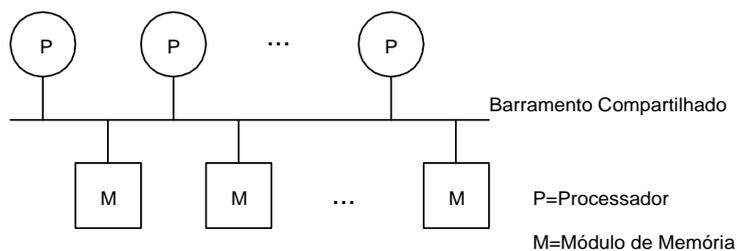
Data de aceite: 21/11/97.



**Figura 1.** Sistema Multiprocessador.

Todo tráfego entre processadores e memórias é coordenado pela rede de interconexão. Dessa forma, a performance do sistema está relacionada a um projeto eficiente da rede que interconecta os processadores aos módulos de memória.

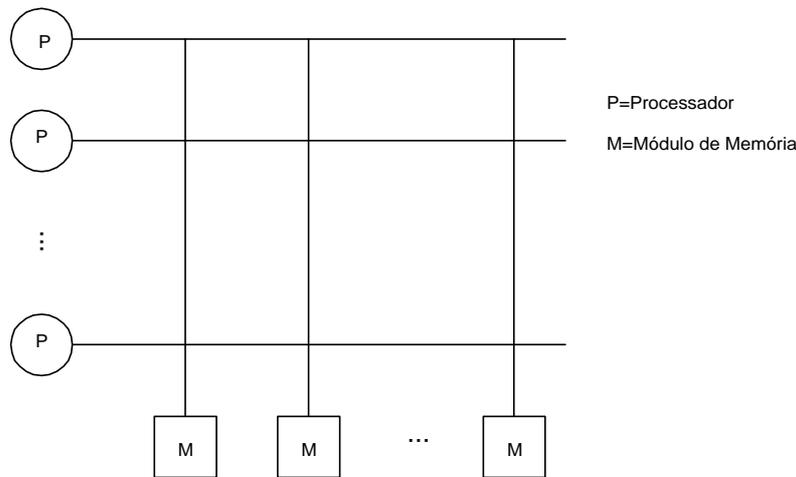
A Figura 2 mostra uma interconexão através de um barramento compartilhado, que é a forma mais simples de interconexão, entretanto, é ineficiente e pouco confiável.



**Figura 2.** Interconexão por Barramento Compartilhado.

A interconexão por barramento compartilhado não permite mais de uma transferência por vez entre processadores e memórias. Esse tipo de rede se torna mais ineficiente conforme o sistema cresce. Para um grande número de processadores, o tempo de espera para obter o controle do barramento é demasiado longo.

A Figura 3 apresenta um sistema interconectado através de uma chave crossbar, a qual suporta todas as possíveis conexões distintas entre processadores e memórias simultaneamente. Essa chave apresenta boa performance, mas seu custo se torna proibitivo para um sistema grande, pois o custo cresce rapidamente com o aumento do tamanho da chave.



**Figura 3.** Chave Crossbar.

Como alternativa em termos de custo e de performance, as redes de interconexão multiestágio (RIM) encontram-se no meio do espectro em cujas extremidades estão o barramento compartilhado e a crossbar. As RIMs apresentam um balanço entre custo e performance. Um exemplo de RIM, a rede delta, é mostrado na Figura 4<sup>1</sup>.

Muitas pesquisas têm sido desenvolvidas na área de redes de interconexão (Park e Bose, 1997; Wang *et al*, 1995) e alguns trabalhos focalizam a análise de performance de redes (Wu e Lee, 1992; Patel, 1981; Ligon e Ramachandran, 1997; Bhuyan *et al*, 1989).

#### **Elemento de Chaveamento**

O elemento básico empregado na operação de uma rede multiestágio é denominado de elemento de chaveamento (EC). Esse dispositivo

---

<sup>1</sup> Elementos de Chaveamento (ECs) serão definidos na próxima seção.

corresponde a uma pequena chave crossbar, como, por exemplo, uma chave 2x2 (Figura 5).

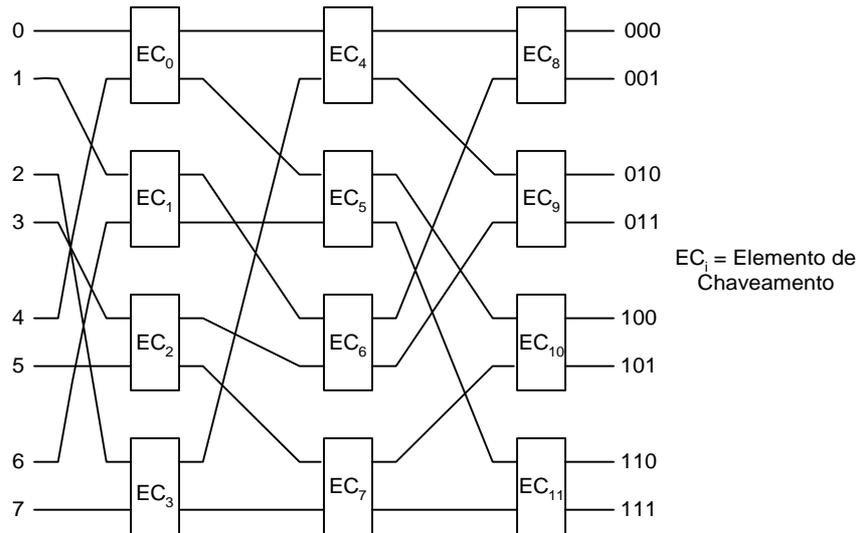


Figura 4. Rede Delta.



Figura 5. Elemento de Chaveamento.

O EC opera de acordo com um bit de controle C; quando C=0, a entrada é conectada à saída 0 e, quando C=1, é conectada à saída 1. Em geral, a política de resolução de conflitos, usada para definir entre duas entradas simultâneas destinadas à mesma saída, baseia-se em dois métodos:

- determinístico, escolhe sempre a entrada A ou sempre a B.
- randômico, a escolha é aleatória.

O EC tem seu princípio de operação baseado num algoritmo simples, descrito a seguir; no caso de conflito, uma alternativa ao descarte da informação bloqueada pode ser a utilização de buffers.

<pre> Se <math>C_A = C_B</math> então se <math>C_A = 0</math> então { <math>S_0 = [A]</math> buffer = [B] } senão { <math>S_1 = [A]</math> buffer = [B] } senão se <math>C_A = 0</math> e <math>C_B = 1</math> então { <math>S_0 = [A]</math> <math>S_1 = [B]</math> } senão { <math>S_0 = [B]</math> <math>S_1 = [A]</math> } </pre>	<p>Endereços de destino: <math>C_A</math> e <math>C_B</math> (bits de controle para as entradas A e B)</p> <p>Saídas: saída 0 (<math>S_0</math>), saída 1 (<math>S_1</math>)</p> <p>Entradas: entrada A=[A], entrada B=[B]</p> <p>Buffer: armazena entrada bloqueada</p>
---	--

Figura 6. Algoritmo do EC.

### Padrão de Interconexão

Uma rede multiestágio pode ser construída a partir da interligação de vários ECs distribuídos em diversos estágios (“colunas”), seguindo um determinado padrão de interligação. A partir de um EC  $2 \times 2$ , é possível obter-se um demultiplexador  $1 \times 2^n$  (Patel, 1981), fazendo-se uma árvore binária desse módulo, conforme mostra a Figura 7. Assim, pode-se construir uma rede  $8 \times 8$  com ECs  $2 \times 2$ , seguindo os passos mostrados na Figura 8.

A Figura 8a mostra o início com uma árvore demultiplexadora. Esse processo prossegue da Figura 8b até a 8d, em que obtém-se a rede  $8 \times 8$ . Cada EC que recebe entradas vindas de outros ECs deve ter ambas as entradas provenientes dos terminais de saídas superiores do EC anterior, ou ambas provenientes das saídas inferiores. A rede delta é uma rede *bit-controlled*, isto é, cada estágio da rede é controlado por um dígito (bit) do endereço de destino. Por exemplo, com os destinos marcados em binário, uma fonte A requer conexão com o destino ( $d_2d_1d_0$ ). No primeiro estágio os ECs são controlados pelo bit  $d_2$ ; no segundo, por  $d_1$  e, no terceiro, por  $d_0$ .

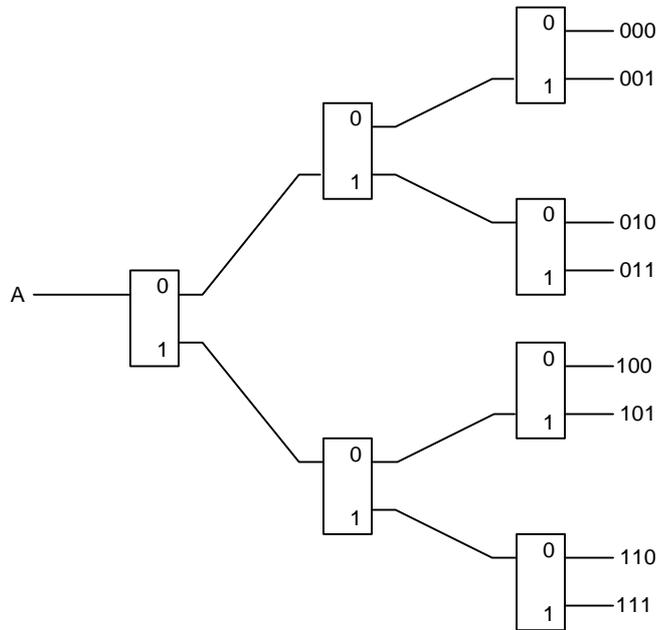


Figura 7. Demultiplexador 1x8.

A Figura 4 apresenta as entradas da rede rotuladas de forma diferente da Figura 8d. Esse renomeamento das entradas é feito para possibilitar uma permutação identidade, ou seja, a conexão de 0 para 0, 1 para 1 e assim por diante, simultaneamente.

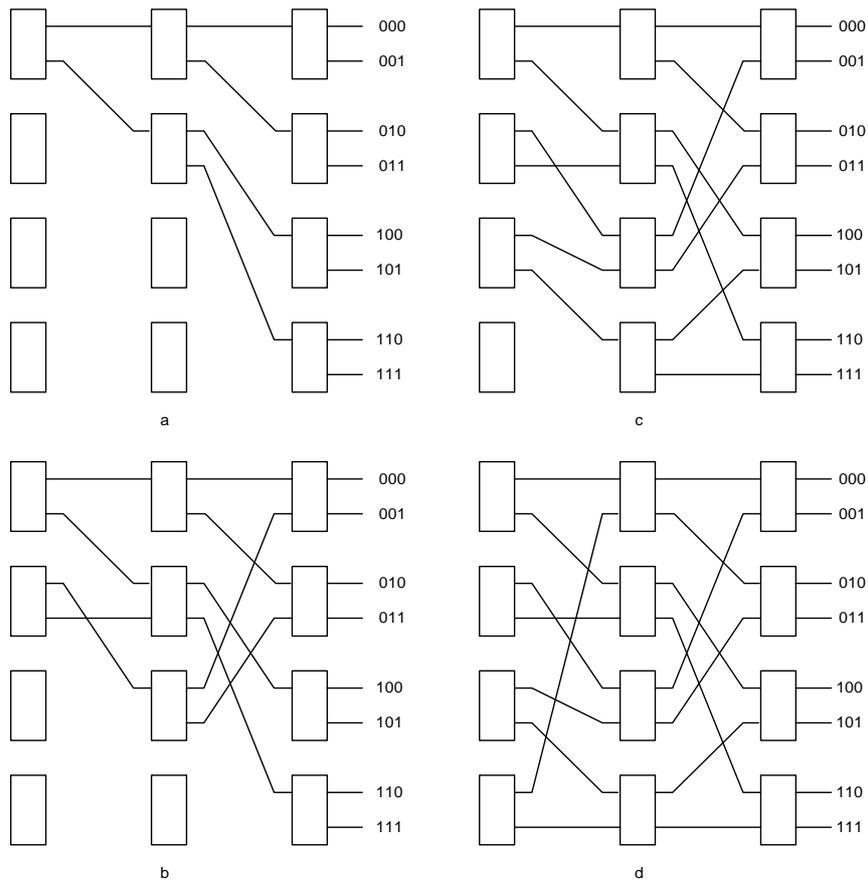
#### Análise de Performance

Considerando a capacidade de conexões (CC) como sendo o número máximo de conexões entrada-saída simultâneas que podem ser efetuadas pela rede, para uma chave crossbar  $N \times N$ ,  $CC = \{N\}$  e, para uma RIM  $N \times N$ ,  $CC = \{N\}$ , ainda que algumas permutações não sejam possíveis.

No caso de chaves crossbar assimétricas, em que o número de entradas é diferente do número de saídas, a capacidade de conexão é  $CC = \min\{M, N\}$ , sendo  $M$  o número de entradas e  $N$  o de saídas. Numa RIM assimétrica,  $CC = \min\{M, N_i\}$ , com  $i$  correspondendo ao  $i$ -ésimo estágio da rede.

Num sistema multiprocessador, a Largura de Banda (LB) é definida como o número de pedidos de acesso à memória aceitos por ciclo; e a

Probabilidade de Aceitação (PA) é a relação da LB esperada pelo número esperado de pedidos gerados por ciclo, isto é, a probabilidade de que um pedido arbitrário seja aceito.



**Figura 8.** Construção de uma rede delta 8x8.

Para analisar as redes, são consideradas algumas hipóteses:

- os pedidos gerados por cada processador são independentes, gerados randomicamente e uniformemente distribuídos sobre todos os módulos de memória;
- as mensagens começam e terminam simultaneamente, operação síncrona;

- cada processador gera um novo pedido com uma probabilidade  $p$ , no início de um ciclo. E  $p$  é o número médio de pedidos gerados por ciclo por cada processador;
- os pedidos bloqueados são ignorados, e os pedidos emitidos em um ciclo são independentes dos pedidos emitidos no ciclo anterior.

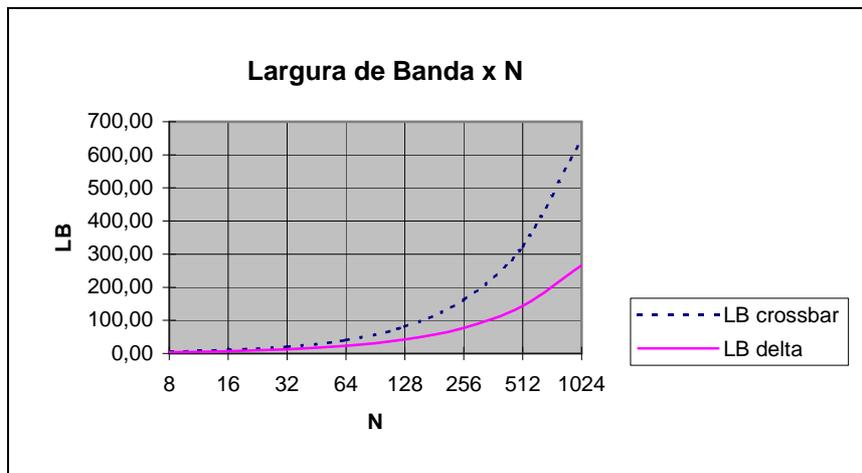
Para uma chave crossbar  $M \times N$ , a Largura de Banda e a Probabilidade de Aceitação são dadas por (Patel, 1981; Bhuyan *et al*, 1989):

- $LB = N - N (1 - p/N)^M$
- $PA = LB/pM = (N/pM) - (N/pM) (1 - p/N)^M$

E para a rede delta:

- $LB = b^n p_n$ , em que  $b$  corresponde ao tamanho do EC  $a \times b$ , nesse caso EC  $2 \times 2$  e,  $p_n = 1 - (1 - m_{n-1}/b)$  e  $p_0 = p$
- $PA = b^n p_n / a^n p$

Computando LB e PA para crossbar e rede delta com tamanhos de  $8 \times 8$  a  $1024 \times 1024$ , temos os seguintes resultados:



**Figura 9.** Largura de Banda x N.

Essa análise considera uma rede delta  $2^n \times 2^n$  ( $N=2^n$ ,  $n \geq 1$ ) que utiliza ECs  $2 \times 2$ , chaves crossbar  $N \times N$  (todos os inteiros  $N \geq 1$ ) e uma taxa de geração de pedidos para cada processador igual a 1.

A Figura 9 mostra a LB em função de  $N$  para a chave crossbar e rede delta, com tamanhos de  $8 \times 8$  a  $1024 \times 1024$ . A LB da crossbar apresenta

diferenças mais significativas, com índices melhores que a da delta, a partir de  $N=32$ . Para valores mais elevados de  $N$ , a crossbar apresenta desempenho muito melhor, chegando a ter LB 143,19% maior que a da delta para  $N=1024$ .

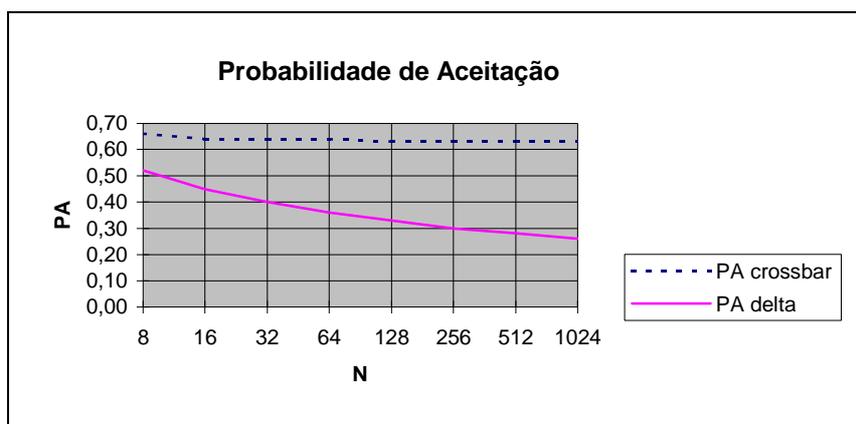


Figura 10. Probabilidade de Aceitação x N.

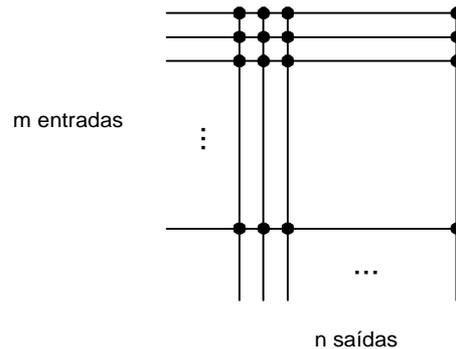
A Figura 10 mostra PA em função de  $N$  para a chave crossbar e rede delta, para  $N$  variando de 8 a 1024. PA para a chave crossbar apresenta valor quase constante, com uma diferença de apenas 4,76% entre o maior e o menor valor. A rede delta apresenta PA decrescente conforme  $N$  cresce, apresentando uma diferença de 100% entre o maior e o menor valor. Como PA corresponde à probabilidade de que um pedido arbitrário seja aceito, dá uma medida do tempo de espera. Quanto maior PA, menor será o tempo de espera.

### Modelo de Custo

Definindo um modelo de custo “uniforme” que possa ser aplicado às duas redes, é possível fazer uma análise comparativa em termos de custo e de performance dessas redes. A Figura 11 mostra o modelo de custo para uma chave  $m \times n$ , com  $m$  entradas e  $n$  saídas.

Conforme esse modelo, o custo ( $C$ ) de uma chave crossbar  $m \times n$  é de  $m.n$  unidades. Cada unidade corresponde a um ponto de interconexão, que para efeito de comparação pode ser considerado como tendo custo

igual a 1. Para chaves crossbar  $M \times N$  em que  $M=N$ , o custo tem um crescimento quadrático.



**Figura 11.** Modelo de Custo para chaves.

Para uma RIM  $N \times N$ , usando ECs  $2 \times 2$ , o custo é calculado da seguinte maneira:

- para um EC  $2 \times 2$ , o custo é de 4 unidades;
- a RIM  $N \times N$  possui  $\log_2 N$  estágios;
- cada estágio tem  $N/2$  ECs.

Então, o custo total é: Custo de 1 EC  $\times$   $N^0$  de ECs por estágio  $\times$   $N^0$  de estágios. Numa RIM  $8 \times 8$ , temos o seguinte custo:

- custo de 1 EC  $2 \times 2 = 4$ ;
- $N^0$  de ECs por estágio  $= 8/2 = 4$ ;
- $N^0$  de estágios  $= \log_2 8 = 3$ ;
- Custo Total  $= 4 \times 4 \times 3 = 48$ .

O custo para uma crossbar  $8 \times 8$  ( $N=8$ ) é  $N^2=64$  unidades; portanto, 33,33% maior que o da correspondente RIM. Através desse modelo de custo, obtêm-se os valores para a chave crossbar e rede delta de  $8 \times 8$  a  $1024 \times 1024$ , mostrados na Figura 12.

A curva de custo em função de  $N$  mostra que a crossbar apresenta um custo extremamente elevado em relação à rede delta, conforme o sistema cresce. A diferença de custo entre as duas redes se acentua rapidamente com o crescimento de  $N$ , devido ao custo quadrático da crossbar. Já no início, com  $N=8$ , o custo da crossbar é 33,33% maior que o da delta; para  $N=128$ , essa diferença é de 814,29%, chegando a 5020% para  $N=1024$ . A

rede delta apresenta uma curva de custo bastante razoável em relação à crossbar, que tem custo proibitivo mesmo para sistemas de porte médio.

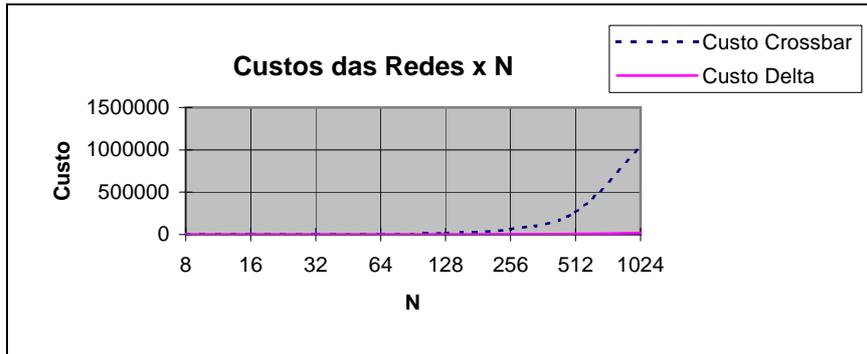


Figura 12. Custos das Redes x N.

**Fator de Custo**

O Fator de Custo (FC) é definido como a relação da largura de banda (LB) pelo custo (C) da rede. Esse fator retrata a relação custo-eficiência da rede, sendo que uma rede ideal deve proporcionar a maior largura de banda com o menor custo possível. A Figura 13 mostra os fatores de custo para as redes em análise.

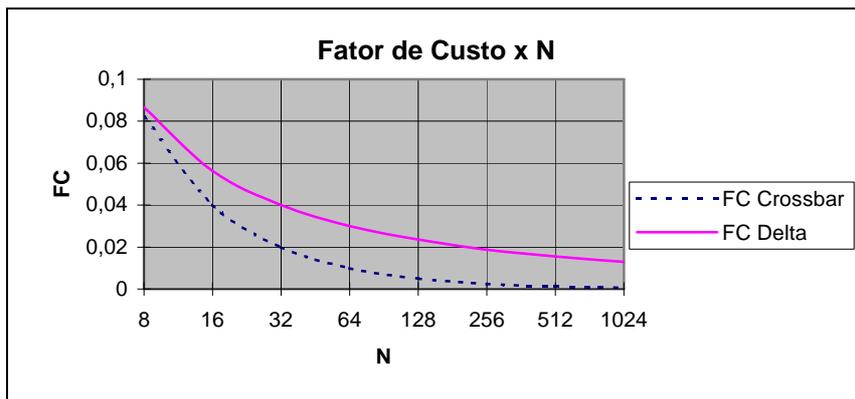


Figura 13. Fator de Custo x N.

A curva de fator de custo mostra que a rede delta apresenta melhor relação custo-eficiência que a crossbar para todo o espectro de  $N$  considerado na análise. Para  $N=8$ , o FC da rede delta é 12,5% maior que o da crossbar, e esse fator melhora com o crescimento de  $N$ , atingindo a diferença de 1996,77% em favor da delta para  $N=1024$ . Isso mostra que a delta apresenta uma melhor taxa de desempenho por custo.

### Aplicações de Redes de Interconexão

A fim de situar as redes de interconexão no contexto de arquiteturas paralelas, podem-se citar algumas das máquinas mais intensamente pesquisadas que empregam tais redes: MTTDA (Arvind e Nikhil, 1990), EM-4 (Sakai *et al*, 1995; Sakai *et al*, 1993) e WOLF (Martini, 1994).

A máquina MTTDA foi proposta pelo grupo do Prof. Arvind do MIT, e possui 64 elementos de processamento e módulos de memória interconectados através de uma rede  $n$ -cúbica. Uma rede  $n$ -cubo é um subconjunto da rede delta.

A arquitetura EM-4 foi proposta no ETL, no Japão, e emprega uma rede ômega, que também é um subconjunto da rede delta, para interconectar seus elementos de processamento.

A arquitetura WOLF, proposta pelo Grupo de Instrumentação e Informática do IFSC-USP, utiliza duas redes de interconexão para interligar as unidades que compõem a máquina. Essa arquitetura foi intensamente analisada para diversas combinações de pares de redes; entre as redes simuladas estavam: crossbar,  $n$ -cubo, ômega e delta.

### Conclusão

Esse artigo apresentou uma análise comparativa das redes crossbar e delta, segundo alguns parâmetros como: capacidade de conexão, largura de banda, probabilidade de aceitação, custo e fator de custo.

Dentre esses parâmetros, o de maior impacto na avaliação é o fator de custo, pois abrange tanto a performance da rede como também seu custo. Considerando ainda um outro aspecto de sistemas paralelos que é a expansibilidade, o fator de custo tem seu papel ainda mais reforçado nesse contexto, visto que dele pode-se analisar a inviabilidade da chave crossbar quando se considera a expansão do sistema. Nesse quesito, a rede delta apresentou um comportamento muito melhor.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ARVIND & NIKHIL, R.S. Executing a program on the MIT tagged-token dataflow architecture. *IEEE Transactions on Computers*, 39(3):300-318, 1990.
- BHUYAN, L.N., QING, Y. & AGRAWAL, D.P. Performance of multiprocessor interconnection networks. *Computer*, 22(2):25-37, 1989.
- LIGON, W.B. & RAMACHANDRAN, V. Toward a more realistic performance evaluation of interconnection networks. *IEEE Transactions on Parallel & Distributed Systems*, 8(7):681-694, 1997.
- MARTINI, J.A. Estudo de desempenho de redes de interconexão para a arquitetura a fluxo de dados Proto-WOLF. São Carlos, 1994. Dissertação (Mestrado em Ciências Física Aplicada) - Instituto de Física de São Carlos.
- PARK, S. & BOSE, B. All-to-all broadcasting in faulty hipercubes. *IEEE Transactions on Computers*, 46(7):749-755, 1997.
- PATEL, J.H. Performance of processor-memory interconnection for multiprocessors. *IEEE Transactions on Computers*, 30(10):771-780, 1981.
- SAKAI, S., KODAMA, Y., SATO, M., SHAW, A., MATSUOKA, H., HIRONO, H., OKAMOTO, K. & YOKOTA, T. Reduced interprocessor-communication architecture and its implementation on EM-4. *Parallel Computing*, 21(5):753-769, 1995.
- SAKAI, S., KODAMA, Y. & YAMAGUCHI, Y. Design and implementation of a circular omega network in the EM-4. *Parallel Computing*, 19(2):125-142, 1993.
- WANG, M.-C., SIEGEL, H.J., NICHOLS, M.A., ABRAHAM, S. Using a multipath network for reducing the effects of hot spots. *IEEE Transactions on Parallel & Distributed Systems*, 6(3):252-268, 1997.
- WU, C. & LEE, M. Performance analysis of multistage interconnection network configurations and operations. *IEEE Transactions on Computers*, 41(1):18-27, 1992.