Acta Scientiarum 20(4):477-481, 1998.
ISSN 1415-6814.

Influência das redes de interconexão sobre os tempos de execução na arquitetura Wolf

João Angelo Martini[1]*, Álvaro Garcia Neto[2] e Marcos Antônio Cavenaghi[3]

[1]Departamento de Informática, Universidade Estadual de Maringá, Av. Colombo, 5790, 87020-900, Maringá-Paraná, Brazil.
[2]Departamento de Física e Informática do IFSC-Universidade de São Paulo, São Carlos-São Paulo, Brazil.
e-mail: alvaro@ifsc.sc.usp.br [3]Departamento de Computação da Universidade Estadual Paulista, Bauru-São Paulo, Brazil,
*e-mail: marcos@dco.bauru.unesp.br *Author for correspondence.*

RESUMO. Este artigo apresenta os resultados das simulações das redes de interconexão da arquitetura Wolf. Tendo por base os requisitos de comunicação da arquitetura, selecionamos algumas redes de interconexão para simulação por meio do simulador SAW e, utilizando alguns testes amplamente difundidos em pesquisas de fluxo de dados analisamos a influência dessas redes sobre os tempos de execução.

Palavras-chave: redes de interconexão, fluxo de dados, arquiteturas paralelas

ABSTRACT. Interconnection Networks influence on execution time in Wolf architecture. This study presents simulation results of Wolf architecture interconnection networks. Based on architecture communication requirements, we selected some interconnection networks for simulation through Wolf simulator. Using some widely know tests, we analyzed the influence of this interconnection networks on execution time.

Key words: interconnection networks, dataflow, parallel architectures.

Redes de interconexão desempenham papel relevante no contexto de arquiteturas paralelas. Elas constituem a estrutura básica de comunicação de um sistema de processamento paralelo. O projeto de redes é um compromisso entre custo, desempenho e complexidade de controle. A solução ideal possibilita que todas as unidades do sistema se comuniquem simultaneamente sem conflitos, mas essa rede ideal tem custo e complexidade muito elevados. Busca-se, então, uma alternativa que não se encontre nos extremos do espectro de custo e desempenho, isto é, uma rede com custo acessível e desempenho satisfatório. Outro fator importante é a independência de aplicações: a rede ideal deve satisfazer aplicações de propósito geral e não endereçar-se apenas a um tipo específico.

Além da escolha da topologia mais adequada, outras decisões também têm papel crucial no projeto de redes de interconexão. Características como modo de operação, estratégia de controle e método de chaveamento, podem determinar diferentes comportamentos do sistema.

O modo de operação trata da maneira como os dados são transmitidos e existem dois tipos básicos: síncrono e assíncrono. O modo síncrono é caracterizado pela existência de um relógio global que emite sinais para todos os componentes do sistema a fim de sincronizá-los. O modo assíncrono opera sem a presença do relógio global, possibilitando melhor expansibilidade e modularidade, mas com maior complexidade. A estratégia de controle corresponde à maneira como são gerados os sinais de controle que coordenam as funções de roteamento da rede. O controle pode ser centralizado ou distribuído. No tipo centralizado, os sinais são gerados por uma única unidade de controle, o que pode comprometer a confiabilidade do sistema, além de requerer um controlador bastante complexo. No distribuído, os sinais são originados localmente nos componentes do sistema através de controladores mais simplificados. Assim, elimina-se a dependência de um controlador global complexo. O método de chaveamento diz respeito à utilização física das chaves na realização de uma função de roteamento. Há duas principais metodologias de chaveamento: por circuito e por pacotes. No chaveamento por circuito, estabelece-se um caminho físico entre fonte e destino antes de se iniciar a comunicação e mantém-se o caminho durante toda a transmissão. No chaveamento por pacotes a mensagem a ser transmitida é dividida em pequenos pacotes, para ser emitida por meio da rede

através dos elementos de chaveamento que os recebem e enviam a um novo elemento num estágio subseqüente.

Redes de interconexão têm sido abordadas apenas superficialmente no contexto de arquiteturas a fluxo de dados, sendo relegadas a um plano secundário, o que certamente não condiz com a função primordial que elas desempenham. Esse aparente desinteresse torna-se um fator agravante uma vez que o custo e o desempenho estão intrinsecamente relacionados com a estrutura de comunicação da arquitetura, demandando uma avaliação criteriosa e profunda dos requisitos de comunicação.

Para contextualizar a utilização de redes de interconexão em arquiteturas a fluxo de dados, podemos citar as máquinas desenvolvidas no MIT (Arvind e Nikhil, 1990), Universidade de Manchester (Gurd *et al*., 1985), NTT e ETL (Hiraki *et al*., 1987; Sakai *et al*., 1993) entre outras que empregam tais redes para efetuarem a comunicação. No MIT, a máquina MTTDA (*MIT Tagged Token Dataflow Architecture*) utiliza uma rede n-cúbica de chaveamento por pacotes para interligar PEs e *I-structures* e a máquina VIM emprega uma hierarquia de redes de interconexão que efetuam a comunicação entre os diversos módulos do sistema. No Japão, a máquina Sigma-1 utiliza uma rede hierárquica de dois níveis, e a EM-4 utiliza uma rede multi-estágio ômega com controle distribuído. A Universidade de Manchester desenvolveu uma máquina a fluxo de dados dinâmica rotulada por fichas denominada MMDM (*Manchester Multi-Ring Dataflow Machine*), que possui uma estrutura em anel e também utiliza uma chave para conectar os anéis e o hospedeiro.

Arquitetura Wolf

A arquitetura Wolf (Cavenaghi, 1992) baseia-se no modelo de fluxo de dados dinâmico com granularidade variável e é derivada da MMDM. A Figura 1 apresenta as unidades que compõem a arquitetura. A comunicação entre as unidades é feita por meio da troca de fichas que representam os dados e constituem pacotes de informações.

Descrição Funcional. Uma descrição funcional mais detalhada é discutida em Garcia Neto e Ruggiero (1992). A seguir, é apresentado de forma suscinta o funcionamento da arquitetura. O mecanismo de Entrada e Saída (E/S) é responsável pela conexão da máquina com o exterior. A Chave de Coleta (CC) recolhe fichas provenientes da Entrada e das demais unidades conectadas em suas entradas e as transmite para a Memória de Dados (MD) ou Saída. Caso a MD esteja indisponível, a CC transmite para a Fila de Fichas (FF) que as armazena e as distribui posteriormente quando a MD estiver disponível. A FF regula o fluxo de fichas na máquina evitando que os anéis (formados pelo par Memória de Instruções (MI) e Unidade Funcional (UF)) sejam sobrecarregados. A Unidade de Emparelhamento (UE) realiza o emparelhamento de operandos utilizando o endereço recebido da CC para procurar a ficha parceira. As fichas emparelhadas na MD são enviadas à Chave de Distribuição (CD) que as distribui para as seguintes unidades: anéis, Processador Vetorial e Memória Vetorial (PVMV) e Unidade de Controle de Paralelismo (UCP), de acordo com o tipo de ficha. A MI armazena as informações que serão usadas para produzir uma ficha executável (EX), a qual será executada pela correspondente UF, e os resultados produzidos pela UF são enviados à CC. Dados estruturados são tratados pelo PVMV e a Unidade de Controle de Paralelismo (UCP) controla o número de instanciações distintas de uma mesma instrução (Ruggiero, 1987; Magna, 1992).

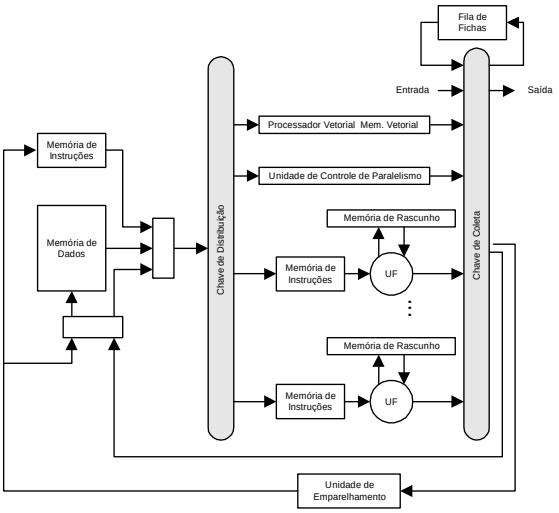

Figura 1. Arquitetura Wolf

Requisitos de Comunicação. A arquitetura Wolf é uma máquina de propósito geral não vinculada a um determinado padrão de comunicação de uma aplicação particular. Isso implica na escolha de uma topologia dinâmica que devido a reconfigurabilidade, permite atender a padrões arbitrários de comunicação. A comunicação entre as unidades é feita através da troca de fichas como em arquiteturas de passagem de mensagem. Como os anéis entre as chaves CD e CC podem trabalhar independentemente e em paralelo, a transferência de fichas através das redes deve ser realizada de modo assíncrono, permitindo suprir os anéis que se

encontram disponíveis para processamento. A estratégia de controle distribuído foi adotada por questões de simplicidade e confiabilidade evitando-se a dependência de uma única unidade complexa geradora de sinais de controle (Ibbett e Topham, 1989). Com base nesses requisitos, temos as seguintes características operacionais para as redes CD e CC: chaveamento por pacotes, modo de operação assíncrono e controle distribuído.

Simulador da Arquitetura Wolf (SAW)

A primeira versão do SAW (Cavenaghi, 1992) usa duas redes de interconexão do tipo Crossbar (Ward e Halstead Jr., 1990), denominadas Cross1 e Cross2, para CD e CC, respectivamente. Com base nos requisitos de comunicação, foram selecionadas três redes para simulação e análise através do SAW: Crossbar, multi-estágio derivada da Delta (Hwang e Briggs, 1984; Ibbett e Topham, 1989) e uma rede derivada de uma GSN (Bhuyan e Agrawal, 1983).

A chave Cross1 identifica o tipo da ficha através de seu endereço. Se for EX, envia-na para o PVMV, caso esteja livre, senão bloqueia na entrada. Caso o tipo seja par (PA), envia-na para a primeira UF livre, dando maior prioridade à ordem crescente de endereços. Se todas as UFs estiverem ocupadas, a ficha será bloqueada.

A Cross2 identifica o tipo através do endereço da ficha e a envia para uma de suas três saídas desde que haja disponibilidade. Somente as UFs enviam fichas para a Saída porque são elas que produzem fichas do tipo SA. Há possibilidade de re-roteamento quando a ficha é destinada à MD e esta não se encontra disponível. Nesse caso, envia-se para a FF e, se ela também estiver indisponível, ocorre o bloqueio. Na ocorrência de conflitos, as fichas são transmitidas obedecendo-se a ordem crescente dos endereços de entrada da rede.

Chave de Coleta (CC). A CC foi implementada utilizando-se as seguintes estruturas de interconexão: Crossbar 8x2, rede 8x2 derivada de uma GSN 8x4 e uma rede 8x2 derivada de uma multi-estágio Delta 9x4.

Na CC implementada através da crossbar as entradas provenientes da FF, Entrada e PVMV somente endereçam fichas à MD e não à Saída, e o tipo de ficha é identificado pelo seu endereço. Através do uso de um Arbitrador, que coordena a transmissão de fichas destinadas à MD, pode-se simplificar a rede reduzindo-a a duas saídas. Isso simplifica o controle e descarta a necessidade de re-roteamento como na Cross2. O Arbitrador verifica a possibilidade de transmitir as fichas à MD, podendo enviá-las à FF, caso a MD esteja indisponível ou bloqueá-las, caso a FF também esteja indisponível. A política de resolução de conflitos determina que será transmitida a ficha cuja porta de entrada na rede tenha o endereço de menor ordem, obedecendo a uma ordem crescente de endereços, a partir do topo da rede.

Na implementação com a rede 8x2 GSN-derivada, há um estágio a mais que a anterior. A identificação do tipo da ficha e a política de resolução de conflitos são iguais. Os elementos de chaveamento do primeiro estágio somente transmitem para o estágio seguinte se os respectivos buffers de entrada estiverem livres, senão ocorre o bloqueio. O Arbitrador realiza as mesmas funções já descritas na crossbar.

A CC implementada através de uma rede derivada a partir da Delta também possui dois estágios e segue os mesmos princípios de transmissão e resolução de conflitos descritos para as demais implementações.

Chave de Distribuição (CD). A CD trata somente as fichas dos tipos PA e EX. O tipo EX deve ser enviado apenas ao PVMV e o PA somente aos anéis. Assim podemos separar a CD denominada Demux em dois blocos: um Elemento de Chaveamento (EC) semelhante ao Arbitrador da CC e um Distribuidor cuja função é detectar qual dos anéis está habilitado a receber a ficha PA. O EC identifica o tipo de ficha e a envia para a saída correspondente. As MIs que se encontram disponíveis podem receber fichas do Distribuidor; em caso de conflitos, a transmissão é feita por ordem crescente de endereços.

Avaliação das Chaves de Coleta e de Distribuição

Para investigar a influência das redes de interconexão sobre os tempos de execução foram empregados alguns testes (Chambers *et al*., 1984; Gurd *et al*., 1985) bastante difundidos em pesquisas de fluxo de dados. Foram escolhidos quatro testes codificados em Sisal (MacGraw, 1985) e que exploram características distintas da máquina:

- Incremento (Inc) que é o mais simples dos testes utilizados e é estritamente seqüencial.
- Função de Ackermann (Ack) que explora características de recursividade e gera alto grau de comunicação.
- Integração Binária (IB) também apresenta alto grau de recursividade, efetua o cálculo da área de uma função por integração binária utilizando a regra do trapézio.
- Multiplicação de Matrizes (MM) explora as características do PVMV.

Resultados e discussão

A Tabela 1 mostra os resultados obtidos para cada teste para as diversas combinações possíveis dos

pares CD-CC em função do tempo de execução, o que indica a eficiência do par de redes, pois elas podem melhorar ou degradar a performance da máquina, conforme esse parâmetro seja menor ou maior. Os resultados dessa tabela apresentam diferenças significativas de um teste para outro. Por exemplo, podemos notar as diferenças dos tempos de execução para a melhor combinação de redes para os testes Incremento (a) e Integração Binária (b): o tempo de execução para (a) é 1178 e para (b) 56004.

Tabela 1. Tempos de execução

Redes	Incremento	Ackermann	Int. Bin	Mult. M
Cross1-Crossbar	1502	218202	136001	89002
Cross1-Cross2	1178	137403	57000	69002
Cross1-Delta	1716	260604	116001	84539
Cross1-Gsn	1715	273400	180000	82008
Demux-Crossbar	1715	246802	136000	93004
Demux-Cross2	1390	154802	56004	70002
Demux-Delta	1925	275605	114004	87002
Demux-Gsn	1936	284804	184003	84011

Com o intuito de identificar o par de redes com melhor desempenho, os tempos de execução foram classificados em ordem crescente. O processo de classificação enquadrou na mesma ordem os pares cujos tempos eram insignificativamente distintos. A Tabela 2 mostra os resultados em função dessa classificação, permitindo-se identificar facilmente os pares de redes com melhor desempenho.

Tabela 2. Classificação das redes

Redes	Incremento	Ackermann	Int. Bin.	Mult. M.
Cross1-Crossbar	3	3	5	6
Cross1-Cross2	1	1	2	1
Cross1-Delta	4	5	4	4
Cross1-Gsn	4	6	6	3
Demux-Crossbar	4	4	5	7
Demux -Cross2	2	2	1	2
Demux-Delta	5	6	3	5
Demux-Gsn	5	7	7	4

A seguir, são discutidos os resultados obtidos em função das diversas combinações de pares de redes de interconexão para CD-CC:

- **Incremento:** A combinação que apresentou o melhor tempo de execução foi o par Cross1-Cross2, enquanto os pares Demux-Delta e Demux-Gsn executaram no maior tempo, sendo os piores casos para esse teste.
- **Ackermann:** O melhor tempo novamente foi obtido pelo par Cross1-Cross2. O pior caso foi apresentado pelo par Demux-Gsn que executou o programa no maior tempo.
- **Integração Binária:** O melhor tempo foi apresentado pelo par Demux-Cross2 e o pior pelo par Demux-Gsn.
- **Multiplicação de Matrizes:** O melhor tempo para esse teste foi obtido pelo par Cross1-Cross2, enquanto o par Demux-Crossbar apresentou o pior índice de tempo.

O par de redes de interconexão Cross1-Cross2 apresentou os melhores índices de tempo para três dos quatro testes empregados nessa avaliação. Ficou em segunda colocação apenas para um dos testes, mesmo assim com uma diferença pequena (1,8%), em relação ao primeiro classificado para o referido teste. O par Demux-Cross2 também apresentou performance muito boa, com índices próximos aos obtidos pelo par Cross1-Cross2. Já o par Demux-Gsn apresentou os piores índices para três dos quatro testes, mostrando ser a combinação mais ineficiente de redes CD-CC para a arquitetura em termos de tempo de execução.

Referências bibliográficas

Arvind; Nikhil, R.S. Executing a program on the MIT Tagged-Token Dataflow Architecture. *IEEE Transact. Comput.*, *39*(3):300-318, 1990.

Bhuyan, L. N.; Agrawal, D.P. Design and performance of generalized interconnection networks. *IEEE Transact. Comput.*, *32*(12):1081-1090, 1983.

Cavenaghi, M.A. *Implementação de um simulador para a arquitetura fluxo de dados Wolf*. São Carlos, 1992. (Master's Thesis in Applied Physics) - Universidade de São Paulo.

Chambers, F.B.; Duce, D.A.; Jones, G.P. *Distributed Computing*. Academic Press, Inc.. Orlando, Florida 32887, United States of America, 1984.

Garcia Neto, A.; Ruggiero, C.A. The proto-architecture of Wolf, a dataflow supercomputer. In: CONFERÊNCIA LATINO AMERICANA DE INFORMÁTICA, 18, 1992, Las Palmas de Gran Canaria, 1992, p. 531-539.

Gurd, J.R.; Kirkham, C.C.; Watson, I. The manchester prototype dataflow computer. *Communic. of the ACM*. *28*(1):34-52, 1985.

Hiraki, K.; Nishida, K.; Sekiguchi, S.; Shimada, T.; Yuba, T. The Sigma-1 dataflow supercomputer: a challenge for new generation supercomputing systems. *J. Inform. Process.*, *10*(4): 219-226, 1987.

Hwang, K.; Briggs, F.A. Computer architecture and parallel processing. McGraw-Hill, *McGraw-Hill series in computer organization and architecture*, United States of America, 1984.

Ibbett, R.N.; Topham, N.P. Architecture of high performance computers. *Macmillan Computer Science Series*. London, England, 1989. v.2.

MacGraw, J.R. *Sisal:* streams and iteration in single assignment language, language reference manual, version 1.2. Lawrence Livermore National Laboratory, 1985.

Magna, P. *Redução dos bits de endereçamento da máquina de fluxo de dados de Manchester*. São Carlos, 1992. (Master's Thesis in Applied Physics) - Universidade de São Paulo.

Ruggiero, C.A. Throttle mechanisms for the Manchester Dataflow Machine. Department of Computer Science, University of Manchester. *Technical Report Series*, Nº UMCS 87-8-1, Oxford Road, Manchester M13 9PL, England, 1987

Sakai, S.; Kodama, Y.; Yamaguchi, Y. Design and implementation of a circular omega network in the EM-4. *Parallel Comput.*, *19*(2):125-142, 1993.

Ward, S.A.; Halstead Jr., R.H. Computation Structures. *The MIT Electrical Engineering and Computer Science Series*. The MIT Press, Massachusetts: Cambridge, 1990.

Received on October 23, 1998.
Accepted on December 20, 1998.