

ArMa - GAPP: Estudo e Aplicação de Arquiteturas Matriciais

CÉSAR A. F. DE ROSE¹
RICARDO MENNA BARRETO¹
ALEXANDRE S. CARISSIMI¹

GERSON G. H. CAVALHEIRO¹
PHILIPPE O. A. NAVAU¹
JOÃO P. KITAJIMA¹

¹Curso de Pós-Graduação em Ciências da Computação
Universidade Federal do Rio Grande do Sul
Caixa Postal 15064
91501-970 Porto Alegre, RS, Brasil
navaux@inf.ufrgs.br

Abstract. The purpose of the ArMa-GAPP (Array Processor - Geometric Arithmetic Parallel Processor) project is the study and utilization of array architectures based on the GAPP chip. This work presents the current state of the GAPP research in CPGCC/UFRGS. It also describes the use of this architecture in the field of image processing.

1 Introdução

Arquiteturas matriciais [Hwang and Briggs, 1984] vêm sendo pesquisadas como uma opção para construção de máquinas com alto grau de paralelismo. Suas características básicas são regularidade e modularidade. Suas aplicações são voltadas ao mapeamento eficiente da solução de problemas onde verifica-se independência de dados, tais como processamento de imagens, transformação rápida de Fourier (FFT) e manipulação de matrizes.

No curso de Pós-Graduação em Ciências da Computação da Universidade Federal do Rio Grande do Sul, desenvolve-se o projeto **ArMa-GAPP** (Arquiteturas Matriciais - *Geometric Arithmetic Parallel Processor*) que estuda a viabilidade do emprego de arquiteturas matriciais e o desenvolvimento de aplicações. Como arquitetura base utiliza-se o circuito GAPP, composta de 72 elementos processadores (EPs). Encontra-se a disposição do projeto a placa NCR GAPP [NCR, 1986] composta de 2 circuitos GAPP num total de 144 EPs.

Até o momento, já foram produzidas ferramentas de suporte para a utilização da placa matricial, implementação de aplicações na área de processamento de imagens e também análises de desempenho, tanto do circuito GAPP, como da placa.

Discute-se a seguir características gerais das arquiteturas matriciais e, em particular, da placa NCR GAPP. A seção 3 discute o trabalho desenvolvido no projeto e a seção 4 é dedicada ao trabalho na área de processamento de imagens. Como conclusão (seção 5) é apresentado o estado atual e perspectivas futuras.

2 Arquiteturas Matriciais

Arquiteturas matriciais são máquinas compostas por um conjunto homogêneo de EPs conectados entre si por uma malha de comunicação uniforme. A característica básica destas arquiteturas é a execução síncrona, por todos os EPs, de instruções enviadas por uma unidade de controle. Segundo a classificação de Flynn [Flynn, 1972], arquiteturas matriciais são máquina SIMD (*Single Instruction flow, Multiple Data flow*).

O ganho de desempenho que pode ser obtido com a utilização de arquiteturas SIMD se dá, sobretudo, em aplicações onde um conjunto de dados deva sofrer o mesmo conjunto de operações. Desta forma, os EPs são inicialmente carregados com os dados originais, e, em seguida, a unidade de controle envia as instruções a serem processadas aos EPs. Após a operação ser completada, os dados manipulados encontram-se disponíveis nas memórias locais dos processadores.

2.1 A Arquitetura GAPP

O circuito GAPP é composto de 72 EPs, conectados em uma malha *mesh* 12x6. Em uma malha *mesh*, cada EP é conectado aos 4 vizinhos mais próximos: Norte, Sul, Leste e Oeste. Os elementos processadores trabalham com palavra de 1 bit. Possuem 4 registradores de 1 bit, memória interna de 128 bits e uma ULA somadora com capacidade de realizar operações lógicas simples (NOT, AND, OR, XOR e XNOR). A figura 1 apresenta um esquema do elemento processador.

Todos os EPs recebem a mesma microinstrução, que pode ativar em paralelo operações com a ULA,

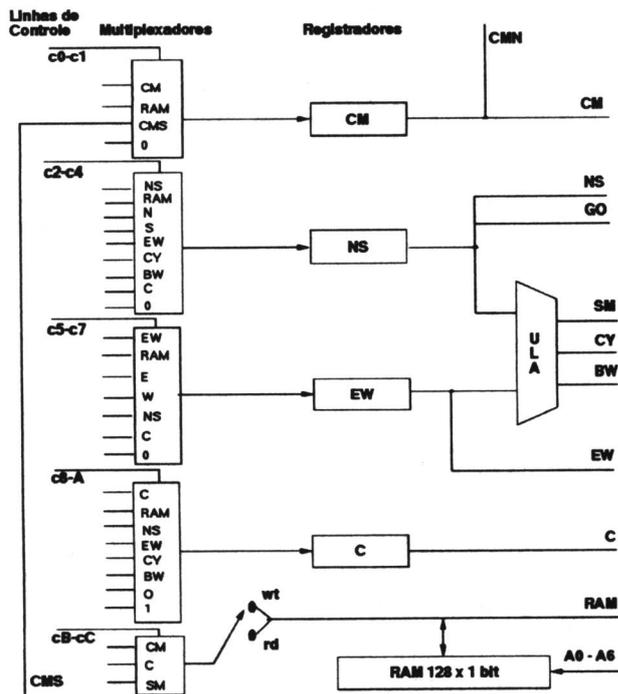
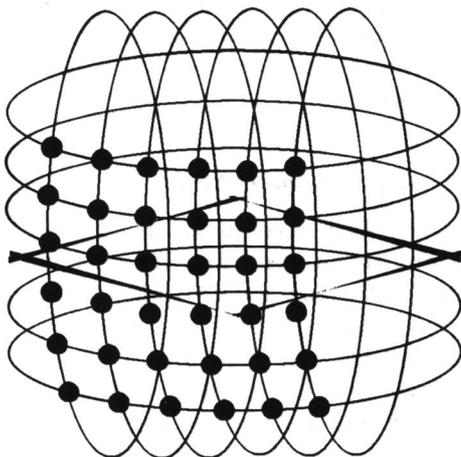


Figure 1: Esquema do EP GAPP

registradores, memória interna e até mesmo comunicação entre com os vizinhos.

É possível agrupar vários circuitos GAPP obtendo um sistema de dimensões múltiplas de 12×6 elementos. A placa NCR GAPP possui um sistema com 2 circuitos GAPP totalizando 144 EPs. Os EPs estão conectados em uma malha *mesh* 12×12 configurada em *wraparound*. Uma malha *mesh* 6×6 em *wraparound* é apresentada na figura 2.

Figure 2: Malha *mesh* 6×6 em *wraparound*

A placa NCR é conectada a um computador IBM-PC, o qual serve como máquina hospedeira, fa-

zendo as vezes de uma unidade de controle necessária as arquiteturas SIMD. Sua utilização é bastante simples, uma vez que a comunicação com o computador hospedeiro se dá através do barramento interno do IBM-PC.

Alguns trabalhos foram desenvolvidos para analisar o desempenho do GAPP, [Wong and Lua, 1990] e [Cavalheiro et al, 1992].

O trabalho de [Wong and Lua, 1990] discute o desempenho do circuito GAPP e o compara com outras arquiteturas matriciais (CLIP4, ICL DAP e MPP) semelhantes ao GAPP. Em suas conclusões, afirmam que, de modo geral, é necessário um maior número de instruções para o GAPP executar uma operação que as demais. Em contra-partida, devido a sua reduzida área, o GAPP é capaz de fornecer até 9 vezes mais resultados que as outras arquiteturas por área ocupada.

Em [Cavalheiro et al, 1992] é analisado o desempenho do sistema 12×12 EPs providos pela placa NCR e também aplicações que podem tirar melhor desempenho deste sistema.

3 Diretrizes do projeto

As principais linhas de pesquisa do projeto ArMa-GAPP são descritas abaixo apresentando suas proposições, o que já foi feito, o estado atual e os projetos futuros de cada linha.

É importante ressaltar que, apesar de cada uma das linhas de pesquisa ter seus objetivos próprios, existe um grande interação entre as áreas, ocorrendo um forte compartilhamento dos resultados e produtos. Exemplos desta interação são o aproveitamento das ferramentas criadas para suporte, no desenvolvimento de novas aplicações, e o aproveitamento dos resultados de desempenho das aplicações geradas, na avaliação de desempenho do circuito GAPP.

3.1 Aplicabilidade do circuito GAPP

Estuda as características estruturais e funcionais do circuito GAPP [Kitajima, 1990] e analisa o seu desempenho associado a diversas arquiteturas.

Os resultados obtidos permitem que se tenha uma orientação em relação as áreas em que a aplicação do circuito GAPP seria viável e eficiente. Ou seja, é possível prever em que classes de algoritmos o circuito matricial traria bons resultados quando aplicado [Carissimi, 1989].

Os algoritmos matriciais desenvolvidos podem, com as informações detalhadas do circuito GAPP, serem adaptados a estas características, obtendo-se assim uma racionalização de processamento e consequentemente uma melhora de desempenho.

3.2 Suporte ao desenvolvimento de aplicações

Procura racionalizar o desenvolvimento de aplicações paralelas matriciais para a placa NCR que se utiliza de circuitos GAPP. São estudadas as fases necessárias para o desenvolvimento de uma aplicação para este ambiente e confeccionadas ferramentas para automatizar este processo [Carissimi, 1987].

Dentro da idéia de facilitar a comunicação do hospedeiro com a placa NCR, foi desenvolvida na linguagem "C", uma biblioteca com rotinas de entrada e saída (E/S) [De Rose et al, 1992].

Atualmente se estuda uma forma de facilitar a ativação da placa a partir do hospedeiro, e de auxiliar na geração do código necessário para esta operação. Está em fase de desenvolvimento um filtro para converter código da linguagem GAL para a linguagem "C".

As ferramentas e bibliotecas geradas são colocadas à disposição para a linha de desenvolvimento de aplicações. Todos os documentos e ferramentas geradas são colocados à disposição dos alunos do curso de mestrado, que na disciplina de Arquiteturas Especiais de Computadores para Processamento Paralelo têm a oportunidade de experimentar a programação de um ambiente matricial.

3.3 Desenvolvimento de aplicações

Investe no desenvolvimento de aplicações para a placa NCR com circuitos GAPP, aproveitando-se as informações sobre as características do circuito GAPP, das ferramentas e na documentação de suporte.

A análise do desempenho das aplicações geradas é utilizada na avaliação do desempenho do circuito GAPP. As dificuldades encontradas no desenvolvimento são repassadas para o suporte.

A principal área de desenvolvimento até agora foi o processamento gráfico. Algoritmos matriciais para convolução de imagens foram estudados e implementados no ambiente GAPP.

Como projeto futuro se encontram o desenvolvimento de operações gráficas mais complexas e algoritmos associativos.

4 Processamento de Imagens

No setor de desenvolvimento de aplicações, investiu-se inicialmente na área de processamento de imagens, por possuir algoritmos com características de baixa dependência de dados e operações homogêneas em todos os pontos da imagem. Algoritmos desta natureza, como por exemplo o de convolução de imagens, têm seu mapeamento facilitado neste tipo de arquitetura.

A distribuição dos dados a serem processados é feita de forma igual entre os elementos processadores da matriz, recebendo cada um uma parte da imagem ou até mesmo um único ponto da imagem.

O algoritmo de convolução [Mascarenhas e Velasco, 1988][Paz e Cunha, 1988][Dawson, 1987] é facilmente implementado na arquitetura GAPP. Cada EP recebe um ponto a ser processado, obtendo dos elementos vizinhos os valores dos pontos relevantes para o cálculo da convolução. A figura 3 mostra como se processa a convolução na visão de um processador. Pelo sincronismo de execução das instruções, o fluxo de comunicação mostrado se processa simultaneamente em todos os EPs.

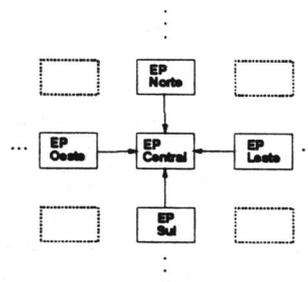


Figure 3: Implementação da convolução no GAPP

Foram implementados diversos algoritmos de convolução de imagens com diferentes características. Uma análise detalhada do comportamento do ambiente matricial GAPP em relação a esta classe de algoritmos pode ser encontrada em [Cavaleiro et al, 1992]. Com as várias implementações fez-se um estudo do desempenho da arquitetura GAPP.

Os resultados de performance obtidos até agora não foram satisfatórios, obteve-se muitas vezes um desempenho pior do algoritmo paralelo em relação ao seqüencial. O principal problema encontrado foi a fraca taxa de E/S da placa NCR aliada a baixa complexidade dos algoritmos implementados e ao número reduzido de EPs disponíveis na placa NCR.

Atualmente procuram-se formas de diminuir os efeitos do gargalo que se constituiu a E/S da placa. Também estão sendo pesquisadas e implementadas operações de processamento de imagens de maior complexidade.

5 Estado Atual/Perspectivas Futuras

As medições efetuadas com a placa GAPP mostraram que o poder de processamento do circuito GAPP é bastante alto. A principal razão deste desempenho é sem dúvida o alto paralelismo de dados aliado a uma velocidade de 10Mhz.

Entretanto, o desempenho global medido nas operações de convolução de imagens com o uso da placa não foi satisfatório, uma vez que este foi inferior ao obtido de forma seqüencial no hospedeiro. Os resultados mostraram uma queda no desempenho de em média 50% em relação ao processamento seqüencial dos filtros implementados. Somente a partir de um aumento na carga de trabalho dos algoritmos de aproximadamente 8 vezes é que o algoritmo paralelo superou o seqüencial.

A razão deste fraco desempenho com os filtros implementados se deve a forma que a placa efetua a troca de dados com o hospedeiro. A operação de entrada e saída na placa GAPP é feita de forma seqüencial e bloqueante, interrompendo tanto hospedeiro como placa. Como se não bastasse, esta operação ainda é muito lenta, se comparada com a velocidade de processamento do circuito GAPP. A operação de carga (ou retirada) de dados na placa tem um custo tão grande que só o tempo de comunicação com a placa nos filtros implementados já supera o algoritmo seqüencial. Com a operação de E/S como gargalo do sistema, a placa só consegue um bom desempenho quando o trabalho a ser efetuado pelos EPs compensa o alto custo de E/S.

Outra razão para o fraco desempenho obtido é o baixo grau de paralelismo de dados encontrado na placa. Com uma matriz de 12×12 elementos processadores apenas 0.04% da imagem (de dimensão 640×400) pode ser processada a cada iteração do algoritmo. Com um número maior de elementos processadores o paralelismo de dados seria maior e a placa forneceria resultados melhores, mesmo para algoritmos com menor carga computacional.

Os resultados encontrados confirmam as conclusões de [Wong and Lua, 1990] em relação ao desempenho do circuito GAPP. Porém sua análise não levou em consideração o problema de entrada e saída de dados, que mostrou-se um ponto fraco da placa NCR. Desta forma, nas operações de convolução implementadas não foi possível obter resultados tão animadores quanto os de [Wong and Lua, 1990]. A conexão da placa GAPP com o hospedeiro compromete o desempenho do conjunto, uma vez que a operação de E/S, realizada na forma serial, constitui um gargalo para muitas aplicações, em que a quantidade de processamento não justifica a carga de dados.

Atualmente trabalha-se na implementação de algoritmos de maior complexidade buscando um melhor desempenho da placa NCR. Estuda-se também a possibilidade do aumento do paralelismo através da confecção de uma placa com um maior número de elementos processadores.

Com as rotinas de E/S entre a placa e o hospedeiro, as ferramentas de auxílio no desenvolvimento de aplicações, e toda documentação gerada pelos estudos realizados, estão caindo gradativamente as principais barreiras que impediam uma melhor utilização dos recursos da placa matricial GAPP.

6 Referências

- A. S. CARÍSSIMI, **Implementação de Arquiteturas SIMD**. Porto Alegre, CPGCC da UFRGS, 1989. (Dissertação de Mestrado)
- A. S. CARÍSSIMI, **Montador e Gerador de Código Objeto para GAPP** Porto Alegre, CPGCC da UFRGS, 1987. (Relatório Interno)
- G. G. H. CAVALHEIRO, C. DE ROSE e R. MENNA BARRETO, **A Operação de Convolução de uma Imagem em uma Arquitetura Matricial**. 1992. (Trabalho submetido ao IV SBAC-PAD)
- B. M. DAWSON, Introduction to Image Processing Algorithms. *Byte*, v. 12, n. 3, pp. 169-186, Mar. 1987.
- C. DE ROSE, G. G. H. CAVALHEIRO, R. MENNA BARRETO, **Rotinas de Comunicação para a Placa GAPP**. Porto Alegre: CPGCC da UFRGS, 1992. (Relatório de Pesquisa)
- M. J. FLYNN, Some computer organizations and their effectiveness. *IEEE Transactions on Computers*, New York, v. C-21, n. 9. pp. 948-160, Sept. 1972.
- J. W. KITAJIMA, E. HENS e P. NAVAU, **GAPP - Um Processador Matricial**. Porto Alegre, CPGCC da UFRGS, 1990. (Relatório Interno)
- K. HWANG, F. BRIGGS, **Computer Architecture and Parallel Processing**. New York: McGraw Hill. 1984.
- N. MASCARENHAS, F. VELASCO, **Processamento Digital de Imagens**. Escola Brasileiro-Argentina de Informática, 4., 1989, Termas de Rio Hondo. **Proceedings...** Buenos Aires: Kapelusz, 1988.
- NCR Microelectronics. **GAPP Personal Computer Development System User's Manual**. Fort Collins:[s.n.] 1986. 99p.
- E. P. PAZ, T. N. CUNHA, **Iniciação ao Processamento Digital de Imagens**. Congresso da Sociedade Brasileira de Computação, 8., 1988, Rio de Janeiro. **Proceedings...** Rio de Janeiro: SBC, 1988. 55 p.
- W. F. WONG, K. T. LUA, **A Preliminary Evaluation of a Massively Parallel Processor: GAPP**. *Microprocessing and Microprogramming*, North-Holland, v.29, n.1, p. 53-61, July 1990.