AVALIAÇÃO DE DESEMPENHO DE UM SISTEMA MULTIPROCESSADOR

Ricardo Zelenovsky
Instituto Militar de Engenharia
Seção de Eletricidade - SE/3
Praça General Tibúrcio, 80 - CEP 22290
Rio de Janeiro - RJ

RESUMO

É proposto o modelamento de um sistema multiprocessador. É feita análise de seu desempenho com relação a decomposição e alocação de subproblemas aos processadores, tamanho do grão e eficiência do "cache". São estabelecidos critérios para adequar o sistema aos algorítmos de forma a maximizar o desempenho.

ABSTRACT

The proposition is the modeling of a multiple-processor system. An analysis of its performance is done respectively with the decomposition and allocation of subproblems to processors, grain size and hit-ratio of cache. Criteria are established in order to tune up the system with the algorithms so as to rise up the performance to its maximum.

1. INTRODUÇÃO

A avaliação de desempenho de sistemas multiprocessadores não é uma tarefa simples, pois além da grande quantidade de parâmetros a serem analisados, o inter-relacionamento entre eles é complexo e difícil de ser quantizado.

Segundo Cvetanovic [1] os parâmetros de influência mais significativa na performance de um sistema multiprocessador são:

- quantidade de paralelismo intrínseco do problema,
- método utilizado na decomposição do problema em subproblemas,
- método utilizado para alocar os subproblemas aos processadores,
- tamanho do grão,
- método de acesso aos dados,
- estrutura de interconexão (acesso às memórias),
- velocidades de processadores, memória e rede de interconexão.

Esse trabalho visa o desenvolvimento de vários modelos simplificados para um sistema multiprocessador, avaliando-os segundo os parâmetros listados, a fim de determinar suas interações.

2. MINISUPERCOMPUTADOR MS8701

Nos laboratórios da Escola Politécnica da Universidade de São Paulo, realiza-se, sob orientação do professor João Antonio Zuffo, um trabalho sobre arquitetura paralela, pioneiro no Brasil. Está em desenvolvimento o MINISUPERCOMPUTADOR MS8701 cuja unidade básica de processamento paralelo é constituida de 4 processadores 68020 da Motorola reunidos em uma "Placa de Processamento Geral", PPG. O sistema completo constitui-se de 16 dessas PPGs ligadas via 2 barramentos VME. A figura 1 descreve o esquema de uma PPG onde pode ser notado o processador 68020, a memória local e os 2 barramentos "VME".

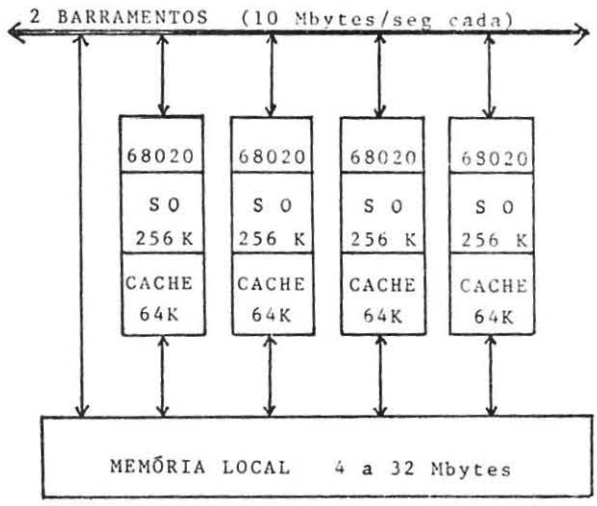

Figura 1. Placa de Processamento Geral (PPG).

A figura 2 mostra a ligação de até 16 dessas PPG via barramentos. A memória local de cada PPG pode ser acessada pelos processadores de outra PPG, sendo nesse caso chamada de memória externa. Existem pois, 3 níveis de hierarquia de memória: memória cache, memória local, memória externa.

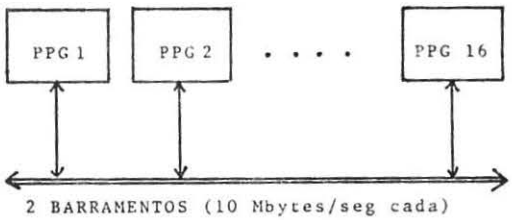

Figura 2. Esquema do MS8701 com as 16 PPG.

O esquema de arbitração dentro de uma PPG é conhecido como "round robin", onde a hierarquia de prioridade circula entre os elementos concorrentes (a cada instante um processador é o mais prioritário); 64 processadores disputam acessos aos barramentos sendo o esquema de arbritação novamente o "round robin".

3. MODELO E SUPOSIÇÕES

A carga de processamento e de comunicação resultante da decomposição do algorítmo, é igualmente distribuída entre os processadores("comunicação" aqui designa acesso à memória). O processamento é caracterizado por uma "unidade de processamento", up e a comunicação pela "unidade de comunicação", uc. A seguir são designadas as principais variáveis.

Tp → quantidade de "up" de um dado algorítmo,
tp → tempo gasto para executar uma "up",
Tc → quantidade de "uc" de um dado algorítmo,
tc → tempo gasto para executar um "uc",
N → número de processadores,
BW → banda passante da memória,
HR → taxa de acertos do "cache" (hit ratio).

Análogo à definição dada por Vrsalovic [2], definem-se: "Função Decomposição de Processamento", Dp, como a razão entre o tempo de processamento para um sistema uniprocessador e o tempo de processamento para cada processador em um sistema multiprocessador e - "Função Decomposição de Comunicação", Dc, como a razão entre o tempo de acesso a dados em um sistema uniprocessador e o tempo de acesso a dados para cada processador em um sistema multiprocessador. Os parâmetros tc, tp, N, BW dependem das características do sistema, já Dc, Dp, Tc, Tp dizem respeito ao algorítmo. A adequabilidade de um algorítmo ao processamento paralelo é representada por suas funções Dp e Dc,calculadas para cada algorítmo. Por exemplo, a transformada rápida de Fourrier tem Dc = O(N/LogN).

Como o interesse do presente trabalho é o estudo dos esquemas de interconexão entre processadores e memória, utiliza-se Dp = N para evitar que a decomposição do processamento não mascare os resultados.

"Ganho de Velocidade", S (speed up) é dado pela relação entre o tempo gasto em um sistema uniprocessador e o tempo gasto em um sistema multiprocessador. O sistema uniprocessador básico é constituido por um processador e uma memória idênticos às da PPG sem "cache".

4. AVALIAÇÃO DO DESEMPENHO DO MS8701

Segue-se uma série de casos onde o desempenho do MS8701 é analisado em função da carga de trabalho.

4.1. Caso 1. PPG Isolada

Este é o caso mais simples onde toda carga de trabalho é distribuida dentro de uma única PPG. Nesse esquema N processadores (O < N < 5) disputam a memória cuja arbitração é feita segundo o "round robin".

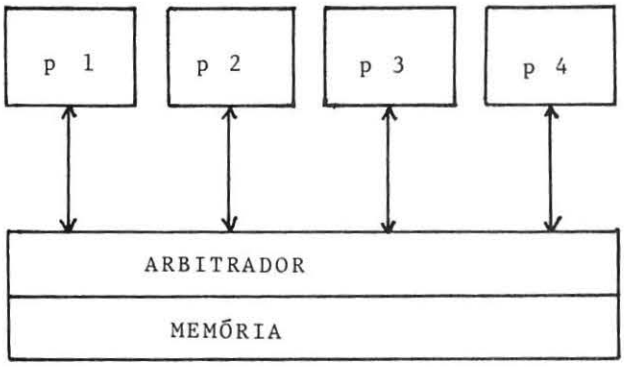

Figura 3. PPG isolada (caso 1).

O tempo para execução serial é dado pela soma dos tempos gastos em processamento e comunicação, (Tp tp + Tc tc); já o tempo para execução paralela é dado por:

$$\frac{Tp}{Dp}\,tp + \frac{Tc}{Dc}\,tc\,\frac{N}{BW}$$

O "cache" é caracterizado pelo seu "hit-ratio", sendo este dependente da estrutura do algorítmo e do tamanho do cache. O tempo de ciclo de acesso ao cache (tcache) é de 180 nseg e como o tempo do ciclo de acesso a memória local é de 480 nseg, resulta a relação xl=tcache/tc=0.37. Os Tc acessos realizados podem ser divididos em (HR Tc) acessos no cache e ((1-HR) Tc) acessos fora do cache. Assim o tempo gasto para execução paralela é dado por:

$$\frac{Tp}{Dp}\,tp + (1-HR)\frac{Tc}{Dc}\,tc\frac{N}{BW} + HR\frac{Tc}{Dc}\,tcache$$

Logo o ganho de velocidade é dado por:

$$S = \frac{1 + K}{\frac{K}{Dp} + \frac{(1-HR)}{Dc}\,\frac{N}{BW} + \frac{HR}{Dc}\,xl} \qquad (1)$$

$K = \frac{Tp\ tp}{Tc\ tc}$ → relação entre processamento e comunicação.

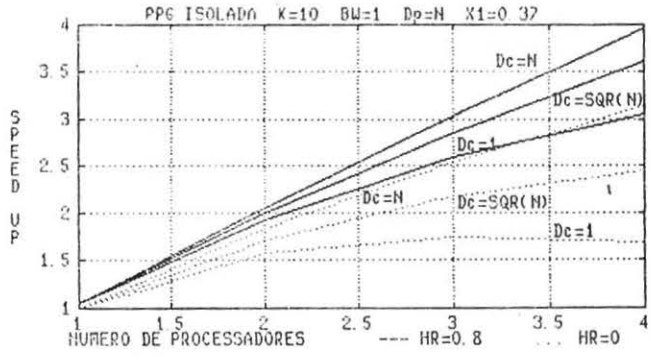

Figura 4. Caso 1.

A figura 4 mostra o ganho de velocidade para diversos valores de Dc e HR, pode-se notar a grande dependência em relação à função decomposição de comunicação e a melhora trazida pelo "cache"; valores de K menores do que 1 implicam em mais comunicação que processamento, gerando grande quantidade de conflitos; para valores elevados de K, o ganho de velocidade aproxima-se de Dp. Para analisar a melhoria trazida pelo "cache", utiliza-se um sistema, hipotético sem "cache", cuja banda passante, BW', é aumentada de forma a atingir o mesmo ganho de velocidade do sistema com "cache". Cria-se assim um sistema multiprocessador hipotético onde a variação da banda passante é utilizada para capturar o efeito do "cache" (ou também para capturar as alterações trazidas por uma nova arquitetura, como será visto adiante).

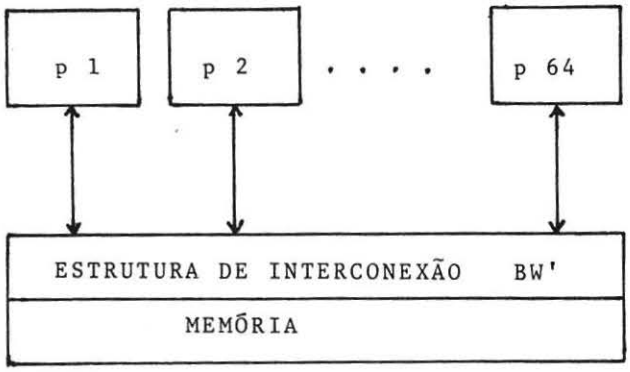

Figura 5: PPG hipotética com até 64 processadores.

O ganho de velocidade hipotético é dado por:

$$S' = \frac{1 + K}{\frac{K}{Dp} + \frac{1}{Dc}\,\frac{N}{BW'}} \qquad (3)$$

Igualando-se as equações (2) e (3) obtêm-se BW' em função de HR, N e BW. Na figura 6 está a banda hipotética, onde pode-se notar a grande sensibilidade de BW' para valores elevados de HR.

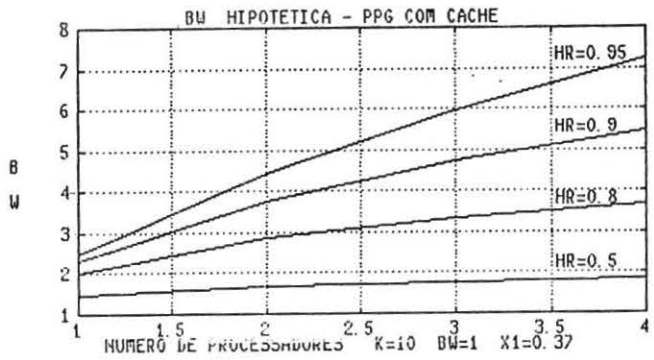

Figura 6. BW'.

$$BW' = \frac{N\ BW}{(1-HR)N + HR\ BW\ x1} \qquad (4)$$

4.2. Caso 2. PPG Mais N-4 Processadores

Neste caso pode-se utilizar até 64 processadores, sendo os 4 primeiros alocados dentro de uma PPG e todos os demais chegam via barramento. Todos os dados estão localizados na memória local da PPG.

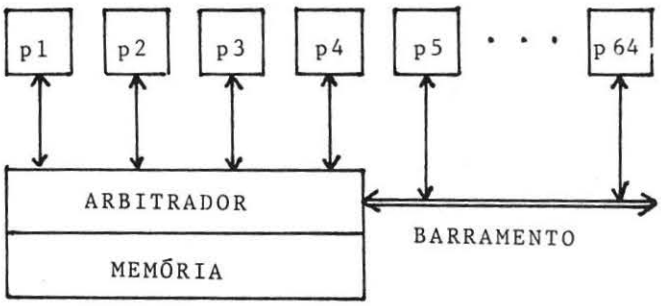

Figura 7. PPG mais N-4 processadores.

Como a arbitração dentro da PPG é feita segundo o "round-robin", o processador externo só tem acesso à memória após um ciclo de 4 acessos. A figura 8 mostra a análise de um caso onde os 7 processadores, 4 da PPG e 3 externos, realizam 4 acessos cada um (X simboliza acesso: é utilizado Tc=28, Dc=7 e N=7).

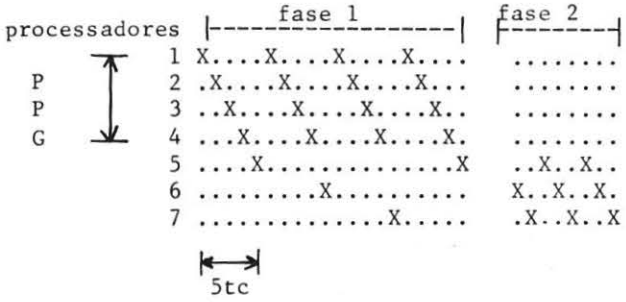

Figura 8. Esquema de acessos a memória para o caso 2.

Pela figura acima pode-se ver que a sequência de acessos à memória 12345123461234712345675675 67. Como os processadores internos da PPG encerram seu processamento antes dos externos, existem 2 fases de processamento: uma com os processadores da PPG presentes e outra sem os processadores da PPG.

O tempo gasto na fase 1 é (5 Tc tc)/(Dc BW). Na fase 2 os processadores da PPG estão inativos, restando apenas os processadores externos, porém (Tc/Dc)/(N-4) acessos externos já foram atendidos durante a primeira fase restando:

$$\frac{Tc}{Dc} - \frac{Tc/Dc}{N-4} = \frac{Tc}{Dc}\,\frac{N-5}{N-4}$$

Como cada acesso leva (N-4)(tc+tarb), o tempo gasto em comunicação na segunda fase é dado por:

$$\frac{Tc}{Dc}\,\frac{N-5}{N-4}\,(N-4)\frac{tc+tarb}{BW} = \frac{Tc}{Dc}\,(N-5)\,\frac{x3\ tc}{BW}$$

onde x3 = (tc+tarb)/tc = (480+480)/480 = 2

tarb = tempo de arbitração do barramento

O tempo total gasto no processamento paralelo é a soma do tempo gasto na fase 1 e na fase 2. Assim o ganho de velocidade é dado pela equação (2) para N < 5 e pela equação (7)para N > 4.

$$S = \frac{1 + K}{\frac{K}{Dp} + \frac{(1-HR)}{Dc}\,\frac{5}{BW} + \alpha + \frac{2HR\ x1}{Dc}} \qquad (7)$$

onde $\alpha = \frac{(1-HR)\ x3}{Dc}\,\frac{N-5}{BW}$

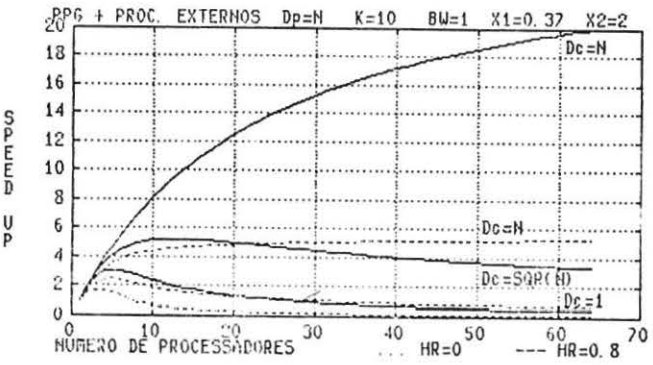

Figura 9. Caso 2.

A figura 9 mostra esse ganho de velocidade onde nota-se que um grande número de processadores só deve ser utilizado com valores altos de Dc e HR. Esta variação na performance pode ser traduzida em termos de ganho de banda passante para o sistema hipotético, que é calculada através das equações (7) e (3), resultando na equação (4) para N < 5 e na equação (8) para N > 4. ELa é mostrado na figura 10, onde pode ser novamente sentida a grande sensibilidade para valores elevados de HR. Os picos iniciais do gráfico são devidos à alocação inicial dentro da PPG.

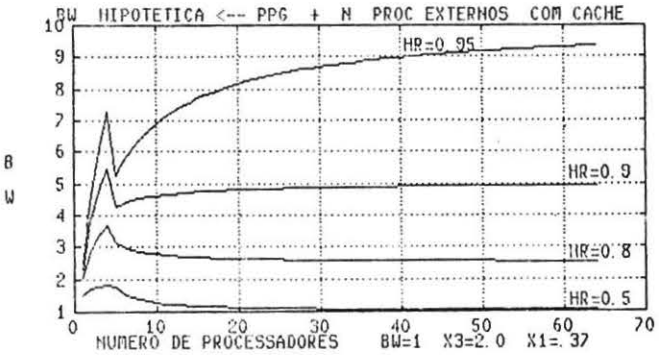

Figura 10. BW'.

$$BW' = \frac{N\ BW}{(1-HR)5+(1-HR)x3(N-5)+2\ HR\ BW\ x1} \qquad (8)$$

4.3. Caso 3. Todas PPGs

Existem agora até 16 PPGs (até 64 processadores), o total de acessos a memória (Tc) é dividido em acessos dentro da PPG (Tci) e acessos a memória de outra PPG (Tce), a letra p simboliza a percentagem de acessos externos.

$$Tc = Tci + Tce, \text{ com } p = \frac{Tce}{Tc} \text{ tem-se } (1-p) = \frac{Tci}{Tc}$$

A figura 11 ilustra esse caso onde o processador hipotético 5 representa o acesso a memória dessa PPG através do barramento,com tempo de acesso tcbus (usa-se X4=tcbus/tc=2).

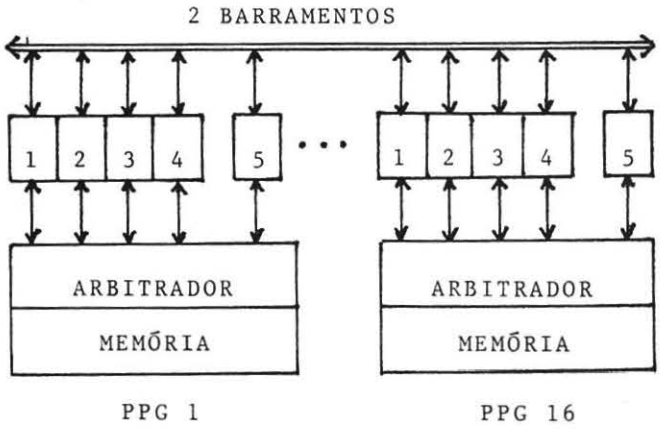

Figura 11. Todas PPG ligadas pelo barramento.

O tempo total para processamento paralelo é:

$$\frac{Tp}{Dp}\,tp + (1-HR)\frac{Tci}{Dc}\,tc\frac{5}{BW} + HR\frac{Tci}{Dc}\,tcache +$$

$$+(1-HR)\frac{Tce}{Dc}tcbus\frac{N}{BWb} + HR\frac{Tce}{Dc}\,tcache$$

onde BWb representa a banda passante do barramento, no presente caso BWb=2, pois existem 2 barramentos e considera-se desprezível a ocorrência de processadores diferentes disputando a mesma memória de uma mesma PPG. O ganho de velocidade é dado pela equação (2) para N < 5 e pela equação (9) para N > 4.

$$S = \frac{1 + K}{\frac{K}{Dp} + (1-HR)\frac{(1-p)}{Dc}\,\frac{5}{BW} + \alpha + \frac{HR}{Dc}\ x1} \qquad (9)$$

onde $\alpha = (1-HR)\,\frac{p}{Dc}\,x4\,\frac{N}{BWb}$

Na figura 12 nota-se que uma alta performance só é conseguida para valores elevados de Dc e HR, sendo nos demais casos aconselhável a utilização de um número pequeno de processadores.

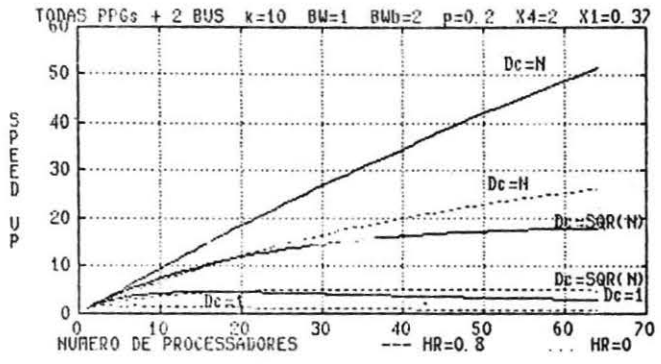

Figura 12. Caso 3

A conversão em banda passante da PPG hipotética, dessas alterações de velocidade, é feita pelas equações (3) e (9), resultando na equação (4) para N < 5 e na equação (10) para N > 4.

$$BW' = \frac{N\ BW\ BWb}{(1-HR)(1-p)5\ BWb+\alpha+HR\ x1\ BW\ BWb} \quad (10)$$

onde α = (1-HR) p x4 N BW

4.4. Decomposição do Processamento

Algumas vezes o processamento não pode ser totalmente decomposto entre os N processadores, pois parte dele precisa ser executado serialmente por um único processador. Neste caso divide-se a quantidade de processamento em uma porção paralela (Tpp) e uma porção serial(Tps).

$$Tp = Tpp + Tps, \text{ se } q = \frac{Tpp}{Tp} \text{ então}$$

$$1-q = \frac{Tps}{Tp}$$

Esse efeito pode ser facilmente traduzido na função decomposição de processamento como é mostrado na equação (11). A figura 13 mostra o caso 3 com uma porção serial onde pode ser notado seu custo elevado.

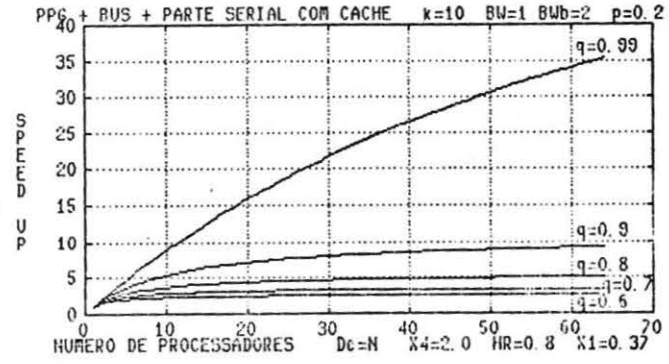

Figura 13. Caso 3 (serial).

$$Dp = \frac{Tp}{\frac{Tpp}{N} + Tps} = \frac{1}{\frac{q}{N} + (1-q)} \quad (11)$$

5. CONCLUSÕES

No presente trabalho são fornecidos meios para determinar qual a melhor arquitetura do MS8701 frente a uma certa carga de trabalho, caracterizada pelo Tp, Tc, Dp e Dc. Existe ainda dificuldade para estimativa do "hit-ratio" e dos acessos a memória externa. As diferenças de desempenho das arquiteturas são convertidas em banda passante BW' do sistema hipotético, permitindo a quantização da comparação.

De uma forma geral, uma grande eficiência é conseguida com até 4 processadores dentro de uma PPG. O aumento do número de processadores só traz bons resultados se a eficiência do"cache" e a decomposição de comunicação forem elevadas. Os algorítmos com trechos seriais apresentam custo elevados.

A ausência de resultados práticos deve-se ao fato do MS8701 estar ainda em construção.Espera-se sua disponibilidade em futuro próximo.

REFERÊNCIAS

[1] CVETANOVIC, Zarka, "The effects of problem partitioning, allocation, and granularity on the performance of multiple - processor systems", IEEE Transactions on Computers, vol c-36, número 4, abril de 1987,pag 421-432.

[2] VRSLOVIC, D. e GEHRINGER, E.F. e SEGALL,Z. Z. e SIEWIOREK, D.P., "The influence of parallel decomposition strategies on the performance of multiprocessor systems", Proc. 12th Ann. Int. Comput. Architect., Boston, MA, IEEE Comput. Soc. and ACM,June 1985, pag 396-405.