Parâmetros Arquiteturais Críticos em Clusters de SMPs

Edward David Moreno[1]
[1] Faculdade de Informática da Fundação Eurípides Soares da Rocha -
Av. Hygino Muzzi Filho 529, CEP 17525-901, Marília, S.P.
{edmoreno@fundanet.br}

Resumo

Este trabalho explora o impacto de mudar a velocidade das redes de interconexão e barramento interno em cada nó SMP, no desempenho e impacto de agrupar vários processadores/nó em clusters de SMPs. Estuda-se também o impacto de considerar caches especiais para acessos remotos e, a inter-relação que há entre esse cache e a velocidade das redes de interconexão, junto ao número de processadores por nó. Na avaliação de desempenho usou-se simulação comandada por programa, sendo necessário implementar um simulador chamado por nós de SIM-SMP, o qual é estimulado com programas do benchmarks SPLASH-2. Foram usadas como métricas de performance o NET: Tempo de Execução Normalizado, URE: Utilização da Rede e o UBA: Utilização do barramento. Os resultados obtidos permitem concluir que ainda com rápidas redes de interconexão, é possível obter benefícios na clusterização, isto é, agrupar vários processadores por nó. Além disso, os benefícios de caches remotos ainda são mantidos pois eles sempre conseguirão diminuir a utilização da rede, apesar de aumentarem a sobrecarga no barramento interno.

Abstract

We explore variations in the interconnection network and bus speed, as modern and future CC-NUMAs architectures, which are based in SMPs clusters. We also study the effects of the second level cache size. Using as metrics: Normalized execution time, cache hit rate, usage of interconnection network and bus, and execution driven simulation with six programs from SPLASH-2, we show remote caches will improve the overall performance in SMP-based CC-NUMAs, when advances in interconnection network, bus, second level cache and processor's speed occur.

1. Introdução

Motivados pelos contínuos e rápidos avanços na microeletrônica, na construção de circuitos integrados, nos sistemas digitais e na arquitetura de sistemas computacionais, hoje existem arquiteturas cc-NUMA de alto desempenho com 2 ou 4 processadores por nó, cada processador executando em até 2 GHz. Inclusive, em um futuro muito próximo, é possível a existência comercial e de baixo custo de *chips* multiprocessadores com 2 ou 4 processadores/nó [1, 2, 3, 10, 11].

Além disso, sabe-se que a velocidade dos processadores aumenta 60-80 % a cada ano e que a velocidade das memórias também cresce, porém não no ritmo dos primeiros, pois tal melhoramento é de 5-8 % por ano [8, 9 17]. Essa grande diferença produz então um *gap* entre essas tecnologias, que com certeza continuará crescendo. Nesta situação, é possível que os caches remotos (netcaches) ofereçam um bom desempenho, tornando-se indispensáveis nas futuras arquiteturas baseadas em nós de SMPs. Portanto, neste trabalho estuda-se o efeito dos caches remotos, como uma maneira de diminuir o problema dos acessos remotos, produzidos por esse gap entre as tecnologias mencionadas acima.

A figura 1 mostra uma arquitetura muito usada, a qual é composta de vários nós SMPs interconectados por uma rede de interconexão de alto desempenho [12]. Nessa figura destacam-se alguns parâmetros considerados como críticos na avaliação do desempenho das arquiteturas multiprocessadas baseadas em SMPs. Consideram-se como sendo importantes a velocidade dos processadores, a velocidade da rede geral de interconexão (que liga os diferentes nós), a velocidade do barramento interno em cada nó e o tamanho do cache L2. Outro parâmetro também importante é o cache remoto, útil para armazenar dados pertencentes a outros nós SMPs.

Essa arquitetura tem sido grandemente estudada, principalmente analisando-se o seu desempenho (*speedup* e escalabilidade) para diferentes programas (SPLASH-2, TPC, NAS e etc.) [4, 5, 6, 7, 8, 9, 13]. O impacto de vários parâmetros também tem sido objeto de estudo e pesquisas, por exemplo: os caches remotos, os caches locais (primeiro e segundo nível, L1 e L2), número de processadores por nó, impacto e desempenho de novos processadores, protocolos de consistência de memória, análise de diferentes interfaces de rede e etc.

Não obstante essa quantidade de pesquisas, o presente artigo motivou-se com a seguinte questão: O que pode acontecer quando a velocidade das redes aumentar, pois

sendo assim os acessos remotos serão rapidamente resolvidos. Nessa situação, será que ainda compensa usar várias processadores por nó ? Será que ainda é bom inserir caches remotos ? O fato de inserir caches remotos, não sobrecarrega demais o barramento interno, ao ponto de diminuir os benefícios de rápidas redes de interconexão ?

Dessa maneira o artigo pretende analisar algumas situações para responder qualitativamente essas questões. Através de simulações (com o SIM-SMP, ferramenta projetada pelos autores deste trabalho) e usando 06 programas do *benchmark* SPLASH-2 [16], obteve-se a seguinte conclusão: O fato de agrupar vários processadores em um único nó, ainda será opção válida de projeto, pois o aumento do gap de velocidades entre processadores e memória será ainda mantido e possivelmente até crescerá. Dessa maneira, o possível uso de rápidas redes de interconexão não inibirá essa tendência. Além disso, o uso de caches remotos, também será de grande utilidade, apesar da sobrecarga produzidas ao barramento interno de cada nó SMP.

O artigo está organizado em 05 seções. A seção 02 focaliza a arquitetura em estudo e o ambiente de avaliação de desempenho, enfatizando nos parâmetros críticos dessa arquitetura. Já a seção 03 mostra resultados e analisa o fenômeno da clusterização, isto é, considerar vários processadores em um único nó. Isso é estudado, mudando também a velocidade da rede de interconexão. A seção 04 mostra também o impacto de se inserir caches remotos. A seção 05 tenta mostrar o impacto de se ter barramentos mais rápidos e a sua inter-relação com caches remotos e redes rápidas. Finalmente a seção 06 apresenta algumas conclusões.

2. Arquitetura SMP e Ambiente De Simulação

Nesta seção apresenta-se, brevemente, a arquitetura em estudo (ver fig. 1), os seus principais parâmetros e valores usados e a descrição do simulador SIM-SMP (fig. 2).

ARQUITETURA: Cada aglomerado pode conter um ou vários processadores (de 1 até 4 processadores) com os seus próprios caches locais (tanto de primeiro nível e/ou segundo nível: L1/L2), memória principal local, unidades de entrada e saída e a interface de acesso à rede de interconexão. Todos estes elementos são conectados através de um barramento compartilhado. Vários destes aglomerados são interconectados através de uma rede de interconexão de alta velocidade (baixa latência e alta largura de banda). Em termos de coerência de cache, estes sistemas podem usar um protocolo *snoopy* de 4-estados interno ao nó (para trabalhar com o barramento interno do nó) e um protocolo baseado nos esquemas de diretório (para coerência externa) [9, 14].

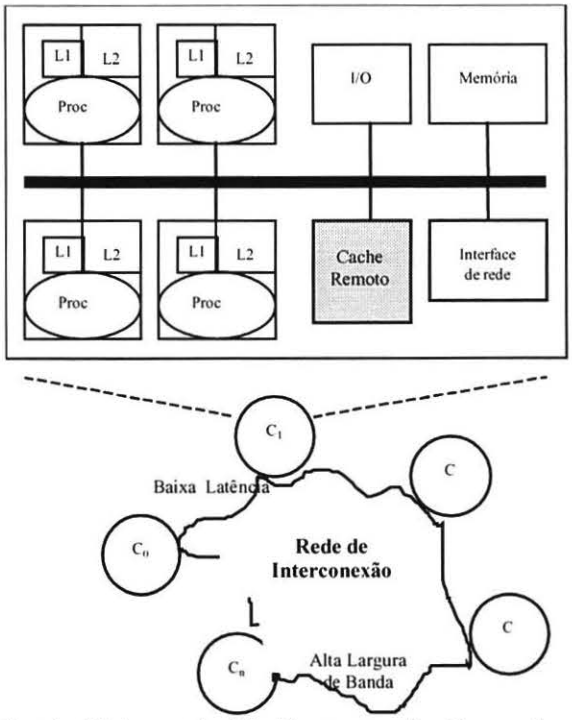

Fig. 1. *Sistema de Alto Desempenho Baseado em Clusters de SMPs.*

TABELA 1. *PRINCIPAIS PARÂMETROS E VALORES USADOS NA ARQUITETURA SMP*

Elemento	Valor
No. CPUs	1-32, cada um a 2 GHz. Cada nó SMP pode ter 1, 2 ou 4 processadores.
Bus	400 Mbytes/s, latência de arbitramento: 20 ns.
Interconexão	2.4 Gbit/segundo. Tamanho de 8 bytes.
Cache L1	I-cache=D-cache= 128 Kbytes, Bloco de 32 bytes, Mapeamento *Direto*
Cache L2	I-cache=D-cache= 1 Mbyte, Bloco de 64 bytes, *Mapeamento Direto*
Memória	*Read time= Write time=* 80 ns, *Tag time* = 45 ns, Bloco de 64 bytes.
Cache Remoto	Valores dependem do seu tamanho e organização. Esses valores são calculados usando o software CACTI [15]. Por exemplo, 4 Mbytes, Bloco de 128 bytes, *4-way*: Tempo de acesso = 30.96 ns, *Tag* = 24.98 ns.
Sucesso no cache L1	Depende do software CACTI, em função dos tamanhos dos caches.
Miss em L2	*NUMA-time*, depende do software CACTI e da arquitetura do cache L2 e Memória
Sem conteção	Um processador pode acessar um bloco de cache em 200 ns. Se o dado estiver localizado no mesmo cluster ou 500 ns se o dado estiver em outro cluster.

O SIMULADOR SIM-SMP: Para analisar a arquitetura, foi construído um simulador, SIM_SMP, em linguagem C++, que associa cada objeto a cada componente da arquitetura. O SIM_SMP, além do funcionamento, considera também os tempos de atraso e de execução própria de cada bloco.

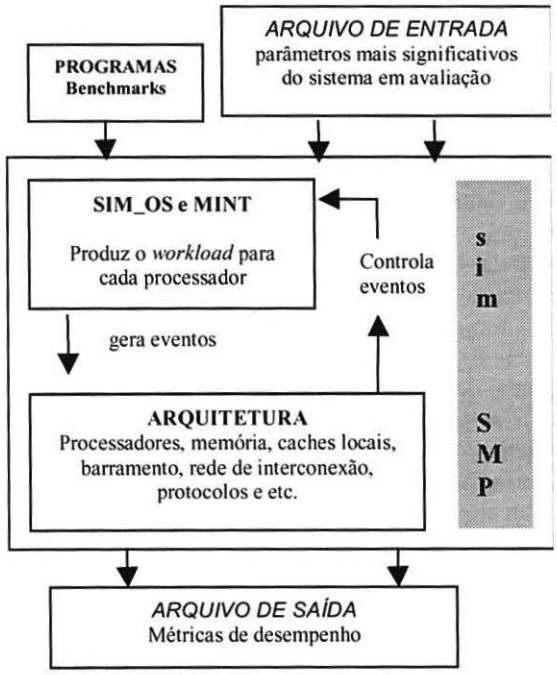

Fig 2. *O Simulador SIM-SMP*

O simulador SIM-SMP usa o MINT (versão 2.7) para gerar os eventos de interesse para o módulo da arquitetura [9]. O SIM_SMP foi projetado seguindo a metodologia de orientação a objetos, onde cada módulo em *hardware* ou em *software* é representado por uma classe do mesmo nome. Os principais módulos são os processadores, os caches locais, a memória, a rede de interconexão, o barramento, os caches remotos e o protocolo de coerência de cache. Para estimular o simulador, necessita-se de programas aplicativos e de um arquivo de entrada que define a organização do sistema total e, também, os valores principais dos parâmetros de entrada em cada módulo da arquitetura. O estímulo foi realizado com 06 programas do benchmark SPLASH-2 [16]: Barnes, Water, Ocean, FFT, Radix e LU.

3. EFEITO DA REDE DE INTERCONEXÃO E DA CLUSTERIZAÇÃO

Importante lembrar que os processadores aumentam a performance em 60-80% a cada ano e que as memórias reduzem o seu tempo de acesso em 5-8%/ano.

Não obstante não existe uma regra experimental em relação aos melhoramentos da tecnologia do barramento e das redes de interconexão. Por esse motivo, supõem-se diferentes opções nestes parâmetros, especificamente avanços da ordem de 50%, 100% nas velocidades da rede de interconexão e avanços de 50%, 75% e 100% na velocidade do barramento interno em cada nó. Toma-se mais um ponto de avaliação para o barramento, uma vez que existem menos avanços e menos expectativas de melhoramentos do que o caso das redes. Estas percentagens são tomadas em relação aos valores básicos indicados na tabela 1.

Considerando seis aplicações do SPLASH-2 [16], a tabela 2 mostra os efeitos de melhoramentos da rede de interconexão na clusterização. Isto é, qual é o efeito de se ter uma rede rápida de interconexão no número de processadores por nó.

Tabela 2 *Efeito da "clusterização" (sem caches remotos). Comparação 4P/nó e 2P/nó*

	Barnes 4P	Water 4P	Ocean 2P	FFT 2P	Radix 2P	LU 4P
Base	2	7	1	0.7	1	2
PR1	2	2.6	1	0.6	0.3	2
PR2	1	0.5	0.2	0.1	0.3	2.1

Nessa tabela aparece o impacto medido através do ganho no tempo de execução, o qual foi normalizado considerando como base o tempo de execução do mesmo sistema, porém com um único processador por nó, isto é, sem clusterização. A linha indicada como BASE refere-se aos parâmetros descritos na tabela 1. Já as linhas PR1 e PR2, relacionam-se com aumentos de 50 e 100% na velocidade da rede de interconexão, usando-se como referência a contenção indicada na tabela 1.

Em cada coluna aparece o programa usado e a indicação 4P ou 2P. Isso significa que a arquitetura sempre obteve melhores resultados, para esse programa, para 4 ou 2 processadores por nó. Importante salientar que o sistema foi testado sem usar caches remotos.

Os dados dessa tabela permitem dizer que agrupar vários processadores em um único melhora a performance. Em nosso caso, a arquitetura base teve benefícios máximo de 7 %. Quando se aumenta a velocidade da rede de interconexão, também se obtém benefícios, em nosso caso, máximo de 2.6 e 2.1 % para avanços de 50 e 100 % na contenção da rede de interconexão.

Esse resultado é significativo, pois quer dizer que ainda com redes rápidas de interconexão, serão obtidos benefícios na clusterização. Esse resultado é ainda mais significativo, pois o nosso estudo supõe que as redes melhoram as suas características. Na verdade, quando isso acontecer, também deve aumentar a velocidade dos processadores e poucas

PÁGINA FALTANDO

Esse trabalho foi disponibilizado online como parte do esforço de publicação de todo o histórico do evento, entretanto, a versão do trabalho recuperada não continha essa página.

PÁGINA FALTANDO

Esse trabalho foi disponibilizado online como parte do esforço de publicação de todo o histórico do evento, entretanto, a versão do trabalho recuperada não continha essa página.

da rede), indiretamente o monitoramento dos caches dos outros nós também é realizado mais rapidamente.

Já programas com maior comunicação entre os diferentes processadores do mesmo nó, mais que a comunicação inter-nós, possuem uma alta contenção do barramento já que este não acompanhou a velocidade da rede geral. Isto significa que, novamente, o barramento interno de cada nó SMP é o gargalo do sistema pois há um aumento significativo na sua contenção.

5. EFEITO DA TECNOLOGIA DE BARAMENTO

Nesta seção apresentam-se e analisam-se os resultados obtidos do impacto dos melhoramentos da tecnologia do barramento na eficiência de caches remotos, tanto em multiprocessadores modernos quanto em sistemas com rápidas redes. Os resultados são mostrados na tabela 7.

Tabela 7 *Efeito do Barramento na Eficiência dos Caches Remotos*

	Barnes	Water	Ocean	FFT	Radix	LU
REDE ^{Max}	47	28	55	50	27	32
REDE ^{Min}	23	11	34	20	18	24
BASE ^{Max}	51	31	61	49	34	40
BASE ^{Min}	25	14	35	19	19	26

Foram realizadas simulações supondo que o barramento interno melhora a sua velocidade em 50, 75 e 100 %. Nesses casos se manteve o cache remoto em 4 Mbytes, associatividade 4. Nesses casos se manteve a velocidade da rede de interconexão (arquitetura base) e se mudou também a velocidade da rede de interconexão, isto é, as mudanças da rede foram consideradas também.

Nessa tabela, observa-se uma descrição REDE (MAX e MIN), que significa que os dados foram obtidos supondo redes rápidas 100% (MAX) e 50%, com relação à base (que usa os dados da tabela 1).

Interessante perceber, por exemplo, na arquitetura base, que tendo melhoramentos do barramento, consegue-se sempre um impacto positivo na utilização dos caches remotos. Essa observação se mantém ainda quando as redes também aumentam a sua velocidade, porém com valores menores.

Interessante notar o impacto que há no tempo de execução. Esses resultados são mostrados na tabela 8. Ali se supõem um melhoramento do barramento em 50, 75 e 100%, indicados, respectivamente como B1, B2 e B3. Além disso, se considera o nó SMP com caches remotos e sem eles, para 2 e 4 processadores por nó. Lembrar que o sistema em estudo possui 32 processadores.

Tabela 8 *Efeito do Barramento em Clusters de SMPs, sobre o tempo de execução*

Cache Remoto	B1: 50%	B2: 75%	B3: 100%
SIM ^{RR}	$(4, 25)^4$	$(0, 8)^2$	$(0, 3)^2$
SIM ^{Base}	$(4, 14)^4$	$(3, 9)^2$	$(1, 4)^2$
NÃO ^{RR}	$(0, 21)^2$	$(1, 9)^2$	$(1, 3)^2$
NÃO ^{Base}	$(0, 6)^2$	$(1, 8)^2$	$(1, 5)^2$

Os resultados mostrados nas tabelas 7 e 8, permitem fazer as seguintes observações:

(i) Neste caso, como era de se esperar, constata-se maiores desempenhos na utilização dos caches remotos: entre 14% e 61% para os clusters modernos.

(ii) Em sistemas com rápidas redes, o ganho de desempenho é menor do que o apresentado nos sistemas modernos (ver o conjunto de tabelas 7). As diferenças encontradas oscilam entre 14% e 26% nos sistemas modernos e entre 8% e 24% para sistemas com rápidas redes, exceto para o programa FFT onde a diferença foi de 30% uma vez que nesta aplicação os maiores impactos devem-se mais aos melhoramentos na rede de interconexão do que no barramento. Essa diminuição no desempenho e eficiência dos caches remotos ocorreu por se ter barramentos mais rápidos e, consequentemente, tal recurso oferece uma menor sobrecarga na utilização desse elemento.

(iii) As diferenças no nível de "clusterização" novamente continuam dependendo das características da aplicação. Os resultados da tabela 8 permitem visualizar o intervalo de diferença (mínimas e máximas)^p entre 2P/nó e 4P/nó com um melhor desempenho para o sistema com "p" processadores/nó. Na tabela, "SIM" indica que o sistema possui caches remoto e "NÃO" representa um sistema de vários processadores/nó sem esses caches. Além disso, se indica como RR se houver redes rápidas de interconexão sendo usadas em conjunto e, Base se os valores da tabela 1 são mantidos na arquitetura.

Assim, quando os caches remotos são levados em conta e com um melhoramento máximo de 50% na velocidade do barramento, é claramente visível que uma boa opção é considerar 4 processadores/nó. Se acontecem avanços que produzem melhoramentos superiores a 50%, é melhor agrupar 2P/nó. Sem caches remotos, é melhor agrupar 2 processadores/nó independentemente dos avanços na tecnologia do barramento.

Esse resultado explica-se pelo fato das altas sobrecargas geradas nos sistemas com vários processadores/nó pois existe um maior número de processadores competindo por esse recurso. Isto é, há uma sobrecarga maior no nó quando os caches remotos são inseridos, porque produz uma maior contenção no barramento.

Em sistemas com redes e barramentos rápidos, onde possivelmente os barramentos serão muito mais rápidos, a

melhor opção é agrupar 2 processadores/nó, seja com caches remotos ou sem eles. Este resultado é surpreendente, pois acreditava-se que com tais melhoramentos a opção viável deveria ser maior do que dois. Acontece todavia que a análise está considerando também o aumento e melhoramentos na velocidade dos processadores e na memória. Portanto, pode-se dizer que o barramento, como esperado, continuará sendo o maior ponto de engarrafamento desses sistemas. Ainda assim, pode-se afirmar que caches remotos serão de grande ajuda nesse impasse tecnológico entre os diferentes recursos do sistema.

O fato 2 processadores/nó (2P/nó) oferecerem maior desempenho, tanto em clusters com caches remotos ou sem estes, deve-se fundamentalmente aos melhoramentos esperados na rede de interconexão. Os nossos resultados permitem afirmar que a eficiência de caches remotos é altamente dependente tanto da rede quanto do barramento e muito mais suscetível aos melhoramentos na rede de interconexão geral do que nos do barramento interno a cada nó SMP. Ainda assim, é necessário investir em avanços da tecnologia do barramento, pois a sua sobrecarga é significativa e pode diminuir os ganhos do uso de caches remotos.

6. CONCLUSÕES

Neste trabalho apresentaram-se os efeitos da variação de alguns parâmetros, considerados como críticos nas arquiteturas baseadas em clusters e SMPs, sempre visando recursos hardware mais rápidos possibilitados pelos contínuos avanços tecnológicos. Com base nos resultados obtidos via simulação comandada por programa e usando programas do *benchmark* SPLASH-2, encontrou-se que caches remotos continuam e continuarão sendo um elemento que alivia e aliviará a grande e contínua diferença entre as velocidades e desempenho dos processadores e memórias. Assim, fizeram-se avaliações quando mudanças nas velocidades das redes de interconexão inter-nós e intra-nós aconteçam e observou-se que ainda com essas redes mais rápidas, os caches remotos serão de grande utilidade na diminuição do problema conhecido como "The Memory Wall" [17].

7. REFERÊNCIAS BIBLIOGRÁFICAS

[1] BARROSO, Luiz André et Al. Piranha: A Scalable Architecture Based on Single-Chip Multiprocessing. In Proc. of the 27th ACM International Symposium on Computer Architecture. (ISCA-2000), June 2000, Vancouver, CA

[2] BARROSO, A.L.; Gharachorloo, K.; Bugnion, E. Memory System Characterization of Commercial Workloads. *In Proceedings of ISCA-99*, June, 1998.

[3] BHANDARKAR, D.; Ding, J. Performance Characterization of the Pentium Pro Processor. *In Proceedings of Intl. Symp. on Computer Architecture (ISCA)*, Jun., 1997.

[4] CAO, Q.; Trancoso, P.; et Al. Detailed Characterization of a Quad Pentium Pro Server Running TPC-D. *In Proceedings of ICCD-99*, US.A., Set., 1999.

[5] CVETANOVIC, Z.; D. Bhandarkar, "Performance Characterization of the Alpha 21164 Microprocessor Using TP and SPEC Workloads," *Proc. Int. Symp. High-Performance Computer Architecture*, February 1996.

[6] DESOTA, D.; Forester, R. Effectiveness of Remote Cache in a NUMA System. *In Proc. of Workshop on Computer Architecture Evaluation using Commercial Workloads*. Feb., 1999.

[7] LENOSKY, Daniel E.; Weber, Wolf-Dietrich. Scalable Shared-Memory Multiprocessing. *Morgan Kaufmann Publishers*, San Francisco - California, 1995.

[8] LOVETT, Tom.; Clapp, Russell. STiNG: A CC-NUMA Computer System for the Commercial Marketplace. *In Proceedings of the 23rd Annual Intl. Symp. on Computer Architecture*, p. 308-317, May, 1996.

[9] MORENO, E.; Netcaches on Engineering and Commercial Applications. In Book: *High Performance Computign Systems and Applications*. Kluwer Publishers, Dec. 2000.

[10] NAYFEH, Basem A.; Olukotun, Kunle; Singh, Jaswinder Pal. The Impact of Shared-Cache Clustering in Small-Scale Shared-Memory Multiprocessors. *In Proc. of HPCA-2 (High Performance Computer Architecture*, p. 74-84, 1996.

[11] PATTERSON, D. et Al. A Case for Intelligent RAM: IRAM. *In IEEE Micro,* April, 1997.

[12] PFISTER, Gregory F. In Search of Clusters. *Prentice Hall PTR, 2 Edition, New Jersey*, 1998.

[13] TRANCOSO, P.; Larriba-Pey, J.L.; Zhang, Z.; Torrellas, J. The Memory Performance of DSS Commercial Workloads in Shared-Memory Multiprocessors. *In Proc. of HPCA-97*, Feb. 1997.

[14] VRANESIC, Z. et Al. The NUMAchine Multiprocessor: Performance and Experiences. In Proc. of Intl. Congress on Computing Parallel, Tokyo, Japan ICCP-2001.

[15] WILTON, Steven J.E.; Jouppi, Norman P. CACTI: An Enhanced Cache Access and Cycle Time Model. *In IEEE Journal of Soolid-State Circuits*, Vol. 31, No.5, p. 677-688, May, 1996.

[16] WOO, S.C.; Ohara, M.; Torrie, E.; Singh,J.P.; Gupta, A. The SPLASH-2 Programs: Characterization and Methodological Considerations. *In Proc. of the 22nd Annual Intl. Symp. on Computer Architecture*, p. 24-36, June, 1995.

[17] WULF, Wm.A. and McKee, S.A. Hitting the Memory Wall: Implications of the Obvious. *ACM Computer Architecture News, Vol. 23, No.1*, March, 1995.