「スーパーテクニカルサーバ HITACHI SR8000 LINPACK 性能のご紹介」

（株）日立製作所

1．はじめに

科学技術計算用計算機の性能を評価するベンチマークテストの１つに LINPACK があります。LINPACK 性能は、計算機システムの実行演算性能を示す１つの指標になっていることに加え、Tennessee 大学の Dongarra 教授等によって、WWW を通して性能の順位が世界中に紹介されていることから、各ハードウェアーベンダーは、LINPACK プログラムの最適化を実施し、積極的に性能を公表しています。

前回の報告（1998 年 1 月）では、SR2201 向け LINPACK の最適化方式についてご紹介しました。今回は、次元数を任意に選択できる LINPACK の Highly Parallel Computing に対するスーパーテクニカルサーバＨＩＴＡＣＨＩ SR8000 向き最適化方式及び性能測定結果についてご紹介いたします。最適化方式につきましては、SR2201 向け最適化と共通する部分がありますので、ここでは SR2201 向け最適化と方式が異なる部分についてご紹介させていただきます。

2．LINPACK 性能の推移

Tennessee 大学の Dongarra 教授等がまとめている「Performance of Various Computers Using Standard Linear Equations Software」（**http://www.netlib.org/benchmark/performance.ps**）に登録されている、1996 年 4 月 29 日、1999 年 4 月 19 日時点での Highly Parallel Computing に対する上位 10 機種を、それぞれ表 2.1、表 2.2 に示します。

表 2.1　1996 年 4 月 29 日時点での Highly Parallel Computing に対する上位 10 機種

順位	機種名	ノード数	理論性能 GFLOPS	Rmax 性能 GFLOPS	実行効率 %	次元数
1	Intel Paragon XP/S MP	6,768	338	281.1	83.2	128,600
2	Intel Paragon XP/S MP	6,144	307	256.2	83.5	122,500
3	Intel Paragon XP/S MP	5,376	269	223.6	83.1	114,500
4	HITACHI SR2201/1024	1,024	307	220.4	71.8	138,240
5	Intel Paragon XP/S MP	4,608	230	191.5	83.3	106,000
6	Numerical Wind Tunnel	140	236	170.4	72.2	42,000
7	Numerical Wind Tunnel	128	216	157.9	73.1	40,960
8	Intel Paragon XP/S MP	3,648	182	151.7	83.4	95,000
9	Fujitsu VPP500/128	128	205	149.7	73.0	40,960
10	Intel Paragon XPS-140	3,680	184	143.4	77.9	55,700

(注)　実行効率　＝　(Rmax 性能)／(理論性能)

表 2.2　1999 年 4 月 19 日時点での Highly Parallel Computing に対する上位 10 機種

順位	機種名	ノード数	理論性能 GFLOPS	Rmax 性能 GFLOPS	実行効率 %	次元数
1	ASCI Red	9,472	3,154	2,121.3	67.3	251,904
2	SGI ASCI Blue	5,040	2,520	1,608.0	63.8	374,400
3	ASCI Red	6,912	2,302	1,533.6	66.6	207,360
4	Intel ASCI Option Red	9,152	1,830	1,338.0	73.1	235,000
5	CRAY T3E-1200	1,488	1,786	1,127.0	63.1	148,800
6	Intel ASCI Option Red	7,264	1,453	1,068.0	73.5	215,000
7	CRAY T3E-1200E	1,080	1,296	891.5	68.8	259,200
8	HITACHI SR8000/128	128	1,024	873.6	85.3	120,000
9	CRAY T3E-900	1,320	1,188	815.1	68.6	134,400
10	SGI Origin 2000	2,048	1,024	690.9	67.5	229,248

(注)　実行効率　＝　(Rmax 性能)／(理論性能)

1996 年 4 月から 1999 年 4 月までの 3 年間で、第 1 位の LINPACK 演算性能が約 7.5 倍に向上し、上位 6 機種が 1TFLOPS を超えるというように、計算機の性能が飛躍的に向上していることが分かります。しかしその一方で、理論性能に対する Rmax 性能の比を示す実行効率の低下現象があります。1996 年 4 月の上位 10 機種すべてが実行効率 70%を超えていたのに対し、1999 年 4 月の時点では 7 機種が実行効率 70%を超えていません。

一般に、並列計算機の性能向上は、接続するノード数を増加する方式を採っていますが、通信処理が隠蔽でき並列化性能の向上が比較的容易な LINPACK でさえ実行効率の劣化が見られることから、ノードの数を増加する方式にも限界があると考えられます。

表 2.2 を見ると、HITACHI SR8000/128 のみが実行効率 80%を超える性能を実現しています。これは、単体のノード性能を 8GFLOPS にまで高めることによって、128 ノードという比較的小規模な並列化で理論性能 1TFLOPS を実現していることに起因しています。

3．最適化方式

3.1　SR2201 向け最適化方式との相違点

SR8000 向け最適化と SR2201 向け最適化の相違点を表 3.1 に示します。

表 3.1　SR8000 向け最適化と SR2201 向け最適化の相違点

最適化項目	SR8000	SR2201
LU 分解手法	外積形式 GAUSS 法	ブロック形式 GAUSS 法
同時消去段数	8 段 4 列	5 段 2 列
データー分割方式	ブロック形式サイクリック列分割	ブロック形式 Scattered Square 分割
軸列の列方向転送	１対１通信	マルチキャスト

3.2　LU 分解手法

SR2201 では、プロセッサーから SC（Storage controller）にデーターを送出する線と SC からプロセッサーにデーターを送出する線の 2 つの単方向データー・アドレス共通線があります。アドレスとデーターをプロセッサーから SC に送出するストアー処理の場合、アドレス、データー共プロセッサーから SC（Storage controller）にデーターを送出する線を使用し、 SC からプロセッサーにデーターを送る線は使用しない状態になります。一方ロード処理の場合、アドレスはプロセッサーから SC、データーは SC からプロセッサーにデーターを送出しますので両方の線を使用します。したがって、SR2201 ではストアー処理よりもロード処理の方が性能的に優位になります。以上のことから、SR2201 向け最適化では、主消去演算をロード処理のみで構成する内積演算にすることを目的に、LU 分解処理にブロック形式 GAUSS 法を採用していました。

SR8000 では、データー用とアドレス用の線を物理的に分けていますので、ロード処理に対するストアー処理の性能差がありません。したがって、SR8000 向け最適化では、ループ長の長大化が可能な外積形式 GAUSS 法を採用しています。

外積形式 GAUSS 法による LU 分解処理を図 3-1 に示します。

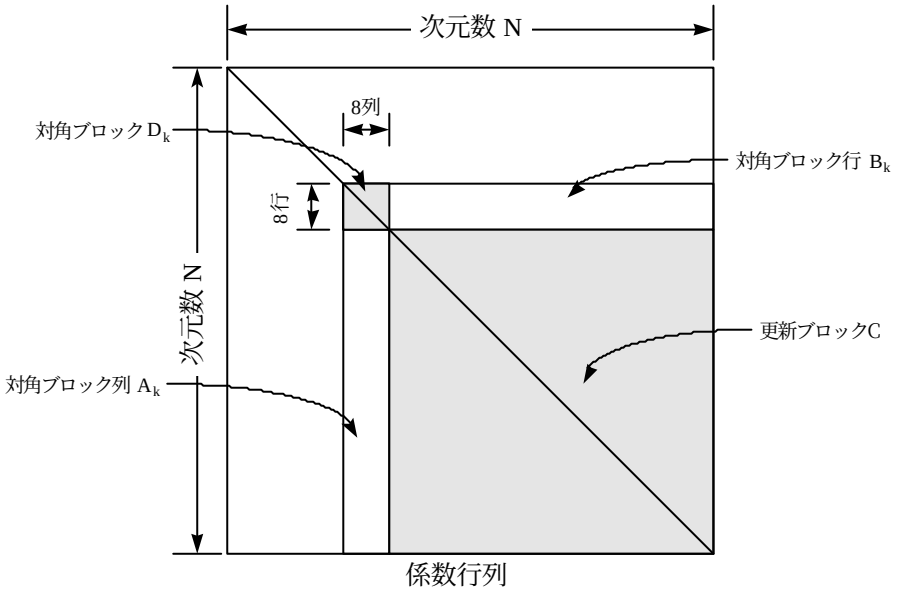

図 3-1　外積形式 GAUSS 法による LU 分解処理（第 k ブロック段消去）

3.3　同時消去段数

SR2201 向け最適化では、物理的に 128 個存在する倍精度浮動小数点レジスターのうち、同時に参照できる個数が 28 個であるため、同時消去段数 5 段 2 列が展開の限界になっていました。しかし SR8000 では、物理的に 160 個存在する倍精度浮動小数点レジスターのうち同時に参照できる浮動小数点レジスターを 128 個に拡張しているため、SR2201 に比べて同時消去段数の増加が可能になります。SR8000 向き最適化で採用している 8 段 4 列同時消去の処理を図 3-2 に示します。

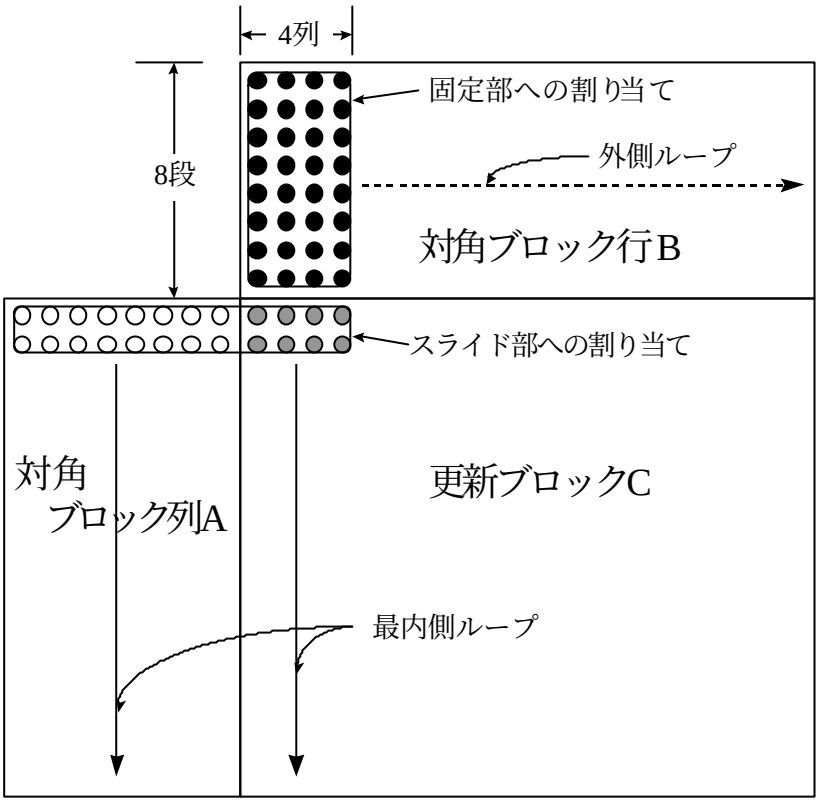

図 3-2　8 段 4 列同時消去の処理

8 段 4 列同時消去の場合、図 3-2 の対角ブロック行 B で参照する最内側ループ内固定の浮動小数点データーは 32 個です。最内側ループ内固定の浮動小数点データーは、固定部の浮動小数点レジスターに割り当てますので、8 段 4 列同時消去の場合、固定部の浮動小数点レジスターを 32 個使用します。SR8000 の 1 プロセッサー当たりの固定部の浮動小数点レジスター数は 32 個ですので、8 段 4 列同時消去が最大の展開数になります。

3.4　データー分割方式

SR2201 では、60 行 60 列のブロックを行方向、列方向共サイクリックにノードに割り当てるブロック形式の Scattered Square 分割を採用していました。これは、SR2201 が数千ノードという大規模構成の並列化を対象としているため、通信相手となるノード数を $\sqrt{NPU}$（NPU：ノード台数）に削減することを目的に採用した分割方式です。

SR8000 向き最適化では、ノードの最大構成が 128 であること、LU 分解方式に外積形式の GAUSS 法を採用していることから、1 ブロック 8 列のブロック形式のサイクリック列分割を採用しています。ブロック形式サイクリック列分割によるデーター分割方式を図 3-3 に示します。

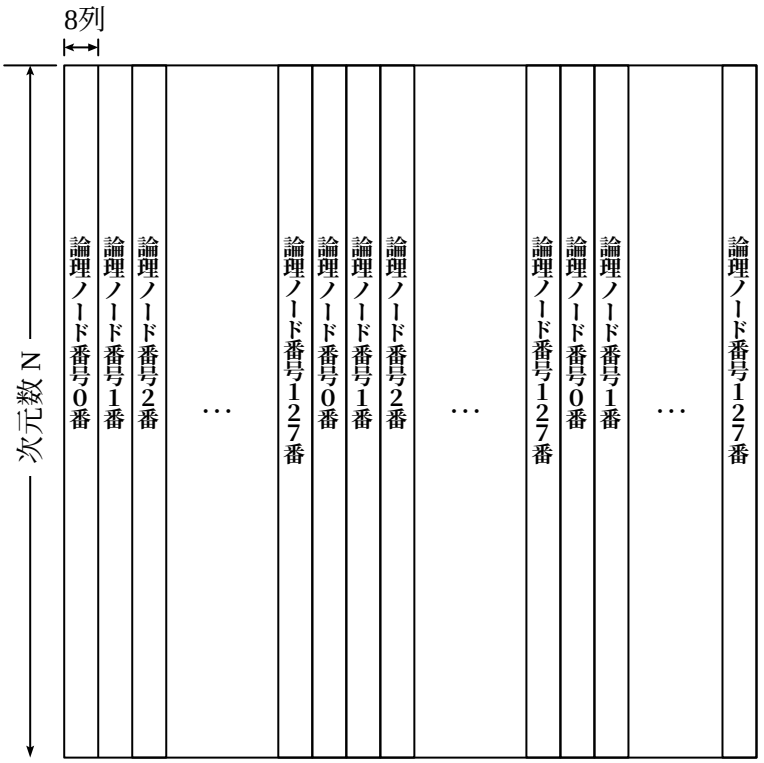

図 3-3　1 ブロック 8 列のブロック形式サイクリック列分割（128 ノード使用時）

ブロック形式サイクリック列分割には、次に示す点で並列化性能の向上に有利になります。

・部分軸選択及び軸ブロック列の更新処理で通信処理が発生しない

・外積形式 GAUSS 法を採用することによってループ長の長大化が可能

3.5　軸列の列方向転送

列方向の通信処理に 2 進木のマルチキャストまたはブロードキャストを使用すると、列方向通信の通信起動回数は $\log_2 NPU$ に削減できます。しかし、マルチキャストまたはブロードキャストは同期型通信ですので、ノード間の処理時間のブレが通信時間に含まれる結果となります。

LU 分解処理の並列化では、部分軸選択を含む軸ブロック列の更新処理と、更新結果を他のノードに転送する通信処理を主消去演算処理に隠蔽できます。したがって、通信時間が各消去ブロック段の主消去時間を超えない範囲であれば 1 対 1 通信で逐次にデーター転送する方がノード間の処理時間のブレの影響を受けない分性能的に優位になります。SR8000 の 20 ノードで比較した 1 対 1 通信とマルチキャストでの Rmax 性能の差を表 3.2 に示します。

表 3.2　SR8000 の 20 ノードでの 1 対 1 通信とマルチキャストの Rmax 性能比較

①1 対 1 通信	②マルチキャスト	①／②
144.5 GFLOPS	139.8 GFLOPS	1.034

表 3.2 に示すように、対角ブロック列の転送に 1 対 1 通信を適用した場合、同期型通信のマルチキャストを適用した場合に比べて 3.4%の性能向上になります。

図 3-4 に部分軸選択処理及び対角ブロック列の主消去への隠蔽方式を示します。

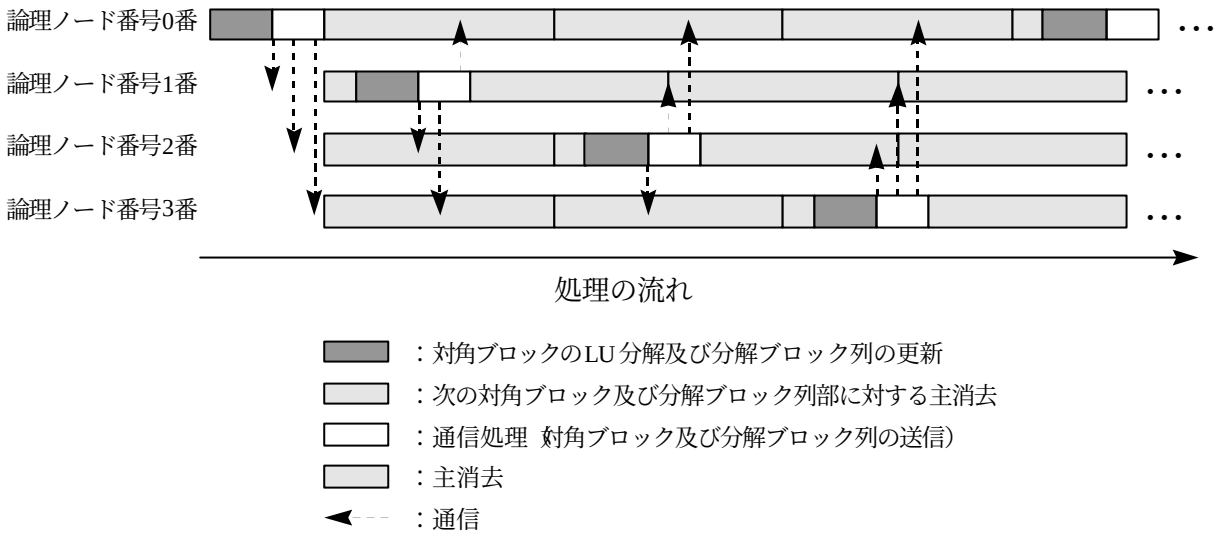

図 3-4　部分軸選択処理及び対角ブロック列の主消去への隠蔽方式（4 ノード使用時）

4．性能測定結果

4.1　Highly Parallel Computing 演算性能

SR8000 の Highly Parallel Computing (Rmax)の性能測定結果を表 4.1 に示します。

表 4.1 SR8000 の Highly Parallel Computing (Rmax)性能

ノード数	①理論性能 GFLOPS	②Rmax 性能 GFLOPS	実行効率 (②/①)%	並列化効率 %	次元数
1	8.0	7.50	93.8	100.0	10,728
2	16.0	14.6	91.3	97.3	15,176
4	32.0	29.1	90.9	97.0	21,464
8	64.0	58.3	91.1	97.2	74,880
16	128.0	115.9	90.5	96.6	42,928
32	256.0	229.5	89.6	95.6	65,000
64	512.0	449.7	87.8	93.7	92,000
128	1,024.0	873.6	85.3	91.0	120,000

また、Rmax 性能に対する S-3800 と SR2201 との比較を図 4-1 に示します。

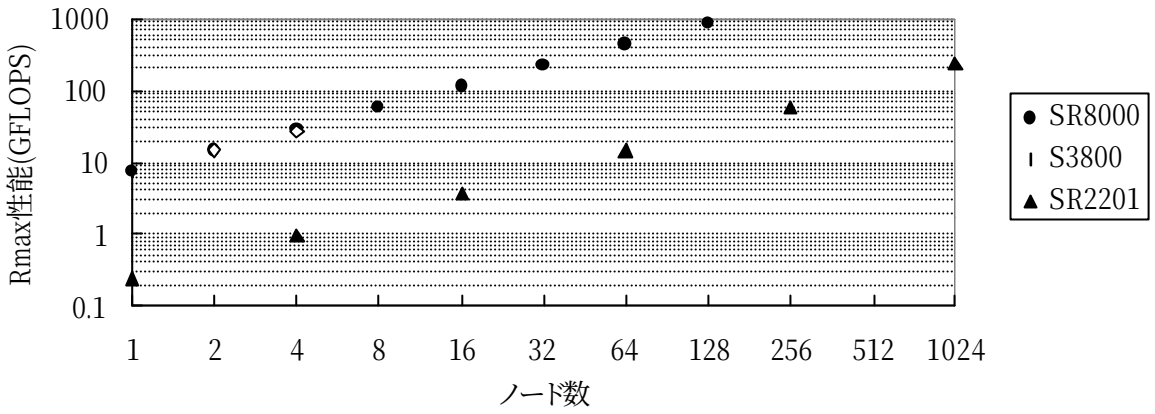

図 4-1　SR8000、S-3800、SR2201 の Rmax 性能比較

4.2　Nhalf 次元数の比較

Dongarra 教授等がまとめている LINPACK 性能には、Rmax 性能の他に、Nhalf 次元数を紹介しています。Nhalf 次元数とは、Rmax 性能の 50%の演算性能を実現する次元数を示すものであり、並列計算機の演算性能と通信性能のバランスを示す指標になっています。

表 4.2 に 1999 年 4 月 19 日に発表になった Rmax 性能上位 10 機種の Nhalf 次元数の比較を示します。

表 4.2　1999 年 4 月 19 日の Rmax 性能上位 10 機種の Nhalf 次元数の比較

順位	機種名	ノード数	Rmax 性能 GFLOPS	①Rmax 次元数	②Nhalf 次元数	②／① %
1	ASCI Red	9,472	2,121.3	251,904	66,000	26.2
2	SGI ASCI Blue	5,040	1,608.0	374,400	138,000	36.9
3	ASCI Red	6,912	1,533.6	207,360	41,700	20.1
4	Intel ASCI Option Red	9,152	1,338.0	235,000	63,000	26.8
5	CRAY T3E-1200	1,488	1,127.0	148,800	28,272	19.0
6	Intel ASCI Option Red	7,264	1,068.0	215,000	53,400	24.8
7	CRAY T3E-1200E	1,080	891.5	259,200	26,400	10.2
8	HITACHI SR8000/128	128	873.6	120,000	16,000	13.3
9	CRAY T3E-900	1,320	815.1	134,400	26,880	20.0
10	SGI Origin 2000	2,048	690.9	229,248	80,640	35.2

LINPACK は、次元数の小規模化に伴って演算負荷に対する通信負荷の割合が高くなります。したがって、次元数を小さくしていくと並列化効率が低下し演算性能が低下します。Nhalf 次元数は Rmax 性能の 50%の演算性能を維持できる問題規模を示しており、Rmax 性能を実現した次元数との比率の小さい方が演算性能と通信性能のバランスが良いと言えます。

5．まとめ

以上、LINPACK ベンチマークテストの SR8000 向け最適化方式及び性能測定結果についてご紹介いたしました。今回ご紹介いたしました LINPACK 性能の最新情報は次の WWW サイトから入手できます。

(1) LINPACK ベンチマークテスト
http://www.netlib.org/benchmark/performance.ps
(2) TOP 500
http://www.netlib.org/benchmark/top500.html

6．参考文献

[1] 東京大学大型計算機センター：「センターニュース」、Vol.30 、No.1、45-53、（1998 年 1 月）
[2] 小国力、村田健郎、三好俊郎、ドンガラ,J.J.、長谷川秀彦 ： 「行列計算ソフトウェア WS、スーパーコン、並列計算機」、 丸善、（1991 年）
[3] 村田健郎、小国力、三好俊郎、小柳義夫 ： 「工学における数値シミュレーション」、丸善、(1988 年)
[4] Jack J Dongarra : 「Performance of Various Computers Using Standard Linear Equations Software」、April 29, 1996. （**http://www.netlib.org/benchmark/performance.ps**）
[5] Jack J Dongarra : 「Performance of Various Computers Using Standard Linear Equations Software」、April 19, 1999. （**http://www.netlib.org/benchmark/performance.ps**）

以　上