

産業界の技術動向

スーパーコンピュータ「京」¹ 搭載CPUの開発について

富士通株式会社 プロセッサ開発統括部

清 田 直 宏

1. はじめに

独立行政法人理化学研究所と富士通株式会社が共同開発したスーパーコンピュータ「京」が、LINPACK ベンチマークにより世界のスーパーコンピュータ演算性能をランク付けする TOP500 において、2011 年 6 月に 8.16PFLOPS (FLOPS:1 秒間に処理する浮動小数点演算数) を記録して世界一を奪取し、更に同年 11 月には「京」の名の由来となった 10 ペタを超える 10.51PFLOPS を記録し、二期連続世界一を果たした。

この成果は、全 864 筐体の中に 88,128 個におよぶ CPU を搭載した大規模システム (図 1) の上で、LINPACK ベンチマークのプログラムを走行させ、実行効率 93.2% で、29 時間以上停止することなく動き続け、その演算性能を極限まで絞り出して得られた。これを実現した理由は、CPU「SPARC64TM 2 VIIIfx^[1]」の高性能・低消費電力化設計技術、インターコネクト「Tofu^[2]」の新アーキテクチャ技術、システムボードの水冷技術、筐体の高密度実装技術等を駆使したハードウェアと、オペレーティングシステム、ファイルシステム、コンパイラ等の大規模並列システムを実現するソフトウェアを、それぞれ高いレベルで開発し、かつ、初期の仕様検討時からシステム全体の演算性能を最大限に引き出すことが出来るようにそれらのバランスを最適化し続けたことにある。

本稿ではこれらいずれも重要な各要素技術の中で、私が携わった CPU「SPARC64 VIIIfx」の開発について述べる。



図 1 スーパーコンピュータ「京」

¹「京」は独立行政法人理化学研究所の登録商標です。

² SPARC64 は、米国 SPARC International, Inc. のライセンスを受けて使用している同社の登録商標です。

2. SPARC64 VIIIfx の開発目標

「京」が目標としたシステム性能は 10PFLOPS で、世界最先端の非常に難度の高い数値であった。それと同時に、納入先の制約条件により、システム全体で使用可能な電力や設置面積の上限が定まっていた。このシステム要件から導き出された SPARC64 VIIIfx の開発目標は、前機種比 3 倍以上の高性能化と 2 分の 1 以下の低消費電力化の両方を実現することになり、そこには途方もなく大きな課題が立ちだかっていた。

しかし、これまで脈々と受け継がれてきた技術と経験、ノウハウといったものに、若い技術者の新しい発想が加わり、高いハードルをひとつずつ乗り越えていった。その結果、この目標がただ無謀な盲進ではなく、世界一の CPU を実現するという強い信念と挑戦意欲へと、皆が次第に変わってゆき、誰一人として諦める技術者はいなかった。

こうして確定した SPARC64 VIIIfx の仕様（図 2）と、ダイ写真（図 3）を下記に示す。

項目	諸元
動作周波数	2GHz
コア数	8
プロセステクノロジー	富士通セミコンダクター (FSL) 45nm CMOS
ダイサイズ	22.7mm x 22.6mm
トランジスタ数	約 7 億 6000 万個
ピーク演算性能	128GFLOPS
メモリ帯域	64GB/s（理論ピーク値）
消費電力	58W（プロセス条件 TYP,30℃）

図 2 SPARC64 VIIIfx の仕様

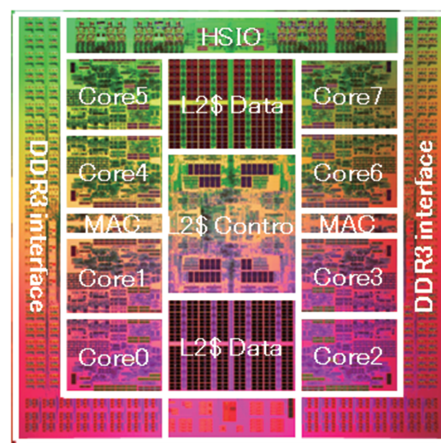


図 3 SPARC64 VIIIfx のダイ写真

3. SPARC64 VIIIfx の特長^{[3] [4]}

この SPARC64 VIIIfx を開発中に常に念頭に置いていたのは、高性能、低消費電力、高信頼性、という三つのキーワードであった。これらの課題解決のためにとった各施策は絶妙なバランスの上で成立した。

3.1. 高性能

SPARC64 VIIIfx の開発においては、理論ピーク性能だけを追い求めた、単なるベンチマークマシンとせず、実際の研究で用いられるアプリケーションでの高い実効性能を実現する汎用スーパーコンピュータとすることを目指した。そのために、SPARC-V9 アーキテクチャの拡張を行い、科学技術計算を効率良く実行可能とする命令セット HPC-ACE (High Performance Computing-Arithmetic Computational Extensions) を新たに開発した。HPC-ACE は、レジスタ数の拡張、SIMD 演算、セクタキャッシュ機構、条件付き実行、三角関数の高速化、除算・平方根近似の機能を有しており、いずれの機能も動作周波数を上げることを要せずに性能向上を可能としている。これにより 2GHz 動作で理論ピーク性能 128GFLOPS となり、SPARC64 VIIIfx の電力あたり性能の向上に大きく寄与している。

またチップ上の 8 つのコアによる並列処理を高速化するために、全てのコアで 2 次キャッシュを共有し、さらにコア間の同期処理をハードウェアで行う機能も備えた。これに富士通の自動並列コンパイラを組み合わせることで、ユーザはプログラミングの際に複数コアであることを特に意識せずに、複数コ

アをあたかも高速なひとつのCPUとして扱うことが可能となる。富士通は、この技術方式をVISIMPACT (Virtual Single Processor by Integrated Multi-core Parallel Architecture) と呼んでおり、SPARC64 VII から継承している。

この高い実効性能を目指した開発を行った結果、LINPACK ベンチマークの記録だけでなく、より総合的な性能を評価するベンチマークである HPC チャレンジにおいても、2011 年には 4 部門すべてで 1 位を獲得し (図 4)、更に、2012 年には TOP500 では 3 位に後退したにも関わらず、4 部門中 3 部門で 1 位を獲得した (図 5)。また、性能ベンチマークでなく実際のアプリケーションでの性能や成果を評価するゴードン・ベル賞も二年連続で「京」を使用したグループが受賞しており、その汎用性と高い実効性能を証明した。

TOP500				
	システム名	設置機関	性能値 (TFLOPS)	実行効率 (%)
1	京	理研 計算科学研究機構	10510	93.17
2	天河1A号	天津スパコンセンタ	2566	54.58
3	Jaguar	オークリッジ研	1759	75.46
Global HPL				
	システム名	設置機関	性能値(TFLOPS)	
1	京	理研 計算科学研究機構	2118	
2	Cray XT5	オークリッジ研	1533	
3	Cray XT5	テネシー大学	736	
Global RandomAccess				
	システム名	設置機関	性能値(GUPS)	
1	京	理研 計算科学研究機構	121	
2	IBM BG/P	ローレンスリバモア研	117	
3	IBM BG/P	アルゴンヌ研	103	
EP STREAM(Triad) Per system				
	システム名	設置機関	性能値(TB/s)	
1	京	理研 計算科学研究機構	812	
2	Cray XT5	オークリッジ研	398	
3	IBM BG/P	ローレンスリバモア研	267	
Global FFT				
	システム名	設置機関	性能値(TFLOPS)	
1	京	理研 計算科学研究機構	34.7	
2	NEC SX-9	海洋研究開発機構	11.9	
3	Cray XT5	オークリッジ研	10.7	

図 4 TOP500 と HPC チャレンジの上位 3 位一覧
(2011 年 11 月)

TOP500				
	システム名	設置機関	性能値 (TFLOPS)	実行効率 (%)
1	Titan	オークリッジ研	17590	64.88
2	Sequoia	ローレンス・リバモア研	16324.8	81.09
3	京	理研 計算科学研究機構	10510	93.17
Global HPL				
	システム名	設置機関	性能値(TFLOPS)	
1	京	理研 計算科学研究機構	9796	
2	Cray XT5	オークリッジ研	1534	
3	IBM Power775	IBM(社内設備)	1344	
Global RandomAccess				
	システム名	設置機関	性能値(GUPS)	
1	IBM Power775	IBM(社内設備)	2021	
2	京	理研 計算科学研究機構	472	
3	IBM BG/P	ローレンスリバモア研	117	
EP STREAM(Triad) Per system				
	システム名	設置機関	性能値(TB/s)	
1	京	理研 計算科学研究機構	3857	
2	IBM Power775	IBM(社内設備)	525	
3	Cray XT5	オークリッジ研	398	
Global FFT				
	システム名	設置機関	性能値(TFLOPS)	
1	京	理研 計算科学研究機構	205.9	
2	IBM Power775	IBM(社内設備)	132.7	
3	NEC SX-9	海洋研究開発機構	11.9	

図 5 TOP500 と HPC チャレンジの上位 3 位一覧
(2012 年 11 月)

3.2. 低消費電力

高性能と低消費電力の両立は非常に困難であったが、それぞれのバランスから動作周波数を 2GHz と設定し、低リークのトランジスタの採用と、冷却方式に水冷を採用してジャンクション温度を通常の 85℃から 30℃まで低下させることで、リーク電力をチップ全体の電力の 10% に抑えた。更に高周波動作する各ラッチへのクロック信号に対する徹底したクロックゲーティングや電力削減に効果的な回路構成・制御方式を採用し、不要回路の動作を可能な限り抑止するといった地道な省電力対策の積み重ねにより、動作時に消費するダイナミック電力も大幅に削減した。

その結果、理論ピーク性能 128GFLOPS という高性能ながら、チップばらつきの平均で 58W という低消費電力を実現し、電力当たりの高い性能も達成した。

3.3. 高信頼性

システムとしての稼働率を高めるためには、8 万個以上搭載されている CPU の信頼性を高めることは、最も重要な事項となるが、SPARC64 VIIIfx には、ミッションクリティカルな社会基盤システムで使用

されるビジネスサーバに搭載する CPU の開発で培った高信頼性技術をそのまま継承している。つまり、CPU 内の大部分の回路にエラー検出機能を備えており、エラー検出時にはエラーが発生した命令をハードウェアが自動的に再実行する命令リトライ機構が働き、また、プロセッサ内で使用している RAM や固定小数点、浮動小数点レジスタの 1 ビットエラー検出時にはハードウェアが自動訂正処理を行っている。これにより宇宙線の衝突などで信号が一時的に変化する間欠エラーが発生しても、システムを停止させずに運用することが可能となっている。更に、これらハードウェアによる自律的な故障回復機能でエラーを修復できない場合であっても、故障箇所を正確に特定し縮退させることにより継続運用を可能としている。また、低消費電力のために採用した水冷方式によりジャンクション温度が 30℃ にまで下がったことは、故障率低減に大きな効果があった。

これら高信頼性機能の効果は数値化しにくいだが、その実力は LINPACK ベンチマークの 29 時間無停止走行が証明している。

4. おわりに

これまで述べたように、SPARC64 VIIIfx の開発は非常にチャレンジングな目標を実現するために数々の困難な課題を克服していく必要があったが、それを乗り越えたときの喜びを原動力に開発を進めていったように思う。

今回取り上げた SPARC64 VIIIfx を搭載しているスーパーコンピュータ「京」の上で行われるシミュレーションや解析により、今後様々な謎や課題が解決されていくことを期待しつつ、我々は次期プロセッサにむけた各要素技術の高性能・低消費電力化を更に進めていきたい。

参考文献

- [1] T. Maruyama et al, "SPARC64 VIIIfx : A New-Generation Octocore Processor For PETASCALE Computing", IEEE Micro, vol.30 Issue2, p.30-40, 2010.
- [2] Y. Ajima, S. Sumimoto, T. Shimizu, "Tofu: a 6D mesh/torus interconnect for exascale computers," Computer, vol.42, no.11, pp.36-40, Nov. 2009.
- [3] "スーパーコンピュータ「京」", 情報処理, Vol.53, No.8, 2012.
- [4] "スーパーコンピュータ「京」", 雑誌 FUJITSU, Vol.63, No.3, 2012.