九州大学学術情報リポジトリ
Kyushu University Institutional Repository

大規模システム評価環境PSI-SIM ： 数千個のマルチコア・プロセッサを搭載したペタスケールコンピュータの性能予測

井上, 弘士
九州大学

薄田, 竜太郎
財団法人福岡県産業・科学技術振興財団

安藤, 壽茂
富士通株式会社

石附, 茂
富士通株式会社

他

https://hdl.handle.net/2324/13247

出版情報：電子情報通信学会技術研究報告, 2008-ARC-178. 2008 (39), pp.51-56, 2008-05-06. IPSJ
バージョン：
権利関係：ここに掲載した著作物の利用に関する注意 本著作物の著作権は（社）情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。

KYUSHU UNIVERSITY

社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE

大規模システム評価環境 PSI-SIM：数千個のマルチコア・プロセッサを搭載したペタスケールコンピュータの性能予測

井上弘士 [1)] 薄田竜太郎 [2)] 安藤壽茂 [3)] 石附茂 [3)] 小松秀実 [3)] 稲富雄一 [1)] 本田宏明 [1)]
山村周史 [3)] 柴村英智 [4)] 于雲青 [1)] 青柳睦 [1)] 木村康則 [3)] 村上和彰 [1)]

1) 九州大学 〒819-0395 福岡県福岡市西区元岡 744 番地

2) 財団法人福岡県産業・科学技術振興財団 〒810-0001 福岡県福岡市中央区天神 1-1-1

3) 富士通株式会社〒211-8588 神奈川県川崎市中原区上小田中 4-1-1

4) 財団法人九州先端科学技術研究所 〒814-0001 福岡県福岡市早良区百道浜 2-1-22

E-mail: inoue@i.kyushu-u.ac.jp, psi@psi-project.jp

あらまし 本稿では，ペタフロップス級スーパーコンピュータの性能予測を可能にするシステム評価環境 PSI-SIM の開発について報告する．また，実在するマシンを対象とした性能予測精度実験，ならびに，仮想ペタスケール・システムを対象とした大規模性能予測実験の結果を示す．実在する計算機システムを利用して，2〜3 桁高性能な計算機システムの性能を予測するためには，これらの間に存在する大きな性能差を埋める必要がある．そこで我々は，この要求を満足し高速かつ正確な性能予測を実現するため，1)プログラム・コード抽象化技術の導入，ならびに，2)仮想超並列実行環境の構築を行った．192GFlops（または 52GFlops）の PC クラスタを用いて実在する 6.5TFlops マシンの性能を予測した結果，性能予測誤差は 10%程度と高い精度を実現した．また，4,096 個のマルチコア・プロセッサを搭載した仮想スーパー・コンピュータ（ピーク性能 2.1PFlops）の性能予測を行った結果，HPL の実行において実効性能 1.01PFlops であった．このペタスケール性能予測に要する時間は 6 時間と現実的な範囲内であり，本研究で開発した性能評価法ならびに各種ツールの有効性を示した．

キーワード スーパー・コンピュータ，ペタスケール，高性能マルチコア，性能予測

PSI-SIM: Performance Prediction for Peta-Scale Supercomputers with Thousands of Multi-core Processors

Koji INOUE[1)] Ryutaro SUSUKITA[2)] Hisashige ANDO[3)] Shigeru ISHIZUKI[3)] Hidemi KOMATSU[3)]
Yuichi INADOMI[1)] Hiroaki HONDA[1)] Shuji YAMAMURA[3)] Hidetomo SHIBAMURA[4)]
Yunqing YU[1)] Mutsumi AOYAGI[1)] Yasunori KIMURA[3)] and Kazuaki MURAKAMI[1)]

1) Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, 819-0395 Japan

2) Fukuoka Industry, Science & Technology Foundation, 1-1-1 Tenjin, Chuo-ku, Fukuoka, 810-0001 Japan

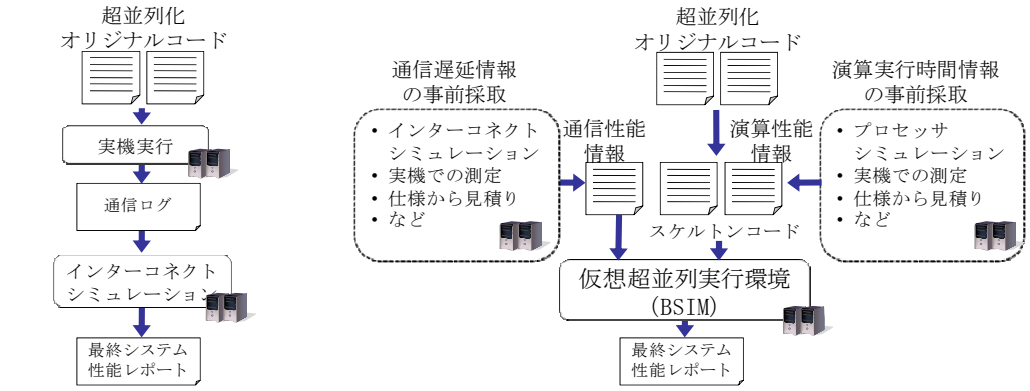
3)Fujitsu Ltd., 4-1-1 Kamikodanaka, Nakahara-Ku, Kawasaki, Kanagawa, 211-8588 Japan

4) Institute of Systems, Information Technologies & Nanotechnologies, 2-1-22 Momochi-hama, Sawara-ku, Fukuoka, 814-0001 Japan

E-mail: inoue@i.kyushu-u.ac.jp, psi@psi-project.jp

Abstract This paper proposes a novel approach to predict the performance of peta-scale supercomputers. It is not easy to accurately predict the performance of large-scale systems such as peta-scale computers. This is because we need to use an existing computer system, the performance of which is exactly lower than the target system. To solve this problem, we have developed a performance evaluation environment called PSI-SIM. The key ideas introduced in PSI-SIM are that 1) we attempt to abstract the application program codes to be used for the performance prediction, and 2) execute them on a virtual parallel execution environment we developed. We have attempted to predict the performance of an existing 6.5 TFlops machine by using a 192 (or 52) GFlops machine. As the result, the prediction error was only 10%. In addition, we have tried to predict the performance of un-existing peta-scale supercomputer, and found that it can achieve 1.01 PFlops for HPL.

Keyword Supercomputer, Peta-Scale, Performance Prediction, Multi-core processor



(A) 従来の性能詳細評価フロー　　(B) PSI-SIM性能詳細評価フロー

図 1：従来ならびに PSI-SIM での性能評価フロー

1. はじめに

高性能コンピュータの開発においては，広大な設計空間を探索し，最適なシステム構成を決定する必要がある．そのためには，高い精度でシステム性能を見積もらなければならない．一般的には，ソフトウェア・シミュレータ上で対象プログラムの動作を模擬することで性能を見積もる．しかしながら，その速度は実機速度と比較して３～４桁以上遅い．そのため，大規模アプリケーションを対象とする場合には極めて長い時間を要する．この問題は，数万から数十万ノードの計算機資源を利用するペタフロップス級大規模システムではより深刻となる．

我々は，ペタフロップス規模のスーパーコンピュータを念頭に置き，その性能予測を可能にする大規模システム性能評価環境 PSI-SIM を開発した．本稿では，PSI-SIM の詳細を述べると共に，実施した大規模性能評価実験の結果を報告する．4,096 個のマルチコア・プロセッサを搭載する仮想スーパーコンピュータ（ピーク性能は 2.1PFlops）の性能を予測した結果，実効性能 1.01PFlops を実現可能であることが分かった．なお，以降本稿では，性能予測対象となる計算機システムを「**ターゲットマシン**」，性能予測を実施する計算機システムを「**ホストマシン**」と呼ぶ．

以下，第 2 節では大規模システム性能予測における問題を整理する．次に，第 3 節において本研究で開発した PSI-SIM の詳細を説明する．第 4 節では，実在する 6.5TFlops マシンを対象とした PSI-SIM 性能予測精度の評価を行う．そして第 5 節では，ピーク性能 2.1PFlops マシンを想定した性能予測実験を実施する．最後に第 6 節でまとめと今後の課題を述べる．

2. 大規模システム性能予測における問題点

これまでに様々な性能評価法に関する研究開発が実施されてきた[1][2][5]．これらの多くは図 1(A)に示すように，1)評価対象となる並列プログラムをホストマシン上で実行し通信ログを採取する．そして，2)これを入力としたインターコネクト・シミュレーションを実施することでシステム性能を予測する．しかしながら，スーパーコンピュータに代表される大規模計算機システムの開発において，ターゲットマシンの性能は，実在するホストマシンと比較して 2～3 桁程度高い場合が多い．そのため，例えばペタフロップス級ターゲットマシンの性能予測をテラフロップス級ホストマシンで実施する場合，この性能ギャップによって以下に示す問題が生じる．

- **ホストマシンのリソース不足に起因するプログラム実行限界**：ターゲットマシン性能を正しく評価するためには，それに見合った大規模問題サイズでのプログラム実行が必要不可欠となる．しかしながら，ホストマシン上で当該プログラムをそのまま実行する場合，ハードウェアリソース不足，特にメモリ容量不足が深刻化する．一般にターゲットマシンはホストマシンより多くの計算ノードを有する．よって，ノード当りの使用メモリ容量が同一の場合でもシステム全体での総使用メモリ容量は大きくなる．その結果，メモリ不足によりホストマシンでのプログラム実行が困難となる．
- **通信ログ採取限界**：数万から数十万ノードを対象としたペタフロップス級アプリケーションの実行では，多くの通信が発生するため通信ログのサイズが極めて大きくなる．その結果，プログラム実行の長期化やハードディスク容量不足などにより通信ログ採取が困難となる．例えば，我々の試算では，4,096 計算ノード(64GB メモリ/ノード)を有するシステムにおいて HPL（High Performance Linpack）を実行した場合には通信ログサイズが 1TB 程度になる．これは計算ノード数の増加に伴い増大するため，数万から数十万ノード規模のターゲットシステムを対象とする場合はより深刻となる．これに加え，開発対象マシンが存在しない場合には実機での通信ログ採取はできない．この場合はプロセッサ・シミュレータ等を利用する必要があるが，その実行には極めて長い時間を要するため非現実的である．
- **インターコネクト・シミュレーションの実行限界**：一般に，インターコネクト・シミュレーションは多くの時間を費やす．例えば，本研究プロジェクトで開発したフリットレベルのインターコネクト・シミュレータ[3]を利用し，計算ノード数 4,096（3 次元トーラス）における Bruck アルゴリズムでの全体全通信 1 回のシミュレーションを実施した結果，その完了に約 9 時間を要している．これは計算ノード数の増加に伴い増大するため，通信ログ生成問題と同様，数万から数十万規模のターゲットシステムを対象とする場合はより深刻になる．

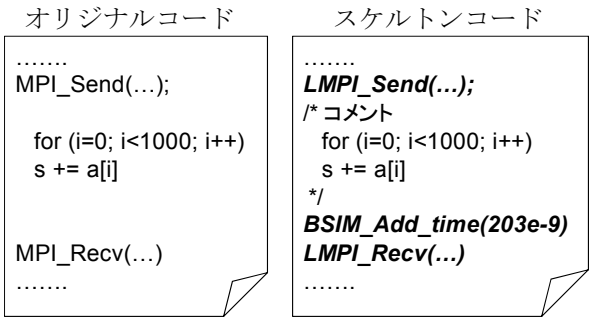


図 2：スケルトンコード

```
  nnn=Nfp*Nep*Nfscale*Nescale
  call LMPI_BCAST(ZPSI2,nnn,MPI_DOUBLE_COMPLEX,Irank,
&       NCOM2(1),ierr)
  mx=(Nep*Nescale-1)/32+1
  nx=((Nep*Nescale-1)/256+1)*8
  jx=(Nfp*Nfscale-1)/32+1
  time_val=7.9d-8+mx*(0.512d-6*jx+1.141d-6*nx
&          +0.666d-6*jx*nx)
  time_val=time_val/2.0d+0*Nblk
  call BSIM_Add_time(time_val)
  nnn=Nep*Nep*(Nblk-nbsn+1)*Nescale*Nescale
  call LMPI_ALLREDUCE(ZCC,ZCCW,nnn,
&         MPI_DOUBLE_COMPLEX,
&         MPI_SUM,NCOM(1),ierr)
```

Scaled size given to LMPI call

zgemm call replaced with BSIM_Add_time call

Scaled sizegiven to LMPI call

図 3：スケルトンコードの実例（PHASE の一部）

3. 大規模システム性能予測環境 PSI-SIM

3.1. PSI-SIM の概要

第 2 節で述べた問題点を解決し，実在する計算機を利用して開発対象の大規模ターゲットマシン性能を予測するためには，1)巨大な通信ログファイルを必要とせず，かつ，2)仮想環境下においてターゲットマシンでのプログラム実行の振舞いを高速に模擬する環境が必要となる．これらの要求を満足するために，我々は大規模システム性能予測環境 PSI-SIM の開発を行った．PSI-SIM の性能予測フローを図 1(B)に示す．

従来方式とは異なり，ターゲットマシンを想定したプロセッサでの演算時間情報，ならびに，通信遅延情報を事前に採取する．これらの値を利用し，ホストマシン上に構築した仮想超並列実行環境にて対象プログラムを実行することでターゲットマシンの性能を予測する．これにより，通信ログファイルの出力が不要となり，かつ，実機マシンでの実行による高速な性能予測が可能となる．これに加え，開発対象マシンが存在しない場合でも，プロセッサ・シミュレータを用いて演算実行情報のみを採取すれば良い．そのため，プログラム全体をシミュレーション実行する場合と比較して劇的に所要時間を短縮できる．PSI-SIM の最大の特徴は，「プログラムコードの抽象化（図 1 におけるスケルトンコード）」と，その実行を可能にする「仮想超並列実行環境 BSIM」にあり，その詳細は第 3.2 節ならびに第 3.3 節で説明する

3.2. プログラムコードの抽象化

ターゲットマシン性能を予測するためには，何らかの方法で対象アプリケーション・プログラムを実行する必要がある．しかしながら，第 2 節で説明したように，ターゲットマシンでの実行を想定したアプリケーション・プログラム（とその問題サイズ）をそのままホストマシンで実行することは極めて難しい．この問題点を解決するために，我々はプログラムコードの抽象化を導入する．本研究の目的は，対象アプリケーションの正しい実行結果を得ることではなく，ターゲットマシンでの実行時間を高速かつ正確に予測することである．そこで，アプリケーションのオリジナルコードに抽象化を施し，プログラムの実行結果は保証しないが，その実行時間を推定可能な性能評価専用コードを作成する．本稿ではこれを「**スケルトンコード**」と呼ぶ．スケルトンコードの概観ならびに実例を図 2 および図 3 に示す．スケルトンコードは主に以下の特徴を有する．

- **演算部分の抽象化**：オリジナルコードを演算部分と通信部分に分割する．そして，図 2 に示すように，演算部分を実行時間という極めて高い抽象度で表現する．つまり，演算部分をコメントアウトし，それに対応するターゲットマシン推定実行時間で置き換える．これにより，プログラム実行の振舞いを損なうことなく，短時間でのスケルトンコード実行が可能となる．図 2 の例では，MPI 通信で挟まれた For ループ全体を推定実行時間で置換している．スケルトンコード中の「BSIM_Add_time」は，第 3.3 節で説明する BSIM に対し当該演算部分の推定実行時間情報を伝達するための API である．
- **疑似通信ライブラリの利用**：実行結果を保証する必要が無いため，ホストマシン上で対象アプリケーション・プログラムを実行する場合には必ずしも実際に通信を行う必要はない．スケルトンコードでは，第 3.3 節で説明する BSIM がサポートする LMPI 通信ライブラリを利用する．これにより，ホストマシンでの通信量を大幅に削減し高速なスケルトンコード実行を可能にする．
- **使用メモリ容量の削減**：実行結果を保証する必要がないため，オリジナルコード実行と同じ容量のメモリを使用する必要はない．そこで，ホストマシンでの実行を可能にすべく，使用メモリ容量を削減する．例えば，巨大配列データを単一のスカラデータに変換する方法や，小規模問題サイズ向けにメモリ領域を確保し演算量や通信量をスケールする方法などがある．ただし，オリジナルコードが有する実行の振舞いを維持するよう注意が必要である．

3.3. 仮想超並列実行環境の構築

仮想超並列実行環境 BSIM は，Multi-Processing Environment(MPE)を拡張して構築された並列プログラム実行環境である．BSIM を利用して第 3.2 節で説明したスケルトンコードを実行することで，小規模ホストマシンを用いた大規模ターゲットマシンの性能予測が可能となる．BSIM が有する主な機能は以下の通りである．

- **実行時間予測機能**：スケルトンコード中に記述された演算部分の推定実行時間（図 2 における BSIM_Add_time 関数の引数）に基づき BSIM 内部の仮想タイマを更新する．また，通信発生時には図 1(B)に示す通信遅延情報を参照し仮想タイマ値へ反映する．これにより，ターゲットマシン性能（ターゲットマシン上での対象プログラム実行時間）を予測する．通信遅延情報を有効にしない場合には，

表 1：テラフロップス・ターゲット性能予測環境

6.5TFlopsターゲットマシン（性能予測対象マシン）	
プロセッサ	インテルItanium2，動作周波数1.6GHz，デュアルコア
計算ノード数	16(32プロセッサ/ノード，128GBメモリ/ノード)，全コア数は1,024
インターコネクト	計算ノード内：クロスバ，計算ノード間：InfiniBand
192GFlopsホストマシン（性能予測実施マシン）	
プロセッサ	インテルXeon， 動作周波数3GHz， シングルコア
計算ノード数	16（2プロセッサ/ノード，7GBメモリ/ノード）
インターコネクト	InfiniBand，Gigabit Ethernet
51.2GFlopsホストマシン（性能予測実施マシン）	
プロセッサ	インテルXeon， 動作周波数1.6GHz， クアッドコア
計算ノード数	2（2プロセッサ/ノード，10GBメモリ/ノード）
インターコネクト	Gigabit Ethernet

全ての通信遅延をゼロと仮定したターゲットマシン性能を予測できる．

- **通信ログ生成機能：**図 2(B)で説明した PSI-SIM 性能評価フローではプログラム実行における通信ログを生成する必要がない．しかしながら，詳細なインターコネクト性能評価/解析を行うためには図 2(A)の従来フローに基づくインターコネクト・シミュレーションを実施する必要がある．これに対応するために，仮想タイマに基づく通信ログ生成機能を有している．
- **疑似通信機能：**スケルトンコード実行における通信時間を削減するために，図 2 で示す LMPI 関数の実行をサポートする．本関数ではホストマシン上での実際の通信は行わない（もしくは極めて短いメッセージの通信のみを行う）．ただし，通信ログ出力や実行時間予測に関しては当該通信の影響が正しく反映される．

4. PSI-SIM 性能予測精度の評価実験

4.1. 実験環境

実在するテラフロップス・ターゲットマシンを対象とした PSI-SIM 適用実験を実施した．これにより，性能予測精度ならびに予測所要時間を評価する．実験環境を表 1 にまとめる．ターゲットマシンに，九州大学情報基盤開発研究センターに設置されたスーパーコンピュータ PRIMEQUEST580（以降 PQ1024 と略す）を利用した．1,024 個のプロセッサコアによる並列計算が可能であり，ピーク性能は 6.5TFlops である．一方，性能予測を実施するホストマシンは PC クラスタを用いる．性能予測対象となるベンチマーク・プログラムとしては，FMO-ERI（分子軌道法における 2 電子積分計算），Open-FMO（フラグメント分子軌道法プログラム），HPL（密行列連立一次方程式），ならびに，PHASE（第一原理擬ポテンシャルバンド計算ソフトウエア）を用いた．実験手順は以下の通りである．

1. 超並列化オリジナルコード開発と実行：PQ1024 において 1,024 個のプロセッサコアを用いる超並列化オリジナルコードを開発し，ターゲットマシンでの実行時間を計測する．
2. スケルトンコード開発：オリジナルコードに対応する超並列化スケルトンコードを作成する．抽象化部分の実行時間推定では，PQ1024 での実行時間を採取し性能モデル式を構築した．
3. 通信遅延情報の取得：PQ1024 における通信サンプル実測値に基づき，メッセージサイズならびに集団通信対象プロセス数でインデックス付けされた通信遅延テーブルを作成する．
4. ターゲットマシン性能の予測：上記 3 で得た通信遅延情報に基づき，ホストマシン上の BSIM 環境にて上記 2 で開発したスケルトンコードを実行する．

4.2. 性能予測結果

PSI-SIM による性能予測結果を図 4 に示す．それぞれの結果に関して，左がターゲットマシンにおける測定実行時間，右側がホストマシンによる予測実行時間を表す．Open-FMO 以外のベンチマークに関しては演算時間と通信時間の内訳を示している（Open-FMO に関しては実験環境の都合上データを採取していない）．また，ホストマシンでの性能予測に要する時間（ BSIM でのスケルトンコード実行時間）を表 2 にまとめる．

まず，性能予測精度に関して議論する．Open-FMO を除く 3 つのベンチマーク・プログラムにおいて，比較的高い精度での性能予測を実現している．実際，ERI，HPL（問題サイズ 320K），PHASE において，ターゲットマシンでのオリジナルコード実行時間との誤差（絶対値）は，それぞれ，4.0%，3.1%，10.5%であった．内訳を見た場合，演算部分に関しては極めて高い精度で性能を予測している．これに対し，通信時間の予測は若干精度が低いため，通信量の多い PHASE においてその影響が顕著に現れている．一方，Open-FMO に関しては，フラグメントサイズが 984 の場合に性能予測誤差が 140%と極めて大きい．この主な原因としては，スケルトンコード開発における抽象化箇所実行時間見積り精度が低かったためと考えられる．

次に，ホストマシンでの性能予測時間について議論する．表 2 で示すように，ターゲットマシンにおけるオリジナルコード実行時間と比較して，ERI と HPL ではその半分程度の時間で性能予測を完了している．また，Open-FMO の場合には 1/3 程度の時間である．ここで，本実験で使用したホストマシンは，ターゲットマシンと比較して 1/130～1/30 程度のピーク性能であることを考慮すると，本研究で提案したプログラム抽象化（つまりスケルトンコード）ならびに BSIM によるその高速実行により非常に短時間での性能予測を実現していることがわかる．

5. 仮想ペタスケール・システムを対象とした大規模性能予測

5.1. ターゲット/ホストマシン構成

本研究の最終目的であるペタスケール・システムの性能予測実現可能性を実証するために，ピーク性能 2.1PFlops の仮想システムを対象とした PSI-SIM 適用実験を実施した．ターゲットマシンならびにホストマシンの概要を表 3 にまとめる．

ターゲットマシンに関しては，文献[4]で提案した高性能/低消費電力プロセッサコアを 8 個搭載する場合を想定した．本プロセッサは，図 5 に示すように，4 命令同時発行可能なアウト・オブ・オーダのスカラ部に加え，コア当り 16 個の SIMD 演算ユニットを搭載する．これを 8 コア搭載することで 512GFlops のプロセッサチップとなる．これを 1 つの計算ノードとし，

表 2：ホストマシンにおける 6.5TFlops ターゲットマシンの性能予測所要時間

	ERI	*Open-FMO(984)*	*HPL(320K)*	*PHASE(16K)*
ホストマシン性能	50GFlops	210GFlops	210GFlops	210GFlops
性能予測所要時間（秒）	505	12,120	2,520	2,400
ターゲットマシンでのオリジナルコード実行時間との比	0.55	0.34	0.58	7.14

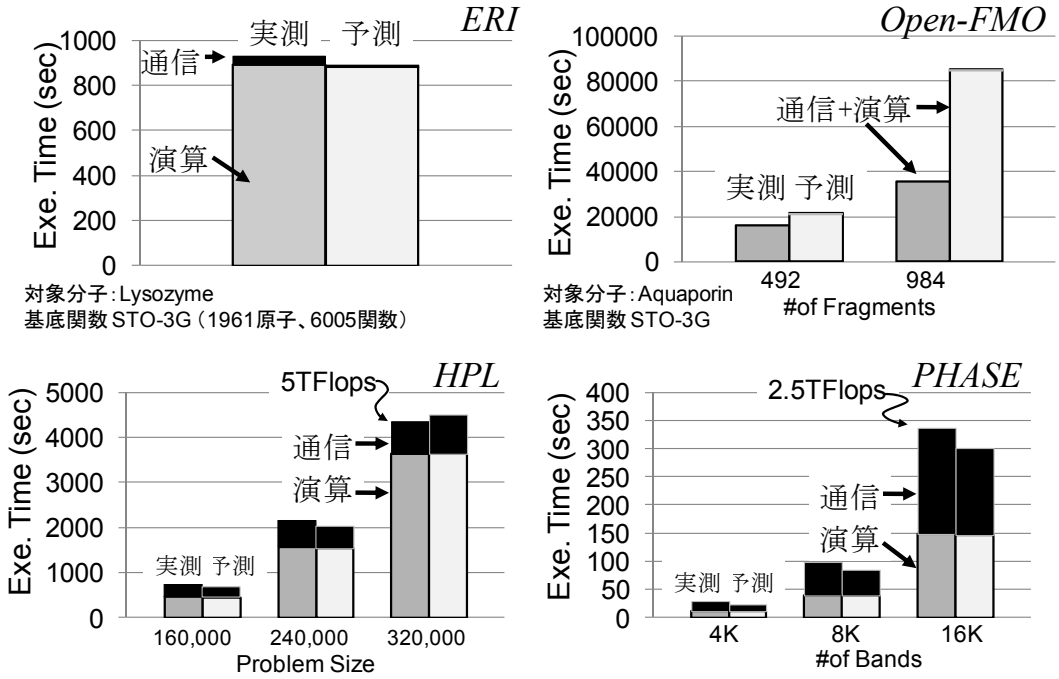


図 4：6.5TFlops ターゲットマシンの性能予測結果

4,096 の計算ノードを 3 次元トーラス・インターコネクトで相互結合する．なお，本インターコネクトにおいては，リンク当りのバンド幅は 20GB/s（片方向当り 10GB/s）を想定しており，インターコネクト/演算ノード性能比は 0.24B/Flop となる．

一方，性能予測を実施するホストマシンに関しては，ピーク性能 1.6TFlops ならびに 192GFlops の 2 種類の実在マシンを使用した．ホストマシンに対するターゲットマシンの性能差は，それぞれ，1,312 倍ならびに 10,937 倍となる．

5.2. 性能予測方法

図 2(B)の性能評価フローに基づきターゲットマシンの実効性能を予測した．ベンチマーク・プログラムには HPL（行列サイズ 131 万元）ならびに PHASE カーネル（バンド数 65,536）を用いた．実験手順の詳細は以下の通りである．

1. スケルトンコード開発：ターゲットマシンは実在しないため，抽象化部分の実行時間推定モデル式の構築においては本研究プロジェクトで開発したサイクルアキュレート・プロセッサ・シミュレータ PSIM を使用した[4]．
2. 通信遅延情報の取得：ターゲットマシンは実在しないため，想定したインターコネクト仕様に基づき通信遅延時間を見積もった．具体的には，仮定したスイッチ・レイテンシとバンド幅に基づき，計算ノード間での平均通信遅延時間を計算する（想定するペタスケール・ターゲットマシンのインターコネクトでは，スイッチ遅延 100ns，筐体内/間のケーブル遅延は 5ns/25ns を想定）．そして，通信対象となる計算ノード番号とメッセージサイズに基づき通信遅延情報を決定した．
3. ターゲットマシン性能の予測：ホストマシン上の BSIM 環境にてスケルトンコードを実行した．

なお，本方式を用いて集団通信時間をモデル化することは難しい．そのため，本評価において通信遅延を考慮した性能予測は集団通信を使用しない HPL のみとした．

5.3. 性能予測結果

性能予測結果を表 4 に示す．ここで，3 列目の「性能予測に要する時間」とは，ホストマシン上でのスケルトンコード実行に要する時間である．実験の結果，想定したペタスケール・システムの性能は，HPL に関して通信遅延が無い場合で 1.07PFlops，3D-Torus ネットワークでの通信遅延を考慮した場合で 1.01PFlops であることがわかった．また，これらの性能予測に要

表 4：ペタスケール・ターゲットマシンの性能予測結果

ベンチマーク	ホストマシン		ターゲットマシン性能（予測）			
	性能	性能予測所要時間	通信遅延無し		通信遅延有り	
			実行時間	性能	実行時間	性能
HPL	1.6TFlops	約6時間	1,397秒	1.07PFlops	1,478秒	1.01PFlops
PHASE	192GFlops	約4時間	165秒	0.65PFlops	---------	---------

表 3：ペタフロップス・ターゲット性能予測環境

2.1PFlopsターゲットマシン（性能予測対象マシン）	
マルチコア・プロセッサ	（4命令発行OOOスカラ＋16SIMD演算ユニット）×8コア，動作周波数2GHz，ピーク性能512GFlops
計算ノード	1プロセッサ/ノード，64GBメモリ/ノード，総ノード数4,096
インターコネクト	3D-Torus（16×16×16），バイセクションバンド幅10.24TB/s
1.6TFlopsホストマシン（性能予測実施マシン）	
プロセッサ	インテルXeon， 動作周波数3.06GHz， ピーク性能6.12GFlops
計算ノード数	128（2プロセッサ/ノード，4GBメモリ/ノード）
インターコネクト	Gigabit Ethernet
192GFlopsホストマシン（性能予測実施マシン）	
プロセッサ	インテルXeon， 動作周波数3GHz， ピーク性能6GFlops
計算ノード数	16（2プロセッサ/ノード，7GBメモリ/ノード）
インターコネクト	Gigabit Ethernet

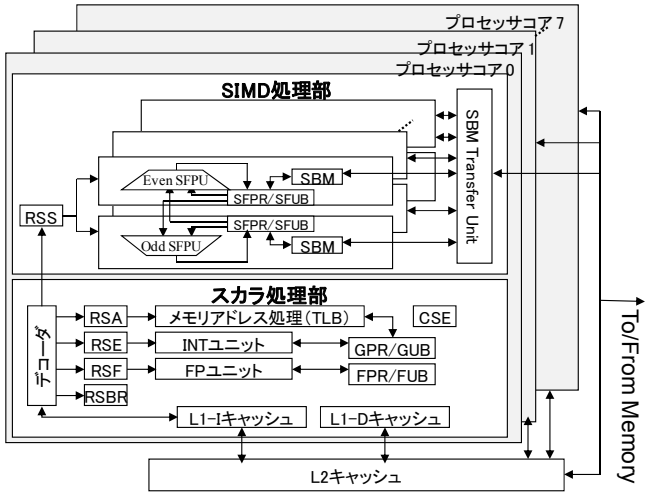


図 5：高性能マルチコア・プロセッサ・イメージ

する時間は約 6 時間であった．PHASE カーネルの場合においても，実効性能は 0.65PFlops と低いものの，その性能予測を約 4 時間で実現している．このように，本研究テーマで開発した PSI-SIM により，性能値では 1,300 倍（2.1PFlops/1.6TFlops）～10,000 倍（2.1PFlops/210GFlops），計算ノード数では 32 倍～256 倍ほど大規模なターゲットマシンの性能を現実時間内で予測することができた．

6. おわりに

本稿では，ペタフロップス級スーパーコンピュータの性能予測を目的に開発した性能評価環境 PSI-SIM の開発について報告した．また，実在マシンを対象とする性能予測精度評価，ならびに，仮想ペタスケール・システムの性能予測に関する実験結果を示した．

今回開発した PSI-SIM によりペタスケール・システムの性能予測実現可能性を実証することができた．一方で，解決すべき課題も多くある．特に，数十万ノードを有する超大規模システムへの対応やより高精度な性能予測，スケルトンコード開発サポート，システム/アプリケーション最適化のための可視化サポート，などが挙げられる．今後，世界に先駆けたスーパーコンピュータ開発を我が国において展開するためには，継続した開発環境の構築とその洗練が必要である．

謝辞

本研究を遂行するにあたり多くのご協力を頂いた富士通株式会社の安里彰氏，松本孝之氏，折居茂夫氏をはじめ，PSIプロジェクト・メンバーの諸氏に感謝致します．なお，本研究は，文部科学省「次世代IT 基盤構築のための研究開発」，研究開発領域「将来のスーパーコンピューティングのための要素技術の研究開発」（平成17 年度～19 年度）における研究開発課題「ペタスケール・システムインターコネクト技術の開発」による．

文　献

[1] 久保田和人，板倉憲一，佐藤三久，朴泰祐： 大規模データ並列プログラムの性能予測手法と NPB2.3 の性能評価，情報処理学会論文誌，Vol. 40, No. 5, pp. 2293- 2304 (1999).

[2] 柴村英智，久我守弘，末吉敏則，“超並列計算機のための相互結合網シミュレータ，”情報処理学会論文誌，Vol. 35, No. 4, pp. 589- 599 (1994).

[3] 柴村英智，薄田竜太郎，本田宏明，稲富雄一，于雲青，井上弘士，青柳睦，″PSI-SIM：大規模並列システムの性能解析に向けた並列相互結合網シミュレータ，″信学技報，vol. 107, no. 276, CPSY2007-32, pp. 45-50, 2007 年 10 月.

[4] 山村 周史，青木 孝，安藤 壽茂，“大規模科学技術計算向け SIMD 拡張スカラプロセッサの提案とその評価，” 情報処理学会研究報告，2007-ARC-174, pp.61-66, 2007 年 8 月.

[5] G. Zheng, G. Kakulapati, and L. V. Kal´e: BigSim: A Parallel Simulator for Performance Prediction of Extremely Large Parallel Machines, *18th International Parallel and Distributed Processing Symposium* (IPDPS’04), p. 78b (2004).