

解説



計算機ベンチマークの最新動向

1. SPEC ベンチマーク†

戸室 隆彦†

1. はじめに

コンピュータの性能指標として SPEC ベンチマークは現在最も広く採用されているベンチマークである。SPEC ベンチマークの結果は、最も一般的にメーカの資料で引用される数値であり、SPEC ベンチマークの結果を示す SPECrate はすでにマイクロプロセッサについての性能指標として共通の言葉となっている。

SPEC の歴史は各メーカがパフォーマンスの向上の目標値として、常に技術革新を図った歴史でもある。また、SPEC ベンチマーク自身もコンピュータ・システムの進歩とともに、より正確な性能計測を目指して何度も改訂が加えられてきている。

2. SPEC とは

SPEC (The Standard Performance Evaluation Corporation) は 1988 年に創設された非営利会社である。その目的は最新のハイパフォーマンス・コンピュータに適用可能なベンチマークの標準化のために、ベンチマークを確立し、保持し、保証することである。この機構の創設者は、アプリケーションに基づくベンチマークによって、ユーザが性能評価を行うに際し、共通の評価基準を提供することで、ユーザに多くの利益を提供することが可能となると信じて SPEC を創設した。1 つのベンチマークで、システムの全性能を明らかにすることは困難であるが、様々な実コードでのベンチマークによって、実際の性能評価に対して価値のある情報が提供可能であることは明らかである。SPEC は、いくつかのグループで形成さ

れ、各グループはそれぞれのベンチマークを提供している。

Open Systems Group (OSG) :

このグループは、オープン・システムの環境でのハイエンドワークステーションやサーバに対する CPU パフォーマンスの計測のためのベンチマークや NFS サーバ・ベンチマークを提供している。

High-Performance Group (HPG) :

このグループは、高性能なコンピュータ・システムにおける基準となるようなアプリケーションでのベンチマークの確立を目的として活動を行っている。

Graphics Performance Characterization Group (GPCG) :

このグループは、業界標準のグラフィック・ベンチマークの提供を目的として活動している。

SPEC の活動は、基本的には次の 2 つである。

1. 性能の計測を目的とする一連のベンチマークの開発

これらのベンチマークのために、ソース・コードとツールからなるパッケージを提供する。このパッケージについては、リリース前にその移植性について、十分にテストを行っている。ライセンス契約によって、SPEC メンバと顧客は各ベンチマークのドキュメントに従って、ベンチマークの実行とレポートを行うことが許される。

2. ニュースとベンチマーク結果を四半期ごとにレポート

このレポートは、SPEC ベンチマークの結果を集大成したもので、SPEC メンバと非メンバは共に、SPEC ニュースに結果を記載することは可能だが、非メンバは費用を払う必要がある。

† SPEC Benchmark by Takahiko TOMURO (Marketing and Sales Support, CRAY RESEARCH JAPAN LTD.).
‡ 日本クレイ(株)マーケティング・セールスサポート

3. 過去の SPEC ベンチマーク

3.1 SPEC89 ベンチマーク

SPEC では 1989 年に最初の CPU ベンチマークを発表している。このベンチマークは the SPEC Benchmark Release 1 として発表され、その後に SPEC 89 と呼ばれるようになった。このベンチマークは 1.2b が最終バージョンである。このベンチマークは高水準言語のソースコード形式で提供される 10 プログラムから構成され、4 つの整数演算 (C 言語) と 6 つの浮動小数点 (FORTRAN) プログラムから成る。計算速度性能としては各プログラムの実行時間の参照時間に対する比率の幾何平均値が示され、スループット性能の評価として、CPU あたり 2 つのベンチマーク・プログラムのコピーを実行してスループット (Thruput Ratios) が計測された (スループット・メソッド A) (表-1)。

表-1 SPEC 89 ベンチマーク評価項目

計算速度性能	スループット性能
SPECint 89	SPECintThruput 89
SPECfp 89	SPECfpThruput 89
SPECmark 89	SPECThruput 89

ト・メソッド A) (表-1)。

SPEC は、移植以外の理由でソースコードの変更を行うことを禁止しており、この移植のためのソースコードの変更に際しても、その変更が性能に影響を与えるものでないことを証明し、SPEC の承認を得ることを要求している。この規則は、現在まで変わらない。

3.2 SPEC92 ベンチマーク

ハードウェアやコンパイラの急速な進歩によって、SPEC 89 での性能評価が問題となったため、SPEC では 1992 年 1 月に SPEC 92 ベンチマークを発表した。SPEC 92 は、CINT 92 (6 つの整数演算性能評価プログラム) と CFP 92 (14 の浮動小数点演算性能評価プログラム) の 2 種類のベンチマークから構成される (表-2, 表-3)。

SPEC 89 に対しては、計測時間の精度の向上、評価対象の増加と不適切なアプリケーションの削除、プログラムの実行パラメータのファイルからの読み込みなどの点を改善している。しかし、SPEC 92 での最大の変更点は、その後の SPEC 95 にも受け継がれるベンチマークの評価方法である (表-4)。SPEC 89 では整数演算プロ

表-2 CINT 92 ベンチマーク・プログラム

プログラム名	参照時間	対象分野	プログラム説明
008.espresso	2270	論理設計	プログラム可能な論理アレイの最適化
022.li	6210	インタプリタ	再帰呼出しによる 9 クイーン問題
023.eqntott	1100	論理設計	ブール方程式を真理表値に翻訳
026.compress	2770	圧縮	テキストファイルの Lempel-Ziv 符号法による圧縮
072.sc	4530	スプレッドシート	Unix の "curses" を使用したスプレッドシート・アプリケーション
085.gcc	5460	プログラミング	コンパイルにより最適化された Sun-3 アセンブラ・コードの生成

表-3 CFP 92 ベンチマーク・プログラム

プログラム名	参照時間	対象分野	プログラム説明
013.spice 2 g 6	2400	回路設計	アナログ回路シミュレーション
015.doduc	1860	シミュレーション	原子炉系の事象系列解析のためのモンテカルロ・シミュレーション
034.mdljdp 2	7090	量子化学	500 個の原子の相互作用に対する運動方程式の解法
039.wave 5	3700	電磁気学	マックスウェルの方程式と粒子方程式の解法
047.tomcatv	2650	メッシュ生成	ベクトル化された 2 次元メッシュ形成
048.ora	7420	光学解析	放射線トレースプログラム
052.alvinn	7690	ロボティスティック	ニューラルネットワークのトレーナ
056.ear	25500	シミュレーション	人間の耳をシミュレーション
077.mdljdp 2	3350	量子化学	034.mdljdp 2 と同様の問題を単精度で解く
078.swm 256	12700	シミュレーション	有限差分モデル
089.su 2 cor	12900	量子物理学	モンテカルロ・シミュレーション
090.hydro 2 d	13700	宇宙物理学	流体力学での Navier-Stokes 方程式の解法
093.nasa 7	16800	NASA カーネル	7 つのベクトル化可能な浮動小数点カーネル
094.fpppp	9200	量子化学	多電子積分計算 (Gaussian 88)

表-4 SPEC 92 ベンチマーク評価項目

ルール	計算性能速度	スループット性能
Aggressive	SPECint 92	SPECrate_int 92
	SPECfp 92	SPECrate_fp 92
Baseline	SPECbase_int 92	SPECrate_base_int 92
	SPECbase_fp 92	SPECrate_base_fp 92

グラムと浮動小数点演算プログラムを合わせてその幾何平均を評価する SPECmark 89 を使用したが、SPEC 92 では整数演算と浮動小数点演算を包括して評価することは行わない。これは、整数演算ベンチマークと浮動小数点演算ベンチマークの平均をとった場合、その結果は浮動小数点演算に大きく依存することになるといった理由や、現在では整数演算、浮動小数点演算のいずれかを強化することがプロセッサの開発傾向となってきたことから両者を包括して評価することは妥当でないと判断されたことによる。

また、SPEC Baseline ルールというベンチマーク実行時の規定が追加された。これは、コンパイラ/プリプロセッサの最適化の実行に際して、その最適化オプションについて、一定の条件を付加するものである。Baseline のルールとして、共通のプログラム言語を使用するベンチマークについては、すべてに同一のコンパイラ・オプションを使用すること、ベンチマーク・プログラムの特定のサブルーチン・モジュール名をコンパイラ・オプションに指定することの禁止などが規定されている。これはユーザがより容易に、アプリケーションの性能向上を図ることができるシステムを選択する指針のためのベンチマークを想定し、一般ユーザにとって不要なまでに過度に最適化されることを避けるためである。一方、Aggressive ルールでは、最大限に性能が発揮可能な最適化オプションの使用が許される。これらのベンチマーク・ルールは、SPEC 95 でも継承されている。

SPECint 92/SPECfp 92 と SPECrate_int 92/SPECrate_fp 92 は、その評価単位が異なることから直接比較することはできない。

4. 現在の SPEC ベンチマーク

4.1 SPEC 95 ベンチマーク

SPEC では、1995 年夏に従来の SPEC 92 に代

わる新しいベンチマークとして SPEC 95 を発表した (Aug. 21.1995)。これは技術進歩によって、プロセッサ、システムおよびコンパイラがその性能を大幅に改善されつつある現状に対応するために、ベンチマーク自身も再度見直す必要が生じたことによる。また、コンピュータ産業の発展にもなって、SPEC ベンチマークが話題になる機会は急増している。このような状況で、新しいベンチマークを発表し、その目的を広く公開することは非常に重要である。

従来 の SPEC ベンチマークが単なる“プロセッサ”のベンチマークであるとの批判に対して、SPEC 95 では、プロセッサとメモリ階層を含むシステムの性能をコンパイラ技術の評価も含んで総合的に評価可能な基準となるベンチマークを目標としている。ただし、SPEC 95 ではグラフィック、ネットワーク、I/O、オペレーティング・システムなどの評価を行うことは意図していない。SPEC 92 に対する主な改善点を次に示す。

実行時間 現在の高性能なプロセッサでは、SPEC 92 ベンチマークのいくつかのプログラムは1分以内でその実行を終了する。このため、SPEC ベンチマークの評価時に、非常に小さな性能データの変化で評価値が大きく変動する場合があった。この点を考慮し、その実行時間をより長くし、将来的なシステムの性能向上に対しても対応可能としている。

アプリケーション・サイズ 従来から指摘されてきたことではあるが、現在のアプリケーション・プログラムはその複雑さを増し、プログラム・サイズも大きくなっている。そのため、SPEC 92 では現在のシステムに想定される一般的な負荷を表現することが困難となっている。このため、SPEC 95 では、ベンチマークの選択において、より規模の大きなものを選択し、アプリケーション・サイズの小さなものとの組合せで、ベンチマークの評価を行うことを意図している。

アプリケーション・タイプ SPEC では、ベンチマークについてより広い分野を対象とする必要性から、イメージ処理とデータベース関連のアプリケーションを追加している。

移植性 現在のコンピュータ環境では、計算機のパフォーマンスの評価は UNIX ワークステーション以外についても重要である。そのため、ベン

チマークはオペレーティング・システムに依存することなく実行できることが重要であった。しかしながら、SPEC 95 の最初のリリースは、UNIX 環境での POSIX または ANSI の開発環境についてのみを対象としている。SPEC は引き続き他のオペレーティング・システム（たとえば WINDOWS/NT など）を対象とするリリースを行う予定である。また、ベンチマーク・ツールも改善されており、より容易にベンチマークの実行も可能となっている。

性能向上-SPEC ベンチマークの性能向上を目的になされた改良・改善は他のすべてのプログラムとユーザの負荷の改善に寄与することを期待してきたが、SPEC ベンチマークのみを対象とした性能向上が意図される恐れがある。この点については、ベンチマークを頻繁に更新することで、その危険性を排除することが可能となる。

SPEC 95 は、整数演算性能の評価を行う CINT 95 と浮動小数点演算の性能評価を行う CFP 95 の 2 種類のベンチマークから構成される（表-5、表-6）。SPEC は、SPEC 95 パッケージとして一枚の CD-ROM でベンチマーク・プログラムのソースコード、そのソースコードのコンパ

イルと結果の検証を行うツール、性能データのレポートを作成するツール、およびベンチマークに関するドキュメントを提供する。CINT 95 のアプリケーションは C 言語で記述されている。CFP 95 のアプリケーションは FORTRAN 言語で記述されている。

性能評価は、SPEC 92 と同じく計算速度とスループットについて計測する。計算速度性能としては、SPARCstation 10/40（40 MHz Super-SPARC 2 次キャッシュなし）での実行時間に基づく数値として設定された SPEC 参照時間に対する個々のベンチマークの実行時間（経過時間）の比率（SPECratios）をもとに 4 つのベンチマーク値を示す（表-7）。SPEC の結果を見れば明らかであるが、各社システムの個々の SPECratio は大きく異なる。SPEC はこれらの個々のベンチマークの結果に基づいて、ユーザは想定されるユーザ自身の負荷と実行するアプリケーションの種類を SPEC ベンチマークの内容と対比し、自身のジョブに対して最も正確にその性能を評価可能なベンチマーク・プログラムに注目することを勧めている。しかし、SPEC では整数演算、浮動小数点演算について、その統計的尺度としてならん

表-5 CINT 95 ベンチマーク・プログラム

プログラム名	参照時間	対象分野	プログラム説明
099.go	4600	ゲーム	囲碁
124.m 88 ksim	1900	シミュレーション	モトローラ 88100 プロセッサ上での Dhrystone ベンチマークのシミュレーション
126.gcc	1700	プログラミング	最適化された SPARC アセンブラ・コードの生成
129.compress	1800	圧縮	テキストファイルの Lempel-Ziv 符号法による圧縮
130.li	1900	言語インタプリタ	Lisp インタプリタ
132.jpeg	2400	画像データ	画像データの圧縮と復元
134.perl	1900	シェルプログラミング	Perl スクリプトによるアナグラムとプライム数の計算
147.vortex	2700	データベース	相互参照データベース

表-6 CFP 95 ベンチマーク・プログラム

プログラム名	参照時間	対象分野	プログラム説明
101.tomcatv	3700	幾何計算	2 次元のメッシュ生成
102.swim	8600	気象予測	有限差分モデルの解析 (1024×1024 グリッド)
103.su 2 cor	1400	量子物理学	モンテカルロ・シミュレーション
104.hydro 2 d	2400	宇宙物理学	流体力学での Navier-Stokes 方程式の解法
107.mgrid	2500	電磁気学	3 次元ポテンシャルのマルチグリッド・ソルバ
110.applu	2200	計算流体力学	ピボット選択を含むマトリックスの解法
125.turb 3 d	4100	シミュレーション	六面体内の乱流計算
141.apsi	2100	気象予測	温度、風速、風向の計算
145.fpppp	9600	量子化学	多電子積分計算 (Gaussian 88)
146.wave	3000	電磁気学	マックスウェルの方程式の解法

表-7 SPEC 95 ベンチマーク評価項目

ルール	計算性能速度	スループット性能
Aggressive	SPECint_95	SPECint_rate_95
	SPECfp_95	SPECint_rate_95
Baseline	SPECint_base_95	SPECint_rate_base_95
	SPECfp_base_95	SPECfp_rate_base_95

の数値を規定することが求められていることも考慮し、各ベンチマーク・プログラムの SPECratios の幾何平均値として SPEC ベンチマークの性能を示している。

スループット性能の評価法としては、与えられたベンチマークを複数個実行する homogeneous capacity method と呼ばれる方法を採用している。この方法はマルチプロセッサ・システムにとって最も適したものとなる。ここでの結果 (SPECrate) は、(個々のベンチマークの性格によって) 特定のタイプのジョブを規定された時間内でいくつ実行できるかを評価したものであり、同じような性格の計算依存のジョブを実行することに対するシステムの処理能力を評価することになる。統計的な尺度としては、各ベンチマーク・プログラムでの SPECrate の幾何平均値として SPEC ベンチマークの性能を示している (表-7)。このスループット性能は、マルチプロセッサ・システムでは、CPU の総合処理能力を評価するものであり、特定のワークロードに対して、システムがどれだけ効果的に使われるかを評価するものではない。

SPEC 92 で高い性能値を示すシステムは SPEC 95 においても高い性能を示すことは予想されるが、SPEC 92 の結果から、SPEC 95 の結果を一意に導く公式はない。

4.2 SPECchpc96 ベンチマーク

High-Performance Group (SPEC/HPG) は、高性能コンピュータの性能評価のための標準ベンチマークとして SPECchpc 96 を発表した。SPECchpc 96 は、科学技術計算分野で使用されるアプリケーションによって性能評価を行うベンチマークである。SPECchpc 96 は、対称型マルチプロセッサ (SMP)、ワークステーション・クラスタ、分散メモリ並列計算機、パラレル・ベクタ (PVP) システムなどの広範囲の高性能コンピュータの性能評価のための標準ベンチマークををめぐして

いる。

SPECchpc 96 の最初のベンチマークとして、地質解析アプリケーションの SPECseis 96 と計算化学アプリケーションの SPECchem 96 がリリースされる。SPECseis 96 は、Atlantic Richfield Corp. (ARCO) が開発したオイル・ガスの探査プログラムを基に開発されたものである。SPECchem 96 は、GAMESS (General Atomic and Molecular Electronic Structure System) を基に開発されたベンチマークである。将来的には SPEC では、気象、流体、構造などの分野のアプリケーションからもベンチマーク・コードを選択することを考えている。

ベンチマーク・コードは逐次実行版とメッセージ・パッシング版が提供される。ベンチマークに際しては、他の SPEC ベンチマークと同様にその実行に際して、一定のルールが規定されている。ベンチマーク値は、これらのベンチマークを一日に何回実行できるかという数値によって示される (86400 (秒)/各ベンチマークの実行経過時間 (秒))。

4.3 システム・レベル ベンチマーク

SPEC ではシステム・レベルでの性能評価ベンチマークとして 2 つのベンチマーク集を提供している。SDM (SPEC Software Development Multitasking Benchmarks) と SFS (System File Server) の 2 種類である。SDM ベンチマークは 2 つのベンチマーク 057. sdet と 061. kenbus 1 から構成される。この両者は、システムの様々な要素 (CPU, I/O, メモリ, オペレーティング・システム, ディスクなど) にストレスを加えることで性能を計測する。SFS Release 1 ベンチマークは、097.laddis ベンチマークから成り、NFS ファイル・システムのもとでイーサネットまたは FDDI を使用してファイル・サーバにアクセスした場合のスループットとレスポンス時間を計測する。

SDM ベンチマーク

SDM ベンチマークでは、2 種類のソフトウェア開発にともなうワークロード (開発環境) を想定している。ベンチマークでは、同時に実行するベンチマークのコピー数を増加することで負荷を増大させる。1 つのベンチマーク負荷は 1 つのスク립トから形成される。性能データは単位時間

内にこのスクリプトを何回実行できたかによって示される (scripts/hr).

057. sdet 商用の UNIX/C の開発環境を想定したベンチマークであり、26 の UNIX コマンド (spell, nroff, diff, make, find など) を実行する。

061. kenbus1 研究・開発環境での UNIX/C の使用を想定したベンチマークであり、18 の UNIX コマンド (cc, cat, grep, mkdir, rm など) を実行する。

ベンチマークでは、スループット値 (スクリプト、時間あたりのユーザ負荷) を、並列に実行されるワーク負荷に対する値としてプロットすることで性能を評価している。ユーザは、種々の並列処理ワーク負荷に対するシステム処理性能と共に、ピーク・スループットを評価することも可能である。これら2つのベンチマークはワーク負荷が異なるため、直接的に比較することはできない。

SFS ベンチマーク

SFS ベンチマークは、現在の 097. laddis と名付けられたベンチマークのみによって構成される。このベンチマークでは与えられた NFS スループット・レベル (operation/sec) に対する NFS ファイル・サーバの応答時間を評価する。

097. laddis オペレーションの負荷とサーバの応答時間をグラフとして示す。

5. SPEC の製品とサービス

SPEC の主な製品は一連のベンチマークである。これらのベンチマークで使用されるコードは、様々なところから提供されたコードに基づいて、SPEC が開発したものである。SPEC は、これらのコードの移植性の検討を行い、また、ツールの開発を行っている。また、これらのプログラムについて、ベンチマークとしてのワークロードの検討を行っている。SPEC で使用されているベンチマーク・コードは同じ名前や似たような名前を持つパブリック・ドメイン プログラムとして一般に広く知られているプログラムとは同一のものではないし、当然その実行時間も異なる。

SPEC ベンチマークのソース・プログラムは一

表-8 SPEC ベンチマーク価格

SPEC 製品	価格	大学向け価格
SDM	\$ 1,450	\$ 725
SFS	\$ 1,200	\$ 600
SPEC 95	\$ 600	\$ 300 *)

*) SPEC 92 ライセンスユーザは 50% ディスカウント

表-9 SPEC 会員費

	入会金	年会費
SPEC Membership	\$ 1,000	\$ 5,000
SPEC Associate	\$ 500	\$ 1,000

般に公開されているが、無料ではない。SPEC では、これら一連のベンチマークについて、個々に値段を設定している。ここでの費用は、会社の運営費用 (テープの作成、ベンチマークに対する質問への返答など) として使われる (表-8)。

SPEC ニュースレターは 3 か月ごとに出版され、様々なシステムでの結果がこの Newsletter にベンチマーク結果として記載されている (1 年 4 回発行分、通常は 50~70 ページ程度)。価格は、米国内については \$550、米国以外については \$575 となっている。SPEC は、一般のメンバの他に、教育機関、または、非営利的な団体に対しては、“SPEC Associate” として準メンバの会員に認定している (表-9)。この準メンバは、投票の権利はないが、SPEC ニュースやベンチマーク・テープなどの SPEC メンバと同等なサービスを受けることができる。おそらくもっと重要な点は、開発中のベンチマークにより早くアクセスすることができ、ベンチマークにおける技術的な面での貢献が可能となることである。SPEC に対する顧問的な立場で行動し、SPEC の事業に技術的なデータを与えることが求められる。SPEC のミーティングは 7 週ごとに開かれ、ベンチマークにおける技術的な作業と決定を行う。すべてのメンバおよび準メンバは、このミーティングに参加し、提議することができる。ベンチマークが SPEC の製品として認められるか否かは、最終的に全メンバによる投票によって問われることになる。SPEC の連絡先を以下に示す。

Standard Performance Evaluation Corporation (SPEC)

2722 Merrilee Drive, Suite 200
Fairfax, VA 22031, USA

Phone : +1-703-698-9604

FAX : +1-703-560-2752

E-Mail : spec-ncga@cup.portal.com

WWW : http://www.specbench.org/

6. おわりに

SPEC ベンチマークは、現在は一般的な標準ベンチマークとなっているが、その性能比較を幾何平均で評価した SPEC 値のみを参照することが多い。しかし、ここに示したように SPEC ベンチマークはタイプの異なるアプリケーションから構成されるため、個々のアプリケーションの性能についても、検討することは非常に重要である。また、SPEC ベンチマークは、プラットフォーム間でのパフォーマンスの比較指標としては、非常に有効な情報を提供するものであるが、完全に自分自身のプログラムをベンチマークしてくれるものではないということを常に認識する必要がある。ユーザは自分自身の環境において、何が重要

であるかを考慮したうえで、最も適切な測定指標を選択することが必要である。

参考文献

- 1) Berry, M., Cybenko, G. and Larson, J.: Scientific Benchmark Characterizations, Parallel Computing, Vol. 17, pp. 1, 173-1, 194 (1991).
- 2) Reilly, J.: SPEC Discusses the History & Reasoning Behind SPEC 95, SPECNewsletter, Vol. 7, No. 3, p. 1 (Sep. 1995).
- 3) SPECNewsletter, Vols. 5-7 (1993-1995).
- 4) Dixit, K.: The SPEC Benchmarks, Parallel Computing, Vol. 17, pp. 1, 195-1, 209 (1991).
(平成 7 年 11 月 28 日受付)



戸室 隆彦

1961 年生。1985 年東北大学工学研究科修士課程修了。同年日本クレイ(株)入社。現在、マーケティング・セールスサポートグループマネージャ。製品性能評価を主に担当。

