解　説

超並列スーパーコンピュータへの移行

―― VPP800 から HPC2500 ――

金澤　正憲*

1　はじめに

京都大学学術情報メディアセンターのスーパーコンピュータは、来年（平成16年）3月1日に、現在の Fujitsu VPP800/63（以後、VPP800 という）から Fujitsu PRIMEPOWER HPC2500（以後、新スパコンという）に置き換えることになりました。ここでは、新スパコンの特徴と概要を VPP800 と比較しながら紹介したいと思います。

2　ベクトル型とスカラ型

スーパーコンピュータには、VPP800 のようなベクトル計算機を並列にした（複数並べた）もの（以後、VPP と呼びます）と、多数の CPU が主記憶を共有した計算機（共有メモリ型スカラ計算機 SMP）を並列にしたものとがあります。

VPP800 では、1 台の CPU で可能な限り多くの計算を速く処理できるように様々な工夫がなされています。例えば、行列積の計算では、積和計算

$$c_{i,j} = \sum_{k=1}^{n} a_{i,k} \times b_{k,j}$$

を次から次へととどまることなく計算を行えるような演算パイプラインが用意されています。即ち、演算パイプラインは、掛算と加算の結果を1マシンサイクル（4ナノ秒、ナノ秒＝10^{-9} 秒）毎にはきだすようになっており、その演算パイプラインが 16 個のデータを同時に吸い込めるので、理論的には、2演算×16÷（4×10^{-9}）＝8GFLOPS という演算能力を持つことになります。

大型計算機センター時代の最初のベクトル計算機 VP100 は1984年4月に導入され、VP200、VP400E、VP2600、VPP500/15、VPP800 と置き換えられてきました。この間、理論性能にできるだけ近い性能がでるように、ハードウェアの面でも、ソフトウェア（コンパイラやライブラリ）の面でも、多くの改善がなされてきました。今では、図1のようなプログラム（1000×1000 の行列積）を実行させれば、何も特別な指定をしなくても、理論性能に近い 7.7GFLOPS 以上がでるようになっています。勿論、実際のプログラムでどの程度の性能がでるかはプログラム、即ち、アルゴリズムによって異なり、理論性能に程遠いこともあります。

```
      DO 30 J=1,MM
        DO 30 I=1,MM
          DO 30 K=1,MM
            C(I,J)=C(I,J)+A(I,K)*B(K,J)
      30 CONTINUE
```

図1．　行列積の計算（MM=1000）

主記憶を複数の CPU で共有する計算機は随分昔からありました。大型計算機センターに最初に導入された富士通 FACOM230-60 も2CPU でしたし、その後導入された計算機は、2～4CPU の計算機でした。マルチ CPU システムと呼ばれていたように、スループットを増大させるために複数の CPU が装備されていました。1つのプログラムは1つの CPU を使ってきました。

*かなざわ まさのり （京都大学学術情報メディアセンター　研究開発部　コンピューティング研究部門）

1 つのプログラムで複数の CPU を使うようになってきたのは、SMP でベクトル計算機と同様な大規模計算ができないかという挑戦が発端となりました。高速ネットワークで接続された複数の CPU を 1 つのプログラムで使用するには、計算しようとする CPU に直結された主記憶にデータをおく必要があるため、プログラミングは複雑なものとなります。しかし、複数の CPU が主記憶を共有するとデータの在る場所を気にすることなくプログラミングできるので、複数の CPU を搭載したシステムが増えてきました。現在、計算サーバ(spp と呼ばれている)としてサービスしているコンピュータは、24CPU で主記憶が 24GB の SMP です。[1)]

初期の SMP は、CPU から主記憶へのアクセスがネックになり、多くの CPU を搭載しても並列の効果が 4 台以上は向上しませんでした。ところが、この３～４年に主記憶の構成法に工夫がなされ、CPU の台数が 32 台、64 台と急激に増大し、新スパコンでは 128 台となっています。

新スパコンは、512GB の主記憶に 128 台の CPU が搭載された SMP、これをノードと呼びます、を 11 台超高速のクロスバネットワークで結合した超並列スカラコンピュータから構成されています。表 1. に VPP800 と比較して諸元を示します。

表 1 新スパコンのハードウェアの諸元

項　　目	新スパコン （HPC2500）	現スパコン （VPP800）
演算能力		
ピーク演算能力の総和	8,785　GFLOPS	504　GFLOPS
1 ノードのピーク演算能力	798　GFLOPS	8　GFLOPS
1CPU のピーク演算能力	6.24　GFLOPS	8　GFLOPS
CPU 数	1,408　台／全体 128 個／ノード×11 ノード	63　台
キャッシュ関係	1 次キャッシュ 命令用　128KB データ用　128KB 2 次キャッシュ 共用　２MB	ベクトルレジスタ 128KB
主記憶		
総容量	5,632　GB	504　GB
ノードあたりの容量	512　GB	―――
CPU あたりの容量	――	８　GB
クロスバネットワーク	４GB/秒（in/out）×４	1.6GB/秒×２
ディスク装置		
総容量	8　TB（RAID5）	1　TB
ネットワーク	1Gbps／ノード	

新スパコンの CPU はスーパースカラと呼ばれるプロセッサで、1 マシンサイクル（1.56GHz）毎に、4 つの命令を実行することができます。表１．の 1CPU のピーク演算能力の算出では、1 マシンサイクルで 4 つの浮動小数点演算が実行されていると計算して、即ち、1.56×4GFLOPS＝6.24GFLOPS としています。意味のある実際のプログラムではそのような高い値は得られません。スーパーコンピュータの演算性能の測定によく用いられる LINPACK で、おおよそ半分程度になるものと推測できます。さらに、一般的なプログラムでは、３分の１であれば良いともいわれています。

本センターとしても、SMP の利用に関しては、spp という小さな規模のものがあるだけで十分な経験がなく、ノウハウも経験として体得していません。センターの教職員は、SMP を使いこなす技術を獲得するために研修などを受けています。さらに、利用者の方々に並列化のトライアル的な利用を、開発計画の臨時公募という形で募集しました。応募が多い場合には、研究開発用に導入された SMP（96CPU の 1 ノードからなるマシン）を一部使うことも考えています。

3　新スパコンの基本ソフトウェア

新スパコンで利用できるソフトウェアは、VPP800 で利用できたソフトウェアは殆ど利用できます。表 2 に主なソフトウェアを示します。

OS（オペレーティングシステム）が、UXP/V というVPP 用の Unix から、広く使われている Solaris８になります。一般の利用者に特に留意していただくことはありません。

言語プロセッサは、Fortran、C、C++が同じように使えます。いずれも自動並列化機能を有しています。大雑把に言えば、SMP（1 ノード内）では、データは共有できるので、各 CPU が担当する計算を分割するだけでよいため、ある程度自動的にコンパイラが並列化してくれます。この自動並列化機能をどれだけ賢いものに成長させていくかが課題であり、利用者のプログラムの特徴を知り、コンパイラ作成者（メーカ）と協同して並列のレベルを上げていくことが重要と考えています。

VPPFortran は、XPFortran に引き継がれます。XPFortran の言語仕様は VPPFortran を包含しますから、コンパイルし直すだけで、VPPFortran で書かれたプログラムを実行することができます。基本的には、ベクトル化されていた部分がノード内自動並列化され、並列化指示行はそのまま活かされて複数のノードに分割されて実行されます。この場合、新スパコンのいくつの CPU で VPP800 の 1 台分の能力を超えられるかということがキーになります。利用者のプログラムではどの程度になるか、早く情報を得たいと考えています。新スパコンの設計目標では CPU8 台で９割程度のプログラムが VPP800 を優ることを目標にしているとのことです。

VPP800 には、HPF（High Performance Fortran）が用意されていますが、HPF の利用と一般での普及度を考え、新スパコンでは導入しないことになりました。HPF の利用者には申訳ありませんが、対応をお願いします。

表 2 新スパコンのソフトウェア

項　　目	新スパコン（HPC2500）	現スパコン（VPP800）
OS	Solaris８	UXP/V
言語	Fortran, C, C++	Fortran, C, C++
自動化機能	自動並列化	自動ベクトル化
ノード間並列化	XPFortran	VPPFortran、HPF
並列化機能、ライブラリ	OpenMP、MPI2	MPI2
科学技術計算ライブラリ	SSL2	SSL2
IBM 互換機能	PFD	M-VPP 連携

変数への代入の依存関係に関する情報が不十分なため、自動並列化がなされないことがあります。そのための情報をコンパイラに与えるための機能がOpenMP です。プログラムの中に、「！＄　OMP　」で始まる指示行を挿入することにより、強制的な並列化を促すものです。現在、spp でも利用できます。

ノード間をまたがる並列化のためのライブラリとしては、現在と同様に、MPI２が利用できます。

4　新スパコンのアプリケーションプログラム

新スパコンで利用できるアプリケーションプログラムは、現在サービスしているアプリケーションプログラムのうちの MSC NASTRAN、POPLAS/FEM5、Gaussian03、MOPAC2002、LS-DYNA、VISLINK、AVS が引き続いて利用できます。FSPICE と MASPHYC は、利用状況から今回は導入を見合わせることになりました。ご了承いただきたいと思います。

アプリケーションプログラムについては、汎用コンピュータで充実することを考えていますので、市販のもの、ライセンス契約のもの、シェアウェアなど本センターのコンピュータ上で実行できる様々なソフトウェアについて、ご希望や情報をお寄せください。できる限りご要望をかなえていきたいと考えています。

5　MSP との関係

MSP の利用者でも VPP800 を簡単に利用できるよう M-VPP 連携機能を提供していますが、新スパコンではこの機能の提供はなくなります。その代わり、MSP のフルスクリーンエディタである PFD 相当の機能を新スパコンで提供します。

PFD コマンドを入力すると、ファイルとディレクトリの一覧が表示され、ファイルを選択すればそのファイルの編集を、ディレクトリを選択すればそのディレクトリ下のファイルとディレクトリの一覧が表示されます。画面を上下に移動するとか、編集を終了させるとかのファンクションキーも、MSP と同様に用意されます。さらに、PFD の中で簡単に翻訳・実行ができるような機能も用意されます。

VPP800 や spp で現在提供されている「je コマンド」と同様のコマンドです。一度、使ってみて、使い易さを試してみてください。[2)]

エディタだけでなく標準的な利用に関して、MSP の利用者が使い易いだけでなく、初心者にも取っ付き易いシステムにしたいと考えていますので、不便さなど感じられたことをプログラム相談室、または、consult@kudpc.kyoto-u.ac.jp 宛、メールでお寄せください。

6　置換えについて

新スパコンの設置面積は、現在の VPP800 の 2 倍弱となります。センター北館にはそのような空面積がありません。従って、VPP800 を搬出してから、新スパコンを搬入することになります。このための作業に来年の 2 月中旬から下旬までかかり、スパコンサービスが 1 週間以上停止すると思われます。学年末をむかえますが、計算機の利用を早めに進めていただきますよう、ご協力をお願いいたします。現在、メーカと全面的なサービス停止日を少なくするように、移行計画を立てようとしています。予定が定まり次第、ニュース（ホームページ）などでお知らせしますので、十分ご注意をお願いします。

さらに、この秋から冬にかけて空調設備の増強も行います。そのため、短期間（土曜、日曜になる予定）のサービスを停止する予定ですので、この点もご注意のほど、ご協力をお願いいたします。

7　移行措置について

新スパコンがサービスできるようになれば、できるだけ早く、並列化処理に慣れていただくことが必要と考えています。

現在、VPP800 の 1CPU で実行させているプログラムは、自動並列でどの程度の実行時間になるか。また、新スパコンの CPU を何台まで使うと実行時間はどのようになるか（並列処理における台数効果といいます）などをチェックする必要があります。また、VPPFortran プログラムでは、ノード内の自動並列と並列化指示行によるノード間並列がどのような実行時間となるかをチェックする必要があります。このために、移行期間に負担金を無償にして利用者に並列化を促進することを考えています。これにもご協力をお願いいたします。

8 おわりに

強力な演算能力と大きな主記憶を有する新スパコンを来年 3 月に導入できるようになりました。今後は、モデルの精度や詳しさを従来にもました大規模な科学技術計算やシミュレーションを実行できるようになるとセンター関係者も期待しています。このために、今後も、新スパコンの使い方について、広報でお知らせしていく予定ですので、ご意見やご質問をお寄せください。

参考文献

1）平野彰雄：計算サーバ SPP 活用ガイド、京都大学学術情報メディアセンター全国共同利用版広報、Vol. 2，No.1，pp.263-268，2003。

2）赤坂浩一、平野彰雄、金澤正憲：MSP ユーザのための UNIX 入門、京都大学大型計算機センター広報 Vol.35，No.1，pp.25-34，2002。

附録　用語の説明

スレッド並列：プログラムのデータ（変数や配列）は共有する（一箇所にある）が、処理（命令）は CPU ごとに分割して同時に複数の処理を行わせようとする並列処理方式。新スパコンのように共有メモリ型コンピュータ用の並列化方式です。どの CPU からもデータを参照できるので、処理をいかに分担するかを考えればよいため、比較的容易に並列化できます。単純な DO ループの場合は、ループ回数を並列数で分割して分担すればよい。数多くの CPU が主記憶にアクセスするので、競合（ぶつかり）が発生し、演算速度が低下することがあります。

プロセス並列：プログラムのデータも命令も複数に分割し、それぞれのコンピュータに分割して割当て、並列処理する方式。命令の実行に必要なデータは命令の実行されるコンピュータの主記憶に存在する必要があるとか、他のコンピュータで更新されたデータを参照するときはコンピュータ間で転送する必要があるとか、複雑な操作が必要で、プログラムでもってすべて指示しければなりません。即ち、プログラミングは難しくなります。共有メモリ型コンピュータでも、分散メモリ型コンピュータでもどちらでも実行できます。