京都大学 学術情報メディアセンター 深沢 圭一郎

2.3.1 はじめに

MHD シミュレーションはプラズマの振る舞いを正確に記述できる Vlasov 方程式の近似から求められ る MHD 方程式を利用したシミュレーションである。MHD 方程式は Vlasov 方程式を近似する際に、速 度空間を落とすことで、計算する次元数を減らし、プラズマの大規模構造を計算する際によく利用され ている。惑星磁気圏 MHD コードは惑星の磁場の勢力範囲である磁気圏をグローバルに解くシミュレー ションコードである。太陽からのプラズマの流れ (太陽風) と惑星の磁場が相互作用することで形作ら れる惑星磁気圏は、名に見える現象ではオーロラに代表されるように様々な電磁気現象が起きている。 これらは宇宙空間で運用されている GPS などの衛星や宇宙ステーションだけで無く、誘導電流などの形 で地上にも影響を与えている。これらの現象は宇宙天気と呼ばれており、世界中で広く研究されている。 惑星磁気圏 MHD コードはこの宇宙天気予報において基本となるシミュレーションコードとして利用さ れている。この惑星磁気圏 MHD コードはベクトル機全盛時代から利用されているが、近年では超並列 スカラ計算機にも対応し、様々な計算機上で利用されている。今回ポストペタスケール計算に向けて、 京コンピュータの後継機である FX10 とさらにその後継機である FX100 において、惑星磁気圏 MHD コ ードの性能評価を行った。

2.3.2 プログラム概要

MHD 方程式は以下のようになる。

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\mathbf{v}\rho)$$

$$\frac{\partial \mathbf{v}}{\partial t} = -(\mathbf{v} \cdot \nabla)\mathbf{v} - \frac{1}{\rho}\nabla p + \frac{1}{\rho}\mathbf{J} \times \mathbf{B}$$
(1)
$$\frac{\partial p}{\partial t} = -(\mathbf{v} \cdot \nabla)p - \gamma p \nabla \cdot \mathbf{v}$$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B})$$

上から、連続の式、運動方程式、圧力変化の式(エネルギーの式)、最後が磁場の誘導方程式となる。 簡単に言えば、電磁場を考慮した流体力学方程式と呼べる。MHD 方程式を解く数値計算法としては、 *Ogino* ら[1992]によって開発された Modified Leap Frog (MLF) 法という差分法を使用する。これは最 初の1回を two step Lax-Wendroff 法で解き、続く (*I* – 1)回を Leap Frog 法で解き、その一連の手 続きを繰り返す手法である。*I*の値は数値的に安定の範囲で大きい方が望ましいので、数値精度の線形計 算と予備的シミュレーションから *I*=8 に選んでいる。惑星磁気圏 MHD コードは直交格子を利用してお り、いわゆる stencil 計算である。並列は MPI と OpenMP を利用し、領域分割を行っている。1、2、 3次元分割を性能評価に利用している。通信は基本的には袖領域の隣接通信のみである。

#### 2.3.3 測定環境

性能測定には、東京大学 FX10、名古屋大学 FX100 を利用した。各計算機の構成は表 1 の通りである。 測定には、最大で FX10 を 4800 ノード(76,800 コア)と FX100 を 512 ノード(16,384 コア)を利用 した。

		PRIMEHPC FX10	PRIMEHPC FX100
CPU	Architecture	16 cores SPARC64 IXfx	32 cores SPARC64 XIfx
	Frequency	1.848 GHz (236.544 GFlops)	2.2 GHz (1126.4 GFlops)
	Cache	L1: 32 KB/core	L1: 64 KB/core
		L2: 12 MB/CPU	L2: 24 MB/CPU
Memory	Band width	85 GB/s /CPU (=node)	240 x2 (in/out) GB/s /CPU
B/F		0.36	0.21 or 0.42
Node	Number of CPUs	1	1
	Memory size	32 GB	32 GB
System	Number of nodes	4,800 (76,800 cores)	2,880 (92,1600 cores)
	Rmax	1.135 PFlops	3.244 PFlops
	Node $\operatorname{comm}_{\circ}$	Tofu Interconnect (5 GB/s)	Tofu 2 (12.5 GB/s)

表 1. FX10、FX100 のシステム構成

#### 2.3.4 測定結果

FX10 において、MHD コードの性 能評価結果を図1に載せる。ここでは 1、2、3次元領域分割と3次元領域分 割において計算する配列並びが f(i,j,k,m): typeA と f(m,i,j,k): typeB の場合を評価した。最も性能が高かっ たのは3次元領域分割 typeB で、約 210TFlops (実行効率19%)となった。 一方で配列の順序が違うだけの typeA では110TFlops と半分近くの 性能になっている。この二つのL2 キ ャッシュミス率などを調べた結果、表 2 のようになっており、L2 キャッシュ



図1 FX10における性能測定結果

ミス率とメモリバンド幅の違いが性能の違いに効いていることがわかる。また、L1 キャッシュスラッシングにより typeA の性能が劣化していることも分かっている。最も性能の高い 3 次元領域分割 typeB に式の共通化、配列の次元削減とループ融合といった最適化を施した結果、32 ノード利用時に実行効率が30%となった。

表2 3次元領域分割 typeA と B における SIMD 化率、L2 demand キャッシュミス率、メモリバンド幅

Array type	SIMD efficiency [%]	L2 demand cache miss [%]	Memory bandwidth [GB/s]
Туре А	65.58	19.81	7.49
Type B	54.26	3.64	24.94

FX100 での性能評価結果を図 2 に 載せる。ここでは FX10 での最適化は 加えていない。最も性能が高くなった のは1次元領域分割(実行効率13%) で、次が 3 次元領域分割 typeA(同 10.6%)であった。FX10 で最も性能の 高かった typeB は FX100 では性能が 最も出なかった。このように FX10 と FX100 はコードに対する振る舞い が異なり、FX100 ではベクトル的な コードが速いという結果になった。今 度は 3 次元領域分割 typeA に対して ループや文の分割、プロセスと計算サ



イズの調整、配列の分割、空間タイリングなどを施したところ、16.7%の実行効率を達成することができた。

2.3.5 まとめ

惑星磁気圏を解く MHD シミュレーションコードを用いて、京の後継機種である FX10 と FX100 の性 能評価を行った。基本性能測定として、1 次元、2 次元、3 次元領域分割 typeA と typeB を用いて、それ ぞれの性能を測定した。最も性能が高くなったのは 3 次元領域分割 typeB (実行効率 19%) であり、typeA と比べて 2 倍程度の性能差があった。これは HPC2500 から続く SPARC64v 系の CPU と同様の傾向で あり、キャッシュヒットを考慮したコードの性能が高いことを示している。最も性能が高かった 3 次元 領域分割 typeB にキャッシュをより効率よく利用できるようないくつか最適化を加えたところ、最大で 実行効率が 30%まで向上した。

一方で FX100 に対して、FX10 と同様の性能測定を行ったところ、1 次元領域分割がもっとも性能が 高く(実行効率 13%)、次に 3 次元領域分割 typeA(10.6%)となった。FX10 で最も性能の高かった 3 次元領域分割 typeBは FX100 では逆に最も性能が低かった。これはキャッシュヒットを考慮するよりも、 ベクトル化を考慮した方が良いことを示している。FX100 では 3 次元領域分割 typeA に最適化を加えた。 ループや文の分割によりコンパイラの最適化が促進され、プロセス分割の調整、配列分割やサイクリッ ク並列化により L2 キャッシュミスが減少し、最終的には 16.7%の実行効率となった。

また、今回の FX100 での結果を他の計算機システムと比較すると、1CPU あたりで SX-ACE の 1.6 倍 程度の性能になっており、KNC の Xeon Phi の 2 倍程度の性能となっていた。

# 惑星磁気圏 MHDコードの性能評価

### **深沢圭一郎** 京都大学学術情報メディアセンター

SS研SS研ポストペタアプリ性能WG



#### 【FX10】 TypeA timer71でL1Dキャッシュスラッシング

			Г				
[秒] 200 180 160 140 120 100 80 60 40 20 0	沙一ス(一部)         340 3 p v do i = is, ie1         341 3 p v x =0.5*hx*float(2*(iss+i)-nx2+2*nxp))         342 3 p v ra2 = x * x + y * y + z * z         343 3 p v ra2 = max(ra2, ra_i2)         344 3 p v ra3 = ra2 * sqrt(ra2)         346 3 p v u(i,j,k,8) = u(i,j,k,8)         8         9						
	timer	71 全体		(省略)	((((,,),),),),))))))	ũ	
	timer71 全体						
ТуреА	L1D ミス率 (/ロード・ストア 数)	L1D ミス dm率 (/L1D ミス数)	L1Dミス数	21フテックスす をプリフェッチす が発生している。	れているか2ライン(イン る影響でL1Dキャッシュ。 	テックス)元 スラッシング 	
time o #71			4.065.10	、 L1D競合ツール結果(一部)			
timer/1	Chelo	33.15%	4.900+10	配列名	アドレス	Index	
配列が連約   が高く(目安 _ 訳が高い(	ア野動小数点       347 3 + dx1*(       &         コードキャッ       348 3 f(i+1,j+1,k+1,4)* f(i+1,j+1,k+1,6) &         シュアクセス       349 3 - f(i+1,j+1,k+1,2)* f(i+1,j+1,k+1,6) &         350 3 + f(i+1,j,k+1,2)* f(i+1,j,k+1,6) &       &         351 3 - f(i+1,j,k+1,2)* f(i+1,j+1,k,6) &       &         352 3 + f(i+1,j+1,k,2)* f(i+1,j+1,k,6) &       &         353 3 - f(i+1,j+1,k,2)* f(i+1,j+1,k,6) &       &         353 3 - f(i+1,j+1,k,2)* f(i+1,j+1,k,8) &       &         354 3 + f(i+1,j+1,k,4)* f(i+1,j+1,k,8) &       &         353 3 - f(i+1,j,k,4)* f(i+1,j+1,k,8) &       &         354 3 + f(i+1,j+1,k,4)* f(i+1,j+1,k,8) &       &         354 3 + f(i+1,j+1,k,4) * f(i+1,j+1,k,8) &       &         355 3 - f(i+1,j+1,k,4) * f(i+1,j+1,k,8) &       &         356 3 - f(i+1,j+1,k,4) * f(i+1,j+1,k,8) &       &         357 3 - f(i+1,j+1,k,4) * f(i+1,j+1,k,8) &       &         358 4 - g(i+1,j+1,k,4) * f(i+1,j+1,k,8) &       &         359 3 - f(i+1,j+1,k+1,5)       &         262% 33.15% 4.96E+10       IDERØA         Index<						
⇒ L1D+	・ャッシュスラッシ	ングの疑い有い	J	f(i,j,k,8)	0xF861D3F8	39	
KYOTO UNIVERSITY         Copyright 2015 FUJITS         ff(i,j,k,7)         0x7FEC2469388							

#### 【FX10】 TypeA timer151でL1Dキャッシュスラッシング

					<u>ノース(一部)</u>			
[秒] FX10 TypeA timer151					1362 3 p do i = is1, ie1 (省略) 1371 3 p f(i i k 8)-f(i i k 8)*x2+ff(i i k 8)*(1 0-x2)			
140					1372 3 p $f(i,j,k,7)=f(i,j,k,7)*x2+ff(i,j,k,7)*(1.0-x2)$ 1373 3 p $f(i,j,k,6)=f(i,j,k,6)*x2+ff(i,j,k,6)*(1.0-x2)$			
100	00 浮動小数 1374 3 p f(i,j,k,5)=f(i,j,k,5)*X2+ff(i,j,k,5)*(1.0-X2) 1375 3 p f(i,j,k,4)=f(i,j,k,4)*x2+ff(i,j,k,4)*(1.0-X2) 1376 3 p f(i,j,k,3)=f(i,j,k,3)*x2+ff(i,j,k,3)*(1.0-X2)						4)*(1.0-x2) 3)*(1.0-x2)	
80 60	80 60						2)*(1.0-x2) I)*(1.0-x2)	
40 20	ジ セ	ヘエアク ス待ち		1380 3 p end do				
0	0 Loop (line 1048-1354 1362-1380) 2インテックスすれているが2ライン(インテックス)先 をプリフェッチする影響でL1Dキャッシュスラッシング が発生している。							
	L1D ミス率	I1D ミス dm率			L1D競合ツ	ール結果(一部)		
TypeA	(/ロード·ストア 数)	(/L1D ミス数)	L1Dミス数		配列名	アドレス	Index	
timer151	9 62%	55.81%	7.51E+10	1	u(i,j-1,K,2)	0x192E29098	$34 \rightarrow 30$	
│配列が連約 │が高く(目5 │ 訳が高い(	配列が連続アクセスであるにも関わらずL1Dキャッシュミス率 が高く(目安:3.125%以上)、L1Dキャッシュミスのデマンドの内 訳が高い(目安:20%以上) ストリーム数が91個と多い					0x7FE08DF5120	36	
⇒ L1D‡	⇒ L1Dキャッシュスラッシングの疑い有り					0x7FCD3DDD198	3 36	

KYOTO UNIVERSITY

### 式の共通化@FX10

計算式が長いとコンパイラが適切にスケジューリングで きない場合がある



京都大学 KYOTO UNIVERSITY





#### 今までの最適化すべてを適用した

4ノード利用の結果(計算サイズは800×800×800)

- Flat MPI1 54  $7 \square t \land (4 \land 4 \land 4)$ asis:192.321sec(22.66%)  $\rightarrow$  135.242sec(28.59%)
- Hybrid MPI:  $4 \mathcal{P} \square \forall \mathcal{A} (2 \times 2 \times 1) + 16 \mathcal{A} \lor \forall \mathsf{F}$ asis: 192.570sec(22.78%)  $\rightarrow$  133.895sec(27.67%)

#### 32ノード利用の結果(計算サイズは1600×400×400)

- Flat MPI $\sharp$ 512 $\mathcal{D}$  $\Box$  $\Box$  $\mathcal{A}(8 \times 8 \times 8)$ asis:12.783sec(21.05%)  $\rightarrow$  8.765sec(27.34%)
- Hybrid MPI: 64プロセス(4×4×4)+8スレッド asis:13.247sec(20.51%)→7.672sec(30.34%)



京都大

KYOTO UNIVERSITY



6

5

## 結果概要



## FX10とFX100の比較



### ループや文の分割による性能向上

ステンシル計算だが、L1D\$ミス率が連続アクセスより高い

- ・データ再利用により、連続アクセス時以下が期待値
   →L1D\$スラッシング?
- FISSION\_POINT指示行の挿入により、ループ分割
- ループ分割: ループ分割を人手で実施
- ・ 文の分割: 右辺が長い代入文を、複数の部分式からなる文に人 手で分割

   <u>最適化による性能向上 (FX100 32ノード)</u>

⇒コンパイラ最適化が促進(2→5)



### プロセス分割の調整、サイクリック並列化、他



### 配列の分割

#### 配列f(nx,ny,nz,8)をf1(nx,ny,nz,4)とf2(nx,ny,nz,4)に分割





### Performance measurement of MHD code

#### **Comparison of FX100 with other computer systems**

	Core/CPU	Rmax [TFlops]	Rpeak [TFlops]	Rpeak /CPU [GFlops]	Efficiency [%]	Suitable domain decomposition	CPU architecture
SX-ACE	1024/256	65.50	29.20	114.0	45	3D_A	Vector
K	262144/32768	4194.30	914.12	27.9	22	3D_B	SPARC64 VIIIfx
<b>FX10</b>	76800/4800	1135.41	234.59	48.9	21	3D_B	SPARC64 IXfx
FX100	16384/512	576.72	91.49	178.7	17	3D_A	SPARC64 XIfx
RX200S6	864/144	10.13	3.51	24.4	35	3D_A	Xeon (Westmere)
CX400	23616/2952	510.11	104.23	35.3	20	3D_A	Xeon (SB)
HA8000	23160/1930	500.26	83.42	43.2	17	2D	Xeon (IB)
XC30-HSW	448/32	16.49	1.37	42.8	8	2D	Xeon (HSW)
SR16000/L2	1344/672	25.27	5.38	8.0	21	3D_B	POWER6
Xeon Phi 5120	60/1	1.00	0.08	84.0	8.4	3D_A	Knights Corner
Tesla K20X	896/1	1.31	0.15	153.3	11.7	3D_A	Kepler
						15	

京都大学 KYOTO UNIVERSITY

13