「メニーコア時代のアプリ性能検討WG」報告

井戸村 泰宏 (日本原子力研究開発機構)

科学技術計算分科会 2019年度会合、2019年10月24日、神戸

担当幹事

荻野正雄(名大、大同大、第9回まで)、松尾裕一(JAXA、第10回)

■ 推進委員(会員機関)

井戸村泰宏(JAEA)、石田崇(JAXA)、三浦英昭、沼波政倫(核融合研)、南里豪志(九大) 渡辺宙志(東大、慶応大)、牧野総一郎、加藤由博(第7回から、豊田中研) 梅田隆行、片桐孝洋(名大)、検崎博生、似鳥啓吾、南一生(理研)

■ 推進委員(富士通)

軽部行洋、青木正樹、井口裕次、山中栄次、三吉郁夫、志田直之、千葉修一、内藤俊也 小田和友仁、樽水康平(第6回から)、渡邊健太、河野匡伸(第6回から)

報告書概要

- 基盤技術
 - ループ変換およびスレッド数を動的変更する自動チューニング(片桐)
 - 非ブロッキング集団通信の通信隠蔽効果に関する調査(南里)
 - アプリケーションの性能予測について(南)
 - 富士通C++コンパイラの最適化機能の改善について(渡辺、千葉、小田和)
- アプリ事例
 - JAXA LBM コードの性能測定および改善(石田、三吉)
 - MUTSU コード高速化の検討(三浦)
 - 生体分子粗視化シミュレータCafeMolのFX100での性能測定とチューニング(検 崎、渡邉)
 - ADVENTURE における多倍長精度演算の利用に向けた検討(荻野)
- Post-K性能推定
 - 核融合プラズマ乱流コードGKV のPost-FX100 性能推定(沼波、片桐、青木)
 - プラズマ流体解析コードGT5D のPost-FX100 性能推定(井戸村、内藤、三吉)
 - 宇宙プラズマ5次元ブラソフコードVlasov5の性能測定および推定(梅田、内藤)
 - N 体カーネルの性能推定(似鳥、青木)





nz

実験環境 名大CX400 Xeon E5-2600v3 (14x2cores/node) 非ブロッキング集団通信の通信隠蔽効果に関する調査(南里)

- FX100とCX400における通信隠蔽テスト
 通信の促進にプログレススレッドが必須
 - FX100におけるプログレススレッド (a)アシスタントコアで実行
 - CX400におけるプログレススレッド
 (b)計算用コアで実行
 (c)通信用コアで実行





(b) CX400:32nodes,32procs Ialltoall(1MB)+GEMM(1024²)



(c) CX400:32nodes,32procs lalltoall(1MB)+GEMM(1024²)



アプリケーションの性能予測について(南)

- 改良型ルーフラインモデルの提案
 - ルーフラインモデル[Williams,ACM(実行性能P=N_c/max{t_M,t_c}
 - 改良型ルーフラインモデル

○·予測11能

+ 実則能

Williamsらの

ルーフラインモデリ

1.0-

0.9

0.8

0.7

0.5

"ປ 0.4

ク性能比 0.6

実行性能P=N_c/max{t_M, t_{L2},t_c}





N1=4000, N2=60, N3=80

do i = 1,N1

2M

do k = 1,N3do i = 1.N2

enddo

3M-2L2-2F

■富士通C++コンパイラの最適化機能の改善について(渡辺、千葉、小田和)

- 富士通C++コンパイラの最適化について
 - クラスの定義場所に依存する最適化
 - STLコンテナを利用した場合の最適化
 - C++11"Range-base for"についての最適化
 - 構造体に対する最適化
 - リンク時最適化(LTO)
 - マルチスレッド環境におけるmalloc性能劣化

→IntelC++コンパイラに見られない性能劣化がたくさんある?

STLコンテナの最適化

■ STLコンテナに対する最適化

可変長配列std::vectorはSWPできない

void
calcforce(std::vector<pos> &q, std::vector<pos> &p){
...

固定長配列std::arrayもSWPできない

void

}

calcforce(std::array<pos, N> &q, std::array<pos, N> &p){ ... } This loop is not software pipelined because the loop contains an instruction not covered by software pipelining, such as function call.

普通の配列を渡すとSWPできる

void
calcforce(pos q[N], pos p[N]){
...
}

This loop is software pipelined.

■ 富士通C++コンパイラの対応状況

```
const int N = 20000;
struct pos{
  double x,y,z;
};
void
calcforce(std::vector<pos> &q, std::vector<pos> &p){
  for(int i=0;i<N-1;i++){</pre>
    const double qx = q[i].x;
    const double qy = q[i].y;
    const double qz = q[i].z;
    double px = p[i].x;
    double py = p[i].y;
    double pz = p[i].z;
    for(int j=i+1;j<N;j++){</pre>
      double dx = q[j].x - qx;
      double dy = q[j].y - qy;
      double dz = q[j].z - qz;
      double df = (dx^*dx + dy^*dy + dz^*dz);
      px += df^*dx;
      py += df^*dy;
      pz += df^*dz;
      p[j].x -= df^*dx;
      p[j].y -= df^*dy;
      p[j].z -= df^*dz;
    p[i].x = px;
    p[i].y = py;
    p[i].z = pz;
}
```

/ 0L	S₹	s <i>=</i> d) gnu" " 03		s €d) 3" " 11		.1
L 京」コンパŒ	22版	23版	son	22版	23版	son
A) s∉l((ve3≓ r	B ^{⋇1}	B ^{‰1}	\bigcirc	B ^{※2}	B ^{※2}	\bigcirc
+) s ᠽ l((2rr2y	—	-	-	B ^{⋇2}	B ^{※2}	\bigcirc
C) 配列	0	0	\bigcirc	\bigcirc	\bigcirc	\bigcirc

61.wd86620-- "calc, o2ce ~ c11" スケj r wリング結果を得られなかったため、I pトdeアo cプラcニングを適用できません。

6 2 .wd86640-- "calc,o2ce"c11" ルップ内に関数呼出しなどの最適化8 象外の命令があるため、I pトdeアo cプラcニングを適用できません。

リンク時最適化(LTO)

Link Time Optimization機能の対応状況



実効a 能: 4T6 o 用前を"1"S しR 9peed-uph



Post-FX100に向けた基盤技術の検討

- コア数、SIMD幅、問題規模の変化による並列最適化のやり直し →自動チューニングによるカーネル最適化
- 演算性能向上(20x)と通信性能向上(2x)のギャップによる通信ボトルネック
 →非同期MPI通信を活用した通信隠蔽
- メニーコア環境の複雑な階層構造における最適化指針の提示
 →修正ルーフラインモデルによる性能ボトルネックの把握
- C++によるモダンなプログラム開発
 →富士通C++コンパイラの機能・性能改善

報告書概要

- 基盤技術
 - ループ変換およびスレッド数を動的変更する自動チューニング(片桐)
 - 非ブロッキング集団通信の通信隠蔽効果に関する調査(南里)
 - アプリケーションの性能予測について(南)
 - 富士通C++コンパイラの最適化機能の改善について(渡辺、千葉、小田和)
- アプリ事例
 - JAXA LBM コードの性能測定および改善(石田、三吉)
 - MUTSU コード高速化の検討(三浦)
 - 生体分子粗視化シミュレータCafeMolのFX100での性能測定とチューニング(検 崎、渡邉)
 - ADVENTURE における多倍長精度演算の利用に向けた検討(荻野)
- Post-K性能推定
 - 核融合プラズマ乱流コードGKV のPost-FX100 性能推定(沼波、片桐、青木)
 - プラズマ流体解析コードGT5D のPost-FX100 性能推定(井戸村、内藤、三吉)
 - 宇宙プラズマ5次元ブラソフコードVlasov5の性能測定および推定(梅田、内藤)
 - N 体カーネルの性能推定(似鳥、青木)

JAXA LBM コードの性能測定および改善(石田、三吉)

- 格子ボルツマン法 (Cascaded LBM) $f_i(\mathbf{x} + \mathbf{e}_i \times dt, t + dt) = f_i(\mathbf{x}, t) + dt \times \Omega_i, \quad i = 1, \cdots, b$ $\rho \widetilde{\mathbf{M}}_{p,q,r} = \sum_i (e_{ix} - u_x)^p \cdot (e_{iy} - u_y)^q \cdot (e_{iz} - u_z)^r \cdot f_i$
 - 離散化したボルツマン方程式によって非圧縮N-S方程式のダイナミクス を再現
 - 高Re数乱流向けにCascaded LBMモデルが提案されている
 - 各格子点で27×27行列として与えられる衝突項Ω_iの計算が高コスト →従来モデル(SRT)とCascaded LBMでは5.5倍程度のコスト比

SRTとCascaded LBMの処理コスト比較(FX100)

- 最適化
 - SIMD最適化
 - 行列要素演算の手動展開
 - 配列参照順序変更による

キャッシュ再利用 →約4.2倍の性能向上により SRTとのコスト比を1.25倍に圧縮



MUTSUコード高速化の検討(三浦)

■ 擬スペクトル法磁気流体力学コードの最適化



第一変数	第二変数	第三変数
FFTx (x方向FFT)		
MPI_Alltoall (xペンシル ->yペンシル)	FFTx (x方向FFT)	
FFTy (y方向FFT)	MPI_Alltoall (xペンシル ->yペンシル)	FFTx (x方向FFT)
MPI_Alltoall (yペンシル ->zペンシル)	FFTy (y方向FFT)	MPI_Alltoall(xペンシル ->yペンシル)
FFTz (z方向FFT)	MPI_Alltoall (yペンシル ->zペンシル)	FFTy (y方向FFT)
通信と転置	FFTz (z方向FFT)	MPI_Alltoall (yペンシル ->zペンシル)
	通信と転置	FFTz (z方向FFT)
		通信と転置

FX100(16thread/proc)におけるMUTSU(20483)の性能比較

- P3dfft 2.7.5 ffte sec 通常版 many版 通常版 nproc 256 2283.016 2138.858 1092.112 1036.223 986.424 512 762.255 1024 482.646 428.720 342.041 2048 198.744 199.216 132.365 4096 88.054 89.913 86.406
- FFTの比較
 - P3dfft (+fftw) many版で一括処理
 - in-house (+ffte)
 パイプライン処理

隠蔽版

821.853

549.588

246.963

93.802

61.901

生体分子粗視化シミュレータCafeMol のFX100 での性能測定とチューニング(検崎、渡 邉)

粗視化MDにおけるカの計算カーネルの処理性能

$$V_{local} = K_b \sum_{i} (r_{i,i+1} - r_{0i,i+1})^2 + K_{\theta} \sum_{i} (\theta_i - \theta_{0i})^2 + K_{\phi}^1 \sum_{i} (1 - \cos(\phi_i - \phi_{0i})) + K_{\phi}^3 \sum_{i} (1 - \cos 3(\phi_i - \phi_{0i}))$$

■ 評価環境(1ノード利用)

Skylake(2.8TF,256GB/s,40core:4MPIx10OMP) FX100(1TF,480GB/s,32core:2MPIx16OMP)



sec	Skylake	FX100	mpi_wtime →system_clock	SIMD 最適化	性能比
bond	0.09	0.48	0.38	0.51	5.3x→4.2x
angle	0.28	1.54	1.41	1.06	5.5x→3.8x

- FX100のmpi_wtime(clock_gettimeシステムコール)をsystetm_clockに変更 cf. Post-FX100でmpi_wtimeの実装を改善予定
- SIMD最適化

スレッド内回帰参照を無くすようにループ分割を変更し!ocl norecurrenceを適用 →性能差が演算性能比2.8xにかなり近づいた ADVENTURE における多倍長精度演算の利用に向けた検討(荻野)

Precision	# of iter.	Time [hr.]
double	24,5105	5.78
pseudo-quadruple (QD)	43,598	<u>5.69</u>
quadruple (libquadmath)	38,643	14.09



Double-Double形式の4倍精度ライブラリQDが有効

富士通環境におけるDouble-Double形式ライブラリfast_dd

		性能 mflops(注1)	備考
QD	C++	68.2	インライン、SWPLのみ
	Fortran	21.9	1要素ずつ関数呼出し
fast_dd 単体関数	C++	13.1	1要素ずつ関数呼出し
	Fortran	22.2	1要素ずつ関数呼出し
fast_dd マルチ関数	C++	43.0	2要素ずつ関数呼出し
	Fortran	114.0	2要素ずつ関数呼出し
fast_ddベクトル関数	C++	805.0	SIMD+SWPL
	Fortran	874.0	SIMD+SWPL

注1) double-double形式の加算1回を1flopとしたときの性能

Post-FX100に向けたアプリ事例の検討

- エクサスケール向けCFD手法の最適化技術 →JAXA LBMコードの衝突項最適化 →基礎方程式に基づく3次元FFTのパイプライン通信隠蔽手法
- エクサスケールアプリにおける新たな性能ボトルネック →演算加速によるmpi_wtimeのコスト顕在化
- エクサスケールアプリにおける精度要求 →Double-Double形式のベクトル計算向け4倍精度ライブラリ

報告書概要

- 基盤技術
 - ループ変換およびスレッド数を動的変更する自動チューニング(片桐)
 - 非ブロッキング集団通信の通信隠蔽効果に関する調査(南里)
 - アプリケーションの性能予測について(南)
 - 富士通C++コンパイラの最適化機能の改善について(渡辺、千葉、小田和)
- アプリ事例
 - JAXA LBM コードの性能測定および改善(石田、三吉)
 - MUTSU コード高速化の検討(三浦)
 - 生体分子粗視化シミュレータCafeMolのFX100での性能測定とチューニング(検 崎、渡邉)
 - ADVENTURE における多倍長精度演算の利用に向けた検討(荻野)
- Post-K性能推定
 - 核融合プラズマ乱流コードGKV のPost-FX100 性能推定(沼波、片桐、青木)
 - プラズマ流体解析コードGT5D のPost-FX100 性能推定(井戸村、内藤、三吉)
 - 宇宙プラズマ5次元ブラソフコードVlasov5の性能測定および推定(梅田、内藤)
 - N 体カーネルの性能推定(似鳥、青木)

性能評価環境



測定環境:名大FX100+128reg/32regコンパイラ

- SPARC64XIfx: 1TFlops, 240+240GB/s, 16x2core, 256bitSIMD, HMC
- 評価単位CMG: 0.5TFlops, 240GB/s, 16core
- mpifrtpx -X9 -Kfast -Kparallel -Karray_private -Kpreex -Kopenmp -Kocl -Kmfunc=2 -Kprefetch_cache_level=all -CcdRR8 -Cpp -Koptmsg=2 -Y0,/center/fjlocal/opt /Simulation_compiler/bin -YI,/center/fjlocal/opt/Simulation_compiler/include

推定対象:ポストFX100

- A64FX: 2.7TFlops, 256x4GB/s, 12x4core, 512bitSIMD, HBM2
- 評価単位CMG: 0.675TFlops (1.35x), 256GB/s (1.06x), 12core

核融合プラズマ乱流コードGKV のPost-FX100 性能推定(沼波、片桐、青木

■ 核融合プラズマ5次元格子コードGKVの主要カーネル評価

■ literm_k:線形移流項

	レジスタ数	スレッド 数	実行時間(sec)	比率
	128reg	16	69.13	1.00
FX100 (PA关1] 时间)	32reg	16	69.09	1.01
Post-FX100 (推定值)	32reg	12	50.46	1.39

■ exb_realspcal: 擬スペクトル法による非線形移流項の実空間計算

	レジスタ数	スレッド 数	実行時間(sec)	比率
	128reg	16	4.72	1.00
FX100 (PA关11时间)	32reg	16	4.57	1.04
Deat FV100 (推宁值)	32reg	12	3.78	1.25
POSI-FA100 (推定10)	32reg-tune	12	3.10	1.54

→ループ分割により性能向上(L1Dミス率~2.65%)

プラズマ流体解析コードGT5D のPost-FX100 性能推定(井戸村、内藤、三吉

- 核融合プラズマ5次元格子コードGT5Dにおける線形ソルバGCRの主要カーネル評価
 - DAXPY

	レジスタ数	スレッド 数	実行時間(sec)	比率
[1100 (四字行時間)	128reg	16	0.32	1.00
FA100 (PA关11时间)	32reg	16	0.32	1.00
Post-FX100 (推定值)	32reg	12	0.24	1.33

メモリ機構 HMC→load/store=1/1設計 HBM2→非対称load/storeをサポート

■ SpMV:移流項の4次精度4次元差分(17ステンシル)

	レジスタ数	スレッド 数	実行時間(sec)	比率
	128reg	16	0.37	1.00
FX100 (PA実行時間)	32reg	16	0.43	0.86
	32reg-tune	16	0.38	0.98
Post-FX100 (推定值)	32reg-tune	12	0.37	1.00

→FX100ではループ分割により性能向上(L2スループット:261.6GB/s)
 →Post-FX100ではルーフラインモデルの時間比~1.15から大きく性能劣化
 L2\$帯域幅 FX100:B/F=2.0 (1TB/s) Post-FX100:B/F=1.3 (877.5GB/s)



- 宇宙プラズマ5次元格子コードVlasov5における主要カーネル評価
 - position: 実空間移流カーネル

		レジスタ数	スレッド 数	実行時間(sec)	比率
	[V100 (D4 宝行時間)	128reg	16	5.40	1.00
	FX100 (PA关门时间)	32reg	16	6.27	0.86
•	Post-FX100 (推定值)	32reg	12	7.43	0.73
		レジスタ数	スレッド 数	実行時間(sec)	比率
	[V100 (四字行時間)	128reg	16	12.39	1.00
	FX100 (PA关门时间)	32reg	16	17.97	0.69
	Post-FX100 (推定值)	32reg	12	23.31	0.53
		レジスタ数	スレッド 数	実行時間(sec)	比率
	[V100 (四字行時間)	128reg	16	11.67	1.00
	「ヘ100 (PA天门时间)	32reg	16	18.98	0.61
	Post-FX100 (推定值)	32reg	12	24.42	0.48

N 体カーネルの性能推定(似鳥、青木)

- 重力多体粒子コードにおける主要カーネル評価
 - gravity_simple: N体重力相互作用計算カーネル

	レジスタ数	スレッド 数	実行時間(sec)	比率
	128reg	16	0.153	1.00
FX100 (PA実行時間)	32reg	16	0.313	0.49
	32reg-tune	16	0.255	0.60
Post-FX100 (推定值)	32reg-tune	12	0.151	1.01

→ループ分割による性能改善は見られるが、FX100と同程度の性能

■ ループ内の命令依存関係を眺めてループ分割点を設定



Post-FX100性能推定サマリ

※Post-FX100/FX100性能比: 演算1.35x メモリ帯域1.06x

コード	カーネル	性能律速要因	性能比	最適化
CKM	literm_k	メモリ	1.39	なし
GKV	exb_realspcal	メモリ+L1	1.54	ループ分割
CTED	DAXPY	メモリ	1.33	なし
GISD	SpMV	メモリ+L2	1.00	ループ分割
Vlasov5	posititon	演算	0.73	ループ分割
	velocity_e	演算	0.53	ループ分割
	velocity_b	演算	0.48	ループ分割
重力多体	grabity_simple	演算	1.01	ループ分割

- メモリ律速アプリでは32regの影響は少ない
- HBM2における機構改善によりメモリ律速アプリの実効性能向上が期待される
- ループ分割による最適化が多くの場合に有効だが、自動最適化が今後の 課題
- ISA、out-of-order等、今回の推定で考慮されていない機能による演算律速
 アプリの性能向上が今後の課題