

FOCUS スーパーコンピュータシステムにおける 並列課金インセンティブの効果 II

西川武志^{†1}

概要： 前回の報告では単一のシステムに対し並列度が向上すると課金単価を下げるという課金インセンティブを毎年度変化させて設けることで年度毎に並列度の向上を計ることができた。本報告では A, D 二つの異なるシステムの存在下でも演算性能あたりの時間単価を基準に適切な課金インセンティブを設計することで一方のシステムに利用が大きく偏ること無く並列度を向上させることに成功したことを報告する。

キーワード： インセンティブ設計, 計算センター運用, 運用統計, 並列度向上

1. はじめに

前回の報告[1]では 2011 年度に導入された産業界向けエントリスーパーコンピュータシステム「FOCUS スパコン (ABC システム: 合計理論ピーク性能 26.8TFLOPS)」の A システム (汎用 CPU のみ, ノードあたりの倍精度浮動小数点演算性能 108GFLOPS, 合計 224 ノード) に関して 1 ジョブで複数のノードを確保して使うとノードの時間単価を割引くという並列課金インセンティブを 2014 年度まで毎年度変化させて設定することでジョブ実行時の並列度を増加させることに成功した。

2014 年 1 月から追加導入された FOCUS スパコン D, E システム (合計理論ピーク性能: 245.3TFLOPS, 汎用 CPU 51.2TFLOPS, コプロセッサ 194.1TFLOPS) の D システム (汎用 CPU のみ, ノードあたりの倍精度浮動小数点演算性能 400GFLOPS, 80 ノード) についても 2014 年度から並列課金インセンティブを導入し並列度の向上を試みた。このとき既存の A システムも同時並行に運用しておりノードあたりの理論演算性能が約 4 倍, ローカルストレージの容量が 12 倍, 入出力性能が 2 倍以上の D システムに全ての利用が移行してしまわないようにして両システムの利用率が同等に高まり, かつ並列度が向上するような並列課金インセンティブの設計が必要という前提条件であった。

複数のシステムを持つ大学の大型計算センターのジョブ統計が公表されること[2]はあっても産業界の利用ジョブ統計の公表というのはほとんど存在しなかった。その後, 2000 年代に入り文部科学省の先端大型研究施設戦略活用プログラム, 先端研究施設共用促進イノベーション創出事業 (産業戦略利用), 先端研究施設共用促進事業, 先端研究基盤共用・プラットフォーム形成事業等で地球シミュレータ, 情報基盤センターの各スーパーコンピュータ, 東京工業大学の TSUBAME の産業利用が進められたがジョブ統計の詳細はほとんど公表されなかった。

スーパーコンピュータ「京」の産業利用の統計情報が高度情報科学技術研究機構によって Web 上で公開されてい

る[3]が「京」以外のシステムは産業利用では「京」の数分の一の利用件数しかないことが示されており, 適切なインセンティブ設計をしないと一部のシステムに利用が集中してしまうということを強く示唆している。

本稿では課金インセンティブを適切に設定することで一部のシステムに利用が集中することなく, かつ並列度が年々向上させることを実現できたことを報告する。

2. FOCUS スパコンシステム

FOCUS スパコンシステムの概要については前回報告[1]に A~E, G システム, ストレージシステム等について詳細に述べているので, ここでは今回の報告の中心である A, D システムについて概要を述べる。

2.1 FOCUS スパコン A, D 概要

FOCUS スパコン A, D システムの概要は次の通りである。

(1) A システム (224 ノード)

高並列化環境 (40Gbps QDR-Infiniband 接続)

CPU : Xeon L5640 (Westmere-EP) 2.26 GHz 6 コア×2
108GFLOPS, RAM : 48GB, HDD : 500GB

(2) D システム (80 ノード)

高並列化環境 (56Gbps FDR-Infiniband 接続)

CPU : Xeon E5-2670 (Ivy-Bridge) 2.5 GHz 10 コア×2
400GFLOPS, RAM : 64GB, HDD : 6000GB

ノードあたりの倍精度浮動小数点演算性能値は D システムが A システムの 3.7 倍であり実際に複数のアプリケーションプログラムを実行させたときの性能比は D システムが A システムに対しクロック比の 1.1 倍から理論演算性能比を上回る 4.0 倍に分布した。この性能比の分布は課金インセンティブ設計が一筋縄では済まないことを示唆している。

2.2 FOCUS スパコン課金制度とインセンティブ

FOCUS スパコンの課金制度には大きく二つあり従量制と予約制がある[4]。従量制は常設キューを利用したノード数と経過時間の積にノード時間単価を乗じて利用課金する。予約制は事前に 1 日単位, 月単位, 年度単位でノード数を確保し従量制とは別の単価で課金する。

表 1 に A システムの 2011 年度から 2016 年度の表 2 に D システムの 2014 年度から 2016 年度の従量課金単価を示す。

^{†1}(公財)計算科学振興財団
Foundation for Computational Science

表1 Aシステムの従量課金単価 (2011-2016年度)

2011年度		2012年度	
ノード数	課金単価 [円]	ノード数	課金単価 [円]
1<=n<=16	100	1<=n<=8	100
17<=n<=32	80	9<=n<=16	90
33<=n	60	17<=n<=24	80
		25<=n<=32	70
		33<=n<=40	60
		41<=n	50

2013年度		2014年度	
ノード数	課金単価 [円]	ノード数	課金単価 [円]
n=1	100	n=1	100
n=2	97.5	2<=n<=4	95
3<=n<=4	95	5<=n<=8	90
5<=n<=6	92.5	9<=n<=16	85
7<=n<=8	90	17<=n<=32	80
9<=n<=12	88	33<=n<=48	75
13<=n<=16	86	49<=n<=64	70
17<=n<=20	84	65<=n<=80	65
21<=n<=24	82	81<=n<=96	60
25<=n<=44	80	97<=n<=112	55
45<=n<=64	70	113<=n	50
65<=n<=84	60		
85<=n	50		

2015年度		2016年度	
ノード数	課金単価 [円]	ノード数	課金単価 [円]
1<=n<=8	100	1<=n<=16	100
9<=n<=16	90	17<=n<=32	90
17<=n<=24	80	33<=n<=48	80
25<=n<=32	70	49<=n<=64	70
33<=n<=40	60	65<=n<=80	60
41<=n	50	81<=n	50

表2 Dシステムの従量課金単価 (2014-2016年度)

2014年度		2016年度	
ノード数	課金単価 [円]	ノード数	課金単価 [円]
1<=n<=80	300	1<=n<=8	300
		9<=n<=16	285
		17<=n<=24	270
		25<=n<=32	255
		33<=n<=40	240
		41<=n<=48	225
		49<=n<=56	210
		57<=n<=64	195
		65<=n<=72	180
		73<=n<=80	165

2015年度	
ノード数	課金単価 [円]
1<=n<=32	300
33<=n<=40	285
41<=n<=48	270
49<=n<=54	255
65<=n<=80	240

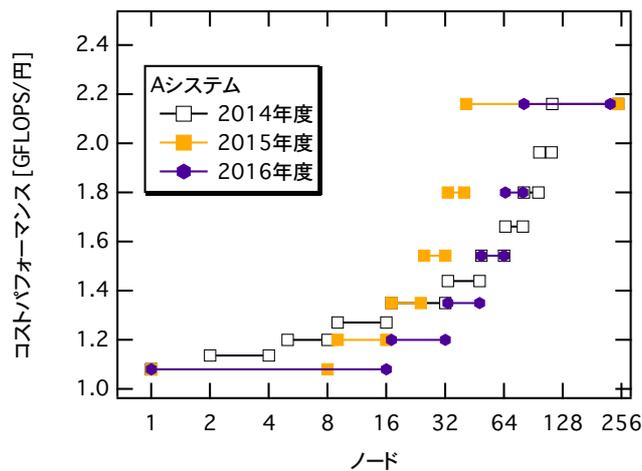


図1 Aシステムのコストパフォーマンス

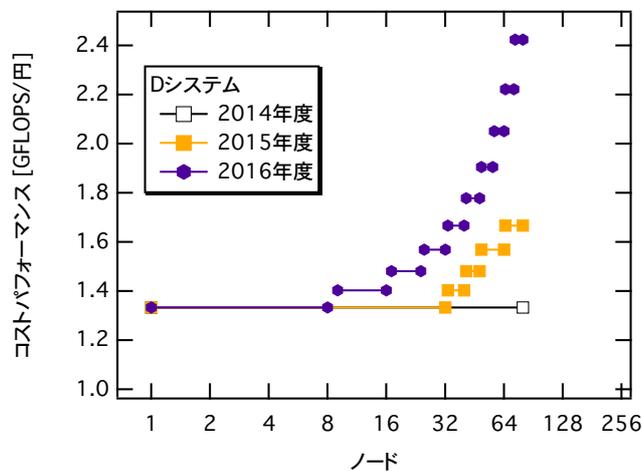


図2 Dシステムのコストパフォーマンス

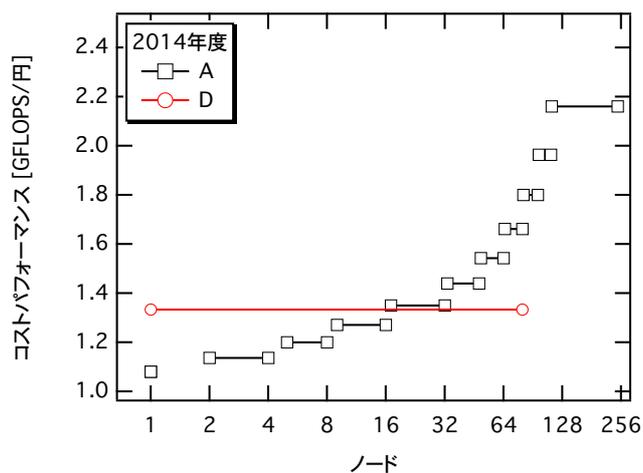


図3 2014年度のA, Dシステムのコストパフォーマンス

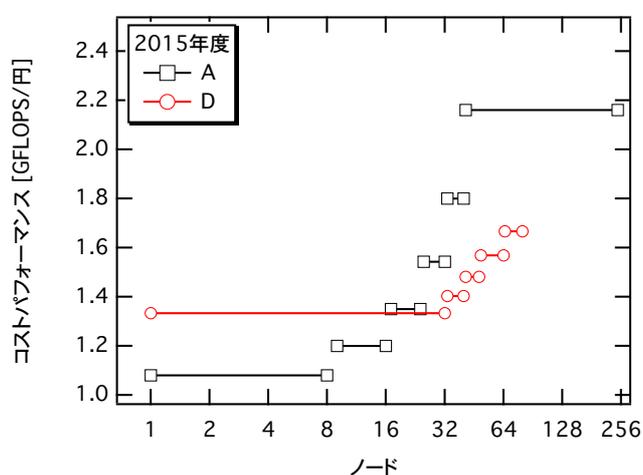


図4 2015年度のA, Dシステムのコストパフォーマンス

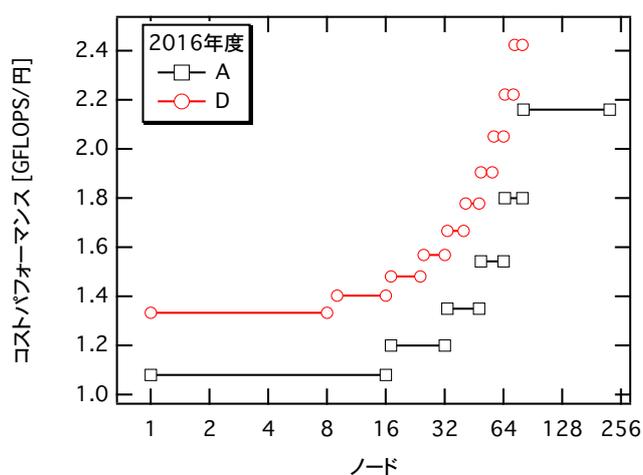


図5 2016年度のA, Dシステムのコストパフォーマンス

従量課金におけるノードあたりの倍精度浮動小数点演算ピーク性能をノード時間あたりの利用単価で除したコストパフォーマンス

トパフォーマンスについて図1にAシステム図2にDシステムを示す。さらにA, Dシステムのコストパフォーマンスを2014年度から2016年度まで比較したものをそれぞれ図3から5に示す。

2014年度から2016年度までの課金インセンティブの設計方針は以下のとおり。

- Aシステムでは前年度に対し翌年度は同じノード数であれば課金単価が上昇、かつ単価が変化するノード数の単位を8ノードから16ノードに拡大
- DシステムではAシステムとは逆に年度が進むにつれ課金単価を低下させ、かつ単価が変化するノード数の単位を8ノードとAシステムより少なくする

結果としてAシステムのDシステムに対するコストパフォーマンス優位性は、図3, 4に示すように2014, 2015年度は16ノード以上に存在するが2014年度から2015年度には減少し図5に示すように2016年度では消滅させた。もちろん理論ピーク性能に基づくものでありアプリケーションによっては2016年度でもAシステムがDシステムに対しコストパフォーマンスが優位な場合も存在しうる。

同様な課金インセンティブは従量制と同様に1日単位の予約制にも同様に適用し月単位と年度単位の予約制には適用していない。

2.3 FOCUS スパコン利用概況

表3に2011年度から2016年度(10月末)までのFOCUSスパコンの利用法人数、課題数、全体、A, Dシステムとの供給と利用された計算資源量とその割合を示す。

2011年度から2016年度まで法人数・課題数は共に年率約2割の増加となっている。Dシステムの利用率も2014年度から2016年に年率約2割の増加となっている。Aシステムの利用率は2014年度から本格稼働したDシステムの影響で2014年度に一旦低下したが2015年度、2016年度と利用が回復している。全体利用率は下がっているように見えるが全体の供給計算資源量が増えており、利用計算資源量も増えている。

表3 FOCUS スパコン利用概況

	2011年度	2012年度	2013年度	2014年度	2015年度	2016年度
法人数	58	98	129	143	160	147
利用課題数	64	108	148	168	197	185
全体供給計算資源量	8.8 PF日	9.6 PF日	16 PF日	28 PF日	28 PF日	34 PF日
A供給計算資源量	8.2 PF日	8.8 PF日				
D供給計算資源量				12 PF日	12 PF日	12 PF日
全体利用計算資源量	6.0 PF日	5.8 PF日	11 PF日	13 PF日	18 PF日	12 PF日
A利用計算資源量	5.6 PF日	5.6 PF日	7.2 PF日	4.6 PF日	6.5 PF日	5.9 PF日
D利用計算資源量				4.0 PF日	5.9 PF日	6.6 PF日
全体利用率	68%	61%	68%	45%	64%	53%
A利用率	68%	63%	82%	52%	74%	67%
D利用率				34%	49%	55%

3. 課金インセンティブの効果

ノード数が増えると課金単価を減少させる課金インセン

タイプの効果について利用結果を示し議論する。課金インセンティブは従量制と1日単位の予約制に適用し、月単位と年度単位の予約制には適用しない結果、適用した制度の並列度を上げる効果があった。課金インセンティブを設定した従量制と1日単位の予約制でも特に従量制の24時間の実行経過制限のあるキューの利用で強く現れた。以下では24時間キューでの利用に絞って議論する。

3.1 利用率分布

図6にAシステム、図7にDシステムの2014年度から2016年度(10月末)までの利用供給計算資源量に対するノード数毎の利用率分布を示す。

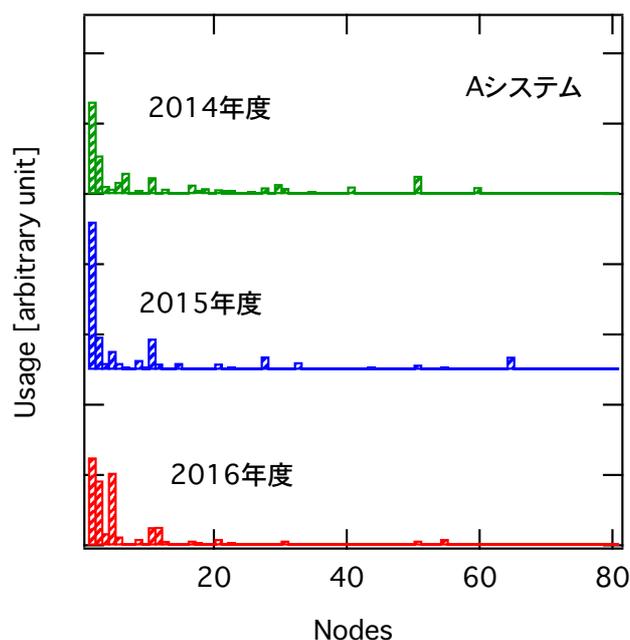


図6 Aシステムの利用分布

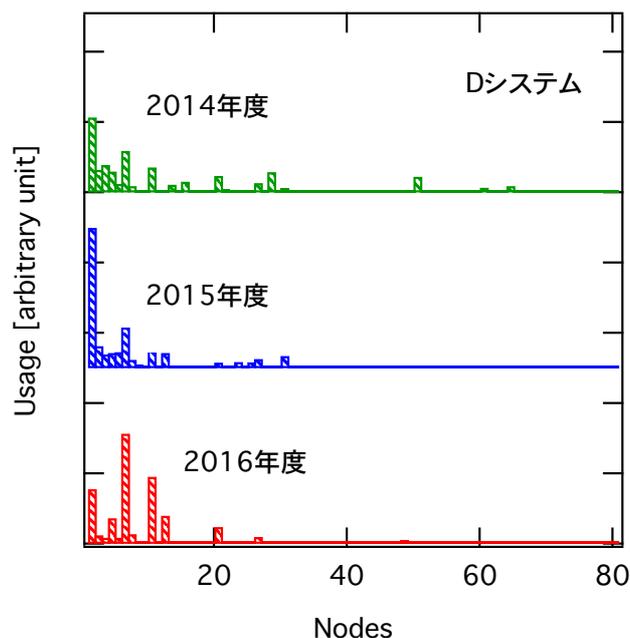


図7 Dシステムの利用分布

A, Dシステム共に40ノードを超えるような大規模並列の利用は少数であり課金インセンティブの影響よりも大規模並列を利用する課題が存在するか否かの影響の方が大きいことが判明した。一方、40ノード以下の領域では課金インセンティブの影響を強く受け課金単価が下がる境目である8ノードや16ノードで利用が増えており課金インセンティブが有効に働いている。

3.2 課題毎利用ノード時間分布

図6, 図7に示した利用率分布では全課題に対する分布を示したため、当該年度に新規に利用を開始する課題の利用ノード数が低い並列度に停滞する影響を強く受けている。そこで図8にA, Dシステムの24時間キューにおける課題毎、年度毎の利用ノード時間のノード数に対する分布および全課題の平均ノード数に対する合計ノード時間を示す。

全課題の平均ノード数は2014年度がAシステム6.8, Dシステム3.1, 2015年度がAシステム2.8, Dシステム1.9, 2016年度がAシステム3.3, Dシステム2.7となった。Aシステムがノードあたり12コア, Dシステムがノードあたり20コアであり、全てのコアを利用したと仮定してコア数に換算し直すと2014年度がAシステム82, Dシステム62, 2015年度がAシステム34, Dシステム38, 2016年度がAシステム40, Dシステム54となった。2015年度からDシステムが平均利用コア数でAシステムを上回ったことがわかった。

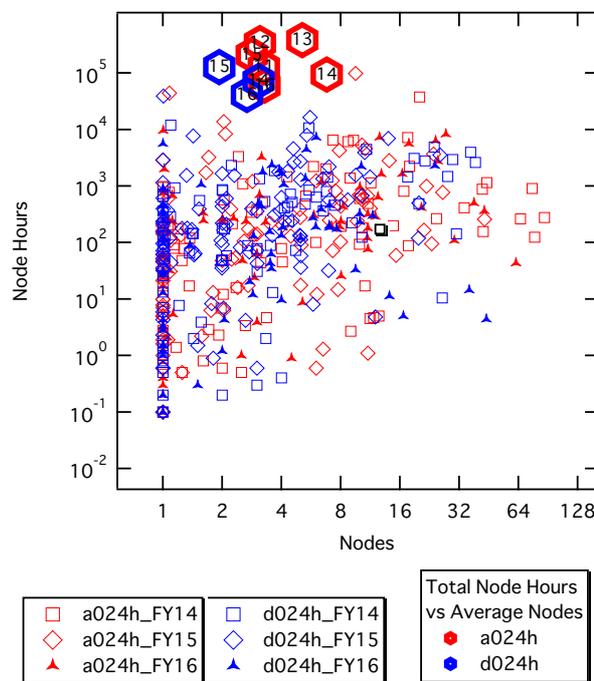


図8 A, Dシステムの24時間キューにおける課題毎、年度毎の利用ノード時間のノード数に対する分布

課題毎の利用状況に着目すると8ノード以下、特に1ノードに多数の利用課題が集中する様子が見られる。これは当該年度に新規に利用を開始する課題が少ないノード数か

ら始まって評価が進むに従ってノード数を増やしているという利用パターンということ、他方で8ノード以上では年度が進むにつれて利用ノード数および利用ノード時間が増加する傾向がみられることが個別課題を詳細に追うことで判明した。このように、A、Dシステムで年度が進むにつれて平均並列度が向上していることから課金インセンティブの効果が現れていると判断できた。

3.3 A、Dシステム間の利用遷移

個々のシステムにおいて課金インセンティブが有効に働いていることは判明したが、新しいシステムDを導入したことでAシステムの利用がどのように変化したかを調べるため図9に24時間キューにおける全課題の平均ノード数に対する月あたり利用ノード時間をAシステムはDシステム導入前の2011年度から、Dシステムは本格稼働開始の2014年度から、またDシステムとAシステムのノードあたり性能比(400GFLOPS/108GFLOPS)をDシステムの月あたり利用ノード時間に乗じたものを(修正Dシステム月あたり利用ノード時間)をプロットした。

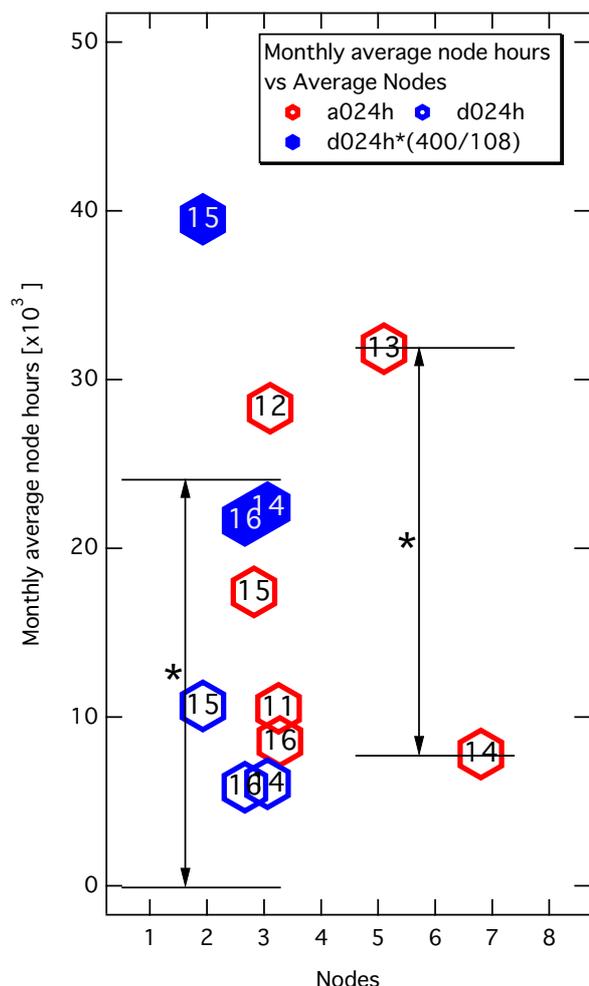


図9 A、Dシステムの24時間キューにおける全課題の平均ノード数に対する月あたり利用ノード時間

図9では2013年度から2014年度にかけてAシステムの

月あたり利用ノード時間は大きく減少(*)した。その減少分と2014年度の修正Dシステム月あたり利用ノード時間に相当する結果となった。そして2015年度、2016年度と修正Dシステム月あたり利用ノード時間はAシステムの月あたり利用ノード時間を上回りAシステムからDシステムに利用が遷移することを促進する課金インセンティブは適切に働いていると言える。

4. まとめ

本稿ではA、D二つの異なるシステムの存在下でも演算性能あたりの時間単価を基準に適切な課金インセンティブを設計することで一方のシステムに利用が大きく偏ること無く並列度を向上させることを示した。

今後の課題として FOCUS スーパーコンピュータシステムの計算資源利用量は下半期が上半期の約2倍という傾向を例年示すため本稿での2016年度は10月末までの実績で下半期1ヶ月分しか反映していない。さらに FOCUS スーパーコンピュータシステムは Broadwell 世代の Intel Xeon CPU を搭載し1ノードあたり汎用 CPU 倍精度浮動小数点演算性能が、1,152GFLOPS の F システム (ノード時間単価 500 円固定) 12 ノード、205GFLOPS の H システム (ノード時間単価 100 円固定) 68 ノードの運用を2016年10月から開始したため、利用状況が大きく変化することが予想される。F、H システムにも並列課金インセンティブを2017年度から導入し4システムが並行に運用され並列度が高くなると課金単価が低下するという複雑なものになり利用者が自分の問題解決に最もコストパフォーマンスが良いシステムはどれなのか、時間が最も短いものはどれなのか直感的にわからない状況となる。よりわかりやすい並列課金インセンティブの説明を提示し所望の運用結果をこれからも得られることを目指す。FOCUS スーパーコンピュータシステムの運用に関して使いやすくするための改良を加えていくつもりであるため運用に関する要望や意見を、是非 request@j-focus.or.jp までお寄せいただきたい。

謝辞 FOCUS スーパーコンピュータシステムの運用や利用者の開拓に尽力されている計算科学振興財団の同僚と利用してくださっている利用者各位に、謹んで感謝の意を表する。

参考文献

- [1] FOCUS スーパーコンピュータシステムにおける並列課金インセンティブの効果, 西川 武志, 研究報告ハイパフォーマンスコンピューティング (HPC) ,2015-HPC-149(2),1-4 (2015-06-19).
- [2] 石田晴久,高橋延匡,斎藤五郎:東京大学大型計算センターにおけるジョブ統計,情報処理,Vol.13,No.10,pp.714-719 (1972).
- [3] 高度情報科学技術研究機構, 産業利用に関する統計情報, http://www.hpci-office.jp/pages/toukei?parent_folder=#toukei_4.
- [4] FOCUS スーパーコンピュータシステム利用料金詳細, <http://www.j-focus.or.jp/focus/fee.html>