3.5世代PCクラスタを中核とする理研RICC：その狙いと現状、今後

姫野 龍太郎
理化学研究所・情報基盤センター

[アブストラクト]

PC クラスタのこれまでの発展は、第一世代：個人のための高性能計算機システムとして生まれ使われた、数十台から数百台規模の時代、第二世代：多くのユーザーを対象とした高性能計算プラットフォームとなり、数千台の規模での運用に使われるようになった時代、第三世代：要素としての PC が 2G のメモリーの制約を超え、マルチコア化し、一万のオーダーとなるコア数のシステムに発展。この第三世代のクラスタに計算用アクセラレータを組み込んで演算性能を高めたシステムを第 3.5 世代と分類した。理研が 8 月から運用を開始した RICC(RIKEN Integrated Cluster of Clusters)は 100 台の GPGPU ボードをアクセラレータとして備え、全体で約 9000 コアを持つシステムとなっている。次世代スーパーコンピュータの稼働を見据え、そのソフトウェア開発プラットフォームとして万に迫るオーダーでの高並列計算のテスト環境を整備すると共に、価格演算性能比と電力演算性能比の優れたアクセラレータの利用普及を行うことが、このシステムの狙いである。運用状況に関しては現在集計中で、当日報告するとともに、その拡張計画に関しても言及する。

[キーワード]

PC クラスタ、アクセラレータ、高性能計算、高並列計算、GPGPU

1. はじめに

日本で初めて大規模な PC クラスタをスーパーコンピュータとしてセンター運用をした RSCC は今年 6 月に 5 年間の運用を終え、RICC にリプレースした。この RSCC や RICC を PC クラスタのこれまでの発展から考えると、次のように分類できる。第一世代：個人のための高性能計算機システムとして生まれ使われた、数十台から数百台規模の時代、第二世代：多くのユーザーを対象とした高性能計算プラットフォームとなり、数千台の規模での運用に使われるようになった時代（RSCC はこの世代）、第三世代：要素としての PC が 2G のメモリーの制約を超え、マルチコア化し、一万のオーダーとなるコア数のシステムに発展。この第三世代のクラスタに計算用アクセラレータを組み込んで演算性能を高めたシステムを第 3.5 世代（RICC はこの世代）と分類した。RICC は 8 月から稼働を始め、現在 3 ヶ月が経過した。そこで、RSCC の 5 年間を振り返るとともに、新しい RICC の狙いとその稼働状況、現状の問題点を報告する。

2. RSCC

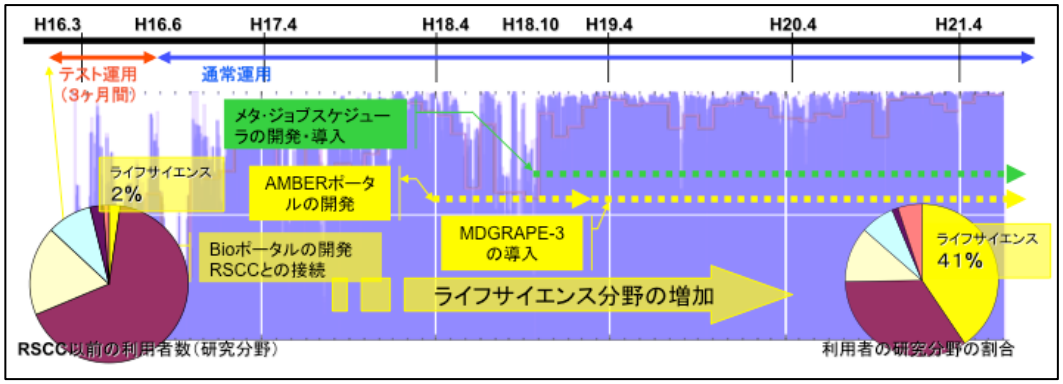
RSCC は PC クラスタ、ベクトル並列コンピュータ、分子動力学計算の専用計算機の 3 種類を複合した計算機システムであった。演算性能の主力には PC クラスタを採用した。これには、次のような狙いがあった。

1） 理研の研究分野が物理・工学から生命科学に広がってきたことに伴い、生命科学の分野の研究者にも使える計算機システムとすること、
2） これまでスーパーコンピュータの利用がほとんどなかった実験研究者にもリソースを提供すること

しかし、この先代のスーパーコンピュータがベクトル並列計算機 VPP700 であったため、従来からの利用者がすぐに PC クラスタに移行できないことを考慮し、ベクトル並列機も用意した。また、価格性能比に優れた専用計算機の可能性をテストするために専用計算機をシステムの中に組み込んだ。

5 年間の稼働の中で、システムの改良を次の 2 点で行った。

a. PC クラスタは全体で 2048 ノードであったが、これはジョブクラスを意識して全体を 5 つのサブシステムに分割していた。これらのサブシステムごとにジョブのキューイングシステムが独立して制御を行っていたため、空いているサブ



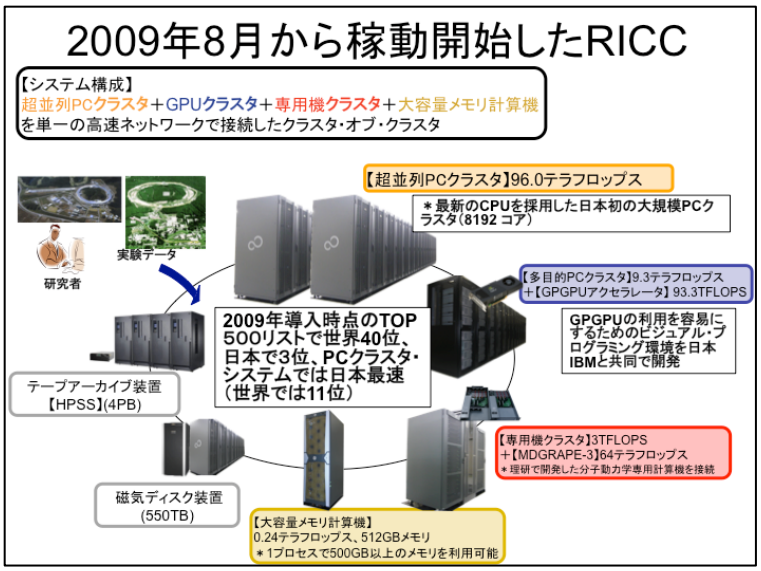
システムがあってもジョブの待ちが長いサブシステムも存在した。そこで、メタスケジューラを富士通と共同で開発、待ち時間の短縮と稼働率の向上に貢献した。

b. 当初 MD-GRAPE2 の小規模な専用計算機であったが、市販アプリに組み込むことで利用が一般化したので、MD-GRAPE3(理論ピーク性能 64TFLOPS)に増強した。５年間の稼働を振り返ってみると、当初の狙いはほぼ達成された。当初心配したような利用者のソフトウェアの移行も円滑に進み、導入当初から稼働 PC クラスタの利用率は高く、稼働後３年目で稼働率は９０％に達した。一方でベクトル並列機の稼働率はあまり高くなく、この５年で、その役目を終えた。特筆すべきは、当初の狙い通り、生命科学や実験研究者という新たな研究者に利用を広めるという狙いは十分に達成されたことである。

3. RICC

RSCC を更に発展させるとともに、次世代スーパーコンピュータの稼働に合わせてソフトウェアを開発している研究者に開発のプラットフォームを提供することを大きな目標としている。このため、RSCC のコンセプトは維持しつつ、最新のコンピュータに置き換えるとともに、

1) 大規模並列のテストに必須の、8000 コアを超えるジョブが実行できる環境を作る
2) 新しい HPC の方向性であるプログラム可能なアクセラレータ(GPGPU)を導入、利用を促進する



という二つのことを狙った。具体的には、１）の実現のため、RSCC の PC クラスタが分割されていて、それらのノードを集めて実行することができなかったのを反省し、RICC の超並列 PC クラスタと多目的 PC クラスタは同一のインターコネクトで結合、全システムを占有すると 8992 コアでの並列実行を実現できるハードウェア設計になっている。また、ジョブスケジューラの機能を強化し、日常的に数千並列のジョブが処理できる見込みである。２）の GPGPU の利用促進のため、CUDA 以外に、ディレクティブによって GPGPU が利用できる PGI コンパイラーの提供をはじめ、日本 IBM と共同で、RIVER(GUI によるプログラミング環境)の開発を行っている。GPGPU は非常に価格性能比が良い上、電力性能比も良い。ただ、ソフトウェアの開発環境が良くないので、今後の HPC の性能向上の上で避けて通れないヘテロなプログラミングの研究開発を情報基盤センターとして行って行く。

4. 稼働状況

８月からテスト運用を行い、１０月から本運用に移行した。この間の稼働率を見ると、テスト運用でも開始２週間で９５％を超える稼働率を記録、RSCC と比べると際立った差が見られる。既に利用者は PC クラスタの利用に関して十分な経験と知識を有し、新システムを待っていたことが見て取れる。今後はこれらのユーザーのうち対応可能なアプリに対して GPGPU を利用するよう誘導する予定である。