加速器実験データ処理の紹介と大規模PCクラスタ

四日市 悟 (理化学研究所 放射線研究室)

放射線研究室では、米国ブルックヘブン国立研究所 (BNL) における RHIC/PHENIX 実験 [1, 2] をはじめとした粒子加速器を用いた実験により、原子核物理学/ハドロン物理学の研究をおこなっている。PHINIX 実験のデータ解析のため、1999 年に RIKEN–CCJ (computing center in Japan) という PC クラスタによる計算センターの建設を開始し、2000 年より同実験の計算センターとして重要な役割を果たしてきた。ここでは、最近の加速器実験のデータ処理の流れと、とくに PC クラスタがどのような役割を果たしているかを紹介し、また CCJ を用いて得たこれまでの成果と将来の展望について述べる。

1 加速器実験のデータ処理の特徴

加速器を用いた原子核/素粒子物理学実験における測定データは、固定標的実験/衝突型実験にかかわらず、現在ではほぼすべて電子的に記録される。測定器には それぞれの原子核/素粒子反応について反応点からの粒子 (群) の飛跡を検出するもの、エネルギー、速度、時間を測定するものなどがある。図 1 に PHENIX 実験の測定器群を示した。同測定器群の channel 数は合計およそ 30 万である。これらの測定器からの波高情報あるいはタイミング情報は汎用/専用の回路モジュールによってデジタイズされ、計算機に送られて、disk または tape 装置に記録される。測定されたひとつひとつの原子核/素粒子反応のことをイベントと呼ぶ。各種の測定器からの情報を組み合わせて、それぞれの粒子を同定し、またそれら粒子の運動量ベクトルやエネルギーの決定をおこなって、どのような反応であったかを明らかにすることをイベント再構成と呼ぶ。こうしたイベント情報を大量に集積して、はじめて必要な物理的情報— たとえば粒子の生成断面積など — を得ることができる。

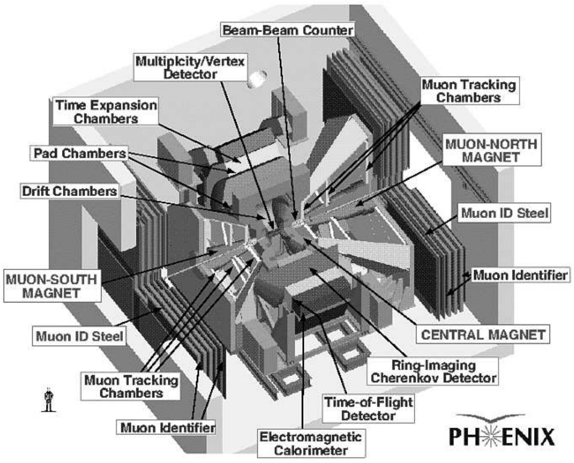

図 1: PHENIX 測定器

計算機による実験データ処理はオンラインとオフラインの 2 種類に大別される。オンライン処理とは、実験中に実時間で—しばしば粗いやり方で—データを処理することを指し、目的はその測定データを記録すべきかどうかそのものを判定したり (トリガー)、測定器の稼働状況をモニターしたりすることである。実時間処理のためトリガーの初期段階では論理回路モジュールの組み合わせあるいは FPGA を用いた論理判定がおこなわれるが、後段では PC クラスタによるイベント再構成情報を判定につかうことがある。また、各種測定器から独立に送られてくる情報を、記録のためにイベント毎に並べなおすこと (イベントビルド) にも PC クラスタが使われることが多くなってきた。

一方、オフライン処理とはこのようにして記録された測定データを後日 (といっても実時間ではないというだけで、実験中からこの種の作業はおこなわれている) まとめて処理することを指す。CCJ が活躍するのはこの段階である。

記録された段階のデータ (raw data) は、一般にデジタイズ回路から出てきたままの byte 列であるが、これをあとで述べる較正情報を使って測定器

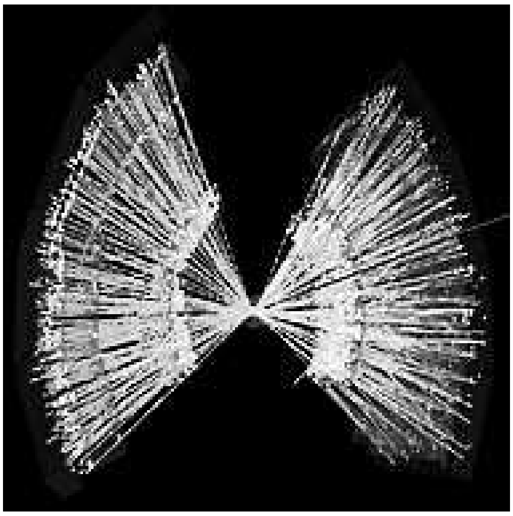

図 2: PHENIX 実験での飛跡再構成の一例

ごとに解析をおこないやすいような位置、時間、エネルギーなど物理的情報に変換 — 例えば 飛跡検出器の hit channel 番号情報を実空間の座標値に置き換えるなど—し、また飛跡再構成などをおこなってその情報を付け加えていく。ちなみに、データ処理のうち、一般にもっとも CPU パワーを必要とするのはこの飛跡再構成である (図 2)。こうして生成されたデータを DST(data summary tape) といい、解析が進行するにしたがってイベントごとに高次の情報が付け加わり、また不要な情報が削られ、場合によっては不要なイベントを削るなどの処理がおこなわれていく。解析段階がすすむに従いサイズを小さくしていくため、各段階を DST/μ(マイクロ)DST/n(ナノ)DST などと呼んだりする。ある段階までは 実験コラボレーションの責任で処理が一括しておこなわれ、公式データが作成されていくが、データサイズがだんだん小さくなってくると、徐々に各人の物理的興味や作業の必要に応じて— 例えば μ 粒子対により J/ψ 粒子を観測したい場合や、μ 粒子測定器の較正をおこなう場合は、μ 粒子がまったく含まれないイベントは捨ててしまう— さまざまな様式の DST がつくられていき、それを用いて効率よく物理の結果を得る。

このデータ処理の流れの一部として、または並行しておこなわれる作業として、キャリブレーション (較正) とシミュレーションにふれておかねばならない。測定器キャリブレーションは、ビーム実験そのものとは別にそれ専用の実験によって取得したデータでおこなわれる場合もあるが、多くは実験データそのものを使用しておこなわれる。この作業には、すべての測定器情報を含む再構成されたイベントデータは必ずしも必要なく、それ自身または少数の測定器の組み合わせでおこなうことができることも多い。この較正情報はイベント再構成に必要であり、また再構成されたイベント情報を用いてさらに較正の精度をあげるなど、数度の繰り返しがおこなわれることが多い。この過程で測定器の精度や検出効率が評価されていく。

シミュレーションはさらに物理シミュレーションと測定器シミュレーションに分かれる。物理シミュレーションでは例えば陽子衝突におけるクォークとジェットのふるまいや、重イオン衝突における核子のふるまいをシミュレートして最終的にどんな粒子がどのようにどれだけ出てくるのかを計算する。測定器シミュレーションは粒子と測定器の物質との相互作用をシミュレートして、測定器そのものによる測定結果のゆがみを評価したり、また、疑似データを生成して、解析方法の性能を評価したりするためにおこなう。たとえば測定器中を通過する粒子飛跡が物質との相互作用で散乱される効果がどれだけ運動量測定精度を悪化させるか、大量に粒子飛跡がある場合にどの程度の確率で正しい飛跡を再構成できるか、などである。

実験データの解析にせよ シミュレーションにせよ、処理をイベント単位まで分割できることが加速器実験データ処理のひとつの特徴で、計算時にイベント間の相関を気にする必要がなく、PC クラスタを用いた並行処理に向いている。イベント単位に分割することで—処理すべきイベント数は百万単位である — 原理的には一度にいくらでもたくさんの CPU を使用することができる。計算のために CPU 間で通信する必要はないので能率は使える CPU の数に比例するが、解析の段階によってはデータストレージ (disk あるいは tape) との入出力が制限要素となる場合がある。実際には、ある程度の数のイベントをまとめたファイルが入出力時の単位となる。PHENIX 実験の raw data では、1 イベントが 300KB 程度で、1 ファイルあたり 1.5GB 程度になるよう作成しているが、DST/μDST/nDST はそれより小さい。ファイルサイズが小さいことはテープからの読み出しや ネットワーク転送などの入出力のオーバーヘッ

ドが大きいことを意味する。

PHENIX 実験は現在 100MB/s のデータ記録速度をもち、一年あたり約半年のビームタイムで 約 88 TB (2002 年から 2003 年にかけておこなわれた Run-3 での実績) の raw data を記録しているが、加速器や測定器性能の向上とともに年間データ量は増える傾向にある。また、2007 年開始予定の CERN LHC 実験においては、実験グループあたり年間 1PB に及ぶデータが記録される予定で、また 2GB/s におよぶ記録速度が必要とされる実験グループもあり、GRID を用いたデータ記録、分配および解析のための技術開発が進行中である。

2　CCJ でのデータ処理

CCJ の現在の構成は 図 3 のようになっている。計算 node は 166 台の PC で、700MHz から 2GHz の PentiumIII/4 の dual CPU 構成、node あたりメモリは 1GB、local disk は 10〜30GB をもち、BNL に設置された、PHENIX 実験の中央計算機センターである RCF(RHIC Computing Facility)[4] にあわせて Red Hat Linux 7.2 で運用されている。データストレージとしては HPSS (High performance storage system) [5] と、RAID を併用している。HPSS は STK のテープロボットとその中の 6 台の 9940B テープドライブ、および 8TB の disk cache を制御し、2003 年 10 月現在で 150 TB(45 万 file) の実験データが格納されている。約 30 TB の RAID disk の上には実験データのほか ユーザのホーム領域や解析ライブラリなどがおかれ、5 台の SUN(Solaris) を NFS データサーバとして計算 node からアクセスできるようになっている。データサーバ同士と HPSS の間は 1000Base-SX, さらに計算 node とは 100Base-T で接続されている。

PHENIX 実験データは RCF では HPSS に格納されており、一部は NFS サーバ上にある。CCJ で必要なデータは、RCF からネットワーク経由で転送するかまたは磁気テープにおとして航空便で輸送され、最終的には CCJ でも HPSS に格納される。Run-3 の 88 TB (62000 file) の raw data のうち、すでに約 40TB はテープ輸送され、解析がすすんでいる。また約 9 TB の Run-2 の μDST などがこの一年間にネットワーク経由で輸送された。典型的には 3〜4MB/s での転送が日米間のネットワークでも可能であった。

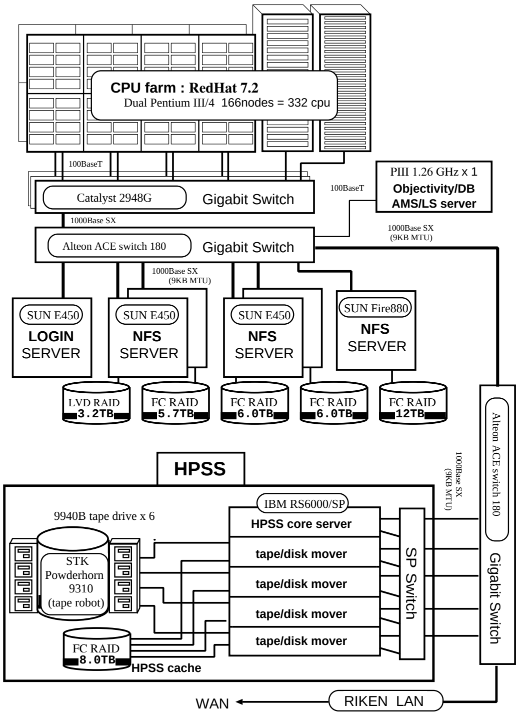

図 3: CCJ の構成 (2003 年 11 月現在)

PHENIX 実験のデータ処理のうち、CCJ でおこなわれる計算は 1) raw data から DST への変換 (イベント再構成を含む)、2) DST から μ/nDST への変換、3) μ/nDST の解析、4) シミュレーションによる疑似データ作成、等である。このうち 3) と 4) はデータ量が比較的小さいためコラボレーション内各大学等の計算センターでも実行できるが、1) と 2) は大量のデータを扱うため、RCF の他には HPSS をもつ CCJ でしか実行できない。そのため公式データ生成の何割かを CCJ が受け持つことが多い。

また、1) と 4) は CPU を大量に消費する job であるが、2) と 3) については入出力がボトルネックとなることが多い。μ/nDST はいまのところ全体のサイズが小さいので NFS サーバ上に置いてあることが多いが、各計算 node 数十台から同時に NFS 経由でサーバ上のデータにアクセスすると総アクセス速度は 10MB/s 程度に低下してしまう。このため サーバー台あたりの最大数を限った rcp コマンドを使って、データをいったん計算 node の local disk に転送してから読み書きするという方法でサーバー台あたり 50MB/s 程度のアクセス速度を得ている。

計算 node での batch queueing システムとして LSF を運用し、CPU 時間で制限される 3 つの汎用 queue(*short/long/bg*、それぞれ 200 min./1 day/7 days) と、プライオリティの高い公式データ生成 job 用 *production* , 大きなメモリやワークエリアを必要とする job のための *largedisk* などいくつかの特殊 queue を設定し、状況に応じて queue ごとの最大同時実行数を変更するなどしている。また、queue にたまった job の実行順序は 'first-in first-served' 原則ではなく、LSF の 'fair share' 機能を用いて 各 user の CPU 使用履歴を考慮にいれた順序の制御がおこなわれている。

理研の新クラスタシステムの計算力は CCJ の十数倍であり、仮に 1/10 使えたとしても計算力は倍以上にふえる。また、HPSS の増強により I/O 性能の向上も期待され、今後、より増加していく PHENIX 実験からの大量のデータをいちはやく解析するための大きな武器となるであろう。

3 これまでの成果と展望

RHIC/PHENIX では 重イオン衝突実験が 2000 年に開始され、重心系エネルギー ($\sqrt{s}$) = 130 GeV および 200 GeV の Au–Au 衝突実験と 200 GeV の d–Au 衝突実験がおこなわれた。クォーク ·グルーオン ·プラズマ (QGP) の検出をめざしていろいろな側面からのデータ解析がおこなわれており、すでに出版された論文は 15 本 (さらに投稿中 6 本)[6] 、国際会議報告は 100 本以上にのぼるが、なかでも Au–Au 衝突で高 P_T(横運動量) 粒子の抑制現象が検出されたことは、RHIC における QGP の生成について肯定的な証拠となっている [7]。

放射線研究室が主力を投入している PHENIX における偏極陽子衝突実験は、陽子のスピンを指標とした核子のクォーク ·グルーオン構造の解明を目指しておこなわれている。電子を用いた深部非弾性散乱実験で判明した、陽子のスピンが 3 つのクォークのスピンの和だけでは説明できないという問題 ('spin crisis') を解決するためには、陽子の中のグルーオン偏極を測定できる偏極陽子衝突実験が不可欠である。世界初の偏極ビーム衝突型加速器ということで運転も困難をきわめている中、Run-3 の $\sqrt{s} = 200$ GeV のデータはいちはやく CCJ で解析され、π^0 粒子生成の非対称度 A_{LL} 測定の予備的な結果が得られている (図 5)。低い P_T でマイナスに出ているこの測定結果は、誤差範囲が大きいながらも従来の理論では説明できない値であり、それだけに解釈は難しいが、陽子中のクォークのスピンにさらにグルーオンのスピンの影響を加えても陽子のスピンを説明できない可能性を示唆しており、大きな話題を呼んでいる。次のビームタイムにおけるデータの蓄積が待ち望まれている。

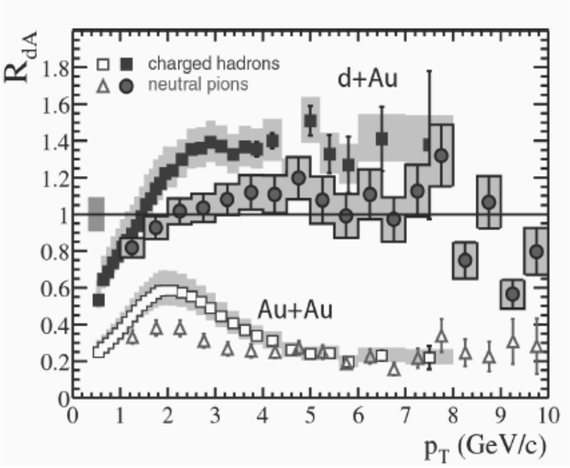

図 4: Au–Au 衝突での 高 P_T 粒子の抑制現象

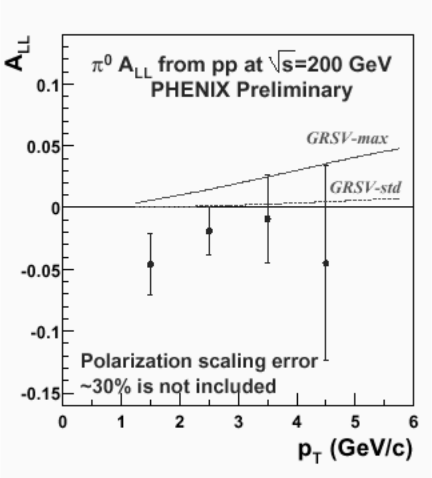

図 5: π^0 粒子生成の非対称度 A_{LL}

その他、KEK–PS では 原子核中でのベクトル中間子崩壊の測定によるカイラル対称性の回復の検出をめざした実験が 2002 年に終了し、CCJ を用いたデータ解析の結果、有限密度環境におけるカイラル対称性の部分的回復の証拠と解釈できる質量分布の変化が検出された [8]。

RHIC では この 2003 年 11 月から Run-4 とよばれる約半年間のビームタイムがはじまり、$\sqrt{s} = 200$ GeV の Au-Au 衝突実験と、陽子偏極度増大のための準備実験がおこなわれる。さらに、半年程度の休みをおいて、Run-5 という偏極陽子衝突実験専用のビームタイムが計画されている。Au-Au 衝突では、QGP 生成のさらなる証拠となる電子対/μ 粒子対測定による J/ψ 抑制現象や、ベクトル中間子の質量変化の検出が期待されている。偏極陽子衝突実験では、次のビームタイムに π^0 粒子生成の非対称度については決着をつけ、また今後 2~3 年のあいだには新測定器の投入などにより陽子中のグルーオン偏極に関する他の channel についても測定ができるようになり、核子スピンの問題を解決しようという計画である。

参考文献

[1] `http://www.bnl.gov/rhic`

[2] `http://www.phenix.bnl.gov`

[3] `http://ccjsun.riken.go.jp/ccj/`
S. Yokkaichi et al.: RIKEN Accel. Prog. Rep. 37
Y. Watanabe et al.: RIKEN Accel. Prog. Rep. 36 (2003) 262

[4] `http://www.rhic.bnl.gov/RCF/`

[5] `http://www.sdsc.edu/hpss/`

[6] `http://www.phenix.bnl.gov/papers.html`

[7] `nucl-ex/0308006`, to be published in Phys. Rev. C
Phys. Rev. Lett. 91 (2003) 072303
Phys. Rev. Lett. 88 (2002) 022301

[8] K. Ozawa *et al.*, Phys. Rev. Lett. 86 (2001) 5019
M. Naruki *et al.*, Nucl. Phys. A, 721 (2003) 297