

招待論文

高エネルギー物理学研究の新しい展開

High Energy Physics Research on the Emerging Network Environment

渡瀬 芳行

高エネルギー加速器研究機構

Yoshiyuki WATASE

High Energy Accelerator Research Organization (KEK)

要旨

大規模な加速器や実験装置を必要とする高エネルギー物理の研究は、施設の国際的な相互利用や共同研究によって進められてきた。そこでは、国際間のネットワーク基盤が必須であり、その恩恵を受けて発展してきた。高エネルギー加速器研究機構の B ファクトリー実験を例に、高エネルギー物理実験のデータ解析において、どのようにネットワークを利用しているかを述べる。また、スーパーSINETをはじめ、最近のネットワーク環境の急速な進歩によって、大規模な共同研究の質的变化が起ころうとしている。グリッド技術により、グローバルな研究資源の共有が将来の姿を予想させている。

ABSTRACT

In recent years High Energy Physics has extended its activities on an international research level, based on its nature of pure science as well as being supported by an continuously developing evolutionary network environment. Typical usage of the network is described here in the case of KEK Belle experiment. In particular, the recent rapid speedup of the network such as SuperSINET is opening up a new research environment which operates as a virtual laboratory or as an open laboratory using the Grid technology on the global network.

[キーワード]

広域ネットワーク, グリッド, データ解析, 高エネルギー物理

[Keywords]

Wide area network, Grid, Data Analysis, High Energy Physics

1 基礎科学とネットワーク

科学は、古くから未知の世界への人間の知的興味から進化してきた。しかし、20 世紀になり、その成果は、未知の世界への精神的な誘い以上に、現実の世界での人間の生きるための知識、技術としての評価が重要になった。その技術は経済を動かし、人間社会の発展に不可欠なものとなった。科学の多様な分野の中で、特に、人間の知的興味に根ざした研究は、基礎科学と言えるものである。しかし、近年では、基礎科学的な成果がより早く、技術や人間の生

活に役立つことが求められるようになっている。それは、科学技術が進歩して来た証左でもある。基礎科学も技術の進歩によって新たな未知の世界へ挑めるので、それらは相乗的に発展している。基礎科学の代表的な分野：宇宙、天文、素粒子などの発展も、技術の進展、社会からの経済的なサポートに依存している。究極の天文台であるハッブル望遠鏡も、宇宙航空技術に基づいているし、大型加速器も多様な要素技術の集大成によって実現している。そのような、技術の一つとして、ネットワーク技術は、多くの人間活動に普遍的に役立つものであるとともに、

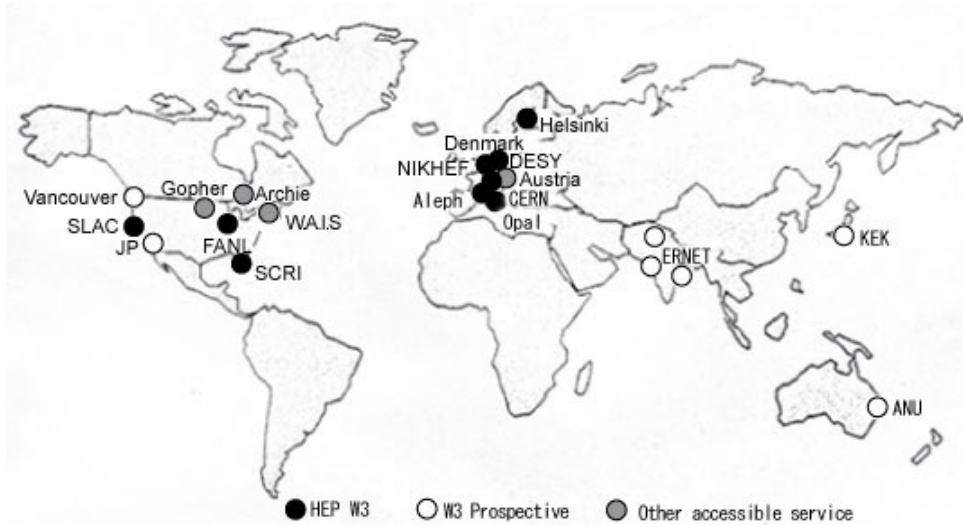


図 1: 初期の世界の WWW サーバの分布(1992 年)^[2]

科学の進展にも計り知れない効果をもたらしている。コミュニケーションは、人間の基本要素であり、文化の広がりを図るメジャーである。20 世紀後半からは、交通手段の広域化と相俟って、情報ネットワークの進展が世界を変えてきたといえる。

こうした中で、基礎科学研究は、経済や社会の障壁を越えて、人類共通の興味に基づく活動として、国際的な共同研究が実施されている。この共同研究に必要なことは、まず、コミュニケーションであり、そのための努力が続けられている。それは、グローバルなネットワークへの飽くなき希求であるとともに、それを利用する技術の開発である。

1960 年代の通信手段は、手紙であり、郵便の時代であった。たまに、緊急の場合はテレックスを使うことができた。国際電話などとても高価で、使うなどということは考えられなかった。それから、およそ 30 年後の 1990 年代に、世の中が劇的な変化を遂げてインターネットの時代になった。国際間を結ぶネットワークとしては、1980 年ころは、米国 IBM が始めた Bitnet が最初であろう。いわゆるメインフレームを使ったネットワークである。その後、DEC 社のコンピュータを結ぶ、DECnet が高エネルギーの世界では、1990 年中頃まで使われた。しかし、1985 年頃から、米国国防省 ARPANET ではじまった、Internet が次第に普及し、現在では、標準になった。

1991 年つくば市の高エネルギー物理学研究所が主催した、「高エネルギー物理学におけるコンピュータ利用国際会議(CHEP91)」^[1]の際、欧州連合原子核研究機構 CERN で、後の World Wide Web(WWW)が作

られているというわきがあった。次の 1992 年秋に開催された同じ会議では、開発者の Tim Berners-Lee から報告された。このときの会議の議事録^[2]には、第 1 図のような世界の WWW サーバの配置図が載っているが、高エネルギー関連の機関を中心に 20 余りでしかなかった。しかし、1, 2 年後には、その有用性が社会に認められて急速に普及し、インターネットの代名詞として WWW が利用されている。すでに、一つのコミュニケーションのメディアとして確立した。最初の頃のことを考えると、想像を絶するような、発展のスピードと社会的な衝撃の大きさである。WWW によって、静的な情報の共有は、国際的な共同研究の多くの問題を解決してくれた。情報としては、図面、プレゼンテーション、論文、ドキュメントなど、それまでは、郵便やファックスで交換していたものが、世界中に散らばった共同研究者によって、即座に共有できることは非常に有効で、WWW の果たした役割は、計り知れないものである。

2 高エネルギー物理学の課題

2.1 素粒子と宇宙

素粒子物理学は、物質の最も原初的な構成要素は何か、また、それらがどのような法則で、相互に作用し合い、この物質宇宙を作っているのかを明らかにする学問分野である。それは、宇宙の始まりから物質がどのように創成されてきたかを研究することにもなり、宇宙物理学と関連している。物質の構造を見る顕微鏡は、光の反射という手法を使うが、原

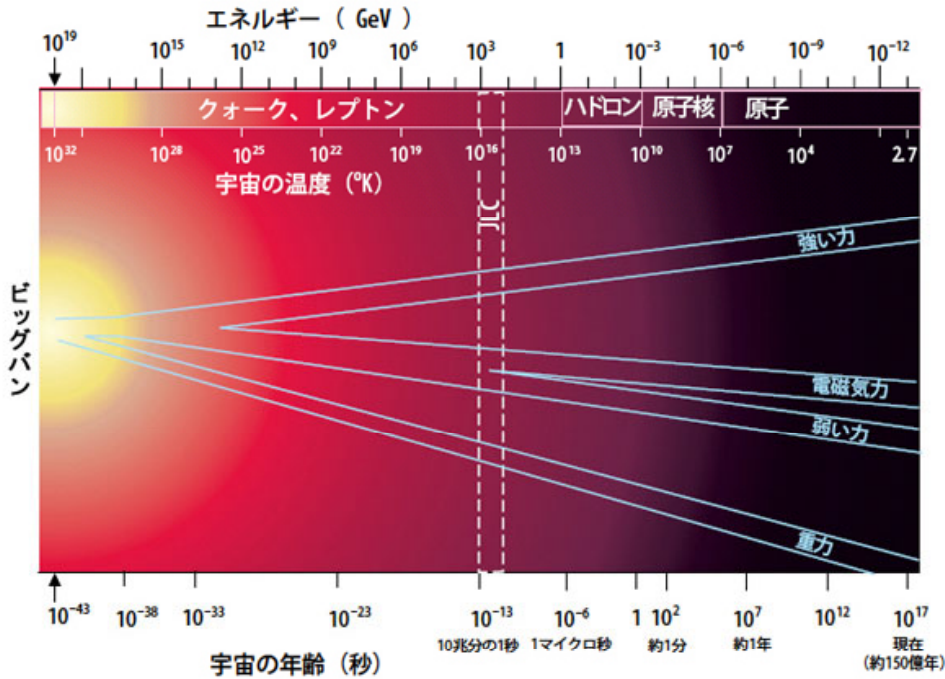


図 2: 宇宙の始まりと素粒子の世界(JLC グループより^[3])

子より小さい構造を調べるには、より波長の短い光であるガンマ線を利用することが必要である。しかし、このような極微の領域では、光と対象物との相互作用を観測して、物の性質を明らかにしている。光の代わりに高いエネルギーに加速した粒子を対象物にぶつけて観測することも同じアプローチである。より詳細な部分の研究には、より短い波長のプローブが必要である。このために粒子ビームを高いエネルギーにまで加速して、そのビームをプローブとして使う。量子力学によれば、粒子は、粒子性と波動性を備えており、その波長λは、次の式で表される。

$$\lambda = 0.1973 / p \quad 10^{-13} \text{ cm}$$

p: 粒子の運動量 (GeV/c)

これまで、到達し得たのは、大きさのスケールでは、 10^{-15} cm で、エネルギーのスケールでは 100 GeV である。個々の粒子がこのようなエネルギーをもつ物質の状態の温度は、 10^{15} 度 K である。このような高エネルギーの状態は、宇宙初期でしかあり得ない。宇宙がビッグバンで誕生したとすると、このようなエネルギー状態は、ビッグバン後、時間のスケールで約 10^{-11} sec である。このように、極微の物質世界の研究と宇宙とが結びついている。これを図示したのが、図 2 である。

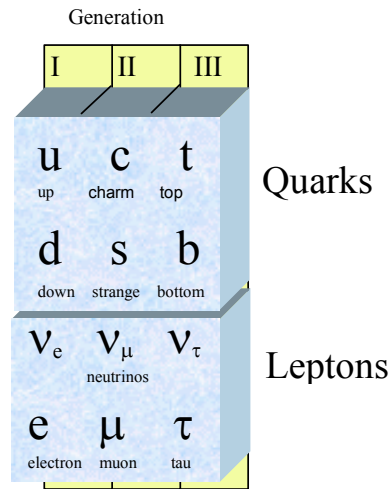


図 3: 物質を構成する素粒子^[4]

2.2 標準理論と統一理論

現在の素粒子物理学では、実験的に確かめ得る事実は「標準理論」と呼ばれるもので、ほぼすべて説明できる。標準理論では、物質の構成要素は、図 3 のように、3 世代のクォークとレプトンであり、それらの間の力を媒介する粒子として、4 つの力（強い力、弱い力、電磁力、重力）に対応して、グルーオン、W/Z ボゾン、光子、重力子がある。しかし、その標準理論がなぜ、そのようになっているのかと

いう標準理論の源についての問いには、いまだ答えは、見当たらない。標準理論では、各種の実験的に測定した物理量が使われており、それがなぜそうなのかは、説明できない気持ち悪さがある。これを解決する研究は、ずっと続けられている。

前節で述べたように、宇宙がビッグバンで創生された時から、 10^{-11} 秒以降は、この標準理論で説明できるが、さらにエネルギーの高い状態、すなわちさらにビッグバンに向かって溯るための理論が求められている。そこでは、図 2 に示したように、4 つの力が同じ強さになり、統一されてひとつの理論ですべてが説明できる。それが「大統一理論とも、すべてのための理論」ともいわれているものである。その理論を組み立てるには、やはり実験的な導きが必要である。標準理論では説明できない事実を発見することが現在の高エネルギー物理学の最大のテーマである。それが、標準理論が唱えられて以来、30 年が経過したが、次のステップに向かうための足がかりである。最近、高エネルギー加速器研究機構での B ファクトリー実験（次節）では、その一部が垣間見えるような実験結果を発表している^[5]。これが真実であるなら、非常に大きな成果であるといえる。

3 大規模実験：Belle

高エネルギー実験の例として、現在、高エネルギー加速器研究機構（KEK）で進行している B ファクトリー実験(Belle)について述べる。この実験は、電子(8GeV)と陽電子(3.5GeV)を衝突させて、2 つの B 中間子^aを生成できる高いエネルギー状態を作る。その結果、生成された B 中間子と反 B 中間子対は、電子と陽電子が非対称なエネルギーであるので、エネルギーの高い電子の方向に飛行する。しかし、両 B 中間子の寿命は短く、直ちに、多様な崩壊の仕方多数の粒子に崩壊していく。この過程が B と反 B とでどのように異なるかを詳細に観測する。その極く小さな相違を見出すことで、宇宙初期での粒子とその反粒子の対称性の差が、我々の宇宙が物質で構成されていて、反物質では構成されていないことの理由を突き止める。

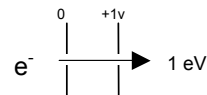
このための加速器は、最高エネルギーを得るためではなく、電子・陽電子ビームの強度を極限まで上げて、生成粒子の量を多くするタイプのものである。

^a B 中間子：bクォークを含む中間子

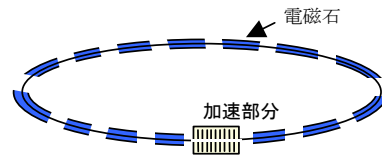
B 中間子を大量に生成することから、B ファクトリーと呼ばれる。

3.1 加速器

粒子加速の基本としては、電場中を荷電粒子が電場に引かれて、エネルギーを得ることによる。図 4-1 のような、電極間に電圧をかけ、図の左側から電子が入ると、電場によって加速される。電位圧が 1 ボルトのとき、電子が受けるエネルギーを 1 電子ボルト (eV) という。線形加速器では、このような電極を多数並べ、可能な限りの高電圧をかけることで加速する。円形の加速器では、図 4-2 のように、円周上に設置した多数の電磁石で粒子を円軌道に周回させ、1 ないし数箇所に、電極を並べた加速部を置く。ここを、何千、何万回も繰り返し通過させるこ



4-1 粒子加速



4-2 円形加速器

図 4: 加速器の概念図

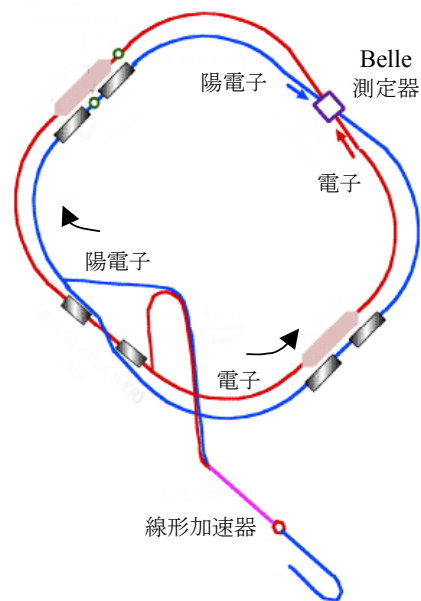


図 5: KEKB 加速器

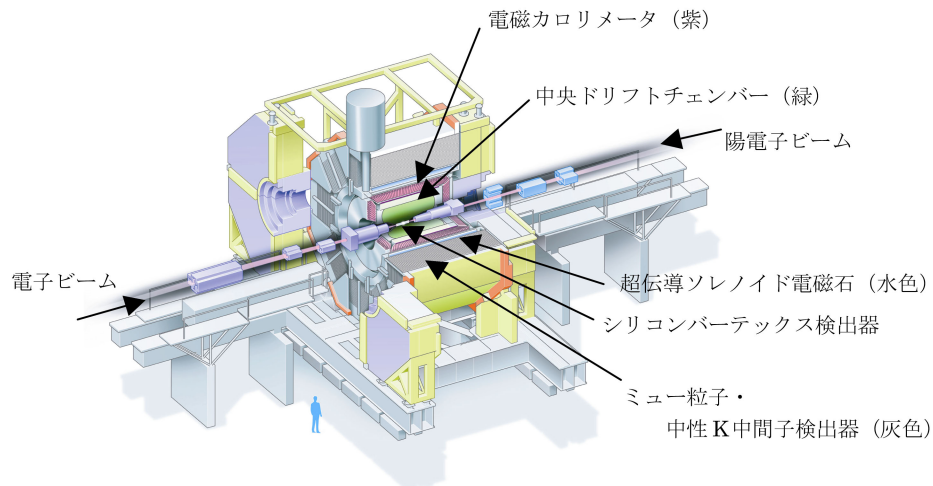


図 6: Belle 実験装置の概要図

とで、エネルギーを上げていく構造になっている。

KEK の B ファクトリー加速器 (KEKB) では^[6]、図 5 のように、線形の加速器で、電子、陽電子をそれぞれのエネルギーまで加速して、陽電子、電子の順で、円形軌道にたくさんの粒子の塊 (バンチ) として入れる。現在は、電子、陽電子とも 1,284 個であり、電子、陽電子による電流としては、それぞれ 1.1 アンペア、1.5 アンペアに達している。バンチは、ほぼ光の早さで周回しており、電子と陽電子のバンチが測定装置の中心で衝突するように制御する。バンチは、次々と毎秒 1 億回衝突する。バンチ内の多数の粒子 (約 500 億個) が含まれていても、実際に正面衝突して、たとえば $e^+e^- \rightarrow BB$ の現象が観測されるのは、毎秒数回程度である。この効率を計るのが、Luminosity (輝度) である。回っている電流に対して、どのくらいの確率で、現象が起こるかを示す。Luminosity は、如何にビームの大きさを小さくして、バンチの数とバンチ内の粒子密度を上げるかで決まる。KEKB では、2003 年 6 月には、世界最高輝度の性能を達成している^[7]。

3.2 測定装置

Belle 実験の測定装置は図 6、図 7 に示したような複合装置である。大きく分けて、内側は、多数の荷電粒子の飛跡を観測する装置として、シリコンバーテックス検出器 (位置測定精度は大略 10 ミクロン)、中央ドリフトチェンバー (位置測定精度 100 ミクロン) から構成され、その外側に、円筒状の超伝導ソレノイド電磁石がある。このソレノイド電磁石は、広い空間にビーム軸方向の様な磁場 (磁場

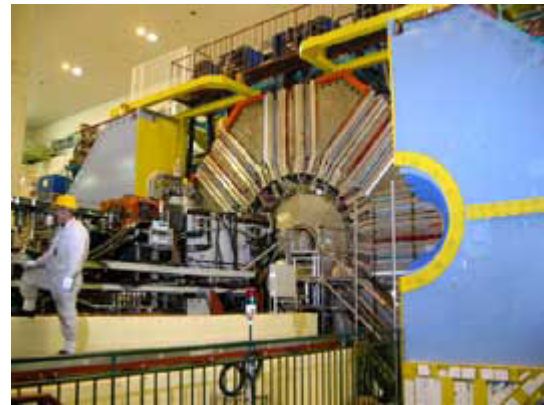


図 7: Belle 実験装置

強度 15kG) を発生させている。この磁場により、荷電粒子は、飛跡が曲げられる。このときの曲率から、粒子の運動量を観測できる。一方、電荷をもたない中性粒子は、主としてガンマー線であるが、衝突点から放射状に飛行して、円筒状に敷き詰めた CsI 結晶 (四角柱) の中で、全エネルギーを失うように作られている。その際、ガンマー線は、電磁シャワー^bを起し、多数の電子を作りながらエネルギーを失っていく。このとき、電子によるチェレンコフ光^cが発生するので、それを光電子増倍管で受けて、電気信号に変えて測定する。光の量が入射した粒子の全エネルギーに比例する。これを電磁カロリメータ

^b 電磁シャワー：電子またはガンマー線が物質に入射すると物質原子の電子と相互作用により、制動輻射、電子対生成、コンプトン散乱を繰り返し次々に起こすこと。

^c チェレンコフ光：物質中で、その中の光速以上の速度の荷電粒子が通過すると、物質分子が受ける電磁力により、振動することにより、発光する現象。

と呼ぶ。こうすることで、可能な限り発生する粒子すべてについて、その粒子の種類、運動量、エネルギーなどを多数のエレメントからの情報を集めて解析する。その荷電粒子の運動量を測定するためには、十分な磁場と粒子飛跡から曲率を求めるために十分な距離が必要で、測定器の概略の大きさが決まる。高エネルギーの実験になるほど、測定すべき粒子の運動量が高く、勢い測定器の外形の大きさは大きくなる。それと同時に、粒子飛跡を測定するための測定点の数も 3 乗で増加する。また、その外側を埋めるカロリメータのカバーする表面積とともに、結晶の数も 2 乗で増大することになる。現在の Belle 実験の場合は、すべての測定点からの信号の出力数は、20 万チャンネルに上る。

電子と陽電子のバンチが激しく衝突しても、稀にしか起こらない事象を見逃さないで、捉えるため、不要な事象は抑えて、必要なものだけのデータをとることが重要である。欲しい事象か否かを判断するロジックが組み込まれたトリガー回路の指令によって、測定器全体からの信号データをデータ収集系に、送りこむ。たとえ収集系に送り込まれても、更に詳しく、データの一部を使って、事象が確かに欲しいものであるかどうか、オンラインファーム^dで解析する。その結果、残されたデータのみを実験データとして、計算科学センターに転送し、磁気テープライブラリーに記録する。この段階でのデータ量は、毎秒 15MB くらいになる。

3.3 国際共同研究の実際

大規模な実験では、国際的に共同して、加速器や実験装置の建設から、実験、さらに実験のデータ解析まで行っている。Belle 実験の場合、13 カ国、55 の研究機関から約 400 名の研究者が参加している。このような共同研究には、国内外の研究機関とのコミュニケーションが容易に出来る環境が必須である。1980 年代から、世界中の主要な研究機関および大学は、専用の国際ネットワークの整備と維持を協力して行ってきた。初期には電子メールのみであったが、1990 年ころより、テレビ会議システムが普及し、国際間での頻繁な打ち合わせをこれで行っている。大きな加速器や実験装置を共同で建設するには、日常的に、会議を持ち、物事を決定して先に進むこと

^d ファーム(Farm)：多数の CPU を高速ネットワークで接続したシステムで、各ノードでは、与えられるデータに対して、同じ処理を行い、結果をネットワークを通して返す。

必要で、その度に海外に出張することは不可能で、このテレビ会議システムの果たす役割は大変大きい。

大規模な加速器の運転は、電源を入れれば即、ビームが出て使えるのではなく、何千台にもなる電磁石の電流の設定、加速部のタイミング、真空度の維持等々、実験のデータ収集が始められるまでに、何日もかかる。したがって、ビーム調整が終わり、いざ実験が始まれば、故障がない限り、一日 24 時間、何ヶ月も連続してデータの収集に当たる。この実験遂行にともなうシフトは、国内外のグループメンバーが平等に割り当てられて、実験室に詰めて、データ収集に当たる。このため、Belle グループの海外のメンバーは入れ替わり立ち代り、KEK にやってくる。

4 データの解析とネットワーク

4.1 実験データの解析^[8]

実験のデータ収集と同時に、データの解析をしていかなければならない。測定器からの信号データから、多数の粒子の空間的な位置情報を得るに当たり、宇宙線の計測やビーム実験のデータを使って、信号データから位置情報を出すための校正データ、校正パラメータを抽出する。これは、実験装置の状態が変われば、その都度、補正していかなければならない。このようなデータ（検出器校正データ）も含め、実験のデータの流れを示したのが図 8 である。

実験中は、トリガー条件により、毎秒 200-500 個の事象データを収集し、磁気テープライブラリーに格納していく。この段階では、平均のデータのサイズは、一事象あたり 30KB 程度である。起こる事象は、時系列的には独立であるが、実験条件の変動について考慮するために、約 30GB 毎に、Run という概念で、ひとつのファイルとして格納する。一日約 400GB のデータとなる。磁気テープに記録されたデータは、校正データに基づいて校正されて、3 次元空間に再構成する。この空間再構成がもっとも計算時間を必要とするプロセスである。このプロセスは、事象ごとに独立であるため、多数の CPU に分散させて、処理できる。Run ごとに処理をして、空間再構成したデータを Data Summary Tape(DST)として、蓄積する。この空間再構成のプロセスでは、飛跡検出器の 1 次データから 3 次元空間での点の座標データを求め、その集合から、荷電粒子の飛跡（スパイラル曲線）を認識する図形認識処理を行う。この飛跡データから、粒子の放出角度、運動量、さらに、他

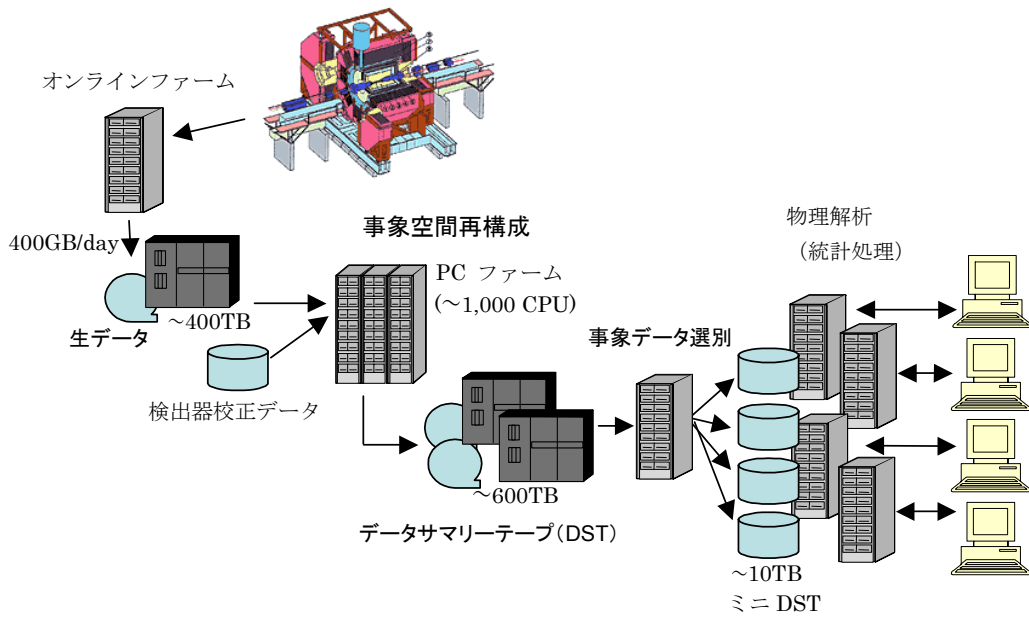


図 8: Belle 実験のデータの流れ

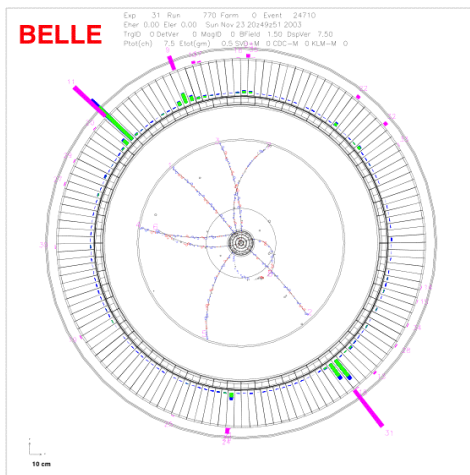


図 9: ビーム軸に沿って見た空間再構成された粒子の飛跡 (曲線) 群および、外縁部の緑、赤色の棒は、カロリメータで測定したエネルギー損失の量を示している。

の検出器からの情報も合わせて粒子の識別などを行う。この結果、図 9 のように、どのような粒子が、どういう運動量で放出されたかと言う、事象の全体像が解る。現在までに蓄積された DST データは、おおよそ 600TB に昇っている。この段階での事象あたりのサイズは、60kB である。こうしたデータも、テープライブラリーに格納する。この後は、多数のグループが、それぞれの物理的な解析テーマに沿って、必要な条件で事象を選択して、新たに事象データフ

ァイルを作る(miniDST と呼ぶ)。この事象選別は、多くの研究グループが頻繁に行うことから、基本的に、DST は、磁気ディスクに格納して、効率よく多くのユーザからのアクセスを処理できることが必要である。

4.2 データの物理解析

高エネルギー実験のデータ解析では、DST を作成するところまでは、実験グループとして、まとめて行う。その後の物理解析にあたっては、DST 全体の中から、特異な指標で、事象を選別し、当面の解析に必要なものだけを集めたデータセット miniDST を作る。大学などの共同研究グループは、この選別された miniDST を使って、それ以降の解析 (物理解析と言う) を行うことになる。したがって、このような miniDST のファイル群や、物理解析の各段階でのデータを共有する必要がある。

また、物理解析にあたっては、既知の理論的なモデルから、実験データに予想される物理量を計算し、実験と比較して、新しい発見に結びつける。そのため、理論的に予想される素粒子事象を計算機の中で起こし、実際の測定器で得られるデータをシミュレーションとして計算する。そのシミュレーションデータを実験生データと全く同じプログラムで、空間再構成し、同じ条件で、事象選別をして、実験値との比較に使う。従って、データの物理解析の流れは、

表 1: 物理解析テーマ (一部) と参加大学など

解析テーマ	参加大学
J/ψ 解析	東北大, 東工大, プリンストン大, 台湾大, ..
τ 解析	名古屋大, 奈良女子大, ..
$\pi\pi$ 解析	東大, 阪大, ..
稀崩壊解析	千葉大, ロシア BINP, ハワイ大, ソウル大, ..
チャーム解析	東北大, 中国 IHEP, シンシナチ大, ..

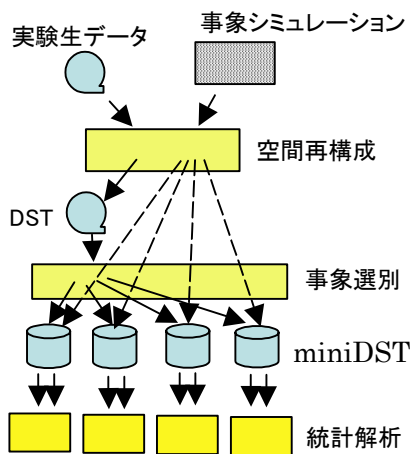


図 10: 物理解析の流れ

図 10 のようになる。Belle グループでは、シミュレーションのデータ量は、実際の実験データ量の 3 倍の数の事象を使っている。このため、実験データの解析以上の計算資源をシミュレーションに充てている。

4.3 データ解析とスーパーSINET

Belle 実験の場合、実験生データの解析は、KEK にある PC ファームで実施しているが、データ量の増大に伴って、参加している大学研究機関での計算機資源をも利用しなければならない状態になっている。さらに、物理解析に関しては、各大学チームで、それぞれ解析テーマをもって、大学院学生も参加して行っている。このような大学の研究室と KEK は、スーパーSINET の高速ネットワークにより結ばれており、必要なデータを相互に転送して解析を分担している。特に、物理解析に必要なシミュレーションに関しては、参加大学での PC ファームを使って、データを作り、KEK に集めて、グループ全体として使うことが行われている。表 1 は、解析テーマと参

加大学などの例を示している。従来は、学生は、研究所に長期間滞在してデータ解析をしていて、先生はたまに来て教育指導を行うことがほとんどであった。しかし、大学グループが大学に居ながら、最新データの解析による研究と学生の教育とを行える環境は、従来できなかったことである。

図 11 は、スーパーSINET を使って行われている Belle 実験グループのデータ転送経路である。現在のデータ転送量は、10TB/月くらいである。一人の研究者が物理解析に使うデータの量は、おおよそ 100GB ~1,000GB 位になる。そのような研究者が常時、100 名くらいが、KEK や国内外の大学に分散して解析しているのである。ここ 2 年間あまり、この環境をフルに活用してデータ解析、物理解析を効率的に進めた結果、物理学での大きな成果が得られている。上述の解析テーマのように、東北大グループが参加している J/ψ 解析では、初期の CP 対称性⁹の破れの成果^[9]を出し、東大、阪大グループは、図 12 のように、 $\pi\pi$ 解析の成果を出している^[5]。これは、8 千 500 万個の B 中間子対の事象データの解析から得られた、わずか 163 個の $\pi\pi$ 崩壊事象からのデータである。いずれも、B 中間子の崩壊では、CP 対称性が大きく破れていることを裏付けている。この事実を説明する小林—益川理論^[10]の正当性を意味する。

それによれば、宇宙初期の段階での粒子反粒子が同数あったとしてもこの CP 対称性の破れで、粒子すなわち物質の世界だけが残って、現在の宇宙を造っている。

また、 τ 解析については、名古屋大グループが自身の大きな計算機資源を使って、進めており、国際会議での論文発表がなされている^[11]。

⁹ CP 対称性：粒子と反粒子とで物理法則が対称であること。

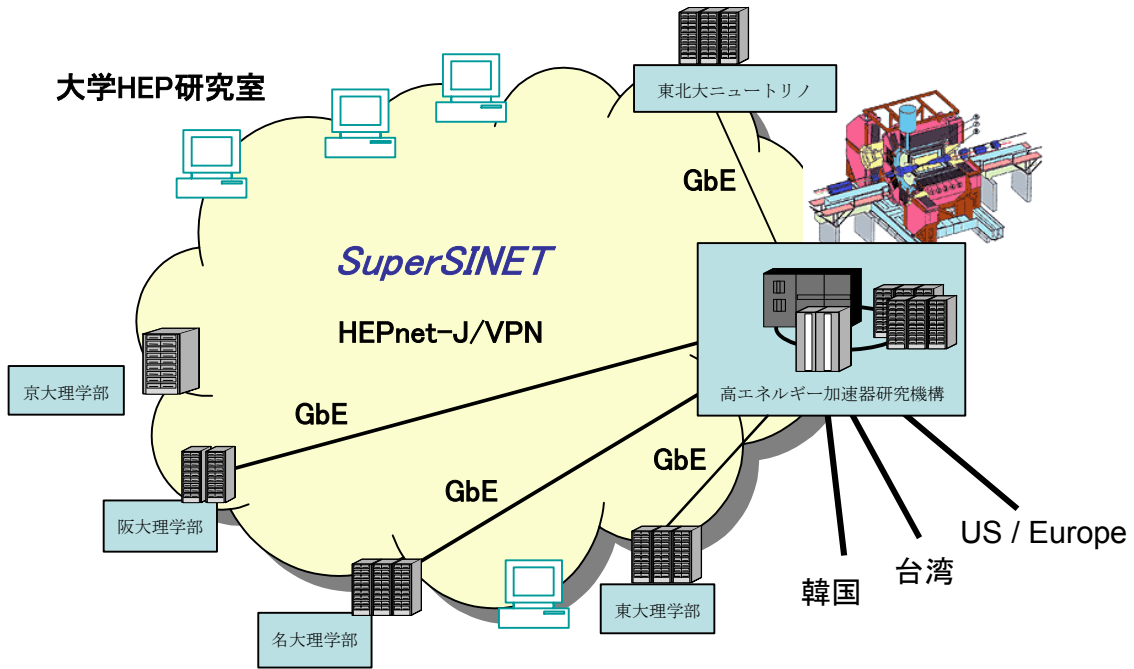


図 11 Belle データ解析とスーパーSINET

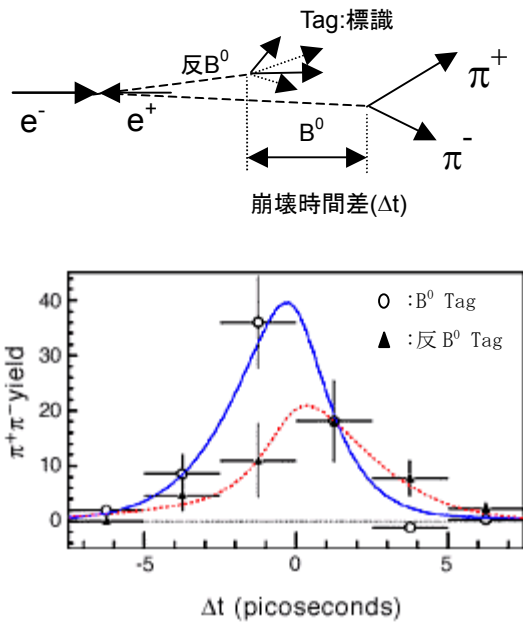


図 12 B^0 , 反 B^0 中間子の崩壊時間差分布

$B^0 \rightarrow \pi^+ \pi^-$ (Δ) と反 $B^0 \rightarrow \pi^+ \pi^-$ (\circ) 事象の崩壊時間分布. 時間の単位はピコ秒(1 兆分の 1 秒). 反 B^0 の方が $\pi^+ \pi^-$ に崩壊し易い.

5 これからの研究とデータグリッド

高エネルギーの場合は, 共同実験グループのメンバー数が数百人から, 多いグループでは千人以上にもなる. 膨大な量の実験データを解析するには, そ

れだけの人員と同時に, 世界の各地の研究機関が持っている計算機資源を活用して解析を分担しなければならない. 従来は, 実験データを磁気テープなどにコピーして, 各地の共同研究グループに配布し, 分散して解析してきた. しかし, ここ 1-2 年の間に, 国内のみならず, 国際的にもネットワークが高速化し, それを使って, 計算機資源を接続することで, データを広域に分散して処理することができる環境になって来た. このようなネットワーク上の資源を共有していくためのシステム技術を, 電力網になぞらえて, グリッド (GRID) という^[12]. このグリッド網にアクセス権があれば, 計算サービス, 特殊なソフトウェアの利用, データアクセス, データ検索, またそれらを組み合わせたサービスが受けられるようになる. 多数の研究機関の共同実験グループを持つ計算資源を協調して運用し, 大規模なデータの解析を行えるシステム (データグリッドと言う) の構築が可能となる. この技術は, ネットワーク上のハードウェア, ソフトウェアを共有するメカニズムを提供するもので, 実験装置の遠隔操作や加速器の共同運転から, 実験データの収集, その解析まで統合したヴァーチャルラボラトリーを構成することが視野にある.

Globus を中心としたデータグリッドの概念図を, 図 13 に示した^[13]. しかし, 現在, Grid 技術は, WEB サービスと連携した OGSA (Open Grid Service

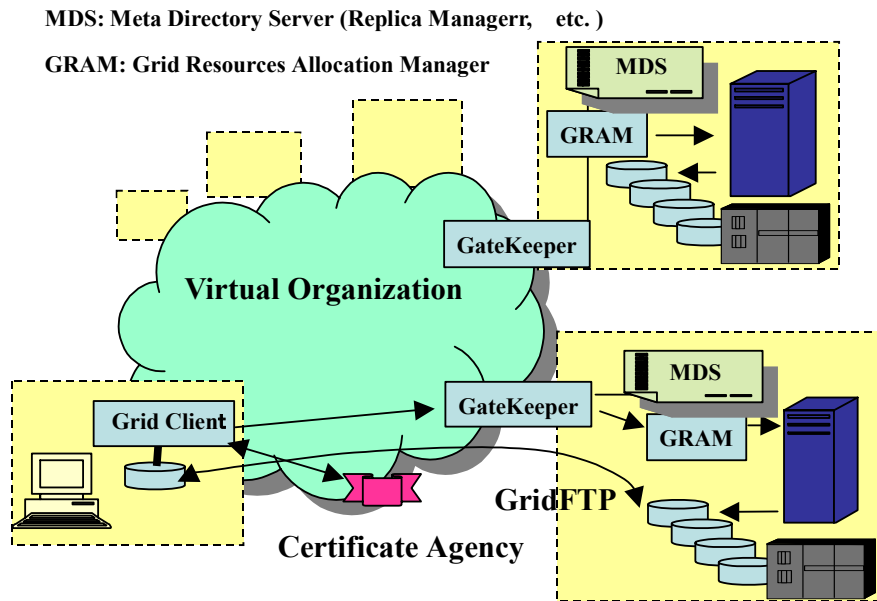


図 13 データグリッド構成概念図

Architecture)^[14]として開発段階にあり、標準化に向けて作業が行われている。

グリッドでは、資源の相互利用を伴うために、相互に信頼して、資源へのアクセスを許すような、バーチャルな組織 (VO:Virtual Organization) を作り上げる必要がある。その許された人に対しては、ユーザとしての証明書を発行して、最初にアクセスする際に、一度、認証を受ける。その後は、その VO に所属する資源にアクセスすることができるようになる。これは、シングルサインオンといわれる機能である。

広域に分散した計算機資源を使って、データ解析を行うには、実験データのファイル、シミュレーションデータファイル、解析用ファイルなど分散した何百万個ものファイルを統一的に管理できなければならない。現状では、基本的な分散管理を人力で行っている。すなわち、データをファイル転送で送ったり、NFS でマウントしたりして共有する。しかし、ファイル管理は、個々のサイト毎に行われている。この大容量のデータを自由自在に、アクセスし、計算機資源の存在場所を意識しなくとも可能にするのが、データグリッドの目指すところである。そのための道具立てとして、グリッド上のすべてのデータのロジカル名とそれが実際存在する物理的なアドレスの対応付けを常に、維持管理できる機構が不可欠である。それと同時に、データが他人にも容易に利用できることが必要で、ひとつのファイルを必要に

応じて、コピーし、もとのファイルが変更されたら、直ちに、そのコピーにも反映されることが必要で、これをレプリカと称している。レプリカの作成は、ユーザからはアトミックなプロセスとして、データの転送、登録、確認をする機能が必要である。データのメタ情報のみならず、多様な計算機資源の情報管理は、MDS (Meta Directory Server) の機能で扱う。

ジョブの実行は、GRAM(Grid Resource Allocation Manager)の機能の中の GateKeeper を通して、ジョブ処理プロセスが起動され、MDS を参照して、資源の割り当てや予約、ジョブスケジューリングなどが行われる。

図 13 で見るように、高エネルギーの場合、VO は実験グループであろうが、そのメンバーは、ネットワークへの接続ができる限り、グループのすべての計算機資源、データ資源を使って、データ解析を行えることになる。

データグリッドは、未だ、開発段階である。現在進行して Belle のような実験において、開発試験を行い、本格的な利用は数年後であろう。グリッド環境を当初から実運用できるのは、準備中の実験である、CERN での LHC 実験[†]、KEK-JAERI の J-PARC[‡]な

[†] LHC: 周長 27km の陽子陽子衝突型加速器で、2007 年実験開始予定。日本も加速器、測定装置建設に資金を出し、参加している。

[‡] J-PARC: KEK と日本原子力研究所が共同で、東海村に建設している大強度陽子加速器施設。2008 年完成予定。

どであろう。LHC 実験については、CERN が中心となり、ヨーロッパ、米国など、世界的にデータグリッドの開発と試験が進行している。そこには、日本からも東大、KEK が参加しており、SINET の海外ネットワークを利用したテストを行っている。

LHC 実験では、まさにグローバルなデータグリッド網を構築し、運用することになる^[15]。

6 まとめ

高エネルギーの実験研究では、大きな加速器を作り、電子や陽子を高エネルギーにまで加速して相互に衝突させ、宇宙初期の非常に高温な状態を再現して、極微の世界での物質の相互作用について実験的に研究する。より高いエネルギー状態を得るには、より大きな加速器や実験施設が必要で、その建設には大きな資金を用意しなければならない。このため、20 世紀の半ば頃から、加速器施設の共同建設や運営、また測定装置の建設から運転までを国際的に共同で行うことが慣例となっている。最近では、1 研究グループが数百人にもなる。そのため国際的なコミュニケーション環境や情報の交換には、専用のネットワークを構築したり、多くの人材を情報分野に当てるなど、大変重要視している。

近年のネットワーク環境の急激な進歩によって、情報の共有から、研究資源すべてに亘って共同できる環境が急速に整いつつあり、基礎科学の世界でも、研究のあり方が変化しつつある。国際的な共同研究はもとより、研究の領域を超えて共同して行う研究も以前より、敷居が低く行い易くなってきた。高エネルギーのみならず、天文やバイオの分野などの大規模なデータを扱うような研究や、大規模な施設を要する研究においても、データや施設のある研究所や大学に居なくとも、ネットワークへ接続できる限り、みな同じ条件で研究に参加できることになる。データや施設をより効率的に使って、多くの大学や研究機関が、独自の研究を展開できることにもつながる。しかし、ネットワークや計算資源などの環境ができて、そこで研究を展開する研究者が、データの公開とか資源の相互提供など、よりソフトの面での協力関係を確立して行かなければ、新しい環境を生かせないことになる。今後の課題である。高エネルギーの次期の計画(2010-2020 年)は、世界で唯一の全長 30km の Global Linear Collider 加速器 (GLC)^[16]を作ろうとしている。そこで得られるデータは、どのように共有するのか、今から検討がなされてい

る。全世界の研究者の共有の財産となり、どの国の研究者もそのデータを利用して研究に参加できることを期待したい。

謝辞

著者は、文章全体を通して協力して頂いた KEK 素粒子原子核研究所の片山伸彦氏、また、図面の制作に協力して頂いた加門弘雄氏に感謝します。

最後に、スーパーSINET はじめ、国内外の学術情報ネットワークの整備運用に日夜努力して頂いている国立情報学研究所に感謝いたします。

参考文献

- [1] Computing in High Energy Physics 91 (Tsukuba, Japan), 1 年半ごとに開かれる国際会議
- [2] T. Berners-Lee and R. Cailliau, Proceedings of Computing in High Energy Physics 92 (Annecy, France 1992), p69
- [3] JLC Project Report, <http://www-jlc.kek.jp/intro-j.html>
- [4] PDG Web Site, <http://www.pdg.org/>
- [5] <http://www.kek.jp/press/2003/belle3.html>
- [6] K. Oide KEKB 加速器, Butsuri(日本物理学会誌) 2003 v.58, no.9 p662
- [7] KEKB World Record, <http://www.kek.jp/newskek/2003/mayjun/luminosity2.html>
- [8] I. Adachi et al., Proc. of Computing in High Energy Physics 03 (San Diego, 2003) <http://www.chep2003.org/ MODT010>
- [9] K. Abe et al., Physical Review Letters 87, 052002 (2002)
- [10] M. Kobayashi and H. Masukawa, <http://www.kek.jp/newskek/2003/mayjun/km.html>
- [11] τ 解析, K. Inami et al. Physics Letters B-551 (2003) p16 K. Abe et al. hep-ex/0310029
- [12] I. Foster and C. Kesselman, The Grid (Morgan Kaufman 1999)
- [13] Globus Project Web Site, <http://www.globus.org/>
- [14] I. Foster et al., "Physiology of the GRID" Grid Computing (Willey 2003) p 217
- [15] LCG Web Site, <http://lcg.web.cern.ch/LCG/>
- [16] GLC Web Site, <http://www-jlc.kek.jp/index-j.html>