核融合プラズマ計測のためのスケーラブルな分散型データ収集保存システムの開発研究

中西秀哉

博士 (工学)

総合研究大学院大学
数物科学研究科
核融合科学専攻

平成 15 年度
(2003)

要旨

核融合プラズマ実験では長い時系列の比較的粒度が大きいプラズマ計測データを，多種多様な計測機器から同時に収集・処理することが求められる．また実験データの統計処理等においても格納データの参照はランダムになる特徴がある．そのため，核融合実験の計測データ処理システムは，大容量データを収集する他分野の実験・観測に比べて，独自のデータ取扱い機能を備えなければならない．

近年，半導体技術，高温プラズマ計測技術の長足の進歩と相まって，データ処理システムはこれまでにない高速で大容量なデータ処理を求められるようになっている．大型ヘリカル実験装置 (LHD) では計画当初でも既に 1 回の放電実験あたり約 600 MB の計測データ量が見込まれ，従来のデータ処理システムに比べて二桁高い性能の実現が要求された．一方，LHD やドイツの Wendelstein 7-X 等，超伝導コイルを用いた核融合装置が世界各地で建設され，定常化プラズマ実験が行われるようになってきている．それに伴いデータ処理システムも，従来の短パルス実験に対応した一括 (バッチ) 処理から，新たに定常化実験計測データの実時間処理に対応する必要に迫られている．

本研究では，このように核融合計測データ処理システムが直面している高速・大容量化と実時間化という二つの大きな課題の解決方法を確立すべく，新しいシステム概念を用いたデータ処理システムを研究開発し，LHD 用として実装，実働させることでその効果を検証した．

約 100 倍にも大容量化し更に増加を続ける計測データへの対処として，計算機資源を集約化する従来の中央集中処理型アーキテクチャを拡張しても，I/O 集中によるシステム性能低下の回避には限界がある．このため I/O 能力を自由に増減できる大規模な並行分散処理型アーキテクチャへの移行と，クライアント/サーバ・モデルによる役割分散，データ収集と計測機器制御の機能分離，データ・ストレージの階層化などによる徹底した分散化により，将来にわたるシステムの柔軟性とスケーラブルな拡張性を実現することが必須である．

分散化されたシステムでは，機能ごとに独立性を保ち主体的なプログラム記述が可能であるオブジェクト指向方法論 (OOM) に則った開発が普及しつつある．反面，オブジェクト指向システムが生来もつ冗長性や処理オーバーヘッドによって I/O 性能が低下する危惧もあり，核融合実験のデータ収集処理では既存システムを全面的に更改する試みはこれまでなされなかった．しかしながら大規模な分散システムの開発では，作業の集約化により効率改善を図る従来手法が効果的ではなくなるのに対して，OOM ではシステム全体の見通しが改善され開発効率向上が可能になるため，核融合分野でも OOM に沿った開発の検

証が早急に必要となっていた．また LHD のような実働する大規模システムでスケーラブルな I/O 性能を実運用に供するには，OOM という新パラダイムに沿った開発だけでは充分ではなく，各機能において性能改善をいかに施すかも極めて重要となる．

そこで，LHD に向けた開発では，徹底したシステム分散化を進めた結果，システムの最小単位である収集エレメントは非常にコンパクトで可搬性がよくなり，システム全体に影響を与えずに，並列数を変えるだけで小〜大規模の構成に対応できるスケーラブルな分散収集システムの実現に成功した．LHD 計測の増設増強にこの特質は遺憾なく発揮され，一実験あたりの収集データを実験開始後の 4 年間連続して年々倍増させることができた．5 年目には当初計画を上回る 740 MB/shot の計測データを 3 分間隔という短い時間で処理しており，1 日あたり 0.15〜0.2 TB の集録量は核融合分野で現在稼動中の他実験装置に比べて一桁大きく，群を抜いて世界一である．

データ処理システムに最も重要な I/O 性能に関しては，オブジェクト指向システムに伴う処理の冗長化を性能改善で補償するため，可逆圧縮ライブラリを組み込んで収集直後に生データを圧縮し以降の取扱いサイズ低減を図ると共に，マルチスレッド化技法により I/O ポート利用効率を上げるなど，システム高速化に取組んだ．加えて，データを格納するオブジェクト指向データベース管理システム (ODBMS) をより高速なものに置換することで，CAMAC でほぼ限界性能に近い 0.7〜1 MB/s の収集・圧縮・格納速度を得ることに成功した．

分散データ収集サーバ群を実験シーケンスに同期動作させるための同期メッセージ授受には，当初利用を計画した仮想オブジェクト共有空間ミドルウェアの処理負荷がノード数増加 (〜10) と共に急上昇するため，より軽負荷の TCP/IP ソケットを隠蔽利用する独自のネットワーク API(Application Program Interface) を開発・実装した．同様に ODBMS からのデータ取出しでも，クライアント/サーバの二階層構造を通信量のより少ない三階層構造に変える専用ネットワーク API を新たに実装して転送速度を改善した．データ検索・転送では，RDBMS によるデータ所在情報検索専用サーバを設けると共に，直前の収集データを次の実験まで主記憶領域の共有メモリ上に保持することによって，参照頻度の多い最新データの応答性を向上させた．これら分散ストレージ対応手法の組合せにより巨大な仮想的 ODBMS 空間を構築・実用化できたことは，OOM の新たな知見といえる．

データの保存の面では，オンライン・ストレージを新しいデータ順に短期・中期・長期の 3 層に分化し，各々相応しい速度性能のランダム・アクセス記憶媒体を用いた．これにより I/O 速度で 2 MB/s 弱，クライアント計算機上への計測データ復元速度で数〜25 MB/s と，非常に高速でストレスのない常時データ取出し環境を実現した．階層化と相互データ移送の構成をとることで，ディザスタ・リカバリ（災害復旧）の機能も必然的に実現され

ている．

プログラム開発及びその効率化の面では，統一モデル化言語 (UML) を仕様分析・設計に全面的に用いると共に，ODBMS を利用してデータ I/O 手順の作成を軽減する等によって，OOM による統一的な開発効率化を図り，従来と同じ開発力でより大規模な分散システムの構築を可能にした．また，対話的データ言語 (IDL) ツールとデータ取出しインターフェースとを併せてエンドユーザ開発 (EUD) 環境とし，ユーザ側にデータ解析・表示開発を解放した．これは各処理の主体的記述という OOM の指針に基づくもので，負担の大きかったユーザインターフェース開発を削減しシステム開発を大きく効率化できた．専らデータを利用するのみだったユーザが，自身で容易にデータ解析・表示機構を作成できるようになった意義も大きい．

実時間化に向けた最大の課題は，高温プラズマの MHD 揺動 (〜1 MHz) が核融合実験の計測には重要であるため実験が長時間化してもサンプリング速度は下げられず，一般化している 50 ch 程度の多チャネル計測で一計測あたり数 10〜100 MB/s の無停止連続収集が必要になる事である．現在依然として主力で用いられている CAMAC デジタイザが実質的に実時間処理に対応できないため，核融合実験では全く未踏の領域となる〜100 MB/s の超広帯域実時間データ収集システムの実現には，高速デジタイザ系と専用処理プログラム双方の研究開発が必須であった．この新デジタイザ系は，CAMAC 系と同様の多チャネル化や増設が容易なモジュラー型フロントエンドとして，大規模並行分散形態をとる CAMAC 系の部分的置換や混在利用が可能になっている必要がある．

そこで本研究では，大規模分散形態に必須な小型計算機と親和性が高い CompactPCI 規格をベースにした ADC モジュールを用いて，高速実時間収集を行う並行分散処理プログラムの開発と試験検証を世界に先駆けて行った．その結果，一筐体あたり〜80 MB/s の連続的な収集・転送・保存に成功し，サンプリング速度抑制や間引きをすることなく短パルス実験時と同じ高速度 (〜1 MHz) 多チャネルでの連続収集を核融合実験分野で初めて実証した．

これは，従来同じ計測信号を二系統に分けて収集処理をせざるを得なかったフィードバック制御用実時間計測と高速プラズマ物理計測とを，一つの超広帯域実時間収集システム上で実現可能にするため，新たな計算機システム設計として今後の核融合実験に貢献し得る．既存の CAMAC や中速デジタイザ類と新デジタイザとの混在組み合わせ利用法とあわせて，定常化実験に対応するコスト・パフォーマンスの高いシステム構築に道を拓くものである．

超広帯域実時間収集を含めたデータ処理システムのオブジェクト指向による統一的開発は，核融合実験分野での開発パラダイムシフトを実証した世界初のケースである．OOM

の開発効率改善により実現されたLHD大規模分散処理システムも，世界の他装置より一桁大きい先駆的な性能とデータ収集実績を示している．音声動画像の収録配信など近年成長普及が著しいマルチメディア分野は，核融合データ処理システムとデータ生成，取出し，粒度等で類似点が多く，本研究で得られた高性能システム技術を含めて，今後の両分野間の知見交換には新たな展開が期待できる．

Abstract

Fusion plasma experiments require the concurrent acquisition and processing of relatively high-granularity plasma diagnostic data observed over long time series using various types of measuring equipment. In addition, the stored data tend to be accessed in a random order even in the statistical analyses of experimental data. Therefore, a diagnostic data processing system for fusion experiments must be equipped with specialized data handling functions in contrast to experiments or observations in other fields that acquire large volumes of data.

The recent rapid advances in semiconductor and high-temperature plasma diagnostic technologies have resulted in a growing need for data acquisition systems to deal with higher-speed and larger-volume data processing. From the initial planning stages, the Large Helical Device (LHD) was expected to handle approximately 600 MB of diagnostic data per discharge, and was required to achieve double-digit performance improvements over existing data acquisition systems. Meanwhile, fusion devices that apply superconducting magnets, such as LHD and Wendelstein 7-X in Germany, have been established in various parts of the world to perform steady-state plasma experiments. Thus, the data acquisition systems in fusion experiments are now required to make steady-state real-time operations possible, although legacy batch processing operations were sufficient in short pulse experiments.

In this study, a data acquisition system based on new concepts has been developed to provide a solution for two major requirements, larger data volume with higher speed processing and real-time operability, confronting fusion diagnostics. Verification was performed by implementing this system and also putting it into practical use for LHD.

To handle the increasing volume of diagnostic data, which has already multiplied by one hundredfold, there will be limits in reducing system performance degradation by I/O over-concentration even by extending the conventional centralized processing architecture that makes intensive use of computing resources. Therefore, both migration to a large-scale parallel distributed-processing architecture that can flexibly increase or decrease the I/O performance and thorough decentralization by functional distribution based on the client/server model, system separation between data acquisition and diagnostic device control, and stratification of data storage are essential to realize long-lived system flexibility and scalability into the future.

The object-oriented method (OOM), which provides independence among all functions and their autonomous programming, is becoming more widely applied for the development

of distributed systems. On the other hand, there are some concerns regarding the I/O performance deterioration by the inherent redundancy and process overhead in OO systems. In the data acquisition and processing systems for fusion experiments, therefore, there have been no previous attempts to reform an existing system completely. However, verification of the development along OOM was urgently required even in this field for large-scale distributed system development, because the established technique, which attempts to increase efficiency by work concentration, is becoming ineffective in large-scale distributed system development and OOM will improve both the prospect of the entire system and development efficiency. Moreover, to provide scalable I/O performance to a large working system, such as LHD, it is crucial to decide not only whether the development will be made along the new OOM paradigm, but is also necessary to know how far the performance can be improved in each function.

As a result of improvements in the complete system distribution in LHD development, the acquisition element, which is the system's minimum unit, has become very compact and even portable. A scalable distributed acquisition system that can deal with from small to large scale configurations has been realized simply by changing parallel number without affecting the entire system. These characteristics have provided full capabilities in expanding and reinforcing LHD diagnostics, and the amount of acquired data doubled each year for four consecutive years since the experiment began. In the fifth year, 740 MB/shot diagnostic data, which is much higher than the original plan, could be processed in intervals of only three minutes. The stored data amount, 0.15-0.2 TB/day, is one order of magnitude larger than those of other currently running experimental devices, and is the best in the world.

As I/O performance is the most important factor in a data processing system, some improvements for better throughput have been applied to compensate for redundancies in the OO system. A loss-less compression library was built in for immediate raw data compression to reduce the handling size, and the I/O port utilization efficiency has been improved using multithreading techniques. Replacement of the data storing object-oriented database management system (ODBMS) with a higher-speed system yielded throughput of 0.7-1 MB/s, which is almost the performance limit of CAMAC-related data acquisition, compression, and storage.

For message passing to synchronize the distributed acquisition servers with the experimental sequence, the first planned middleware that provided the virtual object sharing space was shown to raise the processing load markedly with increases in the number of nodes ($\sim$10). Thus, an original network API (Application Program Interface) that internally uses a light-

weight TCP/IP socket was developed and implemented. Similarly, in data retrieval from the ODBMS, a new dedicated network API that changes the two-tiered client/server structure into a three-tiered structure to reduce network traffic, was implemented to improve transfer speeds. For data retrieval and transfer, the response of the latest data accessed most frequently was improved with both the implementation of a dedicated data location information server using RDBMS and caching the most recent data in the shared region of main memory until the next experiment. This construction and practical use of a huge virtual ODBMS space enabled by a combination of these corresponding methods for distributed storage can be considered as new knowledge in OOM.

The online data storages are separated into three layers by data age; short, medium, and long-term, in each of which random access recording media of an appropriate speed performance has been used. This provides a very high speed and smooth data retrieval environment at all times, in which the fundamental I/O rate is typically a little under 2 MB/s and the diagnostic data restoration rate is up to 25 MB/s on the client computers. The hierarchical storage structure and mutual data migration between them naturally provide some disaster recovery functionality.

In respect of program development and improvement of its efficiency, UML (Unified Modeling Language) was employed throughout the specification analysis and design phases, and also an ODBMS was used to reduce the production burden of data I/O procedures. These efficiency enhancements of OOM-unified development have enabled the construction of a larger-scale distributed system with the same development effort. In addition, both the interactive data language (IDL) tool and the data retrieval interface were used as the end user development (EUD) environment, and it enabled the opening up of data analysis and display development on the user side. This was based on the OOM guidelines of autonomous description of each process, which reduces the large workload that results from user interface development, and significantly improves system-side development. In addition, the end user, who previously could only use the data, can now easily make their own procedures of data analysis and display.

The biggest issue for the real-time operation is that non-stop continuous acquisition requires a speed of several dozens or 100 MB/sec for one multi-channel measurement of typically about 50 channels. This is because MHD fluctuations ($\sim$1 MHz) of high-temperature plasmas are quite significant in fusion plasma diagnostics, and the sampling rates cannot be reduced even if the experiment runs for a longer time. As the CAMAC digitizer, which is still used as the main force in this field, has no substantial capability of real-time processing,

both high-speed digitizers and dedicated program development were required to achieve an ultra-wideband data acquisition system with ∼100 MB/sec real-time transfer, which is never attained in fusion experiments. This new digitizer system has to act as a readily expandable modular front-end that can easily increase the accommodating channels, such as CAMACs, and also have capabilities to partly replace the CAMACs for large-scale parallel distributed configuration or mixed usage.

In this study, ADC modules based on the CompactPCI standard were used because they are quite compatible with the small computers required for a large-scale distributed configuration. Parallel distributed processing programs that enable high-speed real-time data acquisition have been successfully developed and verified ahead of other fusion activities. This system achieved continuous data acquisition, transfer, and storage of 80 MB/sec for one front-end, and, for the first time in the field of fusion experiments, demonstrated continuous acquisition of high-speed (∼1 MHz) multi-channel data without reducing sampling speed or thinning out of the samples.

This allows real-time measurement for feedback control and high-speed data acquisition for plasma physics, which was previously performed by dividing the same measurement signal into two, to deploy into one ultra-wideband real-time acquisition system. It also contributes to a new computing system design for future fusion experiments. Combined with the mixed usage of existing CAMAC and middle-speed digitizers with the new ones, it will allow the development of high cost-performance systems applicable for use in steady-state experiments.

The OOM-unified development of the data processing system including ultra-wideband real-time acquisition is the first case confirming the development paradigm shift in the field of fusion experiments. LHD's large-scale distributed system achieved by the improved development efficiency of OOM also demonstrates pioneering performance and data acquisition rates one-digit higher than those of any other device currently in use throughout the world. The field of multimedia computing, such as sound and motion image recording and delivery, which has expanded dramatically in recent years, uses very similar technology to fusion data processing systems for data generation, retrieval, and granularity. New roll-outs of knowledge exchange of such technologies between both fields in future can be expected, including high-performance system technologies produced by this research.

目次

図目次

表目次

第 1 章

序論

核融合実験におけるデータ処理システムは，各種高温プラズマ計測が高度化・複雑化・大容量化している状況下で，ますますその重要性を増している．例えば外部からの複数センサーの信号処理でトーラスプラズマの断面情報を再構成するコンピュータ・トモグラフィ (CT) 計測では，多チャネル・デジタイザによるデータ収集とコンピュータ計算処理が不可欠である．また 2 次元撮像素子として可視光ばかりでなく赤外線から X 線領域まで幅広い製品群が普及している CCD (Charge Coupled Device) では，素子中にアナログ/デジタル変換器が内蔵されており，信号出力がデータ処理システムに入力・処理されて使われることが前提となっている．このように，多ピクセル・多チャネルセンサー普及の恩恵と，多量のデータの生成・収集・処理を可能にしたデータ処理システムの賜物として，近年，高温プラズマ計測技術が大きな進展をなしとげ核融合プラズマ実験は著しく発達した．

昨今の半導体・電子回路技術の長足の進歩は，核融合プラズマのさまざまな計測器にも大きな影響を及ぼしており，その計測情報は日々高度化・複雑化してきている．そこから生成出力される計測信号に着目すると，半導体センサー等の普及・高精度化に伴った計測器の多チャネル化と，集積回路や半導体記憶素子の爆発的技術革新による計測データの大容量化が，非常に容易に可能になってきている．

また核融合プラズマの生成保持技術そのものも，1980 年代に建造された世界三大トカマク (米 TFTR, 欧 JET, 日 JT-60) などの実験成果によってかなり高度化されてきており，高温・高密度を達成するだけではなく，安定した核融合プラズマをより長い時間にわたって生成・維持できるようになってきている．超伝導技術を導入して定常的な閉じ込め磁場配位の生成を可能にした仏 Tore Supra や九州大学 TRIAM などのトカマク型定常実験装置のほかに，無電流系プラズマ配位によってより安定した定常化プラズマの実現を目指す核融合科学研究所 (NIFS) の大型ヘリカル実験装置 (LHD) や独 Wendelstein 7-X などはまさ

にその典型といえる．こうした核融合プラズマ実験を取り巻く状況の進展は，当然のごとく，計測信号チャネルの増加，計測データの増大という結果をもたらすことになり，それらを取り扱うデータ処理システムは機能・性能の両面で大幅な向上を求められている．

従来，高性能なデータ処理計算機システムを実現するためには，高価な各種計算機資源を一箇所に集約する必要があり，必然的に中央集中型のシステム・アーキテクチャを採らざるを得なかった．この資源集中によって初めて，非常に高価であった高速計算処理能力とある程度のサイズのデータ格納領域とがデータ処理システムとして利用可能になった訳である．このため，オシロスコープやペンレコーダといった個別集録装置から統一されたデータ処理システムへの移行が課題であった中型以上の核融合実験では，多くが資源集中型のデータ処理システムを採用することとなった．その中核となる中型計算機は通称「ミニコン」とよばれ，旧 DEC 社製の PDP-11 に始まる DEC VAX シリーズはその代表例である．現在もなお，世界の中〜大型核融合実験装置のデータ処理システムのほとんどは，部分的・補助的にパーソナル・コンピュータ (PC) やエンジニアリング・ワークステーション (EWS) を導入していても，基本的にこの中央集中型の構成をとっており，VAX ミニコンも多くは現役稼動中である．

こうした中央集中型アーキテクチャでは，計測データは一箇所に集められ高速な演算器を使って処理されるが，データ量を大幅に増やそうとすると 1 〜少数個しかないデータ入出力部 (I/O ポート) がたちまち飽和する，いわゆる I/O 集中が発生する問題がある．これは一つの計算機システムに実装できる I/O ポート数はどのような規模の計算機でも少数に制限されており，大幅な増数が困難なためである．

具体的に新旧の実験装置の例で収集データ量の推移を数値的に比較してみよう．1995 年 9 月にシャットダウンした NIFS のトカマク実験装置 JIPP TII-U のデータ処理システムは，1977 年の稼動以来，データ収集量の増加にあわせて二度のシステム増強を行い，性能実績としては 2 分間隔で 8 MB/shot の処理が可能であった．また最終実験では約 500 の計測チャネルで 24 MB/shot を収集した．これに対して 1998 年に実験を開始した LHD では，当初予定した計測計画として，計測チャネル数が 3,000 強，収集データ量が約 600 MB/shot と算定されていた．この計画値は計測計画の順調な進捗によりわずか 3 年後の 2001 年に早々と達成され，その後も増加を続けて 2003 年初には 750 MB/shot に至っている．(Figure 1.1 参照)

このように最新鋭の核融合実験装置のためのデータ処理システムには，前世代のシステムからくらべて約 100 倍にも迫るデータ処理能力が求められている．最新鋭の LHD 実験装置が要求する従来より 2 桁以上高い処理性能の大容量データ収集を，もし中央集中型の

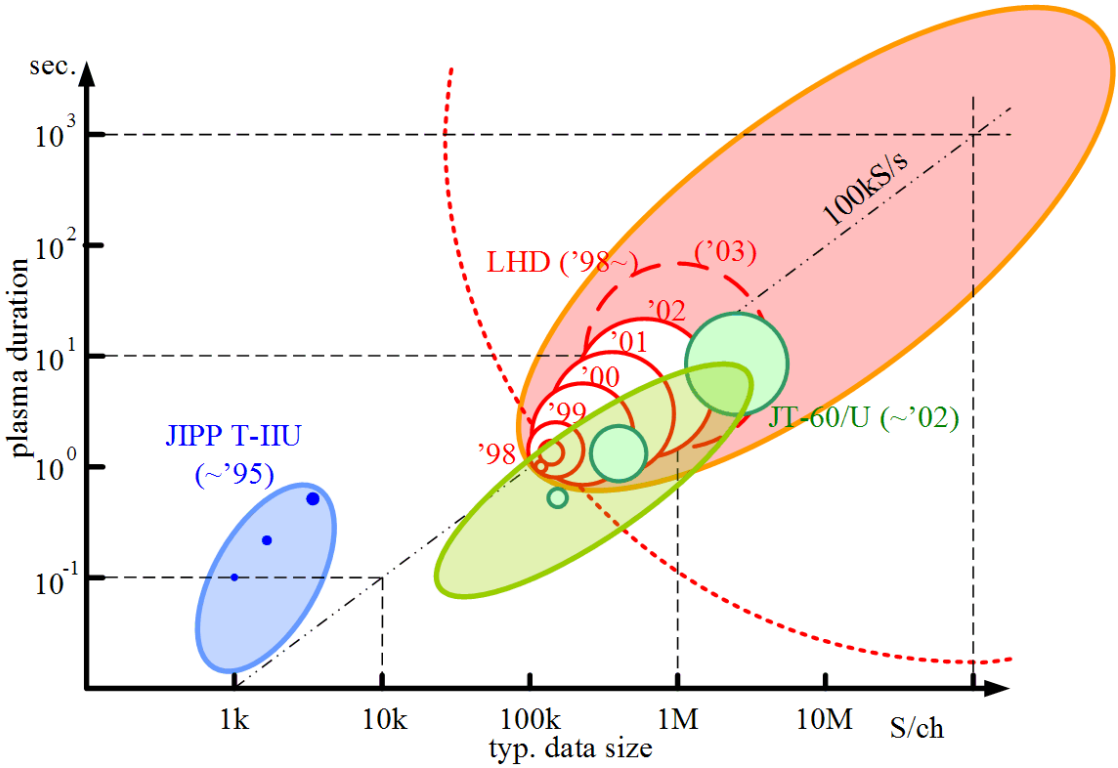


Figure 1.1: Data growth among various generations of fusion exmerimental devices (Circle sizes stand for their data amounts per each discharge, and elongated ones are their operational regions.): As the semiconductor technologies have evolved especially in the computer memories, recent digitizers can apply larger size of them for each diagnostic channel. Because 100 kS/s is the typical sampling ratio for the plasma fluctuation measurements, their data amount would grow larger along with this line. There is the discontinuous threshold near 60 s plasma duration, and above it the data acquisition has to do the real-time operation. The largest dotted circle shows an estimate for LHD's long-pulse experiments.

システム構成をとって I/O 集中を回避しつつ実現させようとすると I/O ポートの帯域を 2 桁以上向上させるしかない．しかし，既に高性能な中〜大型計算機の I/O 性能を更に 100 倍高めるのは事実上不可能である．

つまり 2 桁という大幅なシステム性能の向上は，技術的・手法的な改善改良による対処法では到達できない領域であり，核融合計測データ処理システムとしては，今後，システム設計の基本思想から全く新しいアプローチをとることが求められている訳である．

この結果，LHD データ処理システムの構成として新たなアーキテクチャ設計を模索する

必要が生じ，利用できる I/O ポート数を大幅に増やして対応する，即ち，多数の計算機を並列に用いる方法が想定された．幸い，プラズマ計測の分野に大容量デジタイザや多チャネル計測センサー等を普及させた半導体・集積回路の技術革新は，同時に計算機の劇的な大容量化・小型化をも実現しており，多数計算機を用いた分散形態は実現可能な選択肢となってきている．また，PC 用 HDD の例に見るように，データ保存装置も小型計算機とほぼ歩調をあわせて技術進化を遂げており，大容量化・小型化を伴って普及が進んでいる．

次に重要な問題がシステム開発の効率化である．核融合実験では，実験進行があらかじめ規定されたプリ・プログラムドでシーケンシャルな処理手順であり，大量の計測データが同時一斉に出現するため I/O 処理の非常に高い瞬間最大性能が求められる．このため長い間，核融合データ処理システムではそうした利用形態に合致する I/O オーバーヘッドの少ない構造化システムの開発・利用が続けられてきた．

しかし前述のとおり，従来のデータ処理量から 2 桁高い I/O 性能を要求された LHD データ処理システムでは，中央集中型構成によるハードウェア資源の集約化と同じく人的開発能力の集中投入により要求を達成することは不可能になっており，大規模な分散処理形態への移行とそれに伴う開発負荷の増加に直面することになった．

いうまでもなく開発の対象となるシステムが 10 倍の性能向上を求められても，開発要員や資源が 10 倍提供されることがないのが情報システムの開発部門に課せられる一般的な命題である．換言すると性能向上の要請には開発効率向上の要請が同時に含意されている．LHD データ処理システムに要求された性能向上率は約 100 倍であり，従来の手法をそのまま延用すると 100 倍の負荷が生じて開発不能となる．これを打開するには従来手法を根底から転換して，開発効率を 100 倍改善することができる新たなパラダイム導入が必要であった．

近年，コンピュータ・プログラミングの分野では，構造化手法，第 4 世代言語 (4GL) を用いたイベント駆動型プログラミングに次ぐ新たな開発スタイルとしてオブジェクト指向方法論 (Object Oriented Methodology: OOM) が提起され，利用が確立されつつある．このオブジェクト指向開発 (Object Oriented Development: OOD) は，様々な独立した画面パーツ群をうまく組み合わせて GUI を構築する際などに，従来手法では開発効率が悪く大きな負担を強いられていた状況を大幅に改善する手法として元来普及してきたものである．OOD はプログラムコードを分散化・パーツ化して個々の保守性，再利用性を高めると共に，コード開発自体の分散化分業化を容易にしており，分散コンピューティング・アーキテクチャと極めて親和性が高く，21 世紀初頭においてソフトウェア技術の中心的役割を果たすと目されている．

反面，構造化プログラムの場合に比べると，記憶領域の冗長使用や処理のオーバーヘッドが付随して発生するため，高速大容量のデータ入出力 (I/O) を取扱う大規模データ処理システムの構築には OOD は不向きであると考えられてきた．しかし LHD で求められているような大規模な計測データ処理システムの開発を可能にするためには，構造化システムより処理オーバーヘッドが大きくなっても，開発効率を大幅に向上させる新パラダイム OOM への移行が不可欠である．今まで LHD 以前の核融合実験では，開発効率の大幅向上のためにオブジェクト指向システムへ移行せざるを得ない状況には直面してきておらず，新たに本開発研究を通して，オブジェクト指向開発の効率性と実現可能な計測データ I/O 処理性能とのトレード・オフを具体的に検証する必要があった．

本研究では，データ処理システムを取り巻く環境と要求性能の大幅な上昇に鑑みて，多数の小型計算機により並行して多数 I/O を同時実行できる大規模並行分散 (Massively Parallel Processing; MPP あるいは Massively Parallel Distribution; MPD) アーキテクチャの研究開発を進めて同形態のデータ処理システムを設計，LHD データ処理システムという大規模システムに適用してその有効性を実証した．特にシステムの基本構成単位を可能な限り小さく可搬性良く設計することで，計測器の開発や試験調整時のテストスタンド用途にも対応可能となり，LHD 計測計画の順調な進展に大きく寄与することができた．また共同研究先での単独試験用途や小規模な核融合実験にも容易に導入が可能となっており，構成エレメント数を増やすだけでスケーラブルに小〜大規模のデータ収集系を構成できる優れた特徴をもっている．LHD 計測データは実験開始後も順調に増加を続け当初の計画量を上回っているが，本研究で開発した MPP システムは十分それに対応しており，上限のない I/O 性能が実現できるという特質を実証している．

高価な資源を集約する代表例であったデータストレージ装置についても，本研究では高速ネットワークを介した分散ストレージ構成をとったので，大小のシステム構成にスケーラブルに対応することができた．また利用状況に応じた柔軟性のある増設・容量増強，および技術革新に伴う陳腐化の抑制等も可能になった．これも大規模分散構成を適用した大きな利点であるとともに，小型計算機用の民生品が活用できることで，LHD 大規模データストレージ構築に際しても，大幅コスト削減が可能であったことも示された．

データ収集系の定常運転化は，今までの核融合プラズマ実験にはなかった新たな課題であるが，リアルタイム収集運転が機能的に不可能な現行の CAMAC デジタイザをリアルタイム収集デジタイザに置き換えるだけで，MPP 構成の 1 エレメントあたり 〜80 MB/s もの超高速リアルタイム収集が可能であることも本研究により確かめられた．この成果は，従来 CAMAC 計測で 1 MHz サンプリングで 1〜10 秒の連続計測が上限であったのと比較

すると，1 筐体 40 チャネル程度の平均的な計測で無停止連続 1 MHz サンプリングが可能になったことを意味しており，核融合プラズマ計測に全く新しい展開をもたらすことになろう．

本研究における最も重要なもう一つの成果は，今まで核融合実験データ収集システムの構築で一般的であった構造化プログラミング手法から，新たにオブジェクト指向方法論へと開発パラダイムを移行したことである．オブジェクト指向へのパラダイム・シフトは開発システムの要求分析・仕様策定などの支援ツール群から使用プログラミング言語，開発分業のマネジメントに至るまでの全面的な転換になるため，当初は新パラダイムの学習と準備に開発グループの多くの労力と時間を費消した．しかしその反面，100 倍の性能向上に見合うだけのシステム信頼性向上を達成するなど，従来手法では困難だった堅牢でなおかつ改造への柔軟性が高いシステム構築が可能であることが本研究では検証された．

LHD データ処理のような実働する巨大 I/O システムにおいて，こうしたオブジェクト指向開発システムが世界最高水準の性能を実現した手法は計算機科学にも新たな知見を加えるものである．

本研究がのぞんだ大規模な分散形態と高速大容量な計測データ処理は，核融合実験では全く未踏の新たな領域であり同分野で最初の経験となった．核融合分野で現在世界一のデータ生成量を誇る LHD 実験で 0.15〜0.2 TB/day 程度であるが，こうした大容量の計測データを取扱うその他の分野としては高エネルギー大型加速器・素粒子実験があり，欧 CERN, 米 SLAC, BNL などで 1 桁多い 〜1 TB/day 級の実験が行われている．しかし素粒子計測データは観測事象が非常に高率で発現するものの，各事象毎のデータサイズは 30 kB 程度と比較的小さく［1］，1 事象あたり比較的長い時系列でよりデータ粒度の大きい核融合計測とは取扱いが全く異なるといってよい．

このように核融合計測データは特徴的な取扱いを要するものであるが，データサイズと時系列性，特に読出し時の I/O 要求性能を考えると，近年急激に普及しつつある動画像・音声といったマルチメディア・データとの類似性を見出すことができる．現在，オンデマンド画像 (VOD) 配信やデジタル放送，DVD-Video などでよく用いられる MPEG-2 規格では 2〜80 Mbps の帯域が使用されており[注1]，特に今後 HDTV などの映像素材伝像用に数 10 Mbps の帯域利用が伸びると予想され，本研究の様々な成果を含めた両分野間の相互知見交換により新たな展開の可能性も期待できる．

[注1] MPEG-2 規格は 1994 年に策定された．VideoCD 等の目的で 1992 年に策定された MPEG-1 規格では 1.5 Mbps の帯域を使用する．

第 2 章

核融合プラズマ実験におけるデータ収集システム

核融合プラズマ実験におけるデータ収集系 (Data Acquisition System，DAS) の果たす役割には二つの側面がある．

一つは，プラズマを生成，制御するため電源を含めた各種制御装置の状態を監視し，得られた結果を帰還制御や保護インターロック等の機器制御に利用する場合がある．また，プラズマ自身の平衡条件を満足させたりプラズマの持つ各種不安定性の成長を抑えるなど，プラズマを持続させるために必要なプラズマ状態を監視し，それによりプラズマ自身の制御を行なう場合もある．こうした場合，データ収集系は制御システムの構成部分として対象であるプラズマあるいは機器制御に必要な時間内の動作を順次要求されるため，基本的に実時間処理が必要になるという特徴をもつ．

もう一つの側面は，プラズマを物理計測の対象として捉え，プラズマの各種物理量を測定，数値化することでプラズマの物理的描像を把握するための好適な手段を提供，実施するという面である．この場合は，計測システムとして対象となるプラズマ物理がもつ特徴的時間/空間的長さに対して十分詳細なデータを収集することが優先されるため，より計測対象に特化された収集装置が用いられ，実験中はデータ生成に専念し，終了後一括処理するという動作になることが多い．

今日の高度技術化によって，制御あるいは計測システムの各構成要素は各々飛躍的に技術向上を果たしており，データ収集系もまたそれらと歩調を合わせて向上することが要求されている．このためデータ収集系には，いわゆる実験室の自動化すなわちラボラトリ・オートメーション (LA) の方法を提供する，という狭義の解釈から離れ，学際的計測工学あるいは制御工学の対象として，自立した進化が今後益々求められてくると思われる［2］．

以上のようにプラズマ実験も制御あるいは計測システムの適用分野の一つで，そのデータ収集系は，計測/制御システムを構成する工学要素の集合体であることから，先ず各要素の基本的内容や特徴，関連する技術環境などの現状調査と理解が重要である．このため本章では，データ収集系を構成する各基本要素を主な利用機器に応じて分類し，機能及び関連状況の把握に努める．特にデジタル・データ生成の最前面に位置するデジタイザ・フロントエンド (DFE) や，データ収集系の各要素を連結するためのバスは重要であり，プラズマ実験の計測用バス，およびデータ伝送路については重点的に調査検討を行っている．

また現在プラズマ実験において利用されているデータ収集系の各形態を調査し，比較検討を加えて本研究に至る状況を再確認している．

2.1　データ収集系の基本構成とその現状

計測工学とは,「対象より計測量を検出・伝送・分析・処理・判断を行うための系の構成, 系の制御，ならびに個々の構成要素に関する工学」[3] と定義されており, 計測システムの構成要素を大別すると,

1. (計測対象)
2. 検出
3. 変換
4. 分析
5. 処理・判断
6. 収録
7. 伝送
8. (機器) 制御

のようになる．計測工学の中のデータ収集の分野が対象とするのは，この中でも，変換, 処理・判断，収録，および各構成要素間の伝送，の各分野と, 系全体の構成および制御である．機器制御に関しては，制御工学に依存するところが多く，通常，計測工学とは分けて体系化されており，データ処理の興味の対象からも外れるのでここでは触れない．

以下に，データ収集分野の対象である計測システムの各構成要素の役割とその内容，プラズマ実験におけるデータ収集系での特徴を，最近の計測システムを取り巻く状況と共に述べる．話を具体的にするため，処理・分析を行なう計算機，収録を行なうデータ保存装置，等と利用機器毎に分類している．

2.1.1 アナログ/デジタル信号変換器

物理計測において検出される物理量は，通常, 連続的な数値をとるアナログ信号であるが, アナログ信号は伝送に伴って伝送路の精度による影響を受け, 信号の忠実度が悪化する．こうした測定信号の劣化を無くす伝送方法が，離散値を用いるデジタル信号伝送であり，検出器から出たアナログ信号をデジタル伝送可能なデジタル信号に変換するのが，アナログ–デジタル信号変換部，通称 ADC (Analog-to-Digital Converter) の基本的役割である．

ADC の基本性能は，分解能とサンプリング速度，およびその確度であるが，アナログ–デジタル信号変換時には連続的数値を最寄りの離散的数値に置き換えるデジタル化誤差が必ず伴う．例えば，10 bit 分解能，1 MHz サンプリングの ADC でデジタル化する場合には，測定値で $2^{-10} \cong 0.1\%$，時間で 10^{-6}s = 1μs の誤差が含まれる．通常の ADC ではこの値はそれぞれのデジタル化確度より大きいので, デジタル化誤差がほぼ ADC の精度と考えて良い．

現在市販されプラズマ実験で用いられる汎用の ADC モジュールでは, 分解能は 10，12, 16 bit 等が一般的であり，精度としては 0.1 % 以下になる．物理計測の検出部や前置増幅器に用いられる電子/電気回路の精度を 1 % 以下にする事の難しさを考慮すると, 高分解能 ADC の利用はデータの信頼性を確保するためには必要ない場合もある．

プラズマ実験で良く使われる ADC の動作としては, プラズマ実験そのものの特徴,

1. 予めプログラムされたシーケンスに沿って，実験の全行程が進行する
2. 実験開始イベントが自然現象ではないため，オペレーションによって任意の時刻に実験開始を設定できる
3. 実験の繰り返し頻度が低く，且つ，実験継続時間が物理計測の対象となるプラズマの特徴的時間に対して十分長い

によって，実験開始トリガで A/D 変換を開始しモジュール内の十分大きなローカル・メモリにデータを蓄積，実験終了後, データを外部に転送という手順を踏む．このため，ADC 利用の特徴として,

- ポスト・トリガ・サンプリング
- チャンネルあたり ～数百 kword の大容量メモリ
- ADC からの処理要求は，1 シーケンスで実験終了時の 1 回のみ

をもっており，更に，A/D 変換終了時間が予想可能であることを考えると，ADC からデータ収集計算機への処理要求を待つこと無く, 終了時刻に計算機側から自発的にデータ収集

を開始することも可能となる.

プラズマ実験が予めプログラムされた実験シーケンスであることが, 自然現象を観測する物理計測と異なる大きな特色となっている.

2.1.2 計測用バス

計算機システムの性能では，演算処理ユニット (CPU) の性能もさることながら，全てのデータの伝送路であるバスの転送性能が，大量のデータを取り扱う程重要となってくる. また，データ収集系において ADC などによってデジタル化されたデータは，最終的な格納場所に至るまでの間に，複数の種類のデータ伝送路を経由して転送されるのが一般的である.

プラズマ実験におけるデータ収集では，複数の計測機器から大量のデータが殆ど同じ時間帯に集中して伝送, 処理されるので，データ伝送性能の優劣がデータ処理のスループットを大きく左右する事になる．このため，プラズマ実験のデータ収集系におけるデータ伝送路としては，より多くの複数信号線を複数の機器が共用する事で, 個々のデータ伝送にかかる時間の短縮化を図る場合が多い．この様に双方向通信が可能で, 機器間で共用されたパラレル伝送線を一般をバスと呼んでいる［4, 5, 6］.

バスの信号線の本数は，通常ビット (bit) を単位として数えられ, 1 本の信号線が 1 ビットのデータを伝送するので，8 本のデータ伝送線を持つバスは，8 ビット幅をもつ．よく知られている通り，通常のデータ・バス幅は 8 ビット=1 バイト (byte) の倍数をとる.

バスの標準化は，IEC (International Electrotechnical Comission) や IEEE (Institute of Electrical and Electronics Engineers) によって行なわれており，例えば，IEEE によって規格化された標準には，IEEE Std. の番号が割り与えられる．ここで，一般に良く使われるバスの種類を利用形態の面から分類してみると，ほぼ以下の通りとなる.

内部バス

コンピュータの主基板 (マザー・ボード) 内で使われるローカル・バス. 利用する CPU のアドレス線，データ線に直結して使われるため CPU に依存する. 主記憶 DRAM への接続に通常使われるのでメモリ・バスとも呼ぶ. i486 の VL-Bus や Macintosh の各 PDS 等も該当する.

拡張バス

主基板上に拡張スロット・コネクタとして実装され，主に，各種拡張 I/O 用コントローラ・カードの接続に用いられる．一般に，内部バスとはブリッジとよばれるバ

ス変換器を通して間接的に接続されるため，独自のバス・クロックが使用される．代表的なものに，ISA, EISA, MCA (以上 PC), C, NESA (98), NuBus (Macintosh), SBus (SPARC), PCI, PCMCIA 等がある．

I/O バス

周辺機器接続のインタフェースとして入出力ポート経由で使われるバス．SCSI, IDE/ATA や GP-IB, セントロニクス・インターフェース等があるが，特に最近流行しだしているシリアルバス類 USB，IEEE1394 なども含まれ，ネットワーク物理層の代名詞である Ethernet も該当する．

システム・バス

CPU モジュールを含む任意の用途をもった複数のモジュール (ボード) をプラグイン形式で収納して，モジュール間のバス接続を提供する．互換性のために標準化されたものを特に「標準バス」とも呼ぶ［7］．例としては VME, VXI 等がある．

計測バス

複数の計測モジュールを収納しデータ伝送路を提供する面ではシステム・バスに類似するが，アドレスをスロット番地に (半) 固定として簡略化する等データ・バスを強調した仕様をもつ．CAMAC, FASTBUS 等が有名である．

フィールドバス

使用目的，使用場所，スペックなどは様々に異なるが，一般的には “工場内の機器内または機器間の複数データの通信をデジタル通信技術を用いて行う接続方法” という定義で産業・工業分野で広く普及している［8］．計測バスと共通の目的を持つが，汎用インタフェースではなく産業分野での利用のため機能を特化限定して低コスト化を図った I/O バスの一種ともいえる．自動車の計装系などに良く使われる DeviceNet/CAN バスなどが有名．また Bluetooth はフィールドバスの無線版といえる．

データ伝送路としては，上記の他に，EIA (Electronic Industry Association) 標準の RS-232C (EIA-232-E) 等といったシリアル接続が周辺機器–計算機間の交信に良く用いられる．特に RS-232C は，計算機に標準ポートとして装備されていることや，通信用 IC，ケーブル類が一般に広く普及している事などから広範な用途で利用されてきた．しかし近年までは，データ収集系における伝送路として大量のデータを高速に転送する事を要求される場合，RS-232C では最大転送速度が 20 kbps と速度的に性能不足である事から，実質的にデータ伝送路にシリアル接続が使われる事は少なかった．こうした状況を変えたのが USB, IEEE1394 といったシリアルバスの登場であるが，狭義のパラレルバスの議論とは離れる

ため，データ伝送路の議論と絡めて改めて 2.1.4 節にて述べる．

以下では，実際に計測機器によく使われる各種バスの特徴などについて具体的に取り上げる．

GP-IB

General Purpose Interface Bus の略で, 米ヒューレット・パッカード社が開発した HP-IB が IEEE Std. 488-1978, IEC-625 で規格化されたものである．今日のいわば計測用インタフェース・バスの標準的存在となっている．

バスの信号線構成は，データ・バス 8 本, ハンドシェーク・バス 3 本, 管理バス 5 本と接地線 8 本よりなり, 3 線式ハンドシェークの機構を採用することによって 3 台以上のバス接続機器間の安定した通信を可能にしている［9］．この 3 線式ハンドシェークは，通信相手側の対応を待って次の手続きを行なうという手順を常に取るため，異なる最大通信速度を持つ異機種間通信においても確実に通信が行なえるという特徴があるが, 逆に各機器のうちで最も通信速度の遅い機器の対応可能な速度でしか通信が行なわれないという速度的欠点を併せ持つ．このため，GP-IB の最大転送速度は規格的には 1 MB/s あるところを，通常の異機種接続においては実効的に数十 kB/s 程度しか出せない場合も多い．

同じ I/O バスのセントロニクス・インタフェースや PC/AT 互換機のパラレル・ポートも同様の 3 線式ハンドシェークの信号線を持っているが［4, 10］，通常, 殆どプリンタを対象にした対向接続であるため, 実質的には 2 線式ハンドシェークで使われる．

GP-IB の利用形態としては，0~31 のアドレス番号を機器に固定的に設定し，最大 15 台の機器を総長 20 m 以内で接続して利用できる．また，バスに接続される機器の基本構成は，

1. バス上にコントローラが 1 台と各 1 台以上のトーカ，リスナが存在する
2. バス上にトーカが 1 台と少なくとも 1 台のリスナが存在する

の 2 通りがあり得る．GP-IB では，データ/コマンドの送信者をトーカ (Talker)，受信者をリスナ (Listener) と呼び，前者の場合コントローラが動的に，後者の場合固定的に特定の機器に割り当てられる．コントローラが介在しない後者の形態は, 入出力が固定された限定的な使用法であり, 計算機が介在するデータ収集, 制御系統等では通常前者の形態が用いられ，計算機に接続された GP-IB インタフェースがコントローラになる場合が殆どである．

GP-IB 機器側からの割り込み処理要求は，管理バスのうちの 1 本である SRQ (Service Request) 信号線を用いてコントローラに対して行なわれる. SRQ 線がアクティブである事

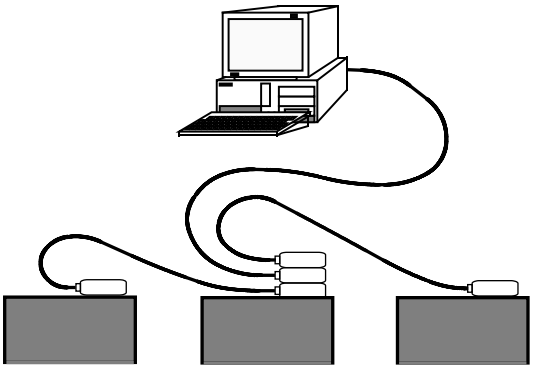


Figure 2.1: Typical GPIB connection between the controller and equipments

を検知したコントローラは，処理要求を出している機器をポーリング (点呼) によって特定する．ポーリングの種類には,

パラレル・ポール　8 本のデータ・バスを 8 台の機器に割り当て同時に要求の有無を判定する

シリアル・ポール　機器を 1 台ずつ呼出し，その機器の状態を示すステータス・バイトを得ることで要求の有無を検査していく

の 2 種類があるが，パラレル・ポールには対応していないコントローラも多く，通常シリアル・ポールが良く使われる．

GP-IB で接続される機器としては，標準的 I/O バスであることから一般に入出力装置が多く，具体的には，デジタル・マルチメータや各種ゲージ・コントローラ，X-Y プロッタなど比較的低速でデータ量の少ないものが対象となっている．また，光エクステンダーで容易に電気的絶縁をとり延長できるので, 小規模のシステム構築時に利用されることが多い.

CAMAC

CAMAC は Computer Automated Measurement and Control の略で, プラズマ実験のみならず，加速器/素粒子物理実験を含めた高エネルギー物理実験におけるデータ収集系の主流である．その歴史は古く，既に 1970 年に CAMAC モジュール製品が発売を開始されている［11, 12］.

多くの機能的追加を重ねながらも，今日まで実験室環境における事実上の標準の地位を保っている理由としては,

Table 2.1: CAMAC Standards

Classification	IEEE Std.	IEC
Instrumentation and Interface	583-1982	482, 516
Serial Highway	595-1982	640
Parallel Highway	596-1982	552
Multiple Controllers in a Crate	675-1982	729
Block Transfers	683-1976	677
Real-Time BASIC Language	726-1982	45
CAMAC Library Subroutines	758-1979	713

- 計算機の仕様に全く依存しない完全に独立した規格である
- 作動プログラムを要しないコマンド動作型の周辺機器である
- N,A,F と呼ばれる制御コマンドが極めて単純でかつ体系化されている

等と考えられる．CAMAC 規格としては，IEEE Std. 583-1982, IEC 482, 516 が基本規格であるが［13］，機能的追加を反映して関連の規格も多い．Table 2.1 に CAMAC の関連規格をあげる．

CAMAC は, 計測バスの特徴として，他のバスとは異なるユニークな機能と名称を持っている場合が多い．例を挙げると，

データウェイ　CAMAC の信号線の総称であるが，他のバスと異なりアドレス線がパラレルに配線されておらず，コントローラと一対一で各スロットに対して直接結線されている．これらはステーション番号線 (N1〜N24) といわれ, このためバス結線は, ほぼ制御線とデータ線のみになるため, この名称がある．VME, FASTBUS 等ではバックプレーンと呼ばれる．

クレート　モジュールのための支持構造および電源，冷却機構を提供するいわば “箱” を示す用語で，CAMAC と FASTBUS でのみ使用される. サブラック (VME)，メインフレーム (VXI) に相当する．

ハイウェイ　クレートと CAMAC を制御する計算機とを遠隔で接続するための伝送路. 規格としては，66 対ケーブルで接続するパラレル・ブランチ・ハイウェイと，光ファイバでリング状に接続するシリアル・ハイウェイとがある．制御は，ブランチ・ドライバあるいはシリアル・ドライバと呼ばれる計算機に接続されたコントローラか

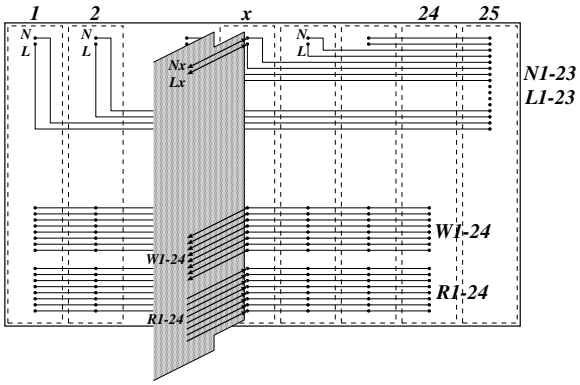


Figure 2.2: CAMAC dataway and module layout

ら画一的に行われる．

等が代表的なものである．このように，CAMAC のデータ伝送路は，データウェイ，ハイウェイとも “シングル・マスタ–マルチ・スレーブ” の設計思想で統一されており, その実装が，ハードウェア的な結線によって固定化されている．このため柔軟な応用が出来ない反面，簡潔な構成になり特にバス制御が単純化され簡便に利用できるという長所を併せ持つ．

その他の実装上の特徴としては,

- 単方向のデータ・バス: Read 24 bit，Write 24 bit
- 最終スロットに位置が固定された単一のクレート・コントローラ
- N 線と同様にコントローラと対向結線された割込み要求 (Look-at-Me) 線 L1〜L24
- Q-response 信号を制御に用いるブロック転送
- “CAMAC サイクル” と呼ばれる 1 μs 中に固定されたバス・タイミング
- カードエッジ・コネクタ
- LEMO といわれる小型同軸コネクタをアナログ信号に利用

等であり，最大 1 Mword/s (1 word = 16 or 24 bit) のデータ転送能力をもつ．このデータ転送速度とカードエッジ・コネクタの信頼性の問題が, 後発のバスと較べた場合の短所と言える．

CAMAC データウェイの制御は, 通常 N,A,F と呼ばれる CAMAC コマンドで完結している．N,A,F はその名前の通りデータウェイのステーション番号線 Nx (x は 1〜24 の内の 1 つ)，サブアドレス番号線 A1,A2,A4,A8, ファンクション番号線 F1,F2,F4,F8,F16 の各信号線出力を指示し, N,A で選択されたあるモジュールのチャンネルに対して F のファンクションを実行させるというものである．F(0)〜F(31) の意味は，例えば F(8) は Test

Look-at-Me というようにほぼ固定されている．

近年発表されるバスは，高速化，高機能化の一途をたどっているが, 複雑さをも同時に増している．プラズマ実験のデータ収集の様な，ある程度頻繁に利用形態や設定が変更される実験室環境での利用の際, 理解しやすく直観的に利用できるという観点で CAMAC はなお十分にその要求を充たしており, 現在も実質的に代替するバスがない存在となっている．

FASTBUS

32 bit 幅のアドレス/データ共用信号線をもったバスで, 通常のバスの TTL と異なりアドレス/データ線に全て ECL を採用することで，データ転送の高速化を図っている．IEEE Std. 960-1986 規格書では，+5 V, -5.2 V, -2 V バス線にそれぞれ 300, 300, 200 A の電流転送能力を推奨していることからも分かる通り, 通常クレート電源が大型化すると共に, 空冷または水冷の強制冷却機構の使用が必要になる［14］.

FASTBUS クレートのバックプレーンには, セグメントと呼ばれるデータ・制御信号線のバスと, アナログ信号やトリガのための補助バックプレーンとがあり［12］, セグメントは，SI (Segment Interconnect) とケーブル・セグメントと呼ばれる接続用ケーブルとによる相互接続で拡張可能であり, 同一のアドレス空間を共有する．アドレス空間が 32 bit 幅あることから，実質的に拡張性の制限はなく, また，全てのアドレスが対等であるため，バス・コントローラの挿入位置も任意である．

FASTBUS の機能的特徴としては，データ転送時のみハンドシェーク転送と非ハンドシェーク転送 (パイプライン転送) が選択可能である点と, 更にアドレッシング機構が,

1. Geographical　セグメントとスロット位置番号
2. Logical　32 bit アドレスを利用した論理番地参照型
3. Broadcast　複数のスレーブに同時参照

と自由度が大きい点などである．また，アドレス線とデータ線が共用であることから,

1. アービトレーション・サイクル
2. アドレス・サイクル
3. データ・サイクル

の順で FASTBUS のシーケンスは実行され, アドレス・サイクルでスレーブがマスタに接続されることでデータ転送相手が指定される．

プラズマ実験における計測用バスとしての位置づけおよび利用法は, ほぼ CAMAC の高

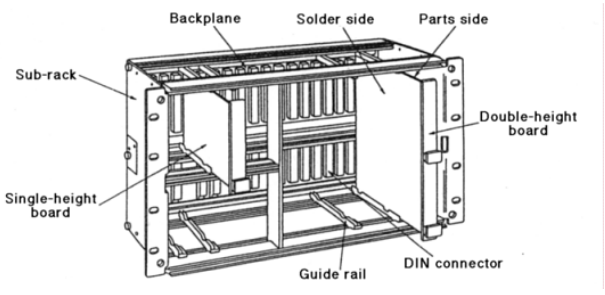


Figure 2.3: Schematic view of VMEbus system［7］

速版というところであるが, CAMAC に比較してより自由度と拡張性が大きい反面, 利用方法が複雑化する欠点を併せ持つ.

VME バス

VME (Versa Module Europa) バスは，モトローラ社が MC68000 CPU のために開発した Versa バスを，ユーロ・カードのサイズでの標準として 1981 年にまとめられたもので, IEEE Std. 1014-1987，および IEC 821 として規格化されている. その本来が，68000 CPU 用システム・バスとして設計された事もあり, 当初よりデータ・バス，アドレス・バスが共に最大 32 bit 幅をもっている．また，バス幅を広く必要としないボードでも, データ・バスでは 8, 16, 32, アドレス・バスでは 16, 24, 32 bit 幅をボード毎に指定して, 上位ビットをデコードしない事で, 他のボードと同時に利用することができる仕様になっている.

更に，VME バス上の I/O 用ポート・アドレスが，メモリ空間上の特定のアドレスに直接マップされるメモリ・マップト I/O が採用されており，アドレス空間を線形で画一的に取り扱うことが可能なため，システム・プログラミングが平易に行えるという長所がある［15］.

このように極めて汎用的な仕様を持っているため, VME バスは計測モジュールのためのプラグイン・バスというよりも, むしろ，汎用バス的な色彩が強く, 歴史的にも EWS のシステム/拡張バスとして一般的に使用され普及してきた経緯をもっている．例を挙げると, EWS のトップメーカーである Sun においても 68010 CPU を使用した Sun2/160 (1985 年) から, SPARCstation における SBus の採用以降も暫くデスクサイド機の標準バスとして使用された［16］.

バスの特徴としては,

- アドレス・バス 16, 24, 32 ビット

- データ・バス 8, 16, 32 ビット, big-endian
- アクノリッジ (Acknowledge) 型ハンドシェイクによる非同期転送
- 理論最大速度 57.2 MB/s, バス・クロック 16 MHz
- 複数のバス・マスタの共存が可能
- 1 つのバス・アービタによる使用権の集中調停 (Arbitration)
- 7 レベルの割り込み要求/処理
- サブバスによる拡張性

等をもち，構造的特徴としては,

- 最大 21 スロットのバックプレーン
- シングルハイト (3U) とダブルハイト (6U) のボード・サイズ
- 96 ピン DIN コネクタの採用

が挙げられる．バス転送形態は，マスタ–スレーブ間のハンドシェークによる非同期転送であり，実効的な転送速度は理論最大速度の半分程度以下と考えて良い.

VME のサブバスは，ダブルハイト・ボードの P2 コネクタ側の空き 64 ピンをユーザ定義として利用したもので，パラレル接続では I/O チャネル，VSB (VME Subsystem bus), VMX (VME Memory Extension bus), シリアルでは VMS (VME Serial bus) がある．それぞれ VME バスと併行して対応するボード間での通信に用いられる.

VME バスは，供給されるボードの市場流通量の多さからみても正に FA や LA における実質的標準と言え，その応用分野は, リアルタイム計測制御から大容量画像データ処理に至るまで極めて多岐にわたっている．しかし，耐実験室環境という側面から見た場合, 静電，電磁ノイズに対する耐性や, 光ファイバによる電気的絶縁が確保できるような仕様を持つ訳ではなく，この点で，通常かなり強いノイズ発生を伴うプラズマ実験においては, 後に述べる計測バスに本流を譲っている.

VXI バス

VXI (VMEbus eXtension for Instrumentation) バスは, 文字通り VME バスをモジュラー型計測器のために機能拡張したもので, 米空軍のモジュラー型自動検査装置 (ATE: Automatic Test Equipment) 計画の成果として 1987 年に世に出た．計測機器をモジュラー化してインタフェースを標準化することで, 個々の計測器間の連携を高め，高収納密度化，高機能化を図っており，IEEE Std. 1155 を受けている.

基本的に VME バスと上位互換性を保持し, データ転送には VME と同じ P1, P2 コネクタの A32, D32 までの信号線を利用する．ボード・サイズも，VME バス互換の A-サイズ (3U), B-サイズ (6U) と, 奥行きを 2 倍強の 340mm に拡張した C-サイズ (高さ 6U 相当), D-サイズ (高さ 9U 相当) の 4 サイズを持っている［17］.

VME バスからの電気的仕様の拡張としては, P2 コネクタの未使用ピンを用いて, ECL レベルの 10MHz システム・クロックと 12bit ローカル・バス, 新たに TTL，ECL 両レベルのトリガ線を提供しているほか, D-サイズ・ボードでは，P3 コネクタにて 100MHz クロックと 24bit ローカル・バス (計 36bit), 4 本の ECL トリガ線を規定し, モジュラー型計測器バスとしての総合的利用性向上を可能にしている［18, 19］.

VXI バスの利用構成としては, 現在以下の 3 つの方法が標準的になっている［19, 20, 21］.

コントローラ (CPU) 組み込み型　最も VME バスの利用形態に近く，バス・コントローラとして VXI モジュール型計算機をスロット 0 に挿して利用する. VXI システムとして最も高速な処理が可能となるが，反面，外部とは全て CPU との通信を経由するため, 通信手順を用意する等利用方法が複雑化する.

GP-IB インタフェース型　GP-IB インタフェースを持った VXI コントローラを用いて GP-IB 経由でバスと接続機器を制御する. GP-IB の転送速度が VXI に較べ低速なので，大量データ転送には速度的に不利となる.

MXI インタフェース型　低速な GP-IB の代わりに，MXIbus (Multisystem eXtension Interface bus) インタフェースを持った VXI コントローラに対して，同じく MXI インタフェースを持った外部計算機から制御を行う. 利用形態は GP-IB とほぼ同様になるが速度的に大きく (最大 20 MB/s) 改善される．更に後続規格の MXI-2 (max. 33 MB/s), MXI-3 (max. 1.25 Gbps) による高速化や光ファイバ接続 (MXI-3) なども可能になっている.

VXI はバス規格そのものが他の計測器関連バスと較べて比較的新しいこともあり，特に，他の装置やバスとの取り合いのための周辺機器，VXI を利用するためのプログラム開発環境，ライブラリやドライバといったミドルウェアの提供，蓄積が VMEbus と較べて少ない．そのため，小規模システム用の市販ソフトウェアはあるものの，中規模以上のプラズマ実験のためのデータ収集装置としては，遠隔制御を行うなどの関連するサポート・ソフトウェアの不足や，デジタイザモジュールの種類の少なさ，製品提供メーカーが非常に小数に限られセカンドソースの確保が困難，などの点で充実度が低いと言わざるを得ない.

WE7000

WE7000 は規格団体などが認定した標準規格ではなく横河電機 (株) が製品化しているいわばプライベート・レーベルだか，最近，高エネルギー素粒子/加速器や核融合などの関連実験分野に普及してきているため特にここで取り上げた［22］.

PC ベース計測器 WE7000 は，各種デジタル測定機器の各機能をモジュール化した「測定モジュール」の種類が豊富で，これらを専用筐体内で自由に組合わせることで複雑な計測データ集録装置を柔軟かつコンパクトに実現できる．また専用のコントロールソフトウェアのほかにも，高度で多様な解析・表示機能をもつ MATLAB や LabVIEW の環境上で WE7000 をコントロールできるインタフェースソフトウェアもあり，信号発生・信号計測から解析まで一貫した作業環境が得られるなど，商用パッケージソフトウェアとの組み合わせで小中規模の計測ステーションを非常に簡便に構築できる長所を備えている.

メーカー 1 社単独による製品規格としては豊富な種類のモジュールが提供されており，ホスト計算機との間を PCI-光インターフェースで比較的高速 (max. 250Mbps) に結ぶ機能もある．しかし，計測用バスとして外部にデータを転送する能力は，古い規格の CAMAC と同程度の最大約 1.5〜1.8 MB/s (実測値) と低い[注1]ため，大容量高速な連続データ転送を伴う用途には不向きである．また，専用の WE7000 コントロールソフトウェアでは，複数のフロントエンドをデイジーチェーンで接続した際にもデータ転送に障害が出ないように，100 kHz/ch の能力を持つ ADC モジュール (WE7272 など．同時 4ch) 等でも最大サンプリングレートは 1 kHz/ch に制限されている[注2].

このためプラズマ揺動計測などの目的ではサンプリング周波数が不十分となっており，次にのべる CompactPCI バス等の利用が検討されうる.

CompactPCI/PXI バス

CompactPCI は PCI バス (Peripheral Component Interconnect bus) [注3]を工業製品/産業分野等に応用することを目的として，米国 PICMG (PCI Industrial Computers Manufactures

[注1] WE7000 のバス規格は非公開のため仕様等詳細は不明.

[注2] WE コントロール API を用いて独自にプログラムを作成すれば，上記の 1.5 MB/s 程度の最大伝送帯域までは利用可能になる.

[注3] 米国 PCISIG (PCI Special Interest Group) が規定化した．PCISIG は PCI 仕様の管理・普及，ベンダ ID の管理などを行っている団体のこと.

Group) [注4]が策定したバス規格である．機構面で高い信頼性を追求しているため工場や実験室などでの利用にふさわしく，FA/LA 分野で次世代バスとして将来普及することが予想されている．

基板外形は VMEbus と同じユーロ・カードサイズの 3U = 100 × 160 [mm], 6U = 233.35 × 160 [mm] で，バスコネクタも VMEbus と同様にピン型の DIN コネクタを J1, J2 の 2 つ利用している．J1 コネクタは 32 ビットの PCI 信号線に使用し，J2 コネクタは 64 ビット PCI 転送かリアパネル I/O 用になっている．CompactPCI 規格には最新のモジュール技術を反映して従来には無かった活線挿抜の機能が新たに入れられたほか，同じユーロカードサイズを用いる VMEbus 規格の欠点のひとつであった対 EMC/ESD 仕様も盛り込まれている．以下に特徴的な点をあげてみる［23］.

ホットスワップ (活線挿抜) 機能　電源が入ってシステムが稼動している状態でアドイン・ボードの挿入・抜除が行える機能．24 時間連続運転が避けられないシステムでは故障の修理，あるいは設備の増強のため，電源が入った状態でモジュールの抜き差しが可能であることが望ましい.

プラグ・アンド・プレイ　(Plug and Play) ISA 以前の拡張カードの場合，カード上のジャンパ・スイッチなどによってカードのアドレスを設定する必要があったが，ハードウェアの設定情報を各カードが保持することによって OS が自動的に検知して I/O アドレスなどを動的に割り付けることが可能になる．ハードウェアの自動検出により手動のドライバ・インストールなどを不要にした.

挿抜ハンドル　CompactPCI ボードをバックプレーンに挿入したり，抜いたりするためのハンドルで，従来のカード・プラは基板を抜くときにのみ働くが、挿抜ハンドルは挿入時にも働く.

ESD/EMC 対策アース　ホットスワップなどの時にも静電気放電などが起きないようにバスコネクタ部，カードエッジ/レール部，フロントパネル/挿抜ハンドル部にケースアースとの接点が設けられている.

CompactPCI は PCI とソフトウェア的に互換仕様であり，PCI で使用したソフトウェアやドライバがそのまま使用できる上に，PCI から機構的に強化されているため，ノイズ・衝撃・振動等に強くより安定した運用が可能である．また，PC 用に大量生産されている最新チップセットを利用することができるため，VMEbus など他のモジュール型バス製品

[注4] 産業用の PCI バス規格の標準化を行っている団体で，PICMG バス，CompactPCI 規格を管理している．日本ではピック・エム・ジーなどと呼ばれる.

に比べると一般に製造コストも低い．これらは PC との親和性の良さを示す側面である．

バス規格としても非常に新しくまだ拡張・更新が続いているが，速度速度もそれを反映しており，33 MHz，32 ビットの CompactPCI システムでも 133 MB/s の伝送帯域があり VMEbus の約 40 MB/s と比べると遥かに高速になっているが，64 ビットの PICMG CompactPCI Rev.2.1 規格や，同 3.0 の 66 MHz 化によりさらに 2 倍 4 倍の高速化が可能になっている．

PXI 仕様に関しては National Instruments 社が提唱した計測用コンピュータの規格で，CompactPCI を計測制御向けに拡張したものである．VMEbus 規格を計測制御向けに拡張した VXI に相当している．

今後の展開と利用の拡大が期待されるバス規格である．

VMEbus 拡張仕様 VME64x, VME320

PCI バスが大きく普及しモジューラー型フロントエンドにも CompactPCI 規格が広がってくると，それまで FA などで多く用いられていた VMEbus の規格上の欠点が浮かび上がり，新しい設計のバス規格にならって解消しようとする動きが出てきてた．

VMEbus との互換性をできるだけ保持して性能改善を目指したのが，通称 VME64x と呼ばれる 1997 年制定の VMEbus 拡張仕様である．これはバースト転送中には使用されないアドレスバスをデータバスと束ねて利用したり，またデータ転送時のストローブ信号の両側エッジを使うことで，元の VMEbus の 4 倍速転送を可能にした［24］．

Table 2.2: Backplane and bus standards of VMEbus and CompactPCI.

bus	clock	max. throughput	max. # slots	bus width
VME	10MHz	40 MB/s (32-bit)	21	32-bits
VME64X	10MHz	160 MB/sec (64-bit)	21	32 or 64-bits
VME320	20MHz	320 MB/sec (32-Bit)	21	32-bits
CompactPCI	33MHz	264 MB/s (Rev.2.1)	8	32 or 64-bits
	66MHz	528 MB/s (Rev.3.0)	8	64-bits
(PCI-X	133MHz	1.06 Gbyte/s (64-bit))		

また VME64x は CompactPCI と同様に，機構面での信頼性向上を目指してホットスワップ機能なども新たに盛り込み，PCI/CompactPCI 陣営への巻き返しを図っている．しかし

圧倒的な利用数を誇る PCI バスは技術革新も速く，既に PCI 拡張仕様 (PCI-X) 対応の製品群も出回っており，性能差が縮まらない状況となっている．

2.1.3 収集計算機

核崩壊や宇宙線観測のような自然現象を観測する物理計測では, 実験開始イベントが能動的に制御できず, 常時イベント待ち状態であることが必要となる．また，加速器物理の計測では実験開始イベントの周期が短く, イベントの実時間処理が要求される場合が多い．これに対してプラズマ実験では，通常, こうした実験開始/終了イベントを含めた ADC からの割込み処理要求が頻繁に起こることはなく, また，ADC からデータ変換終了が割込通知されても, それを規定時間内に処理して速やかに再びイベント待ち状態に復帰することが要求される訳でもない．

プラズマ実験で使用されるデータ収集計算機は, こうした利用条件のため, 割込み処理能力が必要とされることは殆んどないが，反面, 実験終了後に大容量の収集データが同時的に一括して伝送され処理されなければならないので，一括処理時のデータ入出力および演算性能が高いことが重要となる．幸いなことに, 通称パソコンと呼ばれる小型計算機 (PC) やエンジニアリング・ワークステーション (EWS) 等における昨今の記憶容量，入出力および演算性能の驚異的向上は, こうしたバッチ処理性能の要求を十分に充たすまでになっており, 可搬性が良く且つ廉価な PC による収集計算機を実現する環境が整えられている．

以上は，プラズマ実験で最も一般的と思われるパルス運転についての考察であるが，超伝導電磁石を用いた定常的長時間放電実験では, 所謂実験終了後のバッチ処理だけではなく，実験中の実時間監視も必要になってくる．この場合も自然現象観測とは異なり，処理間隔は自発的に決定することが可能であるが，機器制御システムの場合と同様，実時間に則して処理イベントを順次完了していく性能が問われる．実時間処理が通例である機器制御システムでは，FA (Factory Automation) や LA のための計算機として VME 等の利用が多く，オペレーティング・システム (OS) も OS-9, pSOS, VxWorks, VMEexec, VRTX 等のリアルタイム OS が用いられる．

このバッチ処理と実時間処理との二つの動作形態を大型汎用機などで集中して行なうと，大容量の一括データ処理と実時間の割込み処理との相反する二つの計算機性能が要求されるため，システムが複雑化し価格性能比が悪化する事が容易に予想される．広範に普及し，且つ高性能化した PC 技術を十分に活用するためには，一括データ処理を行なう所謂データ収集処理計算機と，実時間監視および機器制御を行なう実時間処理計算機とを分離して分散処理を行なうことが，今日最も有効になりつつある選択肢といえるであろう．

2.1.4 データ伝送路

計算機環境の高性能化にともない，計算機内部および周辺機器とのデータ伝送路も機能分化が進行し, データが経由される伝送路の数，種類とも増加する一方である．

PC が 16 bit CPU であった頃には，CPU，RAM メモリ，拡張スロットが同じバスを共有し同一のシステムクロックで動作していたため，拡張スロットに ADC ボードを挿してデータ収集を行なうとまさしく単一の伝送路で事が足りていた．これに比べて現在の PC では，FSB (Front Side Bus) と呼ばれる CPU 専用バス 1，メモリバス 1，拡張バス複数，I/O バス複数の構成が普及しており，なおも技術革新が続いている．

データ収集と処理/表示とが分離する分散データ処理環境においては，入手容易な EWS や PC を用いたクライアント/サーバ (C/S) モデルによる負荷分散と相まって，収集データの通過伝送路が 10 種以上になる場合も珍しくない．こうしたデータ伝送路を大別すると，

1. 結合が密で複数信号線をパラレルに使って転送帯域が広いが短距離間の伝送に限られるバス
2. 長距離の伝送が可能で，信号線を不特定多数の間で共有して疎結合な対話的通信を行なうネットワーク

に機能的側面から分けることが出来る．

特に，計測システムの構成要素間を連結するバスはデータ収集システム全体の総合性能に対する寄与が極めて大きく，実験の計測分野で利用が多いこと等から計測用バスという観点で既に 2.1.2 節で検討を加えた．

データ伝送路としてのネットワークは，ダウンサイジング/分散処理への潮流と LAN (Local Area Network) 技術の普及に伴って急速にその役割を増してきた．ここでネットワークとは，インテリジェントで且つお互いに対等な不特定多数の計算機間の通信手段を提供する機構である．LAN としては FA 制御関係では回線制御にトークン・パッシングを用いる ARCNET 等も一部使われているが，CSMA/CD を採用した通信帯域幅 10 Mbps の Ethernet および同 100 Mbps の Fast Ethernet が現在の主流である．また幹線としては FDDI［25］や回線交換技術を応用した ATM［26］の利用が一時期広まったが，現在は 1 Gpbs 帯域を提供する Gigabit Ethernet 利用が本格化している．

こうした LAN の相互接続手段として, 通信手順的に遠隔地間通信が可能なインターネットが普及したことにより, 機器制御や収集データの実時間表示を行なうクライアント計算機を遠隔地で利用することが実現しつつある．プラズマ実験の分野でも遠隔実験のためのネットワーク構築の取り組みが行なわれ始めている［27］．

このように, 高速ネットワークをデータ伝送路として計測システムへ導入することによって，一部を除いてデータ収集系の構成要素間の物理的接続距離に関する制約は事実上取り払われたといえ, 今後のプラズマ実験の新たな形態を産み出すものと期待される.

Ethernet とシリアルバス

LAN を構成する最下層の物理層として実質的標準となっている Ethernet も，1 bit 幅をもついわゆる「シリアルバス」規格の一種であることはよく知られた事実である．このため既に述べた「パラレルバス」との区別が曖昧になってきている昨今であるが，高速インターフェースの流れとしては，パラレルバスからシリアルバスへ着実に移行している［28］.

USB, IEEE1394 といったいわば狭義のシリアルバス規格は，旧来の RS-232C シリアル I/F の高速改良版としてパーソナルデバイス向けに普及範囲を広げつつある．接続対象は主に PC 周辺装置や AV 機器である．これに対して，PC 内部で主に HDD などの記憶装置用に使われるバス規格 IDE (ATA) や SCSI のシリアルバス版が各々 Serial ATA, Serial Attached SCSI あるいは FibreChannel にあたる.

こうしたシリアルバスはデータ処理システムの各構成要素間をつなぐバス伝送路として有望であるが，FibreChannel を除き現在はまだ部屋/建物間の伝送距離と電気的絶縁の問題を解消できる規格・製品が完備されていない状況である．この問題は数年のうちに解消されると予想されるが，逆にネットワーク用途ではじまった Ethernet においても，その普及によって狭義シリアルバスの適用場面であるローカルなデバイス I/F として利用する流れも出てきている.

シリアルバスの今後の展開如何では，今まで計算機間の通信であったネットワークと計算機–周辺装置間のデバイス I/F との違いがなくなり，新たなネットワーク・アーキテクチャが成立する可能性も出てきている.

2.1.5 データ保存装置

計測によって得られた実験データの格納場所であり, 計測技術の進歩に伴って大容量化の一途をたどっている計測データを充分格納できると共に，実験後のデータ参照を円滑に行なえることが要求される．しかし，通常流通している記憶媒体においては, 大容量性と高アクセス性との二つの性能は一般的に背反してしまうため,

1. 高速でかつアクセス性の良い短期保存装置
2. 比較的低速でアクセス性が劣るが大容量を実現できる長期 (半永久) 保存装置

に機能を分化して，データ保存装置を構成することが多い．現在流通しているデータ保存媒体を挙げると，紙，磁気テープ (MT)，磁気ディスク (HD)，光磁気ディスク (MO)，光ディスク，等があるが, 紙 (連続紙) に関しては，ペンレコーダ等の低速監視装置での用途のみで，データの保存に使われることは現在では殆んどない．

短期保存装置としては，一時期，磁気テープによる大容量データ・レコーダがプラズマ実験に用いられたこともあったが [29], 現在では殆んど通称ハードディスクといわれる磁気ディスク装置が使われ, ある程度 (〜GB) の容量があり高速かつランダム・アクセス可能な装置として代替する媒体は現れていない. また信頼性や MTBF 時間を向上させるための RAID と呼ばれるディスク・アレイ機構も普及している．

長期保存装置は，テープ等のシーケンシャル・メディアとランダム・アクセス可能メディアとに別れるが, 共に日進月歩の技術革新で大容量化されている．保存媒体からの自動読み出しを検討する場合，自動媒体交換機構が必要になるが，MT カートリッジや MO, CD-ROM 用の物が既に一般に流通している．一媒体当りの記憶容量，単位容量当りの価格，媒体交換機の普及度を考慮すると, プラズマ実験装置の長期データ保存機構としては, 大型装置では DLT, DVD, 中小型装置では CD-R 等が有望視される．

近年のキャッシュ技術の進歩によって，個々の記憶装置は, 既に殆んど半導体記憶によるローカル・キャッシュを備えて入出力の高速化を図っているが，短期および長期保存装置は, 現在まで明確に区別して利用されてきており, 収集データは半自動あるいは手動によって定期的に移しかえ作業が行なわれてきた．プラズマ実験の過去データに対する参照は, 所謂「グッド・ショット」に比較的集中する傾向があり, 長期保存装置に格納された過去データの参照をハードディスクによりキャッシュする手法が確立すれば，実質的に極めて良好なアクセス性が実現できる可能性が高い．

2.2　核融合実験における今までのデータ収集系

本節では，現在までプラズマ実験に実際に用いられてきているデータ収集系の特徴などを調査・比較検討した結果を以下に述べる．

2.2.1　PC/EWS の拡張ボードによる直接収集

PC の拡張スロットに直接 ADC ボードなどを挿してデータ収集を行なうもので，拡張スロット数や同パネル面積の制約から，収集できるデータのチャンネル数は少ない．また，デジタル信号源が多くノイズの巣である計算機の内部に通常あまりシールド性の考慮して製作されていない拡張ボードを挿して使うため，デジタルノイズが計測アナログ信号を汚してしまう可能性が高く，精緻な計測信号の取扱いには向かない．

ノート型 PC の拡張バスである PCMCIA では，カードサイズ的に直接の実装が困難なので，次節に述べるような I/F 接続になり，カード単独での収集は基本的にデスクトップ機以上の計算機に限られる．簡便に独立したデータ収集系が組めて可搬性も良いという特徴があるが，拡張性に乏しい点は否定できない．

大学研究室などでの小規模テストスタンド的なデータ集録で良く使われる．

2.2.2　PC/EWS と I/O バス・インターフェース接続での利用

データ収集にある程度のチャンネル数が必要な場合や，既存の ADC モジュールなどの計測器を使用したい場合あるいは現地調整時などに良く利用される方法である．一台あるいは数台までの CAMAC, VXI といった計測バスや単独の計測機器を，PC の拡張スロットに挿した SCSI, GP-IB 等の I/O バス・コントローラ経由で制御する方法で，PC 側はコントローラ・カードのみで済むのでノート型 PC でも構築が可能であるなど可搬性が高い．

LabVIEW 等の機器制御，データ収集/表示が一体となったソフトウェアが良く利用され簡便に制御，収集が行なえる．このためメーカー独自仕様の I/F なども含め，各応用分野・用途に見合った多種多様な規格が存在しているが，反面，接続可能な対応機器がメーカー製品に限定されたりする場合も多い．

SCSI, GP-IB でもそうであるが，もともと周辺機器とのデータ授受を想定しているため，光エクステンダーなど遠隔利用ができなかったり，あるいは可能であっても速度が低下するなどの制約が生じたりする．

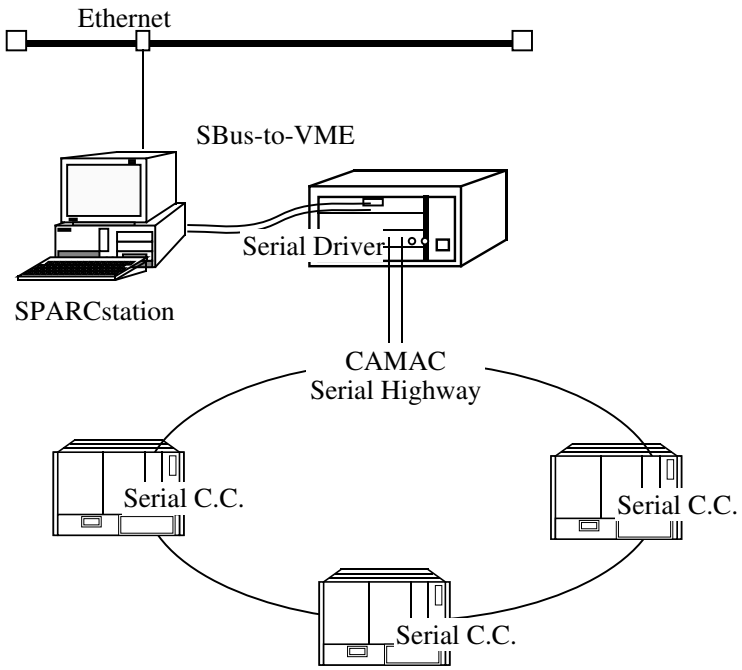


Figure 2.4: Schematic view of CAMAC serial highway controlled by the VMEbus serial driver

2.2.3 EWS/ミニコンと CAMAC ハイウェイの利用

ここ 10 数年で広く一般に普及した EWS をデータ収集計算機として利用し，それまで中規模システムで主流だった専用ミニコンの代替を行なう際によく採られるシステム構成の一つである．CAMAC デジタイザ，クレート・コントローラとハイウェイ構成はそのままで，収集計算機に接続するハイウェイ・ドライバのみを EWS の拡張バスあるいはシステム・バスに接続可能なものに交換して置換えが出来るので，従来構成を最小限の変更に留めて移行できるのが長所といえる．しかし最大の問題点は CAMAC ハイウェイ・ドライバが高価過ぎる点である．

核融合分野では，次節に述べるパラレル接続用 CAMAC ドライバ・ソフトウェアをもとに，日本原子力研究所 (JAERI) の青柳哲雄氏が Sun SPARCstation 上から VME バス版シリアル・ハイウェイ・ドライバを制御してデータ収集を行うソフトウェアを開発されている．このソフトウェアを用いた具体的応用事例として JT-60 や TST (東大理) [30]，WT-3 (京大理) [31] 等がある．

現在入手可能なシリアルハイウェイドライバは対応する I/F の種類が少ないため，バス変換機などを用いる必要のある場合が多い．例えば，上記の Sun SPARCstation シリーズの拡張バス SBus 対応品がないため，Figure 2.4 にも示したような，(SBus) → SBus-VMEbus

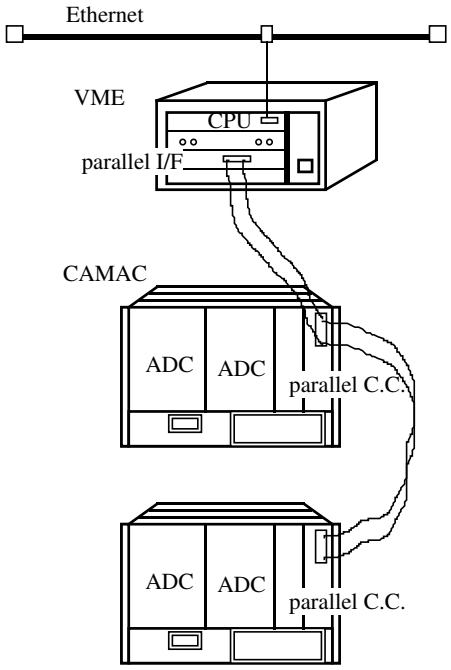


Figure 2.5: Simplified diagram of the parallel connection between VME and CAMAC

変換 I/F → VMEbus 用シリアルハイウェイドライバ → CAMAC シリアルクレートコントローラ，といった接続形態が良く使われる．また同じ CAMAC シリアルハイウェイを利用する形態でも，Figure 2.5 のようにホスト側を VMEbus 型の SPARC CPU ボード等にしてバス変換を少なくすると共に，収納性を向上させる導入事例も JT-60 等では増えている．

2.2.4 EWS/ミニコンとパラレル・インターフェース接続の利用

前項の CAMAC シリアル・ドライバの代りに VMEbus 用パラレル・インタフェースとパラレル・クレート・コントローラの対を使用する方法である．SPARCstation でのデータ収集の他に，VMEbus をシステム・バスとして採用しているデスクサイド型 Sun4 や VME ワンボード型 SPARC 等でも利用可能である［32］．これらの接続形態では，VMEbus-パラレル変換 I/F を用いる事例が多く，Sun SPARCstation(SBus) → SBus-VMEbus 変換 I/F → VMEbus-パラレル変換 I/F → CAMAC パラレルバスクレートコントローラ, という経路になる．Figure 2.5 を参照.

パラレル I/F は，最もよく使用されている KineticSystems 社製 KSC2917-Z1A の場合で最大 1.14 MB/s の転送性能がある．実測値もほぼ同程度あり［33］，VMEbus 上での DMA 転送も可能なので CAMAC データウェイからのデータ転送帯域としては十分である．こ

のドライバ・ソフトウェアは，SunOS 上の Sun4/VME の上位ドライバとして東工大の竹内康雄氏により開発され［34］，現在も KEK にて保守されている［35］.

EWS の OS が UNIX であり実時間処理は機能的には出来ないが，利用時教育が容易で学生が馴染みやすい等の理由から特に大学研究室ほかでの利用が多く，ヘリオトロン-E (京大) や REPUTE-1 (東大) 等で採用されている［36, 37］.

2.2.5 EWS/ミニコンによるグループ構成

ミニコンの主力メーカーである DEC 社製の VAX/VMS システムでは，複数台のホスト間でディスク等のリソース共有を行なって，共同でデータ収集/処理を実行する “VAX クラスタ” と呼ばれるグループ型データ収集系を組むことができる．ネットワークの通信には独自の DECnet プロトコルが用いられていた.

VAX/VMS システムは核融合実験のデータ処理システムとして一世を風靡し今もまだ多くのサイトで稼動している．VAX アークテクチャそのものは既に消滅しているが，その OS であった VMS (現在は OpenVMS) ともども現在は DEC Alpha システム上にその資産が継承されている.

従来型 UNIX との大きな相違点でもあるが, VMS では制御/モニターなどの実時間処理を行なう機能があり, 統合的な制御処理システムの構築が可能である．また，ホスト計算機やディスクなどのリソースを増設するだけで大規模システムにも適応できる拡張性を持っている．物理実験で利用されてきた歴史は長く，CAMAC 関連の対応インタフェースやドライバ・ソフトウェアなどが充実している点も重要である．JFT-2M (JAERI) ［38］や JIPP T-IIU (NIFS), TFTR (米 PPPL), Alcator C-Mod (MIT), CHS (NIFS) 他の多くのサイトで利用されている.

VAX クラスタ上の CAMAC データ処理システムとしては MDSplus が世界の多くのサイトで使用されている［39, 40］．MDSplus は CAMAC Serial Highway Driver を制御して CAMAC データ収集を行う基本機能に，ディレクトリ・ツリー型データファイル管理・取出しと各種の解析アプリケーションおよびグラフィカル・ユーザ・インターフェース (GUI) 等の機能が付け加わったデータ処理システムである．洗練された GUI 等見るべき点が多いが，CAMAC データ収集系の I/O が高価な Highway Driver のみに限定される．また，データ保存が単一のディレクトリ・ツリー上であり，それをネットワーク共有することでデータ共有を行うなど，大量データを取り扱おうとすると I/O ボトルネックが発生しやすい形態になっている．このため大容量のデータ収集系構築にはあまりふさわしいとはいえない.

VAX クラスタはその秀逸な思想がその後の UNIX や PC にも影響を与え，特に近年，多数の数 10〜数 100 台の PC を用いて大型並列計算を行う “PC クラスタ” が脚光を浴びてきている．これは VAX クラスタ等でのグループ内資源の仮想共有から，更に一歩進んだ並列計算処理環境を実現するものであり，今後のデータ処理システムの在り方にも影響を及ぼす重要な技術と予想される．2.4.1 節にて改めて今後への課題として取り上げる．

2.2.6 汎用機を利用した中央集中型

データ収集を中央集中型で行うことの最大の利点は，各計測装置の制御や計測データの収集，保存，表示等を一つのシステム上で行なえ，運用・管理・保守の人的コストが合理化できること，及び設備・資源を集中することで最大処理性能が稼げる点である．PC, EWS 等の小型計算機の処理性能が未だ不十分であった状況においては，大容量の収集データを扱うに十分な処理能力を得るため実効的な選択肢であった．

基本的にシステム構築・保守運用に費用が多くかかり，増強などに対応する場合でも周辺機器に民生普及品を利用するのが困難で，コスト高になる面は否定できない．そのため，PC や EWS が非常に入手容易になっている現在，多数の計算機入出力ポートを提供する基盤としてはコスト・パフォーマンスが悪くなってしまっている．

しかし反面，製造メーカー等が長期にわたる継続的な保守対応を保証するため，過去に導入した装置の最新機器が 20 年後にも入手可能であるなど製品寿命の長さは特筆すべきものがある．JT-60 など実験プロジェクトが長期化する大型装置で使われることが多い [29]．

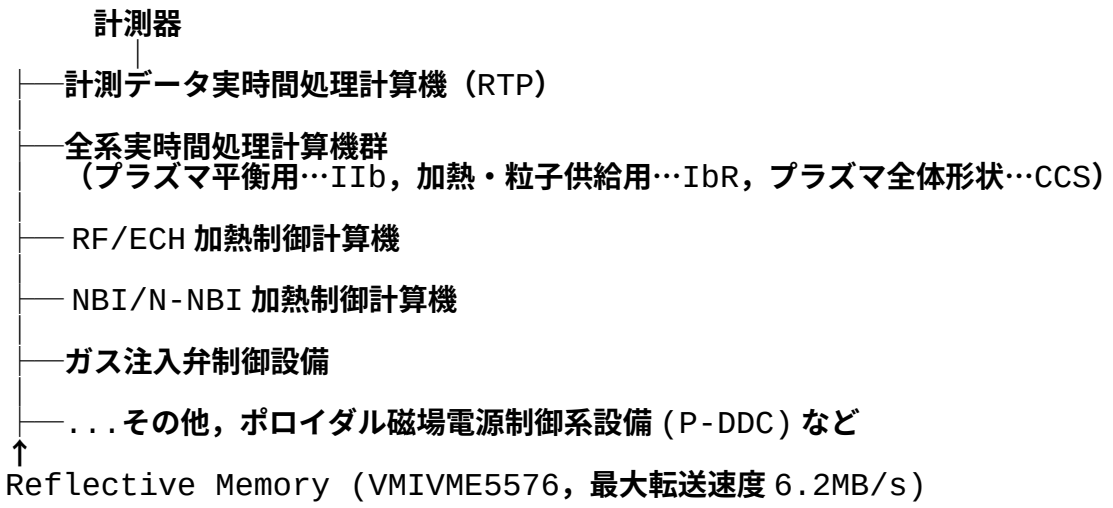
2.3　核融合実験での実時間処理系の現状

核融合実験の計測データ処理システムが対応を迫られている最も新しい課題が定常実時間運転，即ちリアルタイム処理である．ここではリアルタイム処理，長時間運転化のために核融合分野のデータ処理システムが今までとった異なるいくつかのアプローチ例を比較調査して，データ収集量が大容量化した際に取り得る方策等について検討を行う．

2.3.1　JT-60 のリアルタイム計測制御系

JT-60 のリアルタイム処理をつかさどる計測制御系の概要を書くと以下のようになる．ここで全系とは JT-60U の各種運転設備（コイル電源，計測，NBI，RF，ガス注入系等）を管理・統括し所定のプラズマを生成するための中枢部分である．プラズマ実験を遂行するために必要な動作指令を手順どおりに出力し管理制御する放電シーケンス制御と，プラズマ着火から消滅に至るまでプラズマの位置，形状制御や加熱を行うプラズマ制御機能から成る．計測信号のうち Magnetics 計測信号など実時間フィードバック制御に使用する種々の信号もここで収集・処理されている．温度・密度などその他のリアルタイム計測信号は計測データ実時間処理計算機 (RTP) で処理される．JT-60 の放電実験間隔は約 14 分で放電持続時間は最大定格運転時 15 秒となっている．

今まで実績をあげてきているリアルタイム処理の事例としては，全体形状の同定と制御，ECE 計測による磁気島制御や，Wp 計測による NBI 入射ビーム制御でのプラズマ β 値再現性の画期的向上，などがある．以下に計測制御系でリアルタイムデータを共有するために用いられているリフレクティブ・メモリ (RM) の接続系統を模式的に示す．



このように RM で全設備がつながっており，各々運用分担されている[注5]．各部の内容は以下の通り．

- RTP はシステム内に ADC を有しており，その ADC を介して計測器の生信号を取り込み，生信号をプラズマ物理量に換算するための計算やフィルタ処理を行っている．
- 全系計算機（プラズマ平衡用）IIb は，電磁気信号を取り込んでプラズマの位置形状の全体像等の平衡諸量を高精度に求め，それらを他の計算機が使えるように RM に書き込む．またプリプログラムされたプラズマ位置などを維持するように，各ポロイダル磁場コイル用電源への電圧（電流）指令値を計算し電源設備に渡す．これも専用の RM ネットワークで接続されている．
- 全系計算機（加熱・粒子供給用）IbR は，RTP が RM に書き込んだデータを読み込み，場合によっては物理量計算のために更に演算を行い，あらかじめ決められた制御ロジックでアクチュエータ（NBI，RF，ガス注入弁）への指令値を計算する．その指令値を RM 上に書き込む．
- NBI,RF, ガス注入弁等の設備制御用計算機は全系計算機（加熱・粒子供給用）IbR からの指令値を RM から読み込んでそれぞれの設備を制御する．

計測器生信号を物理量に換算する演算は，RTP と全系計算機 IbR の片方あるいは双方で行われる．RTP で全て行わない理由は，特定の制御ロジック処理で必要なパラメータ演算はその制御ロジック内の処理と位置づけて IbR で計算するというポリシーに基づいているが，RTP で二次的な物理量に変換する場合もある．実例としては，RTP は ECE 計測信号から 30 (8) ch 分の電子温度データ 100 サンプル（2 ms 分データ）をまとめて平滑化 (LPF) し RM に書き込む．全系計算機 IbR はそれらのデータを RM を介して読み込み，電子温度データ特定の空間位置での電子温度勾配を計算してプリプログラム指示値に合うようにアクチュエータを制御する．

全系計算機（加熱・粒子供給用）に組み込む制御ロジックは，例えば電子温度勾配のプリプログラム値に比べて実測値が小さくなったら中心加熱用 NBI を必要なユニット数だけ入射するように指令をだす，必要なユニット数は PD 制御の考え方で計算する，等である．

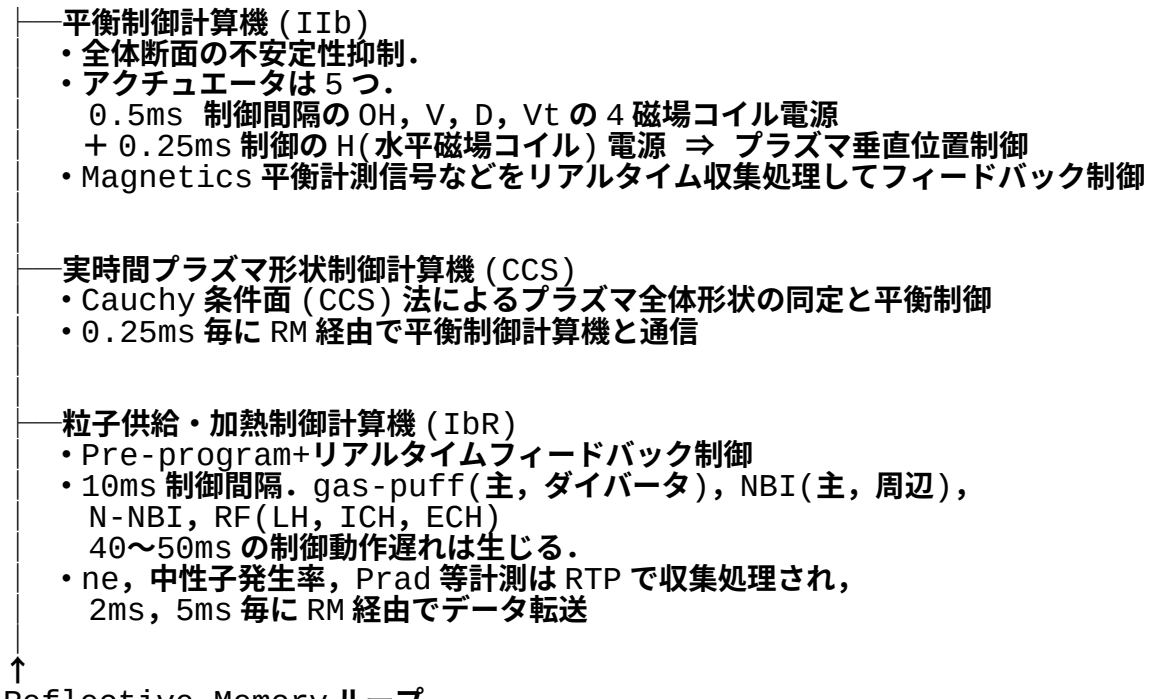
新しくリアルタイムフィードバック系を１つ作るには，まずどのプラズマ物理量をどのようなアクチュエータを使ってどのような制御ロジックにより制御を行うかを決める．そして，RTP での演算アルゴリズム，全系計算機での演算アルゴリズムと制御アルゴリズムを決めて実装が行われる．安全の為のインターロック，異常処理についても，各設備担

[注5] 正確には RM は 3 系統 (3 階層) 存在するため，あくまでも模式的系統図である．

当者と打ち合わせて全系計算機に反映される．例えば，ECE による電子温度勾配が 200 keV/m を超えた場合には RTP で計測データ異常と判定 (計測データ異常ロジックは主に RTP で実施) して，全系計算機では電子温度勾配 FB 制御を停止する等である．

全系実時間計算機群の構成

全系の実時間処理計算機群には，平衡制御 (IIb) 計算機システム，粒子供給・加熱制御 (IbR) 計算機システム，および実時間プラズマ形状制御計算機システム (CCS) 等があり，この間をリフレクティブメモリ (RM) のループでつないでいる．



リフレクティブメモリー中のデータ内容は Table.2.3 の通りで，これらがリアルタイム制御用計算に使用される．IbR で実際に現在行われているフィードバック制御には，ガス注入速度を制御する中性子圧力比 FB 制御や，NBI 入射ユニットを制御する蓄積エネルギー FB 制御などがある［41, 42, 43］.

計測データ処理実時間計算機 (RTP) の構成

RTP の扱う生信号のチャネル数　RTP は 2 枚の 16 ch A/D 変換ボード（max. 32ch）を実装しており 9 ch 入力と 25 ch 入力の二つの運転パターンがある．A/D 変換値は 1 ms

Table 2.3: Real-time data contents transfered through the reflective memory ring in JT-60:

類別	信号種	チャネル数	サンプリング周期 (ms)
命令	PFC 電圧	5 ch	0.25/0.5 ms 更新
	ガス流量	4	10 ms 更新
	NBI on/off	14	10
	RF on/off	4	10
装置状態	NBI 準備完/否	14	10
	NBI on/off	14	10
	RF 準備完/否	4	10
生計測データ	磁気計測	50	0.25/0.5 ms 更新
	コイル電流	6	0.25/0.5
	位置	5	0.25/0.5
	主要計測量	10	10 ms 更新
			… 中性子発生率, 放射パワー, 電子温度など
処理済データ	プラズマ形状	1k	1 ms 更新
	温度分布	15k	10 ms?
	電流分布	15k	10 ms?
クロックカウンタ		2	1 ms, 10 ms

毎 (1 kHz) のソフトウェア読出しで，ADC 自体は読出しタイミングと無関係に自走している．

速度　仕様上は 50 μs/1 6ch だが実測試験で 25 ch 入力時に約 85 μs である．(但しこの 85 μs 中にはアプリケーション・プログラムからのバッファ読み込み時間も含まれる．) アプリケーションプログラムから，複数のデータサンプル方法が設定可能．25 ch 運用時には 1 サンプル/1 ms でダブルバッファ読出し，(8+1) ch 運用時にはリングバッファ読出しを行っている．

系統の収集方法　CAMAC シリアルハイウェイを介した I/F 通信で行う場合と，A/D 変換器で直接アナログ（+/-10 V 入力）信号をデジタル化する場合との 2 種類がある．

リアルタイム計測データの種類・個数　各々にデータ取込・演算処理・取得データ健全性チェック・転送の各処理が伴う．

- FIR 電子密度
- CO2 電子密度

- 中性子発生率
- ダイバータ部放射損失量
- 中心部放射損失量
- 周辺部放射損失量
- ダイバータ領域密度
- ダイバータ部中性粒子圧力
- ダイバータ部中世粒子圧力比
- フーリエ分光器

CAMAC 収集系との連携　RTP の VME 計算機から CAMAC へのアクセスは，KSC2140 シリアルドライバを使っているが，１アクセス (1CAMAC ファンクション発行) に平均 20 μs かかり遅い処理になっている．

RTP 本体はコンカレント日本 (株) 製 MAXION 9000 (MIPS R4400，150MHz，2CPU)，OS は RTU (並列リアルタイム UNIX) を使用しており，本システムの特徴は RTU OS 上のソフトウェア制御である．また CPU 本体が VME モジュールであるため，VMEbus 規格の A/D 変換ボード，RM ボード等を実装することが容易である．本システムへの命題として，

- ２つのプロセスを同時並行に動作させる
- その内の１プロセスは高速リアルタイムに動作させる

があり OS およびハードウェアが選定されている．しかし，昨今の PC あるいは EWS の安価で高速なハードウェア，汎用 OS を利用し，現システムと同様なシステム構築が可能であれば，保守運用にかかる手間が大幅に低減できる可能性がある［44］．

RTP 最大のリアルタイム計測 … ヘテロダイン ECE ラジオメータ電子温度 (Te) 計測

50 kHz 16-bit ユニポーラ (0〜10 V) ADC 8 チャネルを用いて，Te 主成分と Te 揺動成分を実測している．8 ch × 50 ワード/1 ms を 2 回に分けて 256 kB FIFO から読込むため，8 ch × 100 ワードを 2 ms かけて収集していることになる．データ処理ループ手順は以下のプロセスを各々 1 ms 毎で処理し 1.〜4.を終了まで繰り返すものである．3. 4. 処理中の 2 ms はデータを読み飛ばすので 2 ms 毎の読込みになるが，ADC 自体は，放電開始から終了までの間，連続変換動作を続けている．

0. 処理開始命令

1. データ読込み 1/2　…　実測で処理時間 max. 552 μs
2. データ読込み 2/2　…　同 552 μs
3. 演算処理　…　同 640 μs
4. 転送・格納　…　同 500 μs 余

この連続ループと同時並行して以下の処理が進められる.

- 演算結果は 1 ms 毎に 8 ch の出力値を RM を経由して全系 IbR に転送.
- 生データはローカルで保存.
- 各チャネル信号を一次係数とオフセット補正値を用いて一次変換し Te を算出，次に 100 個 (2 ms 分) データで平均を出して，各チャネルごとにそこからの分散 (rms) を計算，Te と共に RM に書込む.

基本的に JT-60U では放電持続時間が max. 15 s なので，リアルタイム表示は殆ど行っていない．例外としては，6 本のフィラメント近似を用いて磁束ループ信号からプラズマ境界や Ip を表示する独立収集＆表示系があるのみである.

このように電流系プラズマの保持を行うトカマク型装置ではアクティブ・フィードバック制御を欠かすことができず，この帰還ループ系では 0.25〜10 ms 間隔でリアルタイム計算処理が組み込まれている．各種アクチュエータを帰還制御するには通常このような時間応答を持ったリアルタイムの計算機割込み処理が必須であるが，同じ動作アルゴリズムでプラズマ揺動計測などが要求する ∼MHz 級の高速サンプリングを処理することはできない．このためリアルタイム計測制御系は，JT-60 の場合と同様，∼MHz サンプリングの計測データ収集系とはシステム的に分離して構築されるのが一般的であった.

しかし昨今の計算機処理能力の劇的向上により，汎用 OS 上でも実効的に ∼1 ms 程度の応答時間での処理が可能になってきている［45］．この問題については 2.4 節で改めて検討する.

2.3.2　TRIAM-1M 長時間データ処理システム

九州大学応用力学研究所で運用されている TRIAM-1M 実験装置は，JT-60 などと同じトカマク型でありながら，超伝導電磁石を用いてプラズマ保持のための定常磁場を発生するユニークな装置である．現在 TRIAM-1M は 10000 秒を超えるトカマクプラズマの最長連続運転記録を保有しており，パルス運転とは異なった長時間運転のためのデータ処理システムが独自に開発され運用されている［46］.

TRIAM-1M データ処理システムで特に注目すべきは，長時間データ処理のデジタイザ・フロントエンドに従来と同じ CAMAC ADC を用いている点である．本来，CAMAC デジタイザには仕様的にプラズマ計測のサンプリング周波数を取り扱えるだけの実時間動作性能はない．しかしここでは，計測信号のケーブリングからデジタイザ，収集ホストへの転送まで完全に対称な 2 系統に分け，それらを一定時間で交互に切り替えて繰り返し動作させることで，長時間の連続運転を可能にした．

この方式は同一信号あたり 2 系統のデータ収集系が必要でデジタイザ・リソースの利用効率は半分に下がっているわけであるが，定常運転にそのままでは対応できない CAMAC デジタイザを使用できるという意味で，核融合実験の現有デジタイザ資産の有効活用に新たに道を拓いている．

LHD の定常化実験に対応する際に CAMAC 資産を活用する方法の一候補として，4.3.1 節でも改めて検討を行っている．

2.3.3 TFTR イベントトリガー収集系

TFTR は米国プリンストン・プラズマ物理研究所 (PPPL) が保有した世界三大トカマク実験装置の一つである[注6]．TFTR のデータ収集系では，中性子発生率など特定の計測アナログ信号強度に対してあらかじめ基準値を設定し，計測信号がその閾値を超えた場合に電子回路的にトリガー信号を生成，他計測にトリガーを配って ADC 等のデジタイザを駆動するいわゆるイベントトリガーシステムを開発した．

プラズマの安定保持技術の進捗に伴って持続時間が延びた状況では，あまり変化しない安定状態時のプラズマを詳細に計測する必然性が薄れてきた．また，有限サンプル数のローカルバッファをもつトランジェント・レコーダー型 ADC では，延長された放電時間をカバーするためにサンプリング速度を落として運転する必要があるが，これによる対応にも限界があった．そこで，プラズマが安定している時間帯の計測を省略し，何らかのプラズマ変動 (イベント) が発生したときのみ，それを検出してデータ収集を行う方式を実現した．

イベントトリガー方式は計測信号を重要な時間時間帯だけデータ収集することで，大容量化の一途をたどっている計測データを適正量に低減する効果があり，今後，核融合実験のデータ処理技術の一課題として更なる研究が望まれる．この方式ではトリガー生成ロジックをいかに構築するかもさることながら，利用できるトランジェント・レコーダー型

[注6] 1997 年に実験プロジェクトは終了している．

ADC にもいくつか制限がある．

1. STOP トリガー動作＋ポストトリガーサンプルの機能がある
2. ローカルバッファを複数領域に分割して使用できる
3. 複数トリガー (マルチ STOP トリガー) を受信して動作できる

トリガーイベントはプラズマ変動の結果として検出された変化なので，その変動原因を解析するためにも，通常イベントが発生した場合その前後の時間帯の状況を計測，データ収集する必要がある．このため ADC 自体は，トリガー待ち状態では常時 AD 変換をし続けて，トリガー受信時にバッファーへの書込みを確定する STOP トリガー動作を行う必要がある．

また CAMAC メモリーからのデータ取出しには数秒以上の時間を要するため，放電中に転送を行うと，この間トリガー受信ができないデッドタイムが生じて必要なイベントに反応できない可能性が出るので，放電中は転送を行わず複数イベント分のデータをローカルに分割格納する機能の必要である．当然，これら ADC はトリガーが発生した時刻を計測する TDC (Time-to-Digital Converter) と併用する必要がある．

イベントトリガー収集系は九大 TRIAM-1M でも研究開発と実証が行われ，定常的なプラズマ状態が続く長時間放電時に有効に機能することが確認されている［46］.

2.4 データ処理システムを取り巻く課題

プラズマ実験のデータ収集を取り巻く環境の近年における著しい変化として挙げられるのは，計測データの大容量化と PC の劇的普及，高性能化であろう．計測技術の進歩に伴って CCD に代表される多次元多チャンネル計測が一般化した結果，計測データは増加の一途を辿っており，特に PC 向け半導体記憶素子の大容量化とそれに伴う単価の低下がそれを可能にしている状況である．また PC, EWS の演算性能もほぼ記憶素子の容量増加に歩調をあわせて高速化してきている．

こうした半導体技術の向上に対して，データ伝送路として主要な役割を果たすバスの伝送性能は必ずしも同期的に進歩しているわけではなく，検出や変換，処理，収録といったデータ収集システムの各構成要素の性能向上を生かしてデータ収集系の総合性能を上げる際のより重要な検討項目となりつつある．データ収集系の設計ではこのバス利用の設計が非常に重要な位置を占めているといえるだろう．

JT-60 に代表されるような電流系プラズマの保持を行うトカマク型装置では，大型装置になるほどプラズマ安定化のためのアクティブ・フィードバック制御装置を欠かすことができず，世界三大トカマクが稼動した 1980 年代後半から核融合プラズマの帰還制御技術が発達してきた．こうした帰還ループ系では，アクチュエータの反応速度に見合った 0.25 〜10 ms 間隔でリアルタイム計測制御の反復計算処理が組み込まれている．

各種アクチュエータを帰還制御するには通常このような時間応答を持ったリアルタイムの計算機割込み処理が必須であるが，同じ動作アルゴリズムでプラズマ揺動計測などが要求する ∼MHz 級の高速サンプリングを処理することはできない．このためリアルタイム計測制御系は，∼MHz サンプリングの計測データ収集系とはシステム的に分離して構築されるのが一般的となり，2 つの計測系を別系統として開発・運用する負担が不可避であった．

この問題の根本は，

	デジタイザ	運転形態	転送速度	サンプリング速度
計測制御系	VMEbus	リアルタイム処理	$<$ 40 MB/s	$<$ 10 kHz
データ収集系	CAMAC	バッチ処理	$\sim$ 1 MB/s	10 kHz $\sim$ 1 MHz

というように計測制御系とデータ収集系との間で，利用デジタイザの運転形態・性能に相互の互換性がない点にあるといえるが，昨今の計算機処理能力の劇的向上により，汎用 OS 上でも実効的に ∼1 ms 程度の応答時間での処理が可能になってきている．このため

1. CAMAC のように高速サンプリング (〜1 MHz) が可能で,
2. VMEbus のようにそれを高速 (〜40 MB/s) にリアルタイムで転送・処理できる,

能力を備えた新たなデジタイザ・システムがあれば，現在まで別々に必要であった 2 系統が 1 系統で済むことになり，人的予算的な負担を大きく低減できる可能性がある．

本研究では LHD 定常化プラズマ実験装置の計測データ処理システム開発の一環として，この課題に取組んでおり，それについては第 4 章にまとめて詳述している．

2.4.1 PC クラスタと Grid コンピューティング

個人用小型計算機いわゆる PC が昨今劇的普及と性能向上を同時に果たしたことにより，PC を複数台同時に用いて大規模な並列計算をする環境が実現しつつある．特に近年の Intel Pentium シリーズに代表される PC 用 CPU の演算能力は数〜10 年前の大型汎用機の演算性能と匹敵するほどであり，これを多数並列に用いることで非常に容易かつスケーラブルに大規模高速演算が可能になっている．

こうした並列処理を実現する基盤としては，PVM あるいは MPI といった並列通信ライブラリがあり MPICH パッケージなどとして公開されている［47］．これらは TCP/IP ベースの Beowulf 型，あるいは高速な Myrinet 等の NIC/プロトコルが使える SCore 型の並列計算環境などとして実際に提供されている［48］．

このように PC クラスタはローカルエリア内に多数存在する PC 群を利用する手法であるが，これを広域に分散した各種資源の集合的利用に拡大して大規模計算を行おうという計算機構が Grid である［49］．

PC クラスタや Grid コンピューティングでは，ユーザ利用環境と共に大規模並列計算の基盤リソースが提供される．特に Grid 等が効力を発揮するのは，数値シミュレーションや構造解析，モデル計算などように計算の入力初期パラメータが少なく，中間/出力結果が大きい SIMD 的な計算処理である．

データ収集系のようなデータ・インテンシブな I/O システムには Grid のような広域分散系はそぐわない点が多いが，PC クラスタに関しては，これを I/O 系に応用して高速大容量な仮想 I/O サーバを構築できる余地はある．これが実現できれば計測データ処理システムの高速化大容量化への課題も解決されうる可能性がある．少なくともデジタイザから一旦収集サーバ上に転送された生データの一次的な解析計算処理については，現在各所で稼動している PC クラスタの性能から見ても容易に実現可能であると考えられる［50］．

次章では，PC クラスタ的構成による高速大容量な仮想 I/O サーバの実現性評価を視野

に入れながら，核融合計測データ処理システムが直面する課題解決に向けた研究の成果を報告する．

第３章

大規模分散系の設計と開発
～LHD 計測データ処理システムの構築～

LHD (Large Helical Device) は，Wendelstein 7-X と並び現在実験プロジェクトが進行しているヘリカル型の大型核融合プラズマ実験装置である［51］．本研究が対象としている核融合プラズマ計測のデータ処理システムは，前章で確認したきたとおり，ヘリカル型やトカマク型といった核融合炉心プラズマの配位といった物理的な側面の違いに依存する点は少ない．しかしその反面，実験の規模や短パルス運転/長時間あるいは準定常放電実験といった実験モード，あるいは計測フィードバック制御などの運転形態によってその内容が大きく異なってくる．

1980 年代に建設された大型プラズマ実験装置である JET，JT-60，TFTR の世界三大トマカクから現在までほぼ十年余の年月が経っており，この間のデータ処理系を取り巻く環境の変化は，特に計算機やネットワーク技術の分野でめざましいものがある．また半導体技術の進歩は，プラズマ計測の分野にもその影響を及ぼし，CCD (Charge Coupled Device) 等の多次元計測素子や大容量半導体記憶素子 (RAM) によって計測データは大容量化，多様化の一途をたどっている．

従来から較べて桁違いに大容量化している実験データを取り扱うためには，汎用機や専用ミニコンを用いる従来からの中央集中型システムが目指した，資源集中によるデータ処理システムの高性能化ではもはや対応が困難となりつつある．このため，LHD に限らず現在稼動中および今後運転開始予定の核融合実験データ処理システムでは，従来の設計とは全く異なる新たな基本設計による開発が緊急に必要とされるようになってきている．特に，最も近年に稼動しはじめた大型実験装置 LHD では，最新鋭のプラズマ計測技術が多数投入されているため，その傾向を顕著に見ることができる．LHD では約 10 秒のパルス

放電 1 回当たり約 600 MB，3000 チャネルのプラズマ物理計測データを実験当初から収集する計画になっており，実験開始後のそれ以上への増加も充分予想されている．

この章では，システム・アーキテクチャ設計の基礎になっている要求仕様分析とアーキテクチャ設計への基本指針の決定について先ず述べた後で，機能的に分離されたデータ収集系および計測制御系のサーバおよびクライアント開発について節を設けて詳しい述べる．それに続いて，計測データシステムからのデータ取出しインターフェースなどクライアント側とのデータ取扱い手法について触れ，それを仲介するネットワークの利用法やアプリケーション構成などを述べる．そして最後には，LHD のような大規模な計測実験で日々増加していくデータに対して，いかにすれば大規模データ・ストレージシステム (Mass Storage System: MSS) がスケーラブルな構成でかつ増強増設が容易な分散形態で構築できるか，大量に保管された計測データの参照サービスを継続的に提供できるか，という大きな課題にのぞんだ LHD 計測データシステムの内容について詳説する．

また本稿末尾に収録した付録 A では，共同研究ネットワークによる LHD 遠隔実験参加環境実現への取組みについて述べている．

3.1 システム要求諸条件の分析

核融合プラズマ計測の進捗状況はデータ収集を行う計測数やチャネル数，総データ収集量等の変遷で観察することができる．Table 3.1 に LHD 実験開始前のプラズマ計測計画の一覧を示す．当然，この表に掲載されている計測器のみで LHD 計測が完遂するわけではなく，実験開始後にも予算などの状況に応じて順次増強され，またその他の計測も増設されることになる．LHD に関しては計測計画は実験開始とともに極めて順調に進行しており，続く Table 3.2 には，最も新しい 2003 年 2 月 7 日の実験番号#41312 の計測データ収集状況も併せて比較のために掲載している．LHD データ処理システムとしては，先ずはこの当初の計測計画に沿って，計測信号約 3000 チャネル分のデジタイザを制御し，1 実験あたり約 600 MB〜 の計測生データを収集・処理できるシステム開発が目標となる．

ここで最新の実験装置の多くに共通する状況として，超伝導電磁石を用いた定常磁場発生装置を導入していることから，従来の短パルス磁場発生装置を持ったプラズマ実験とは異なり，強力なプラズマ閉じ込め磁場を発生するための電源電力を蓄える時間や，あるいは巨大な電磁石が大電力の消費により発熱して高温になるのを除熱する時間をとる必要が全く無いことになる．このため大型の常伝導装置では，たとえば JT-60 では約 14 分など放電実験の間隔を比較的長くあけて運転を行うのが普通であるが，超伝導装置の LHD では，

Table 3.1: LHD data acquisition plan for plasma diagnostics. This estimation was compiled in 1995, about three years before the beginning of LHD plasma experiments.

Diagnostics	channels	data (/MB)	estimated cost (/JPY)
Radiation calorimetry	400	100	50 000 000
VUV spectroscopy	6	60.8	9 000 000
Magnetic probes	300	75	112 500 000
X-ray fluctuation	300	75	112 500 000
MMW interferometry	16	6.5	11 000 000
FIR interferometry	90	18.7	25 000 000
Thomson scattering	1024	2	58 840 000
ECE	337	77	127 400 000
X-ray PHA	24	3	54 396 000
NPA	15	7.8	2 200 000
Crystal spectroscopy	98	8	12 000 000
CXR spectroscopy	68	58	28 600 000
Langmuir probes	228	76	114 000 000
HIBP	26	4.3	9 100 000
Magnetic surface profile	12	17	1 500 000
Neutron ditector	-	-	-
plasma common	96	12	20 600 000
Total	3040	601.1	748 636 000

小型のプラズマ実験装置と同程度の 3 分間隔という非常に短い放電間隔で連続的に運転を行うことが可能になっている．この運転条件はデータ処理システムにとっては，小型装置と比べて 1〜2 桁も大きい計測データを同じ時間で処理する性能を要求されることになり，従来システムにはない高速処理性を実現しなければならないことになる．実際，LHD 装置のプラズマ放電は 3 分間隔で 1 日あたりおよそ 150〜250 ショットを連続して運転することになっており，実験を遂行する上でもショット間の 3 分の間に全ショットのデータを解析・表示して，その結果で次ショットの放電条件を修正する必要から，実際には放電終了から約 1 分以内に全データを収集・処理して表示する能力が求められることになる．

Table 3.2: Recent status of CAMAC data acquisiton in shot #41312. (Feb. 7, 2003)

Diagnostics	channels	raw data size (/Byte)	compressed size (/Byte)
SXfluc	81	19 660 816	6 390 148
Reflectmetry	12	7 864 320	1 914 695
PHA	6	50 331 648	158 305
ICHVOLT	60	15 728 640	961 467
Halpha	126	33 030 144	4 661 543
Brems	120	29 884 512	3 946 725
PHARD	16	85 458 944	236 104
Bolometer	60	15 728 640	2 378 875
FIG	6	1 572 864	321 905
SXmp	81	19 660 816	2 463 082
Fastion	13	35 651 600	274 985
ECH	30	22 020 096	3 852 121
ImpMon	30	7 864 320	895 673
Magnetics	78	20 447 232	3 352 519
Langmuir	96	29 884 416	3 516 689
TESPEL	30	7 864 320	581 629
MMWINT	6	12 582 912	1 108 276
AXUVD	78	20 447 232	3 609 141
SX8O	67	49 545 232	11 680 140
Langmuir2	30	62 914 560	19 569 405
HIBP	24	6 291 456	440 488
RADL	72	47 185 920	16 017 866
GPCRADH	96	62 914 560	8 789 607
SiFNA	1	8 388 608	170 712
MMimg	72	67 633 152	29 885 216
DTS	128	5 632	4 646
Total	1 419	740 562 592	127 181 962

前章でも述べたとおり，汎用機など大型計算機を用いた世界 3 大トカマク装置の一つである JT-60 のデータ処理システムでも 1 実験あたり CAMAC 系～50 MB，VMEbus 系～500 MB 程度のデータ収集量である［52］．しかもデータを処理・解析する時間として使える放電間隔は上述の通り約 14 分と長く，その結果，一日当たりに行われる実験回数も 20～30 回程度に限られている．これに対して放電間隔が 3 分と短く，1 実験あたりのデータ収集量が当初計画で 600 MB，その後の実験の進捗によって～1 GB を超えるような大容量・高速なデータ収集系は，核融合プラズマ実験の分野ではこれまで存在していなかったといえる．

LHD の建設に伴い実験を終了した NIFS のトカマク実験装置 JIPP TII-U のデータ処理システムでも，1977 年の稼動以来データ収集量の増加にあわせて二度のシステム増強を行い，性能実績としては 2 分の実験間隔で 8 MB/shot の処理性能であった．1995 年 9 月にシャットダウンする直前の最終実験では，約 500 の計測チャネルで 24 MB/shot 程度の収集量である．このように既存システムと LHD へのシステム要件との 1～2 桁もの性能差を埋めるためには，全く新しいシステム設計思想によることが LHD データ処理システムの開発には必要になっている．

超伝導電磁石を利用したプラズマ放電実験中は連続して定常磁場を発生し続けている．しかもこの磁場は高温プラズマ閉じ込め用のため，LHD ではプラズマ中心部で約 3～4 T 程度，電磁石コイル中心では約 7 T の磁場強度と非常に強力になっている．こうした強力な定常磁場により，プラズマ実験中は本体装置が格納されている実験室には人の立ち入りが禁止されるため，本体に付属して設置される各種の計測機器へのアクセスも実験中には制限を受けることになる．

このためそれぞれの計測器には，プラズマ実験中に機器運転の状態監視と操作を行うために，遠隔監視および操作が可能である必要がある．また特に LHD の場合，本体を格納している大型ヘリカル実験棟が実験を制御遂行する制御棟とは離れた建屋になっている都合上，各種機器の遠隔監視・操作のために各々の専用線を配することが困難であり，機器の拡張性なども考慮すると，基本的に計測機器はネットワーク経由で監視・制御を行わなければならない．

計測器の遠隔監視・制御性はまた，共同研究を積極的に推進・支援する NIFS および LHD プロジェクトの姿勢からも重要な事項であり，インターネット経由で計測機器が遠隔監視・制御可能になっていることで，遠隔地の共同研究先からもリアルタイムで LHD 実験に参加でき，今後望まれるであろう核融合実験の新たな形態に応えるシステム要件になっている．

上記のような状況は LHD に限らず，今後多くの実験装置に共通となることが予想され，これを新たな計測データ処理システムへ望むべき要件としてまとめると以下のようになる．

1. 0.6〜1 GB/shot のデータを放電終了後約 1 分以内に処理できる高速性
2. 3 分間隔で 150〜200 shot/day の連続運転に耐えられる大容量性
3. 全過去データが 24 時間オンラインになっており常時取出し可能であること．
4. 計測器がネットワーク経由で遠隔操作・監視できること．
   遠隔実験参加が可能なオープンなシステム形態になっていること．
5. 拡張性が高くスケーラブルなシステム性能が出ること．
   計測器の増設・増強に容易に対処でき，かつ性能等に悪影響を及ぼさない．

こうした要件をどのような方針により解決・満足にしていくか，そして具体的にどう実現・実装していくかを次節以降で述べる．

なお，本開発研究の成果でもあり，以降に具体的な実現・実装法を詳説するデータ処理システムに対しては，総称として *LABCOM* **システム**という呼称を与えているので付記する．

3.2 スケーラブルな分散データ処理システムの基本設計

新たに開発される計測データ収集・処理システムは，最新鋭実験装置の計測データを将来の永きにわたって処理が可能であり，また不特定多数の計測データ利用者に対応できるようにという基本要件に基づいて仕様策定と設計がなされなければならない．最新鋭装置ということでは，高度化されてきている高温プラズマ計測技術を反映して，LHD では 30 種類にもおよぶ多種多様な計測器に対応する必要があり，当然，取り扱うべき計測データも同様に多様化している．

反面，ネットワーク上に分散して存在する計測データの利用者に対しては，データ取扱いの便宜上もデータ処理システムは統一したユーザ・インターフェースを提供する必要があり，多種の計測器毎に独立して処理システムを開発するのでは効率が悪すぎて現実的ではない．つまり，計測器の増設増強に随時対応可能であるシステム構成をとり，構成変更が起こっても稼動中の部分には処理速度等を含めて悪影響を及ぼさない独立性の高いシステム設計と，各種計測器やデータ種の違いを吸収して統一された単純でかつネットワーク対応したデータ取扱いインターフェースの実現が求められている．

これらを実現するには，一計測を制御・処理する基本構成要素がシンプルかつコンパクトであり，多くの機能を一箇所で集中的に処理するのではなく，各構成要素が単機能化・分離されていて他機能との連携が基本的に疎結合となっている必要がある．またこれは，制御・収集動作のみならず格納・ユーザ取出しなど他の全ての機能面についても成り立たなければ不十分である．以上をより具体的な設計指針として書き下してみると，

- 並行分散化されたオープンでかつスケーラブルなシステム
- 開発業務の分散化と負荷 (開発量) の大幅低減化
- インターネット技術に基づいた疎結合な連携動作
- 小型計算機を全面的に採用してシステム増強・更新を容易かつ柔軟に
- 汎用 OS/市販ソフトウェアをベースにして開発・保守コストを低減

といった方策を新たに採用することになる．基本的な市販/汎用ソフトウェアをシステムのベースに採用するという最後の条項は，分離分散された単機能処理を担う計算機はコスト的にも小型計算機ベースであるべきという前条項から，同じく開発保守コストを考慮すると容易に得られる結論といえる．特に個人用小型計算機，いわゆる PC (Personal Computer) には現在様々なフリーソフトウェアが提供されており，これらを効率的に用いることで大幅に開発コストの削減も期待できる．

3.2.1 大規模並行分散 (MPP: Massively Parallel Processing)

大容量化した実験データを分割して同時並行に処理することで，処理に掛かる時間を短縮すると共に，提供する機能以外の処理のためにそれぞれの計算機に大きな負荷がかかることのない様に，担当すべき処理毎に計算機を分けて利用し，予め負荷分散を図るのが，並行分散システムの要点である．Table 3.1，3.2 でも見られるとおり，LHD 計測ではおよそ 30 種にものぼるプラズマ計測が予定・準備されており，計測ごとに取り扱うデータ種・量，それに参照形態・参照するユーザが異なっている．このため計測間で相互に処理負荷の影響を与え合うよりも，計測ごとに独立して稼動できるデータ収集処理システムを実装した方が，データ利用者にも開発者にも見通しのよいシステム形態になるといえる．並列に分散された単体の処理エレメントをつなぐ高速ネットワークは，こうした分散配備された各計算機間の相互データ参照や動作を同期するためのタイミングメッセージ伝送路として極めて重要であり，同時並行の分散処理のため時間的に集中して発生するネットワークアクセスを遅延なく十分処理できるだけの広帯域・高速伝送性が求められる．

このように，単機能ごとに分散された多数のサーバ計算機群をまとめて高速ネットワークで接続することで，サーバ計算機群に対して処理要求を送るクライアント計算機からは，あたかもそれが一個の仮想的マクロ・マシンであるかのように見え且つ振舞うことが，LHD データ処理システムの目指す姿といえる．これは，高速ネットワークを内部通信路として使用する疎結合型の並列処理計算機システムに他ならない［53］．Figure 3.1 は仮想的マクロ・マシンを構成するサーバ計算機群の概念構成図である．各種計測を並行分散してデータ収集処理を行うデータ収集計算機群と，分散して収集保存された計測データを統一して参照可能にするためのデータ収集・所在情報をあつめるインデックス・データベースや過去データを保管し参照要求に応じて提供するデータストレージ・サーバ，生データを処理する一次解析サーバ群などで構成される．

3.2.2 機能分散

放電実験終了直後に計測データを一括収集・処理して保存するデータ収集サーバと，24 時間連続で計測機器の状態を監視し且つリアルタイムな遠隔機器操作の機能を提供する計測制御サーバとは，計測データ処理システムの中でも運転・制御する機器の運転形態や利用目的の違いが大きいため，分離して別の計算機システム上に実装・動作させる方が，システム全体が負荷分散され個々のエレメントの独立性・安定性向上に大きく寄与できる．

データ収集サーバ系統は前節に述べた並行分散形態をとったが，後者の計測制御サーバ

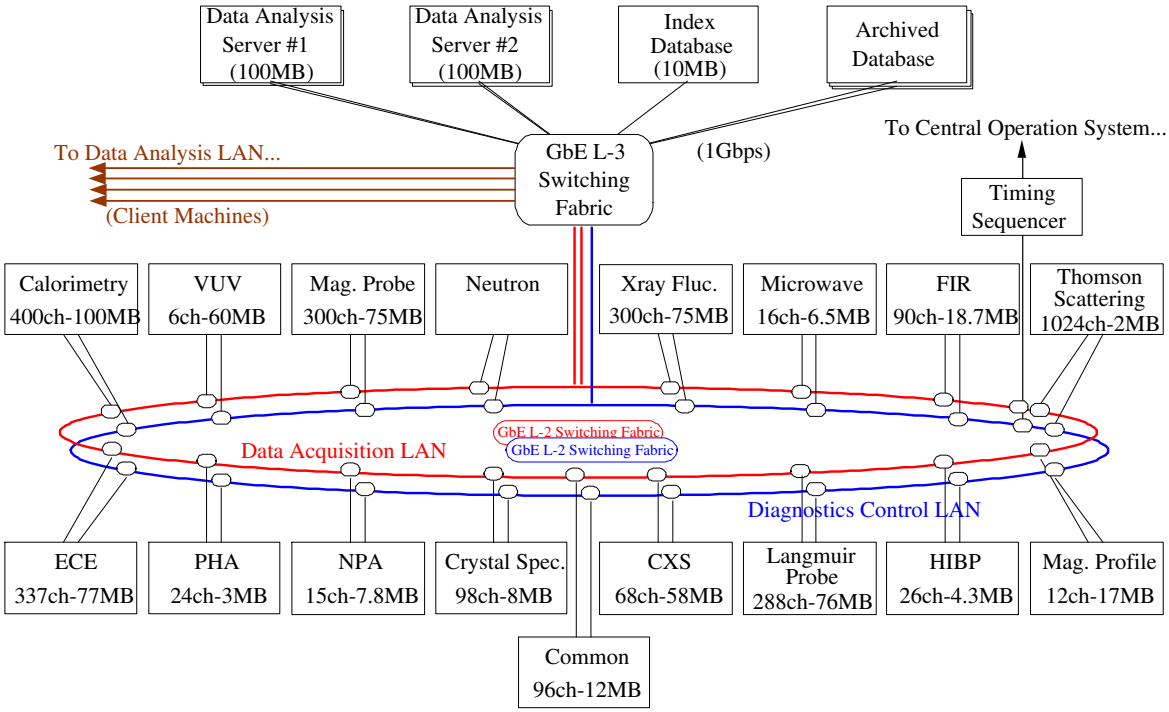


Figure 3.1: Schematic view of LHD diagnostics configuration. The network switching fabrics were at first FDDI based switches and routers, which had been replaced by the Gigabit Ethernet (GbE) multi-layer switches for the throughput improvement in Mar. 2001. Now they are Alcatel OmniCore5052 (L3) and Cabletron SSR8600 (L2), respectively.

系統でもそれに対応して，分散的に配置される計測装置を遠隔制御するためネットワーク分散構成となる必要がある．データ収集系では，計測データを生成するデジタイザから収集計算機まで，それぞれ独立した専用伝送路を並列に用いることで，伝送のボトルネックを下げる効果が期待できる．これに対して，機器の保護インターロックや機器プロセス制御を行う計測制御計算機は，RS-232C や GP-IB，あるいはデジタル接点入出力といった多種の制御用 I/F を取り扱う必要から，結線の都合上も各計測器が設置されている現場で稼動する必要があるため，稼動現場から直接ネットワーク経由で遠隔アクセス可能な動作形態が望ましい．そのため近年のインターネット技術の基礎になっている TCP/IP 通信により，計測制御の各機能が呼び出せるようなシステム実装が求められるとともに，通信路としてはインターネット的なネットワーク線路の共用が必要になる．幸い計測制御系はデータ収集系と異なり大容量データを取り扱う必要がないため，こうした接続形態が可能である．

遠隔にいるユーザが対話的に端末となるクライアント計算機と，データ閲覧・機器制御

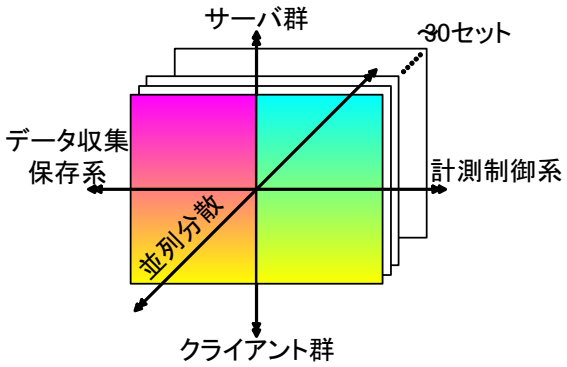


Figure 3.2: Schematic diagram of functional separations. The third dimension can be understood as the parallel distribution.

などの各種機能を提供するサーバ計算機との間のネットワーク利用を介した分散処理の形態としては，クライアント/サーバ (C/S) モデルが一般的であり，C/S の機能毎に処理が分離されることになる．この場合 C/S モデルに沿って，以下のように

クライアント側		サーバ側
処理呼出し	→	処理機能提供
結果受取り	←	処理結果返答

アトミックなリクエストベースで処理が進行し，ネットワーク通信はその間の機能呼出しのメッセージ授受が主になるため，インターネットを用いて線路共用を行っても通信に支障が出る恐れはない．

このような機能的分離にさらに加えて，30 種類に及ぶ計測器を運転する LHD 計測では，計測器毎にデータ収集サーバと計測制御サーバとを 1 セットのデータ処理エレメントとして分散配備することで，個々の計測器のデータ収集を各々独立に同時並行的に実施できる構成とした．概念図を Figure 3.2 に示す．当然，各々の収集系が実験のシーケンス進行に同期して動作する必要から，分散された各エレメントでの同報タイミングメッセージ受信は必要になるが，機能分散や大規模並行分散 (Massively Parallel Processing: MPP) の形態をこのように縦横に用いることで，スケーラブルな構成を保ちながら多種多量のデータ処理装置の取扱いが可能になる．

3.3 大規模分散形態とオブジェクト指向開発

約 100 倍にも大容量化し更に増加を続ける計測データへの対処として，計算機資源を集約化する従来の中央集中処理型アーキテクチャを拡張しても，I/O 集中によるシステム性能低下の回避に限界がある．このため I/O 能力を自由に増減できる大規模な並行分散処理型アーキテクチャへの移行と，クライアント・サーバ・モデルによる役割分散，データ収集と計測機器制御の機能分離，データ・ストレージの階層化などによる徹底した分散化により，将来にわたるシステムの柔軟性とスケーラブルな拡張性の実現を目的とすることは前節までに述べたとおりである．

換言すれば，本研究が目指すスケーラブルな分散系システムは，機能毎の分離を徹底して各々の単機能化分散化を図ることで並列処理数の増減による I/O 性能のスケーラビリティを得る設計方針となっている．これは，各単機能ユニット毎に独立な多数のプログラム群を作成することに他ならないため，その間の取り合いを全て作成しなければならないといった開発の負担増を招き，また分離された各々の機能モジュール開発間の連携も複雑化する．

このソフトウェア工学的問題に対して一つの解決法を示唆するのがパラダイムとしてのオブジェクト指向方法論 (OOM: Object-Oriented Methodology) であり，それに基づいたソフトウェア開発がオブジェクト指向開発 (OOD: Object-Oriented Development) である．

大規模に分散化されたシステムでは，機能ごとに独立性を保ち主体的なプログラム記述が可能であるこのオブジェクト指向方法論に則った OOD が普及しつつある．オブジェクト指向システムは，反面，生来的に冗長性や処理オーバーヘッドによって I/O 性能が低下するという短所も併せ持っている．しかしながら，作業の集約化が難しい従来手法に対し，OOD ではシステム全体の見通しが改善され開発効率が向上するため，将来のスケーラブルな拡張性を持つ大規模な分散システムの開発には OOD の適用はほとんど不可避だといえるであろう．

このような状況においても，核融合実験のデータ収集処理システムでは，予め全て規定されたプリ・プログラムドな実験・データ処理進行という核融合プラズマの短パルス放電実験の固定的処理観念から脱却することが出来ず，オブジェクト指向システムでの I/O 性能低下の危惧もあって，今まで既存システムを全面的に更改する試みはなされなかった．

しかし，特に LHD のように最新鋭実験装置では，現実の問題として従来と同程度の開発力でのシステム構築・プログラム開発作業が困難になってきており，同分野でも OOM

に沿った開発の検証が早急に必要となっている．

以上を踏まえて本研究では，OOD による統一されたシステム開発に取り組み，開発の効率化と I/O 速度などシステム性能とを両立できる手法を LHD 計測データ処理システムの構築・プログラム開発をとおして探る．

3.3.1 大規模分散システム開発の効率化問題

大規模な分散処理システムを構築する際にまず問題になるのが，開発体制とその効率的な業務遂行であり，これはソフトウェア/システム工学の面で常に無視できない重要な研究課題となっている．

従来の C，BASIC，FORTRAN に代表される第 3 世代言語 (高級言語) が，いわゆる構造化プログラミング言語で処理の流れ/手順を記述するものであったが，OOM の世界では，様々な定義 (クラス) を持つ複数の機能主体 (オブジェクト) がそれぞれに関連をもち相互に作用反作用を行うことによって，全体として所定の動作を実現するという全く新しい処理概念を提供した．

従来の構造化言語では，モジュール化されたそれぞれの処理ルーチンの中でも全体の処理の流れやプログラム全域を常に理解してプログラミングを行う必要があり，大規模システム・分散システムの開発においては，多数の開発要員に多くの共通理解を求めることになり，開発をより複雑化し困難なものにしていた．

これに対して OOD では開発の単位が各オブジェクトクラス記述となり，個々の開発をより主体的自己完結的に進めることができるようになった．これは複数での開発業務分担を非常に容易にするため，本研究開発では次節にも述べるように，OOM を全面導入し大規模分散システム構築のために開発業務の効率化を目指した．

こうしたコード開発あるいはソフトウェア保守の業務の分散化は，必然的にその成果であるシステム・プログラムのポータビリティも改善することになり，中長期的な人的負荷の低減を実現する．更に LHD 実験については，今後の共同研究ベースでの共同システム構築をも容易にする効果が期待できるため，こうした分散開発に沿う形態ということでも OOM に基づいて進めることの利点は大きいといえる．

また OOM では統一モデリング言語 (UML) のようなシステム要求分析・設計の各種ツール類がオープン規格で提供されるなど，大規模システムの設計には欠かせない設計支援機構もパラダイムとともに提供されているため，今回の LHD データ処理システムの開発には導入が不可欠と判断された．

3.3.2 オブジェクト指向方法論の全面適用

LHD データ処理システムでは，開発課題の一つとして実験データのオブジェクト指向的取り扱いを目指している［54］．これは個々の実験データの種別を，

- データ取得に必要な各種パラメータ
- 実験データ配列
- データ処理/操作用の関数

を一体として定義するクラスとして定め，実際の実験データをこのクラス定義を雛型としたオブジェクトとして生成，取り扱う方法である［55］．この手法の特徴として，

1. システムの構成単位が実体のあるオブジェクトになるため，システム設定が容易で相互関係が理解しやすい
2. データの生成からデータベース格納，操作まで一貫した実験データに対する取り扱いができる
3. クラス定義がサーバ/クライアント側，或は計算機ハードウェア，OS の別に関係なく利用でき，ソフトウェアの再利用性が高まる
4. データ変換関数等を各オブジェクトと一体で扱え，データ・オブジェクトの抽象化が容易に実現できる
5. 通信をオブジェクト間のメッセージ授受として実現するため，ネットワーク層を意識せずにアプリケーション間の通信を行なう事が出来る

などデータ収集系への採用に利点が多いと考えられるため，多様化する実験データの取り扱いを抽象化し複数計算機上で再利用可能な統一的手法を得る手段として，オブジェクト指向によるシステム構築を LHD に向けた開発では採用した．これは同時に，全ての実験データをオブジェクトとして画一的に扱うことでシステム全体の統一的理解を容易にし，ソフトウェア開発効率の向上を目指すものでもある．

オブジェクト指向データ処理システムでは，データ・オブジェクトは装置の制御パラメータ等を用いて生成，初期化され，また，オブジェクト自身が所属するクラス内で定義されたデータ処理関数によって変換を受けることになる．Figure 3.3 では，こうしたデータ収集や計測制御のサーバ計算機での実験データ・オブジェクトの変換，処理の模式的流れが示されている．

制御する機器の種類が多くなる計測制御に於いては，多様な制御ハードウェアや通信形態の違いを一旦吸収し，ユーザがクライアント計算機上でサーバ側の制御ハードウェア

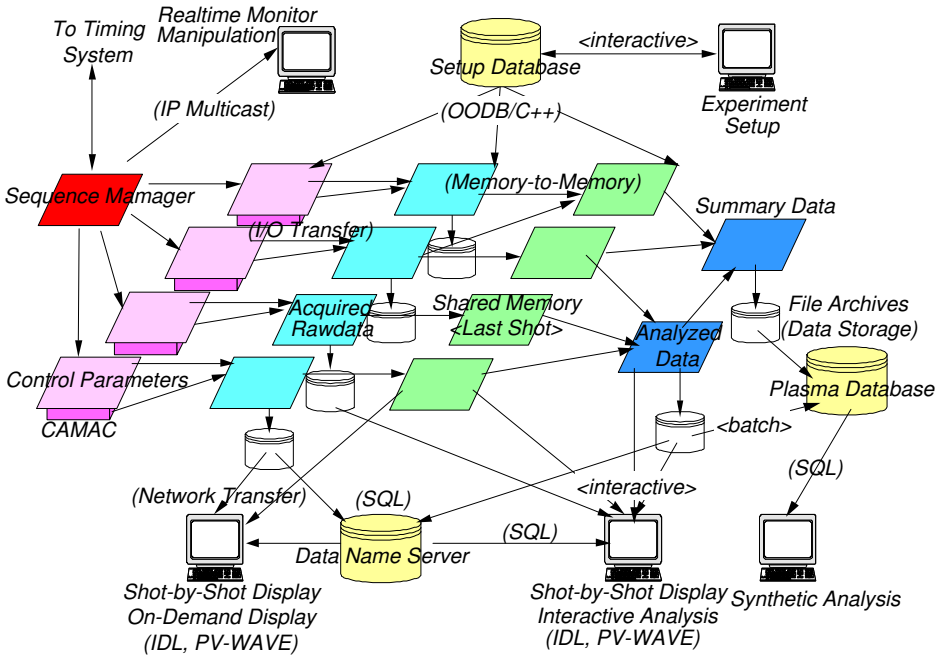


Figure 3.3: Schematic view of data objects' flow in the object-oriented data acquisition system. Within the OO system, data objects will be constructed by using initial conditions like device control parameters, and also transformed themselves with self-defined methods.

の違いを認識しなくて済む機構が重要になる．また，収集データの種類や量が多い LHD データ収集系でも同様の事がいえる．このためには，インタフェース/コントローラを制御する下位層制御オブジェクトから，一旦，計測及び制御の基本単位となるチャネルを表す論理的なチャネル・オブジェクトへの抽象化が必要である．LHD 計測制御系に於いても，クライアント/サーバ間通信は，クライアント/サーバ双方で生成されたチャネル・オブジェクト間で行なわれるメッセージ授受として実現される．このように，取り扱いの対象であるデータ・オブジェクトの抽象化とネットワーク下位層の隠蔽とを行なう事で，クライアント/サーバ・プログラム双方の独立性が高められることになる．

3.3.3 分散系の再結合化手法

前節に述べたオブジェクト指向モデルをネットワーク・アプリケーションとデータ共有とに実際に適用する方法として，LHD データ処理システムでは，最もシステム構築に信頼性のあるオブジェクト指向プログラミング言語として C++を選定した．これは作成された処理プログラムの性能が，他のオブジェクト指向言語と比べると C++コードの処理

速度が抜きんでていることのほかにも，利用実績とその安定性をみても自明の選択といえるが，そのほかにも多 OS/CPU アーキテクチャに対応しており，ソースコードを別の OS/CPU アーキテクチャに移植する際のコード互換性が大きいこともあげられる．コード互換性という意味では専用 Java VM (Virtual Machine) 上で動作する Java 言語も有用性が高い．C++コードに較べてかなり劣るとされてきた処理速度も近年は大きく改善されつつあるが，実メモリー領域や CPU など計算機リソースの消費が大きく起動に時間がかかるなど，大容量の生データを取扱う I/O 系にはそぐわない性質も依然残されている．このため Web ブラウザ上で Java Applet として動作させるなど，最大限の I/O 能力を要求されないクライアント GUI 用として充分利用効果があると判断した．

さて，システム分散化によって分散したコンピュータ上に生成・存在することになった各種データ処理オブジェクトの実体を，改めて再び連携させて処理を進行させるには，以下のような基本機構がその連携動作に必要となってくる．

1. オブジェクトの永続化＝永久保存
2. データ・オブジェクトをネットワーク上で共有/参照/複製する機構
3. オブジェクトがもつデータ処理/操作関数の遠隔呼出し

通常，C++などのプログラミング言語で記述・操作されるオブジェクトは，プログラムが動作する計算機上の主メモリー領域に揮発性の実体として生成・配置されるため，プログラムの動作終了とともに消滅してしまい，半永久的な記録として保存・保管することがそのままではできない．収集された計測データ・オブジェクトは当然保管されて，後日の参照要求にこたえて読み出されるが可能になっていなければならない．しかし，現在入手可能なほとんど全ての計算機 OS は，主メモリー上に存在するオブジェクト実体そのものを，アプリケーションプログラムから永久保存するような機能を提供していないため，何らかのオブジェクト永続化が可能な機構を導入する必要があり，これを提供するのがオブジェクト指向指向データベース管理システム (ODBMS: Object Database Management System) とよばれるソフトウェア・パッケージである．

ODBMS は，C++などの言語バインディング (=API) を通して，別の計算機で動作しているアプリケーション・プログラム (クライアント) のメモリー領域と，自身が永続的に保持している仮想記憶領域との間で，透過的にオブジェクト実体の共有/参照/複製を可能にするため，ODBMS のクライアント・プログラムとなる C++アプリケーション内では，C++で記述したオブジェクト実体をそのまま永続化・保存したり，あるいは永続化されているオブジェクトをメモリー上の C++オブジェクトと同様に取り扱うことができる．Figure 3.4 にその透過的取扱いの概念図を示す．

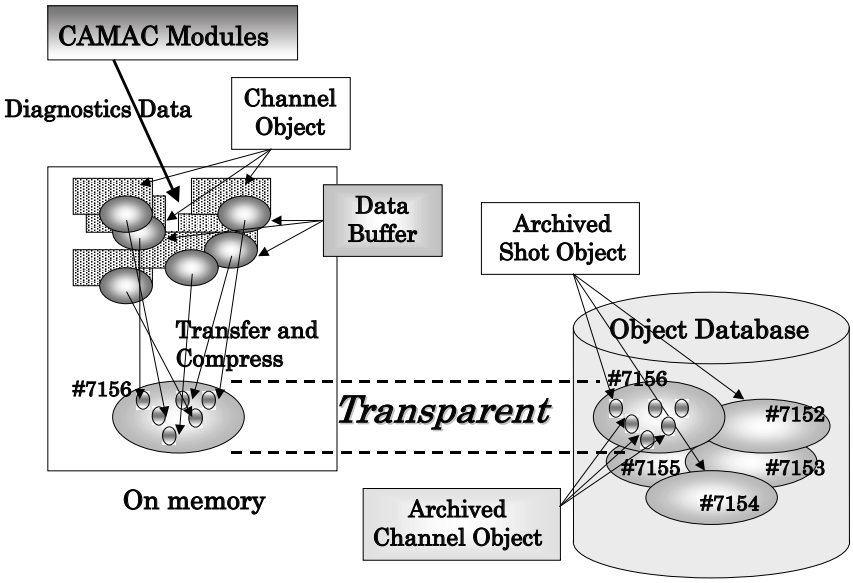


Figure 3.4: Object Transparency between in volatile memory area and in OODB persistent space:

また ODBMS は狭義にはデータオブジェクトの永続化・仮想格納領域を提供するサーバ側プロセスであり，その仮想領域をアプリケーション・プログラムの仮想記憶空間に透過的に接続してアクセスするクライアント・プログラム側との間のやり取りは，提供 API 中で完全に隠蔽されている．しかもほとんどの ODBMS では，クライアント・プログラムがネットワーク上の別ホストで動作するネットワーク・クライアント機能を完備しているため，提供 API を利用するクライアント・アプリケーションと ODBMS 間ではネットワーク経由のデータ・オブジェクト共有/参照/複製が可能になっている．

このように ODBMS の利用によって，半永久記憶媒体への保存の入出力手順を記述するのに費やされていた非常に多くの労力をほとんど完全に削減できることになるため，開発者側に大幅な省力化をもたらすとともに，不良コードを作りやすい入出力関連部分の開発作業を省いて，より不良発生の可能性を減らせるという多大な恩恵を与える．

3.3.4 分散オブジェクトの呼出し

保存の対象となる計測データ・オブジェクトについては，ODBMS がネットワーク上でオブジェクト共有/参照/複製の機能を提供することは以上に述べたとおりだか，分散した

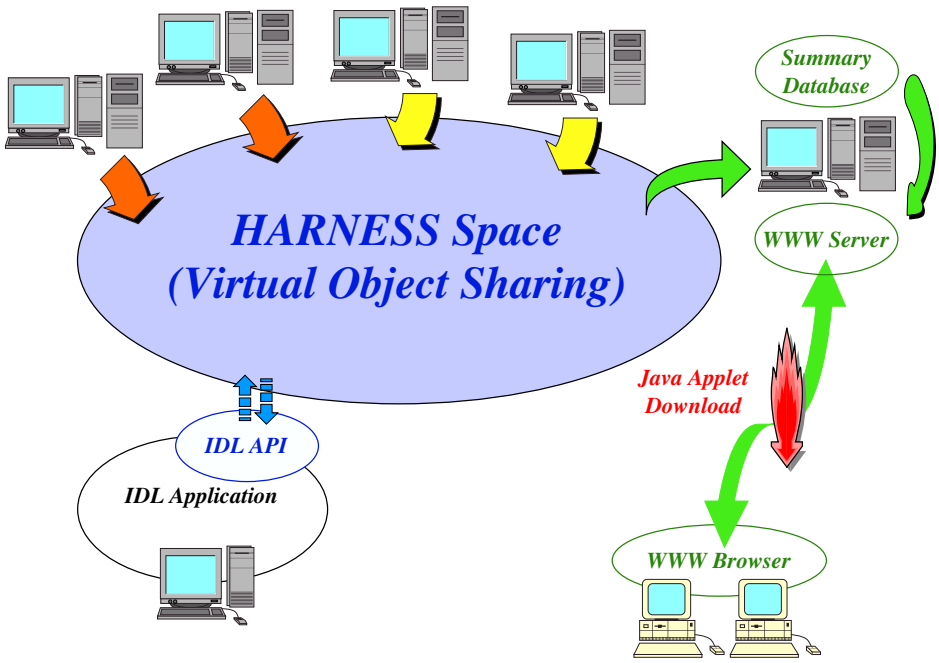


Figure 3.5: Schematic behaviors of network objects sharing by HARNESS.

ホスト上で動作する各プログラム中の揮発性 (＝主メモリ上) オブジェクト間の情報共有や，遠隔オブジェクトの処理機能を呼出すネットワーク対応ミドルウェアも必要であり，次の採用を検討した．

- データ・オブジェクトのネットワーク共有空間を実現する HARNESS
- 遠隔オブジェクトの処理を呼び出せる CORBA，あるいは
  データ処理/操作手続きの遠隔呼出しを可能にする RPC

前者は，データの共有と相互参照の比重が大きいデータ収集サーバ計算機で，後者は，メッセージ授受による対話的制御が主となる計測制御クライアント/サーバでの利用が主たる目的である．これらは，OSI[注1] ネットワーク 7 階層モデルにおける第 5 層 (セッション層)，第 6 層 (プレゼンテーション層) に相当する処理を提供するミドルウェアである．

HARNESS の前身は Nicholas Carriero, David Gelernter によって提起された有名な Linda［56］である．Figure 3.5 で示すとおり，HARNESS はネットワーク上の複数のホスト計算機に対して，共通のオブジェクト共有空間 HARNESS Space を仮想的に実現し，そこへの

[注1] OSI (Open Systems Interconnection): 国際標準化機構 (ISO) によるネットワーク通信プロトコルの標準規格.

オブジェクト書込み・取出しの API を提供するミドルウェアである［57］．揮発性オブジェクトをネットワークをまたがったホスト間で共有したり，あるいは一時的なオブジェクト保存領域としてオブジェクト転送のバッファリング/FIFO [注2] 入出力などが可能であった．このためメッセージ・オブジェクトの出現タイミングを待つことで複数ホスト間の処理を同期させることができるなど，OOM でのネットワーキングシステム開発には非常に好適で，C++アプリケーション・プログラムとネットワーク通信実装の間を完全にシームレスに接続できる可能性と先見性に富んでいる．このように HARNESS は OOD には非常に好適で，活用することが出来れば OOM に沿った開発の効率化に大きく寄与することが期待された．しかし，詳細は 3.4.3 節で改めて述べるが，大規模システムでの利用では充分な通信速度を維持することが出来なくなり，実装的には本格採用することは出来なかった．

遠隔オブジェクトに対する処理呼出し基盤としては CORBA (Common Object Request Broker Architecture) があり［58］，特定のプログラミング言語には依存しない多言語対応が可能になっている．CORBA は，ソフトウェア開発の生産性を向上させるオブジェクト指向モデリングや高い柔軟性を持った分散システム等の標準フレームワークの策定を行なうべく 1989 年に設立された中立の非営利標準化団体 OMG (Object Management Group) が定めた標準仕様で，分散システム環境のインフラを整備，標準化する．あらゆるオブジェクトのインタフェースを IDL (Interface Definition Language) と呼ばれる専用のインタフェース定義言語で定義するが，この IDL による異機種/異言語間の相互運用性の提供が最大の特徴となっている．

同様の機能がオブジェクト指向言語の基本仕様に組み込まれている例としては，Java 言語の RMI (Remote Method Invocation) が有名である［59］．また Windows OS に限定しては，Microsoft 社が提唱している COM/COM+ 環境も一般に普及が進んでいる．いずれもクライアント・プログラムがネットワーク上に存在する分散オブジェクトを呼び出すための基盤を提供するものであり，分散アプリケーションの作成を簡単化することができる．

COM/COM+ 環境は Windows OS 上のみのサポートで異なる OS との間で互換性がなく，また Java RMI は Java 言語独自の仕様であるため C++をベースにしたシステム開発に直接利用することができないが，CORBA については特定のプログラム言語に縛られず，あらゆる言語のプログラムを CORBA で接続することが可能になっている．

また CORBA 以外の既存の分散コンピューティング手法として以下の 2 つもある［60］．

[注2] First In First Out の略．はじめに (First) 入れた (In) データが最初に (First) 出る (Out) という意味で一時的な順次記憶メモリの一種．データ入力側と出力側で動作速度や読み書きタイミングが違う時に，お互いに好きなタイミングでの入力 (書き) と読み出し (出力) ができる．

1. RPC(Remote Procedure Call) を応用する
2. ネットワーク API(Application Programming Interface) を実装する

これらは CORBA や RMI に較べると，より低水準のネットワーク利用プログラミングを余儀なくし，OOM から離れてしまうことで開発負担も上昇するが，CORBA など上部階層のミドルウェアでは OS によって未対応などの問題もある．また，多くを隠蔽されている高水準ライブラリに対して，RPC やネットワーク API 実装では動作時のシステム負荷が見積もりやすいという点もあり，I/O 性能や処理速度を最重視しなければならないデータ処理システムでは，各々の適用箇所ごとに最適な手法の分析と検討・評価を行う必要がある．大規模な分散データ処理システムの各箇所に用いた最終的に実装手法については，3.4 節以降のシステム各部詳細に付随して述べていく．

3.3.5 データ参照インターフェースの仮想化とオープン化

ユーザからの対話的なデータ操作や取り出し・可視化などの処理には，ネットワークを意識すること無くデータ呼び出しが出来る単純で統一されたユーザ・インターフェースが提供される必要がある．

核融合実験においては，実験データは通常ショット番号と呼ばれる通しでつけられる実験番号で一意に管理される場合がほとんどである．実験規模がある程度大きくなると実験データの種類が増えるため，このショット番号にデータ名を加えた２つのキーにより各々の実験データを同定している．これは世界中の全ての核融合実験で共通した認識方法であるため，LHD データ処理システムでも踏襲しているが，多チャネル計測が一般化し各チャネルのデータサイズも大幅に増加している昨今の計測の状況を鑑みると，１ショット x １計測単位での取り扱いデータサイズがかなり大きく読み出し・転送などの処理に時間がかかるため，更にそれを細分化して各チャネル毎の読出しにも対応したほうが，無駄な参照を省くことが出来る．具体的な例で示すと，

```
> retrieve, "Bolometer", 12345, "1:16", dataArray [, ...]
```

というユーザからのデータ参照要求により，ショット番号#12345 の Bolometer 計測データのうち 1〜16ch に相当するデータ配列が dataArray に転送されるというユーザ・インターフェースが最も単純，基本的であり，かつユーザにも解りやすく使いやすいものである．

このデータ参照 I/F が異機種/異 OS 上で全く同様に動作すれば，データ参照ユーザは，計測データがどのように収集され，どこに保存され，どこから取り出して，どのように読むのか，を全く認識する必要がなく，自分の好みのデータ解析環境上に転送・再現された

データの後処理のみ，つまり解析や表示処理などに専念することが出来る．LHD データ処理システムでは，このようにデータ参照 I/F を仮想化しデータ処理システムを完全に隠蔽することで，参照要求に応じてデータを再現・提供するシステム側処理と，それ以降にデータを解析・表示するユーザ側処理とを明確に分離することを目指した．

換言すると，従来は中央集中型システムのメインフレームやミニコン上で，データ解析・表示プログラムがデータ収集プログラムと同じシステムの一部としてシステム開発者からユーザに提供されていたのを，解析・表示タスクを新たにユーザ側の開発課題として解放して，その開発負荷をシステム側と分離する分業体制，即ち，エンド・ユーザ・コンピューティング (EUC) あるいはエンド・ユーザ開発 (EUD) への移行を意図している［61］．これは単に今までお仕着せの GUI を使うことを余儀なくされていたエンドユーザが，より容易かつ自由に解析・表示環境を活用できるということだけではなく，データ処理システム全体の合理化や同システム開発の活性化の観点からも非常に有効と考えられている新たな方策である［62］．EUC/EUD により，情報システムと利用部門との関係／取り合い点を変更し，システムの大規模化に伴う開発のバックログ問題解消をも同時に狙った野心的選択である．クライアント/サーバ・モデルに即したクライアント/サーバ・システム (CSS) は，この EUC と共に採用することで，より効果的な開発活性化を期待できるといえる．

しかしながらデータ解析・可視化の作業環境を全てユーザの手によるプログラム作成に委ねることは，プログラム作成能力上基本的に不可能であり，そうした作業環境を提供するツールの導入・利用が不可欠である．社会一般的な基幹業務等では，こうした場合のアプリケーションツールが Microsoft Office，Lotus Notes，Cybozu Office などいわゆるオフィス製品群やグループウェアとなるわけだが，核融合実験では対話的なデータ解析・可視化環境を提供するツールとして IDL (Interactive Data Language) アプリケーションが DEC VAX 等のミニコン上で歴史的にもよく用いられてきた［63］．CSS 構成のクライアント計算機上で動作する製品としては，IDL より派生・進化した Visual Numerics 社の PV-WAVE/IMSL［64］や Research Systems 社の IDL［65, 66］などがあり，LHD データ処理ではこれらを標準ツールと位置づけて採用している．これらの製品はプロシージャと呼ばれるユーザ作成・定義が可能なスクリプト言語を有しており，データ解析・可視化の自動処理がスクリプト言語ベースで開発できるため，解析・データ可視化環境というだけでなく EUD 環境としても非常に好適である．

このように，統一されたデータ参照インターフェースと IDL アプリケーションという EUD 環境ツールを組み合わせて提供することにより，エンドユーザでも容易に解析・可視化プログラムの開発が可能になる．

3.4　データ収集保存系の設計と開発

プラズマ放電実験におけるデータ処理形態としては，一般的に

1. 物理計測のための高速サンプリングと放電終了後の一括処理
2. 実験中を通して機器制御や監視に必要なリアルタイム処理

の二つに大別されてきた．しかし，LHD 等においては準定常の長時間放電実験が行なわれるため，短パルス放電時にはフィードバック位置制御など限定された用途のみでの利用であったリアルタイム処理が基本的に全ての計測機器で必要になる．

このため LHD のデータ収集系開発では，上記二つの動作モードによって利用する計算機を分割し，計測器制御と密接に関連して同じ実時間動作をするリアルタイム機器モニターを計測制御サーバ計算機で行い，実験終了後一括してデータ転送，処理及び格納を行なうデータ収集サーバ計算機と分離した．リアルタイム機器監視制御に関しては次の 3.5 節で，また実時間運転時のリアルタイム物理計測は第 4 章で改めて述べることにして，本節では CAMAC を用いた高速サンプリングとデータの一括処理を行なうデータ収集サーバについての開発内容を述べる．3.2 節で既に策定したとおり LHD データ処理システムの基本開発方針としては,

1. 機能別及び各計測種類別の完全分散処理
2. クライアント/サーバ・モデル方式
3. 高速ネットワーク利用とインターネット・オープンシステム
4. オブジェクト指向方法論の導入
5. データベース管理システムの活用
6. 階層化されたオンライン・データストレージ

等となっている．

3.4.1　UML によるシステム分析と設計

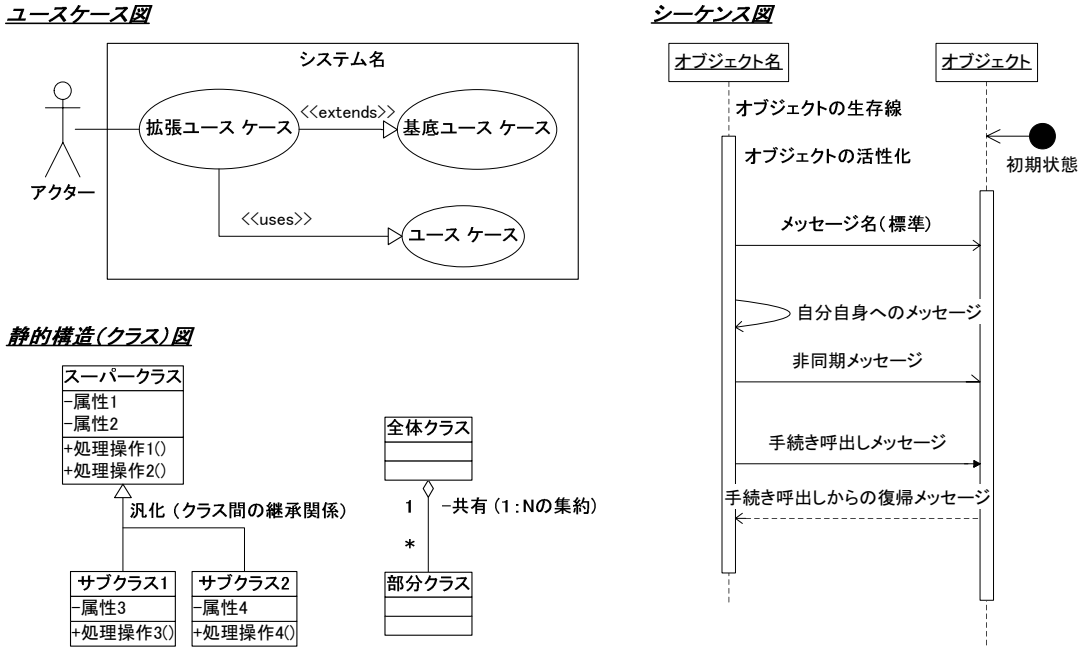
統一モデリング言語 (UML: Unified Modeling Language) は，ソフトウェアシステムの仕様策定，図示，体系化，ドキュメント化の方法を規程する言語で，ビジネス・モデルやソフトウェア・システムなど，大規模で複雑なシステムのモデリングに効果を発揮する各種エンジニアリング手法群を編纂して作られたモデリング技術である．UML が登場するまでは，標準となるような単一のモデリング言語はなく，UML の前身となった OMT、

Booch 法、OOSE など同じようなモデリング言語が多数存在しており，表記法が少しずつ異っていた．こうした標準技法の欠如による OO システム開発の阻害を回避するべく，幅広い用途に適用可能なビジュアル・モデリング言語の共通語として UML が OMG (Object Management Group) の標準化活動によって定められた［67］.

UML は設計技術の標準化を通じて相互運用性と移植性をシステム開発側にもたらすとともに，開発要件の収集，システム分析，ソフトウェア設計における標準化は標準化された実装アーキテクチャが必然的に導き出されることで，開発者はシステム開発に専念できるとともに，その開発を成し遂げるためのパラダイムをも同時に与えれれることになる．このため，少数に限られた要員による高効率な開発が特に求められる LHD のような研究用実験装置のための大規模システム開発では，オブジェクト指向方法論の一環として UML を最大限に活用することが，開発に極めて大きな利益を与えるものである．

Figure 3.6，3.7 に，今回のデータ収集保存系システムの要件分析 UML であるユースケース図を示す．この開発研究のシステム分析・設計においては，UML として主にこのユースケース図と，静的構造 (=クラス) 図 (Figure 3.8)，シーケンス図 (Figure 3.9) を用いた．

＜＜ UML ユースケース図，静的構造図，シーケンス図の凡例＞＞



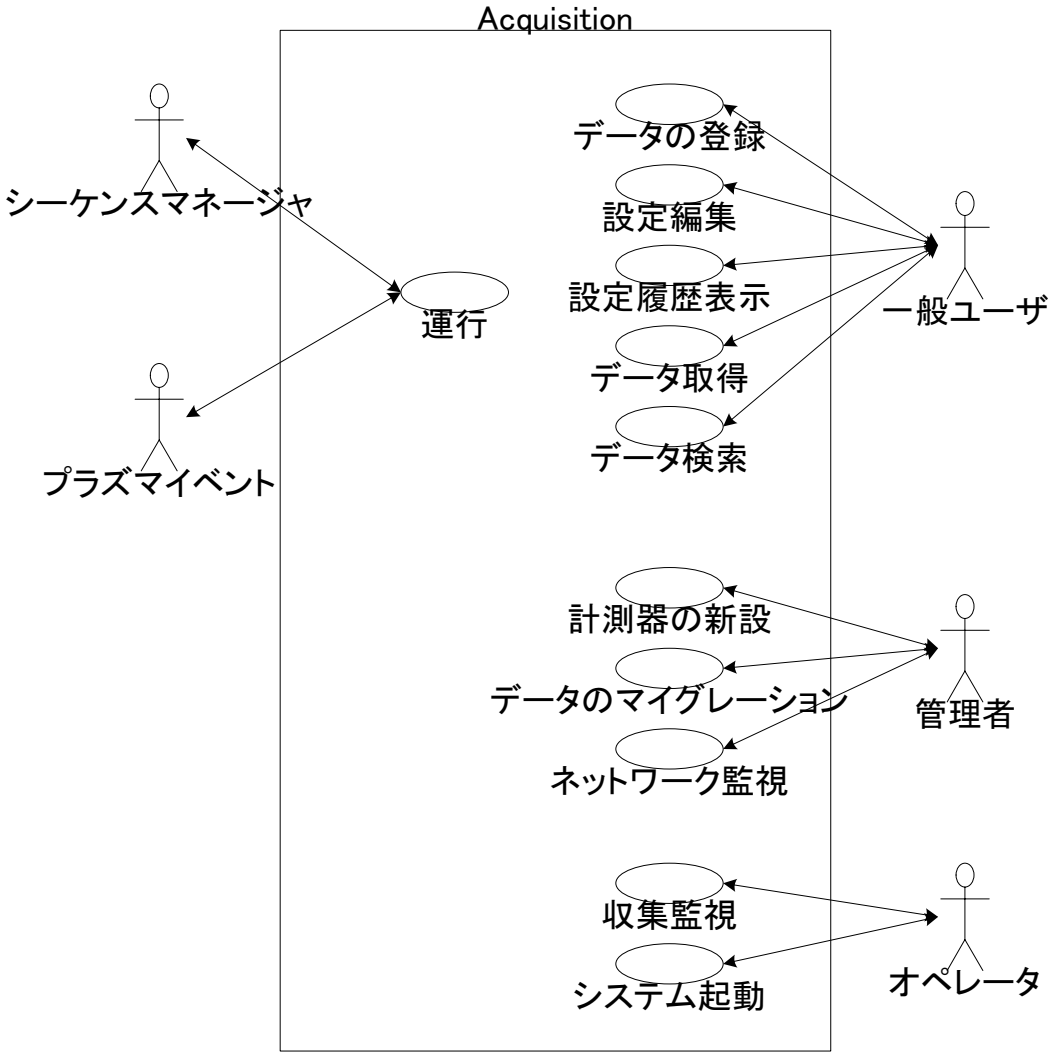


Figure 3.6: ”Use-Case” system requirement analysis for LABCOM system: “Use-Case” chart is one of the UML tools which is quite useful to do an object-oriented system design by making use-cases and their actors very clear.

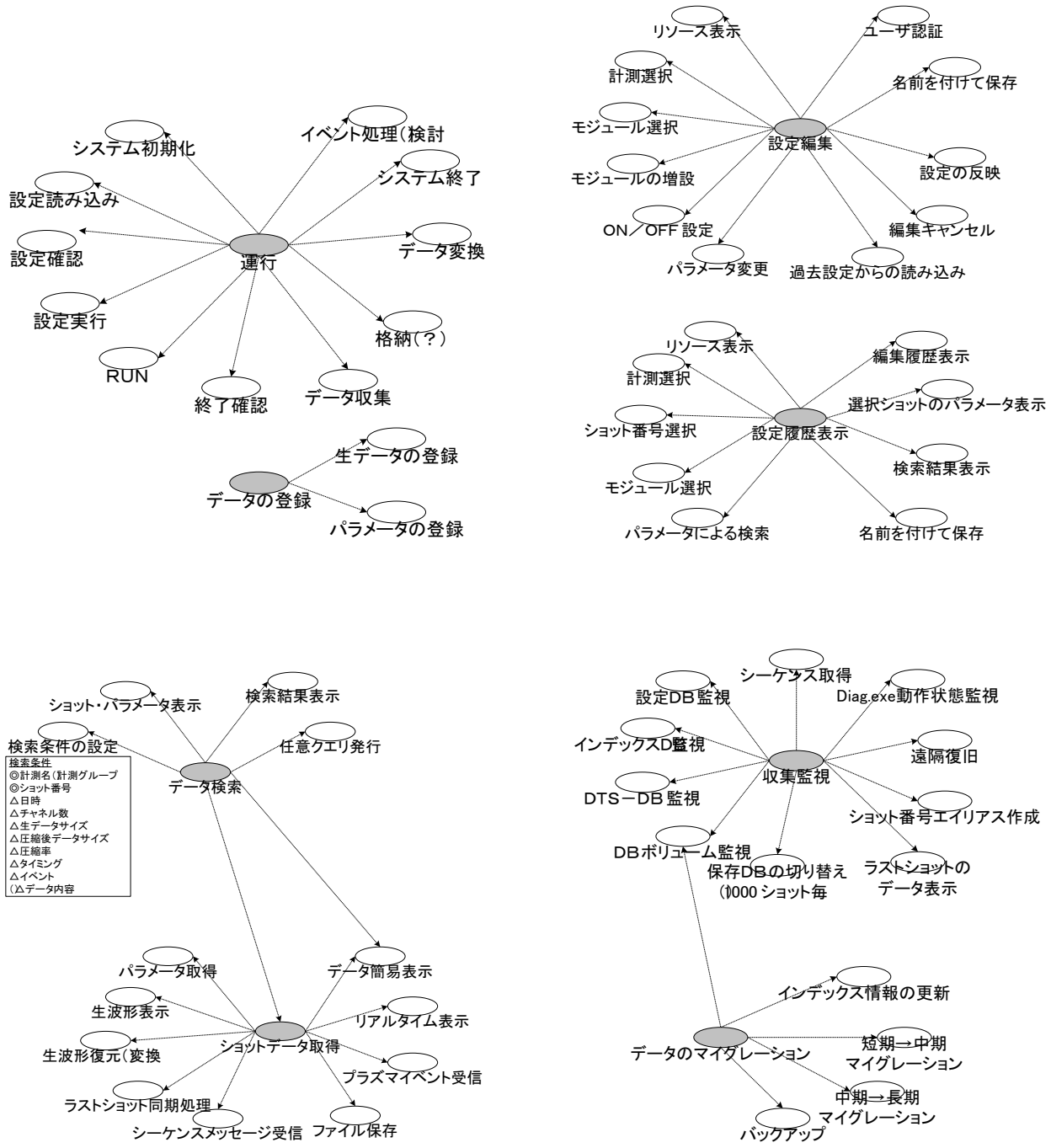


Figure 3.7: Substructures of major use-cases:

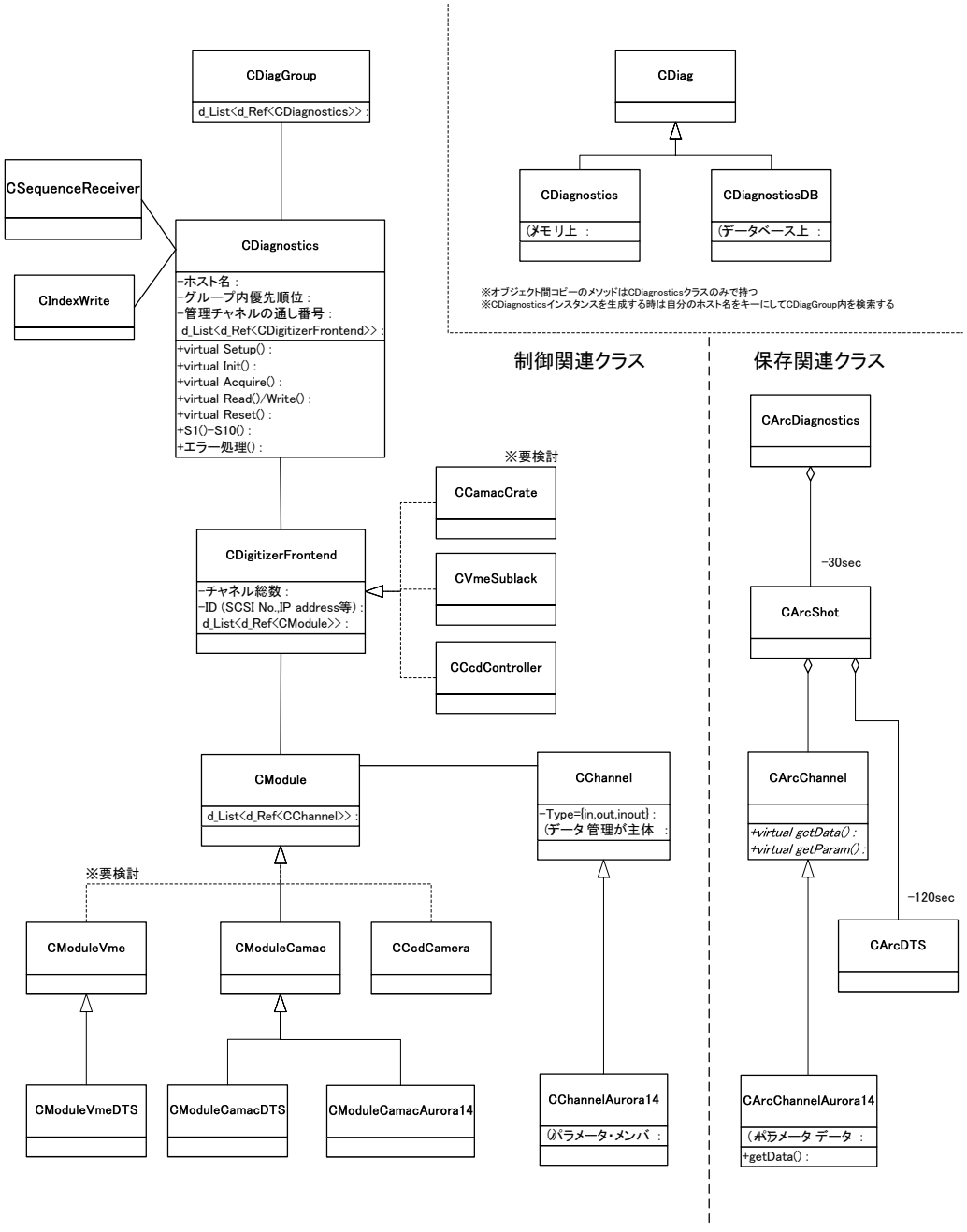


Figure 3.8: UML class chart. These class definitions are shared over the data acquisition programs. Though these classes are described in ODMG-93 manner such as `d_List<d_Ref<Class T>>` persistent pointers, the actual implementation has adopted the persistent class object instead.

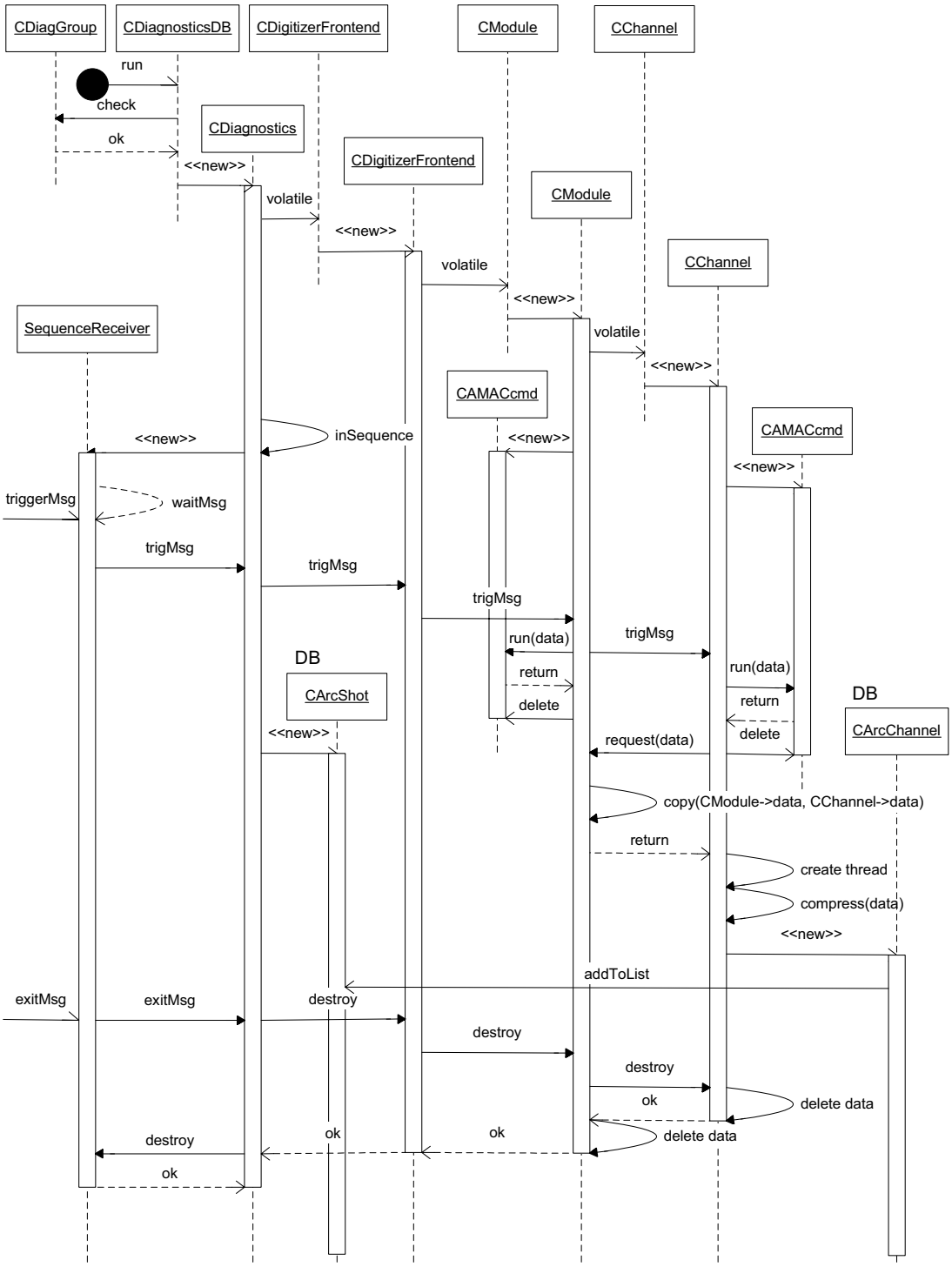


Figure 3.9: UML sequence chart

3.4.2 ポータブルなデータ収集エレメント

ここで開発する一括処理型データ収集システム，いわゆる従来の意味でのプラズマ・データ処理システムの基本構成は以下の通りとなる．

1. CAMAC によるアナログ/デジタル信号変換 (データ収集)
2. SCSI 光エクステンダを利用した SCSI データ転送［68, 69］
3. Windows NT データ収集サーバ計算機の分散配置
4. ネットワーク共有オブジェクト方式によるデータ相互参照
5. オブジェクト指向データベース
6. RAID ディスクと媒体交換機能付長期データ保存装置

アナログ/デジタル信号変換器に関しては，核融合科学研究所で運用されてきた JIPP T-IIU や CHS でのデジタイザ資産を継承するため, LHD データ収集系においても多くの CAMAC デジタイザを継続して利用することになり，デジタイザの主力は先ずは CAMAC 規格のモジュールとなっている．

LHD データ収集系の全体構成を Figure 3.1 に，また分散配備されるデータ収集及び計測制御のサーバ計算機と，それらに接続される CAMAC 等の周辺装置とによる各計測器毎の最小構成単位を Figure 3.10 に示す．このように，LHD の各計測エレメントは，末端ポート 100 Mbps 幹線 1 Gbps のマルチレイヤースイッチで構成されたデータ収集 LAN と計測制御 LAN とによって完全に並列対称的に接続され，各々のエレメントが独立して稼働する水平分散型システムを構成している［53］．こうした分散構成の採用は，集中型システム構成ではプラズマ計測データ量の大幅な増加によって I/O 処理が集中することを回避した設計の当然の結論であるといえる．

一方，近年の所謂パソコン (PC) の性能向上は劇的であり，従来中規模以上のプラズマ実験データ収集に良く利用されたミニコンやメインフレームと呼ばれる大型汎用機を遥かに凌ぐ高性能が十分個人レベルで入手可能になっており，それを活用したコスト/パーフォマンスの高い分散処理系への移行を容易にしている．そうした状況を十分に活用するべく，LHD データ収集系においてもデータ収集サーバ計算機に PC を，そしてその基本ソフトウェアとしてネットワーク OS の Windows NT を導入した．

Windows NT は UNIX や VAX/VMS と同様のマルチタスク/マルチスレッド，マルチユーザの機能を持ち［70］，基本機能としては遜色が無い上に，Win32 というグラフィカルな API (Application Program Interface) を同時に提供する最新 OS である［71］．更に，動作するハードウェアが PC であることから，一般のアプリケーションも EWS やミニコ

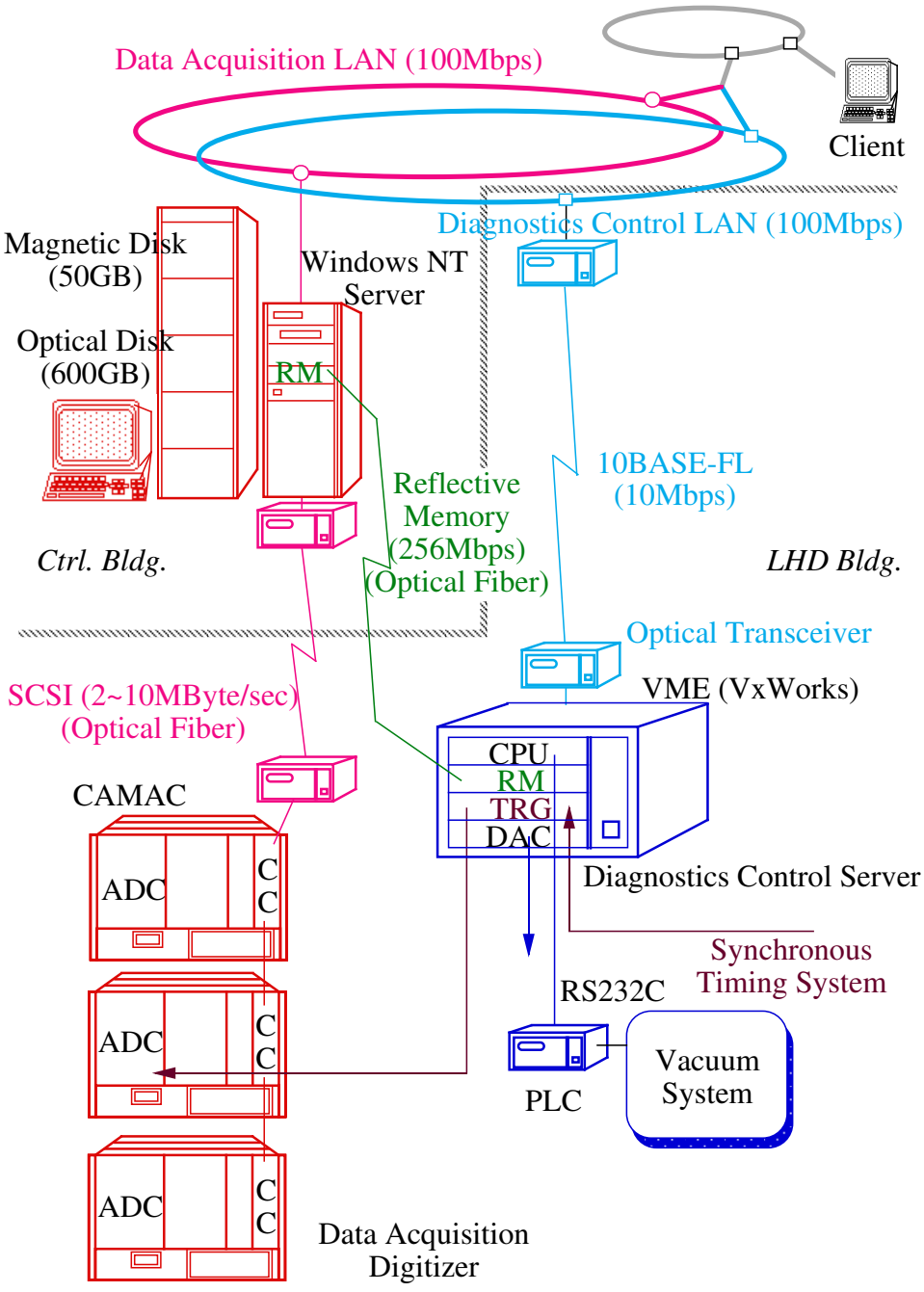


Figure 3.10: Minimum set of the data acquisition and diagnostics control system distributed for each diagnostic device; From the reasons of the distant extension and the electric insulation, every communication link will connect throuth the optical fiber.

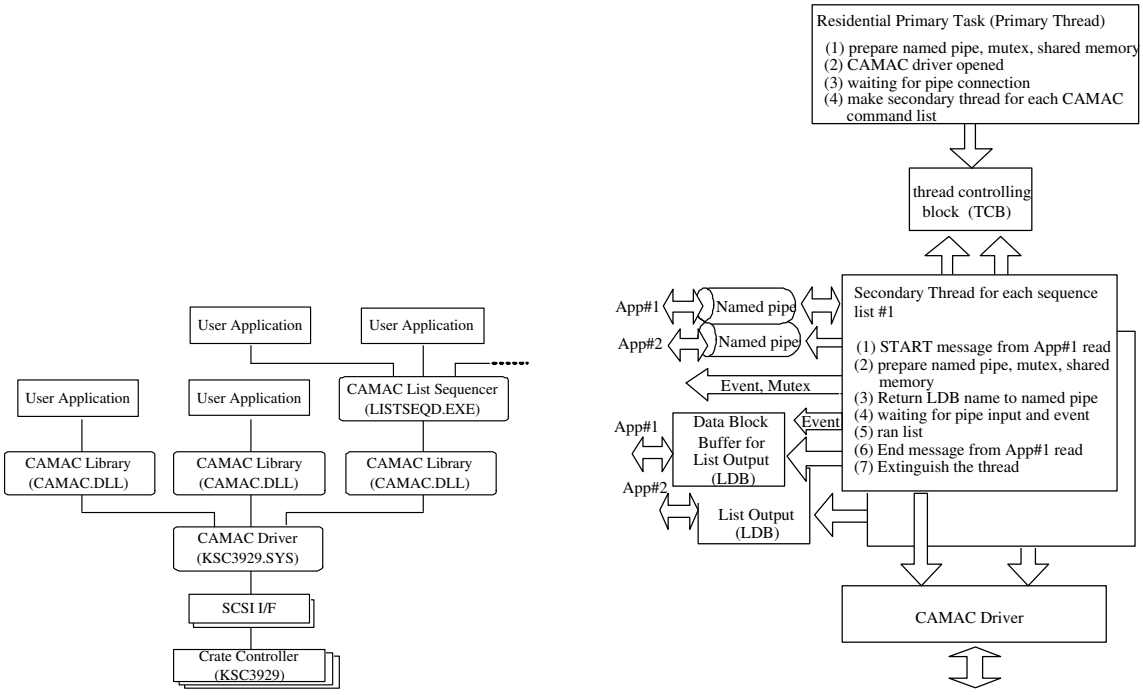


Figure 3.11: CAMAC driver and API layers (left) and command list sequencer (right). As for the crate controller, another product Jorway J73A is also applicable with its own driver and DLL in the same way. Both of the API (CAMAC.DLL) and the service manager task (LISTSEQD.EXE) can be called by the multi-task/multi-user environment.

ンで動作する物よりも選択肢が多く，かつ入手が容易である利点もある．LABCOM システムでは Windows NT で標準提供される SCSI ポート・ドライバの上位クラス・ドライバとして，SCSI CAMAC ドライバを共同開発すると共に，同じく Windows NT 上で動作する CAMAC リスト・シーケンサを開発し，データ収集サーバ計算機の中枢プログラムの一つとしている．この CAMAC ドライバと CAMAC リスト・シーケンサとにより複数の CAMAC 制御コマンド・リストを同時並行で処理可能にすると共に，マルチユーザ利用のためのコマンドリストのキューイングやデータの排他処理機能も提供している［72］.

次節以降でも改めて詳述するように，保存データベースへのデータの格納 (=永続化) は，分散稼動するデータ収集サーバ内で全て並行してローカルに行なわれ，実験シーケンス中は一箇所にデータが集中する事は無い［73］．データ参照クライアントもサーバと同様にネットワーク上に分散していることを考慮すると，このデータ収集システムは構成ばかりでなく収集データの流れまで完全に並列分散化されているということができる．実験シー

ケンスに則ったデータ収集サーバ計算機での収集データの流れは Figure 3.3 に示されているとおりである.

3.4.3 エレメント間の同期とオブジェクト共有

3.3.3 節で述べたとおり，データ収集サーバからのデータ転送手段，および CAMAC 制御オブジェクト間の同期メッセージ授受には，ネットワーク共有オブジェクトの機能をインターネット環境上で提供するソフトウェア・パッケージ HARNESS［57］を当初導入した．これは Figure 3.5 に示したとおり，HARNESS スペースと呼ばれる共有オブジェクト空間をネットワーク上で共有に参加する全てのホスト計算機のメモリ上に再構成するもので，OOD に極めて適したネットワーキング・ミドルウェアとして期待できた.

HARNESS スペースでは，ネットワーク共有という性格上メモリ上のアドレス参照が無意味になり実験データを処理するためのデータ処理関数までは共有できない．しかし計測データをクラス分類して個々の収集データを設定パラメータ等と共にオブジェクトとしてそのまま取り扱えるため，データ収集系で採用したオブジェクト指向プログラミングに際してはプログラム・ソースの再利用性の面でかなりの利点があった．また，収集生データの相互参照によって一次処理，二次処理データを自動計算する場合など，データの参照タイミングが共有オブジェクトの生成出現と同期できるため，参照データの読み込み機構を単純化できるという長所も併せ持っている.

しかしながら，運用初期の段階で，ある程度 (〜10 ノード以上) 多数のホストが同一 HARNESS 空間に接続して同期メッセージ・オブジェクトを授受しようとすると，非常にシステム負荷が上がってしまい実用に耐えない処理速度になる症状が現れた．また，参照の有無に関わらずその HARNESS 空間に接続している全ホストで同様に主記憶領域を占有してしまうため，ある程度大きいサイズのデータ・オブジェクトを共有空間で使用しようとすると，全計算機のリソースを大量に消費する結果となり，実用に耐えなくなってしまった．このため，今回のような MPP 形態のシステム構成には残念ながら HARNESS 利用はそぐわないと判断し使用の継続を断念した.

HARNESS に替わる実装としては，3.3.4 節でも述べたように，CORBA の利用も考えられる．しかし，CORBA では遠隔オブジェクトの生成・消滅をアプリケーション自身で管理する必要があるなどプログラム記述量が多くなり，プログラミング自由度は大きい反面，アプリケーション開発にある程度のスキルが要求されるなど，OOM に沿った開発負荷の低減には必ずしも寄与しないという短所がある．また HARNESS パッケージと同様,

CORBA サーバもいわゆる共有メモリー型のミドルウェアであり，一般的に処理負荷が重く，遅くなる傾向が知られている．

このため本研究では，分散収集エレメント間の同期シーケンス・メッセージ授受の実装として，最終的に，負荷の見積もりが容易で処理も軽快な非共有メモリー型の TCP/IP ソケット通信をベースに独自のネットワーク API を作成・利用することで，速度的な問題を回避している．これら API は OOD に沿ったコード再利用性を考慮して，SequenceSender, SequenceReceiver のメッセージ送受信クラスがもつメンバー・メソッドとして隠蔽実装しており，あくまで同期シーケンス・メッセージ・オブジェクトをこれら送受信の分散オブジェクト群が共有しているように見せている．

3.4.4 オブジェクト指向データベース管理システム (ODBMS) の活用

大型ヘリカル装置 (LHD) の計測データ処理システムのデータ収集系は，計測単位に独立した収集サーバ計算機を配備し，それらを高速なネットワークで接続した分散処理系として構築されている．

データ処理システムのプログラム群は「オブジェクト指向」を全面的に採用して C++言語で開発されており，計測データも収集から保存，その後のクライアント側からの参照まで C++データ・オブジェクトとして一元的に取り扱われる．各計測サーバ計算機で収集された計測データは，圧縮されて計測データ・オブジェクトとして計測サーバ毎にローカルに動作しているオブジェクト指向データベース，いわゆる短期保存ストレージの仮想空間中に永続化格納される．C++で記述された収集プログラム中の計測データ・オブジェクトがそのままシームレスに永続化保存されるため，データの保管のための長い入出力コードをつくる手間が省略され収集システム開発負荷が大幅に低減されている．データを参照するクライアント・プログラムでは，このデータベースに格納された計測データ・オブジェクトを取り出し，元と同じデータ・オブジェクトとして再びメモリー上に復元した後にデータ表示・解析等の処理を行う．また，データ収集 CAMAC モジュールの収集設定情報等もオブジェクト指向データベースに格納され管理されている．Table 3.3，3.4 に基幹業務等で一般的に使用され DBMS の代名詞ともなっているリレーショナルデータベース管理システム (RDBMS) と，今回のデータ収集保存システムに用いたオブジェクト指向データベース管理システム (ODBMS) との機能的な違いを示している．

メモリ上にある揮発性の C++データ・オブジェクトと LHD 計測データ保存系の OODB 中で永続化された同オブジェクトのシームレスな関連性は Figure 3.4 の概念図に既に示さ

Table 3.3: Comparison between RDBMS and ODBMS［74］

	RDBMS	ODBMS
atomic, consistent	Ok	Ok
complex data model	plain	high
user defined type	N/A	Ok
revision management	N/A	Ok
paradigm	4GL (event driven)	Object oriented

れた通りである．本システムでは，ODBMS に価格対性能比が高いものとして米国 Ardent Software 社の O_2 を採用している．データ収集用 CAMAC モジュールの入力チャネルに対応した圧縮されたチャネル・データを１つの保存用データ・オブジェクトとして扱い，データベース入出力の基本単位としている．オブジェクト指向データベースでは，プログラム中でデータベース上のオブジェクトをメモリ上のオブジェクトと同様に取り扱うことが可能なので，ソフトウェア開発にかかる負担が少なくて済む利点があり，プログラム開発を短期間に効率的に進めることができた．

収集サーバから参照クライアントへのデータ転送の HARNESS に替わる手段としては，保存データベースに採用した ODBMS の C++言語バインディング (API) が提供するネットワーク経由での永続性オブジェクト参照機能によって，データ・オブジェクトをクライアント・プログラム中のメモリー上に複製・再現する方法が取り得る．ユーザ側 (クライアント) から見た場合，保存データベース中に格納された実験データの参照はあくまで C++ データ・オブジェクトをネットワーク経由で参照・複製する処理に基づいて行われている．

しかしながら，こうした遠隔オブジェクトの参照・複製処理速度は，データサーバ系の最も重要な I/O 性能であり，高速化については，3.4.5 節および 3.7.2 節にて改めて詳述する．

保存データベースへのデータ書き込み速度としては，O_2 利用を特に最適化していない状態で約 380 kB/s 程度であり，実験開始当初のデータ収集量では特に問題のない範囲であったが，収集データ量の増加に伴って，次節に述べるとおり I/O 高速化のため各種の改善措置が必要となった．

Table 3.4: 実装に関する RDB と OODB との機能的違い

機能	RDBMS	ODBMS
スキーマサポート	スキーマ言語≠プログラム言語. SQL 言語，RDB 設計に習熟が必要.	スキーマ言語＝プログラム言語. アプリケーションコードから自動生成.
入出力コードの追加	データベース内とアプリケーション内との２つのデータ表現を維持する必要がある. DB 内表現をアプリ内部表現に変換するため多量の入出力コードが必要.	DB 内オブジェクトも常にメモリ内にあるかのように参照可能. 入出力コード，同期コードの追加は必要ない.
アプリケーション言語インターフェース	アプリ言語 (C++) とデータベース操作言語 (SQL) との結びつきが弱く 2 言語 C++，SQL の習熟を要する. C++と SQL とは共通のデータモデル/データ型セットに基づいていない.	プログラム言語バインディングにより完全に透過的に DB 内オブジェクトをアプリケーション空間から操作可能.
パフォーマンス	複雑なオブジェクトの再構成には多数テーブルの join が必要でコストがかかる. 全てのオブジェクトは DB サーバ側にキャッシュされる. 同じオブジェクトの繰返し参照には繰返し取得が必要.	インデックス検索には特定の集合クラスを予めスキーマ中で指定して使用する必要がある. クライアント側でオブジェクトがキャッシュされる. 最近参照したオブジェクトはキャッシュに残るため RDB の再検索より 10〜100 倍高速化.
固有の機能		オブジェクトのバージョニングが可能. 設計トランザクションにより複数セッションに跨るロング・トランザクションが可能.

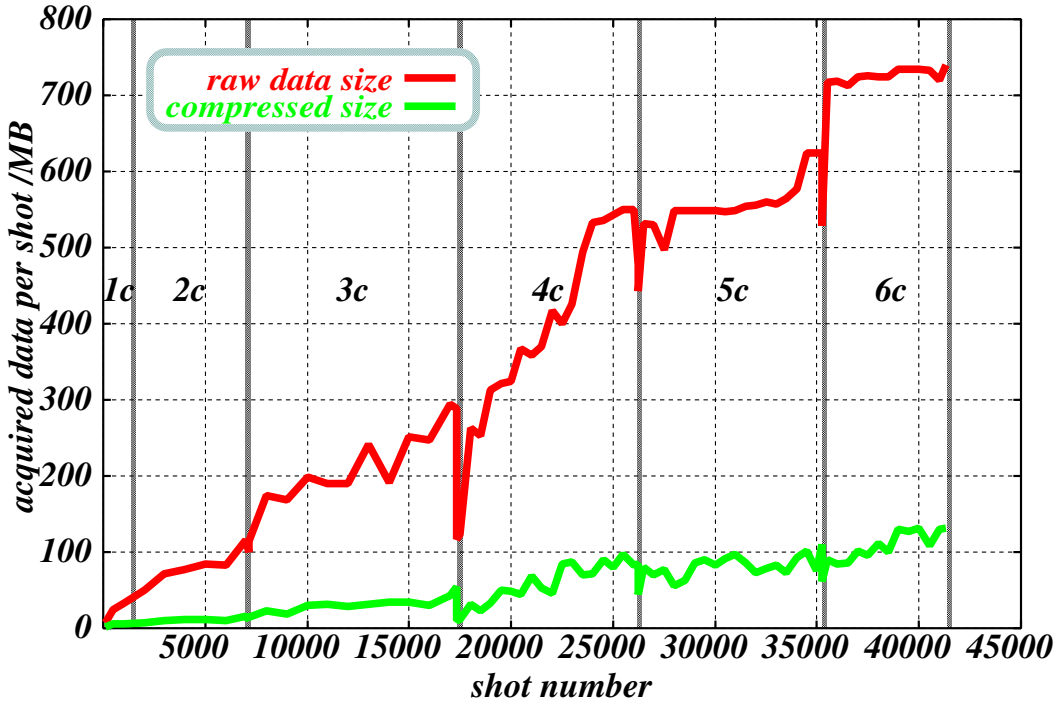


Figure 3.12: Acquired data growth by each shot. At the end of the 6-th experimental campaign, it went up to 750 MB/shot. The initial diagnostics plan of 600 MB/shot raw data had been successfully exceeded. In the storage area, 130 MB/shot compressed data will be stored now.

3.4.5 データ I/O の高速化

LHD 実験開始以降の 5 年間，1 ショットあたりの収集データ量は年々ほぼ倍々のペースで増加を続けており，Figure 3.12 に示すとおり，平成 14 年度の実験第 6 サイクル終了時には CAMAC 系だけで 750 MB/shot の生データの収集・処理を行うに至っている．また，通常の実験モードでは 3 分間隔で一日あたりおよそ 150 回程度の実験を行っており，核融合実験装置の中では群を抜いて世界一のデータ生産量を誇っている．

こうした収集データ量の急峻な増加に対応しながら，かつユーザからのデータ参照にデータ待ちストレスが発生しないよう，LHD データ収集保存系では以下にあげるようないくつかの点で I/O 速度改善を行った．

1. 収集・保存ルーチンの見直しによる CAMAC からのデータ収集速度の改善
2. より高速な I/O 性能を持つ ODBMS への乗り換え
3. データ取出しの下部通信を C/S の二階層⇒三階層化して速度を向上

4. ギガビット級高速スイッチングネットワークの導入

データ参照・取出しの三階層化については改めて 3.7.2 節で述べることにして，以降では主にサーバ内部の処理の高速化について詳述する．

CAMAC デジタイザからホスト計算機へのデータ伝送速度は，CAMAC データウェイの 1MHz システムクロックの制約から，通信制御のオーバーヘッドを考慮すると最大～1MB/s 内外である．LHD で使用している 2 種類の CAMAC クレートコントローラ (C.C.) 用のドライバでの実測最大速度も，KineticSystems KSC3929-Z1B で約 0.7 MB/s，Jorway J73A で約 1 MB/s である．デバイスドライバー類が提供するこれらの最大ポート速度を向上させるのは，既にハードウェア性能の上限に近いため CAMAC 系では困難である．そこで，このポート帯域を最大限に活用してデータ伝送を行うべく，Figure 3.13 のようにデータ収集プログラム中の伝送・圧縮・DB 格納の各処理をマルチスレッド化して，デジタイザの各チャネルデータを切れ目なく伝送させることで，見た目上の伝送にかかる時間の短縮を図った．

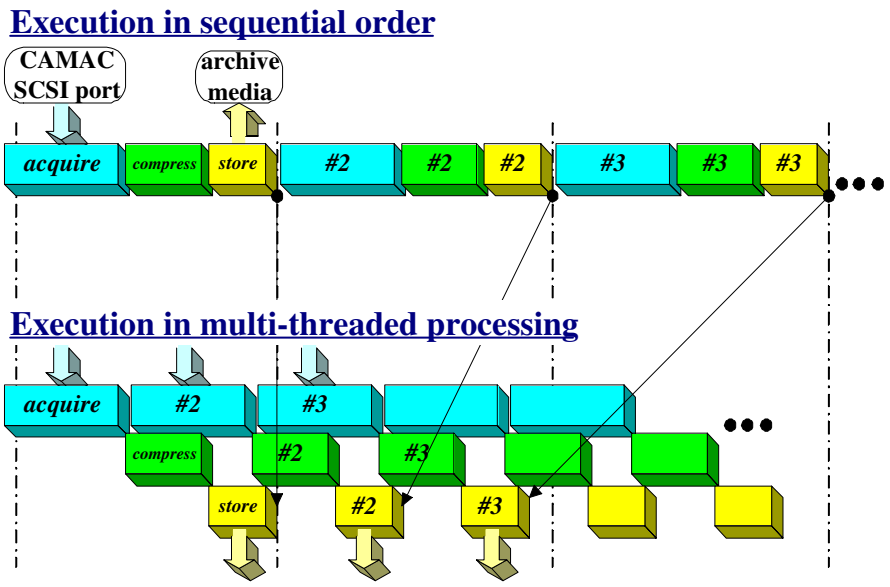


Figure 3.13: Acceleration scheme of multi-threaded data acquisition and processing:

当然この一連の処理では，CAMAC からの伝送と DB 格納という 2 つのデータ入出力を行うため，DB 格納処理もまた同様に速くならなければいけない．ところが当初導入した Ardent Software 社製 O_2 システムでは，オブジェクト・インスタンスの内部表現が O_2C と呼ばれる別の第 4 世代オブジェクト指向言語で記述されているため，C++言語バイン

ディングを使用すると C++インスタンス⇔ O_2C インスタンスの変換が毎度行われていることが調査を進めるうちに判った．O_2 ではこの内部処理のオーバーヘッドが大きすぎるためデータベースとの I/O 速度が上がらず，実測最大 I/O 性能としても 0.4〜0.5 kB/s 程度であった．本システムの収集プログラムの実稼動性能としても，シーケンシャルな処理で 0.38 MB/s，マルチスレッド高速化を施した後でも 0.48 MB/s 程度にとどまっており，O_2 の最大性能に近い値を達成することが出来たものの 0.7〜1 MB/s の CAMAC I/O に追随できるるまでには至らなかった．

同時期に O_2 製品の販売・サポートが終了したことも理由となって，O_2 に替わるより高速処理が可能な ODBMS を新たに選定し直すこととなり，Table 3.5 の通り，実績のある他 ODBMS 製品の比較評価を行った．ソースコードの互換性では，O_2 と同じ永続性ポインターを用いる ODMG 準拠の製品である Objectivity/DB が移植が容易であった．しかし，BLOB を格納・読出しするための大容量記憶領域を動的かつ自動で拡張できる点と，I/O 速度向上のため必要な DB 設計の最適化調査・作業に際してオンサイトで開発支援が受けられる等の理由で，最終的に eXcelon (旧 ObjectStore) 社製の ObjectStore を採用した[注3]．

ObjectStore と O_2 との入出力性能を比較評価した結果が Table 3.6 のとおりである．計測条件として，ObjectStore 6.0 SP1 では 256kB のオブジェクトを 84 個 (21MB) の書込み，読出しを 10 回行いその平均を取っている．前回読出しのキャッシュヒットを避けるため，１回毎のコールド・テストでの実施である．

初代データサーバ (Vectra) 機で O_2 (write) 性能と比較すると，ObjectStore の書込み速度のアドバンテージは 1.5〜2 倍程度である．O_2 (read) のばらつきが大きいが，読出し速度についても同様に ObjectStore の I/O 性能差は O_2 に較べて約 1.8〜5.2 倍と明らかに優位である[注4]．また高速マシンによるホスト性能を変えての試験でも，性能向上率が ObjectStore では約 2 倍になるのに較べると O_2 は 6 割程度で劣っている．O_2 の傾向として，オブジェクトの頭出しに要する検索時間ではそれ程でもないが，I/O レートは 0.4〜1.2 MB/s まで 3 倍近くばらつきが出るという事もある．性能向上率の悪さとばらつきの多さは，やはり先にも述べた C++⇔ O_2C 言語バインディング間の動的変換オーバーヘッドが非常に大きいことが原因と想定される．

また I/O 処理が主になるデータ収集サーバで比較的余裕がある CPU の演算能力を活か

[注3] ODBMS の最大シェア製品でもあり O_2 のように消滅することがなかろう事も考慮の対象となった．

[注4] Objectivity/DB の I/O レートの報告としては max. 3.5 MB/s というものがあり，ObjectStore と同じページサーバ・アーキテクチャを採っていることから考えても，ほぼこの程度の値が得られることが予想される．

Table 3.5: Comparison between popular ODBMS products: O_2, Objectibity/DB and ObjectStore.

	O_2	Objectivity/DB	ObjectStore
developer	Ardent Software Inc.	Objectivity Inc.	eXcelon (ObjectDesign)
license	fixed	free in project	free in project
annual license fee	1 050 000 JPY	1 000 000 JPY	1 500 000 JPY
user support	–	open (e.g. user ML)	abundant
object browser	N/A	ooBrowse (browser)	Inspector (editable)
C++/Java binding	both	good	good (schema sharable)
backup[1]	B, O, I	B, O, I, D	B, O, I, D
persistency	pointer	pointer	object
locking unit	object	page (container)	page
exclusion levels[2]	R, W, L	R, W, L	R, W, MVCC, L
database reference	logical, physical	logical	logical
collection	ODMG	ODMG, STL	inherited class
collection types[3]	L, S, B, A, VA	A, D	L, S, B, A, D
schema evolution	dynamic	dynamic	static
volume extension	static	static	dynamic (automatic)
transaction length[4]	S, L	S, L	S, L, N
multithread/multitask	N/A	Ok	Ok
ANSI SQL	Ok	Ok	(with ToolKit)
MSS support	N/A	OOFS	N/A
market share % (1998)	5.3	7.7	31.2

[1] B, O, I, D ⋯ Batch, Online, Incremental, Distributed
[2] R, W, L ⋯ Read, Write, Lock
[3] L, S, B, A, D ⋯ List, Set, Bag, Array, Dictionary
[4] S, L, N ⋯ Short, Long, Nested.

して，CAMAC からのデータ収集直後に主メモリー上で生データを圧縮処理し以降全て圧縮済データを取扱うことで，DB 格納や読出し等データ I/O サイズの低減を目指した．Table 3.2 に示したとおり，得られるデータ圧縮率は計測毎に大きくばらつきが大きいが，平均すると約 1/6 程度に圧縮されており，見かけのデータ I/O 性能の改善とデータ・ストレージ中の占有サイズの低減の両方で非常に効果的であることが確かめられた．圧縮ルーチンにはフリーウェアの GNU zlib を用い，主メモリー上で生データ・バッファの圧縮演

Table 3.6: I/O performance evaluation between ObjectStore and O_2.

ODBMS (job)	HP Vectra XU6/200[1]		DELL Precision 410[2]		Diagnostics
	rate (MB/s)	search (s)	rate (MB/s)	search (s)	
ObjectStore (write)	0.6 (∼0.8)	-	1.3 (∼1.7)	-	(256kB*84)
ObjectStore (read)	2.1	-	4.0	-	(256kB*84)
O_2 (write)	0.4	-	-	-	–
O_2 (read 52,403Byte)	0.40	0.65	0.63	0.54	FIG
O_2 (read 458,232Byte)	1.07	0.42	1.60	0.16	TESPEL
O_2 (read 1,581,747Byte)	0.41	0.69	-	-	Bolometer
O_2 (read 5,477,682Byte)	0.75	0.75	-	-	Magnetics
O_2 (read 6,322,996Byte)	1.18	0.64	-	-	SXfluc

[1] HP VectraXU 6/200 ⋯ PentiumPro 200MHz Dual 192MB FP-DRAM + WideSCSI RAID
[2] DELL Precision 410 ⋯ PentiumII 450MHz Dual 256MB SDRAM + U2W SCSI Disk

算を行っている．

このように ObjectStore を O_2 に換えて導入し，またデータ圧縮ルーチンを収集直後に組み込んだ I/O サイズの低減を図った結果，CAMAC を用いる収集系として理論性能に近い 0.7〜1 MB/s のスループットを転送・圧縮・DB 保存までを含めて達成することが出来た．PentiumPro 200 MHz Dual の収集ホストを用い KSC3929-Z1B C.C. + Jorway Aurora14 CAMAC ADC の構成で 60 MB のデータ収集に約 90 秒を要するため実測性能値は 0.67 MB/s である．同じく J73A C.C. の場合は約 1 MB/s となった．

平成 13 年度には LHD 共同研究/遠隔実験参加を目的として，新たに SuperSINET の基幹ノードサイトを取得，10Gbps IP 網および 1Gbps Ethernet*3 の導入を果した．特にデータ処理サーバ系など実験 LAN の高速ギガビット化を進めることが出来た．Figure 3.14 にその計測データ収集保存部分の概略を示す．これにより今後は，LHD を用いたより高度な遠隔実験参加形態の実現を目指して，大容量化している計測データの高速遠隔転送や実時間配信なども視野に入ってきている．

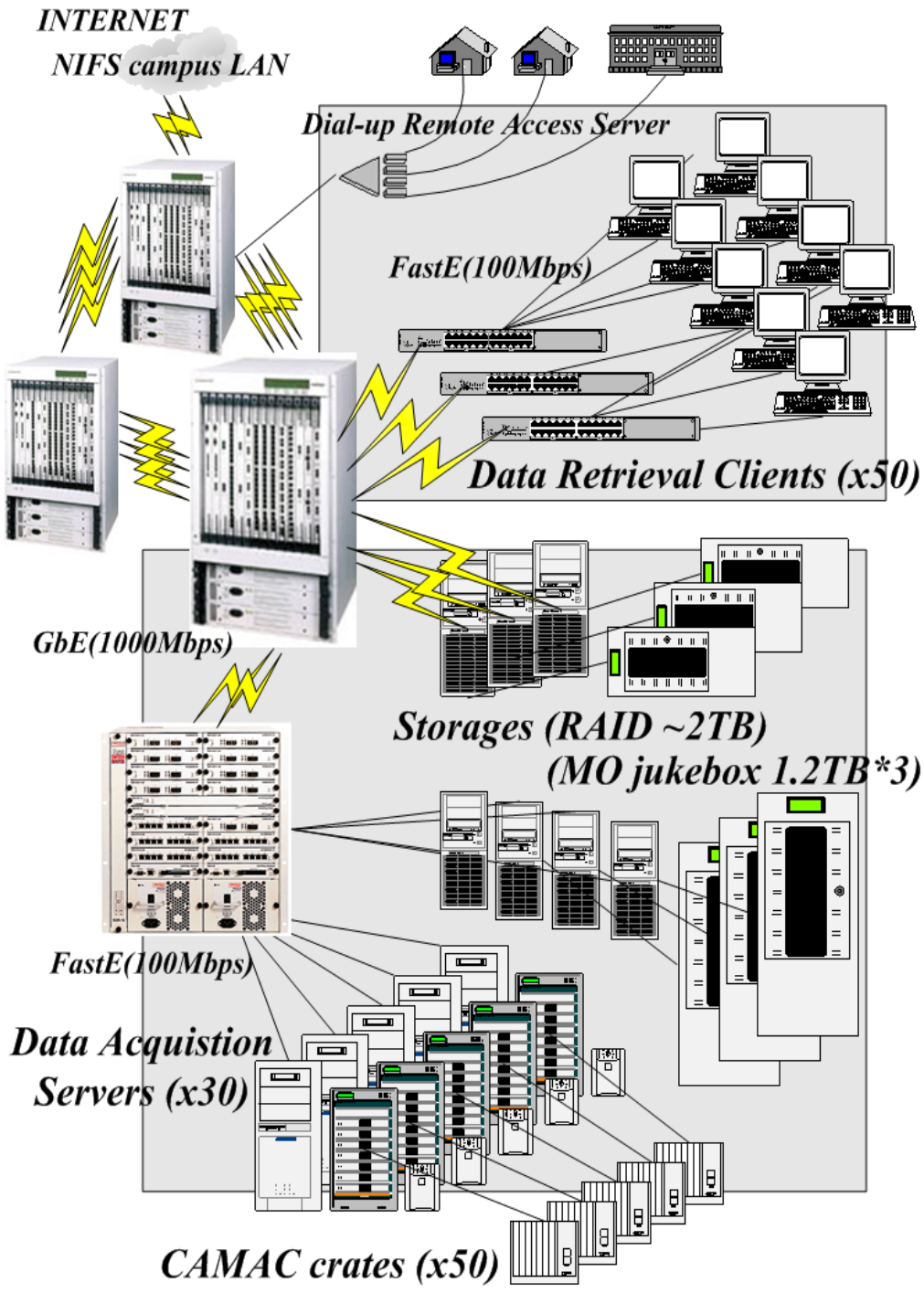


Figure 3.14: Data acquisition network segment based on the Gigabit Ethernet switching fabrics: Faster storage servers are directly collected by the Gigabit Ethernet NIC, and rather slower ones are on Fast Ethernet. Backbone swithes have multiple gigabit links between each other by using the OSPF ECMP load balancing mechanism.

3.5 計測制御系

LHD の計測制御/データ処理系統は，LHD が最新鋭のプラズマ実験装置であることを反映して，取扱う計測は 20〜30 種類を数えており，その多くは遠隔制御や実時間モニターの機構を必要としている．これは，LHD では超伝導磁場コイルを用いて実験中に定常磁場を発生しているため，運転中の本体装置付近への立ち入りが制限され，計測機器の制御には基本的に全て遠隔操作が要求されるためである．また LHD 実験では約 1 時間の準定常プラズマ放電実験が予定されており，計測制御系統にも従来には無かった定常運転の機能が求められている．

ここで前置きとして LHD の制御系統について少し触れておくと，LHD の制御系統は大きくは本体制御 (本体装置，プラズマ)，実験制御 (加熱装置，冷凍機，電源等)，及び計測制御の三つに分かれており，各々に複数の制御サブシステムが属する構成となっているが (Figure 3.15 を参照)，以下，本研究で計測データ処理に関連する計測制御系のみを取り扱っている．この計測制御系の構成は，

- **計測制御分配装置** …… 中央制御装置と連絡し，保護インタロックや実験タイミング・シーケンスの分配，収集を個別計測制御系との間で取り持つ 1 台
- **計測器別制御系** …… 一計測毎に約 30 台設置され，その計測に必要な全ての制御・監視装置の管理・操作を独立して行なう

と単純な二種類のホスト計算機よりなっている．特に後者は，プラズマ計測の種類ごとに独立して計測制御計算機を配して自律的に制御監視動作を行うもので，データ収集保存系と同じ大規模並行分散形態の設計に則っている．しかし，機器の保護や緊急非常措置なども機能的にサポートする必要から，後述する計測タイミングシステムや保護インターロックといった一部のハードウェアは，各計測制御系を横断した独自のワイヤーロジック経路/光ファイバー経路/を持っており，この点がデータ収集保存系とは異なっている．

計測制御系のハードウェアとしては，FA(Factory Automation)，LA(Laboratory Automation) 分野で広く利用される VMEbus システムを採用し，その基本ソフトウェア (OS) としては実時間監視 (モニタ) 及び操作の機能を実装する必要からリアルタイム OS Tornado/VxWorks を用いている［75, 76］．

計測制御系の VME/VxWorks システムに要求される役割は，基本的には各計測機器に属する真空排気系や可動部など全てのモニター/制御装置を監視・制御することであるが，機能別に分類してみると，

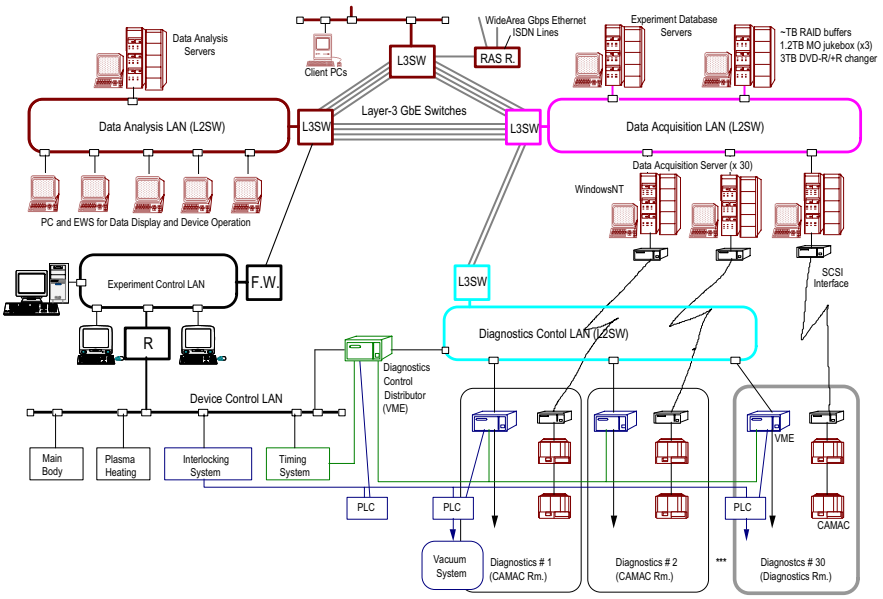


Figure 3.15: Cluster structure of LHD experiment's local network: Windows NT server controls the CAMAC digitizers through the SCSI connection, acquires the rawdata, and then stores them into the local mass-storage. VxWorks system provides the real-time controls for devices.

1. 実時間機器制御
2. 実時間監視 (モニタ)
3. 遠隔操作
4. 実験タイミング・シーケンス (トリガ・システム) 管理
5. 保護インタロック

となる．LHD では磁場配位の生成に超伝導コイルを用いており，プラズマ実験で一般的な短パルス運転の他に準定常的な長時間放電実験が実施される．こうした準定常実験では計測機器に対しても放電中に随時監視および制御操作等が要求されるため，短パルス運転時の放電終了後の一括処理とは全く異なった実時間動作可能な系統が必要であり，LHD では計測制御系の VME/VxWorks システムがこれを担当している．また，超伝導コイルによる定常磁場の発生によって，LHD 本体付近に設置される各計測器は実験中，直接操作を加える事が出来ないため，全ての計測機器に対して完全な遠隔操作方法を実現する事も必須の要件となっている．遠隔操作に関しては LHD が共同研究に供する実験装置であるという位置づけからも，遠隔地からネットワーク経由での制御が必須となることから，オープン・システムの代表である UNIX やインターネット通信技術 (TCP/IP) との親和性の良さ

からも Tornado/VxWorks を選定している．

3.5.1 実時間機器制御と遠隔操作

LHD 計測制御系は，約 30 種類の LHD 計測器毎にローカルに配置された計測制御 VME サーバ計算機と，ユーザの遠隔操作/モニター端末の Windows クライアント計算機により構成される．

- 30 種類の計測器毎に計測制御 VME 計算機を配備
- ADC，DAC，DIO，GP-IB，RS-232C 等の各種 I/O モジュールが利用可能
- VME 計算機にリアルタイム OS Tornado/VxWorks を採用
- 操作端末である GUI クライアント計算機と VME サーバ計算機との C/S 機能分離
- 100Mbps 高速スイッチング・ネットワークによる高速接続

といった特徴を持ち，ネットワーク経由での遠隔操作や最大 10Hz 程度のリアルタイム・モニター機能を各機器ごと独立に提供している．

この各計測毎に設置される計測制御 VME/VxWorks システムと計測機器の利用者端末との連携には，インターネット通信技術を適用し，クライアント–サーバ (C/S) 方式のネットワーク・モデルを応用する事で計測制御サーバの自律性を高めている．C/S モデルによる機能的分離とこの間をつなぐインターネット接続により，A 章でも述べる通り利用者端末 (クライアント) のネットワーク上での位置的制約を払拭し，遠隔地からの機器監視・制御を可能にしている．計測制御クライアント/サーバ間の遠隔処理呼出し機構の実装には，信頼性・処理性能を重視して，主たる UNIX OS やリアルタイム OS Tornado/VxWorks でサポートされているオープンソースの ONC RPC 4.0[注5] ［77］という TCP/IP プロトコル上位層を利用した．これらは 3.5.4 節でも述べるとおり，計測制御アプリケーションからはネットワーク下位層を隠蔽して使用できるように，論理チャネルという抽象化された独自のネットワーク API 用オブジェクトとして統一されており，コード再利用性など開発負荷の軽減が図られている．

また，機器の制御・監視の色彩が強い計測制御のクライアントでは，装置の状態表示にグラフィカルな表示インターフェース (GUI) 利用を欠かすことができないが，この GUI 表示処理の負荷をサーバ側処理とは分離してクライアント計算機側に持たせることで，実時間処理を行うサーバ計算機に必要以上の負荷をかけない機能的な負荷分散を果たしてい

[注5] ONC (Open Network Computing): Sun/AT&T の分散コンピューティングのための製品群. 旧 SunRPC. ソースが無償配布されている．

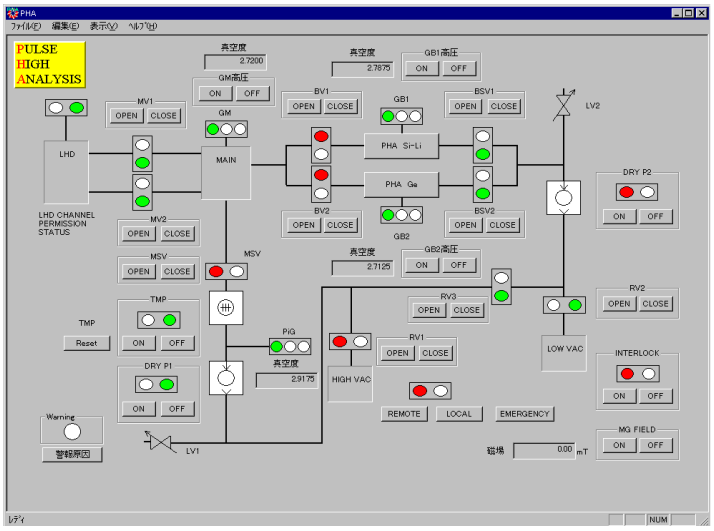


Figure 3.16: Typical GUI program for pulse height analysis (PHA) measurements. It runs on Win32 OS environment.

る．Figure 3.16 は Windows(Win32) 上で動作する PHA 計測用の計測制御 GUI 端末画面の例であり，こうした GUI を Microsoft Visual C++ を用いて個別作成することでユーザの計測機器遠隔操作の支援を行うことができる．尚，後の 3.5.4 節でも述べるとおり，VME サーバとクライアント GUI 間の通信はオブジェクト指向的に抽象化されており，最低限の開発作業で GUI が容易に開発できるように配慮している．

3.5.2 計測タイミング・システムの開発

LHD の様な大型装置に於いては，プラズマ計測の種類や収集チャネル数も多く，それらをサンプリングするデジタイザに必要なクロックやトリガ生成の要求も多様化するのが一般的である．LHD データ処理ではこういったタイミング・システムへの要求を充たすため，複数のローカル・カウンタ出力を備えた VME 型プログラマブル・タイミング・モジュールを新たに自主開発し利用している［78］.

各計測器間で同期した連係動作をさせるために重要な計測タイミング・システム (DTS) は，LHD 中央制御タイミングと共有する高精度発振器 (10MHz) を基準同期クロックとして，エンコードされたトリガー・メッセージをその上に重畳させて変調し，一緒に分配する分散型エンコーダ/デコーダ (変調器/復調器) 方式となっている．この同期ディジタル・メッセージの分配には一芯の光ファイバを用いており，Figure 3.17 に示すとおり，自由度の大きい階層構造になるためシステムの拡張が容易である．

各計測器毎のローカル・トリガは，主トリガから同期クロックをカウント・アップ＆コンペアする事で任意の時刻にプログラム出力が可能であり，又，Figure 3.18 のとおり，このタイミング・モジュールを VME モジュールとして計測制御計算機に組み込む事で，ネットワーク・クライアントから自由に設定が可能という特徴を持っている．

LHD 実験の計測シーケンス中での DTS を介したデータ収集系と計測制御系の連携の様子を Figure 3.20 に示す．データ収集計算機はショット・バイ・ショットでデータ変換/収集/ 処理のシーケンス動作を行なうため，準定常長時間放電実験時は断続的な運転となる．これに対して計測制御計算機は，常時リアルタイム動作で連続運転しているため，短パルス放電時と長時間放電時との動作に違いは生じない．

3.5.3 計測系インターロック網

個々の制御装置の連携に欠かせない保護インタロックには，PLC (Programmable Logic Controller) と呼ばれる半導体リレー回路モジュールを採用した．この PLC の最大の特徴としては，ラダー図と呼ばれるリレー入出力接点の結線回路をダウンロード方式で書き換え可能なため，構成変更や拡張が容易である事が挙げられる．具体的に，従来のインタロック系に利用されてきたリレー盤と比較した場合，

- リレー接点結線回路がプログラマブルで何度でも書き換え可能
- プラグイン・モジュール・タイプなので，拡張が極めて容易
- 光ファイバによる電気的絶縁を取りながらの子機増設も可能
- 小型・軽量化できる
- RS-232C 経由で任意の接点の状態を随時読みだし可能

等の長所をもっている．通信ポートを持ち，計算機との親和性も良好な事から，LHD でも計測制御サーバ計算機に RS-232C で接続して利用し，常時，各制御信号/接点の状態を監視する．通常の具体的用途は，計算機によるソフトウェア制御よりもより信頼性が要求されるハードウェア的な保護インタロックや帰還制御ループ等であり，LHD 計測でも計測制御系への運転許可/停止指令や，重/軽故障などの機器運転インターロック信号や実験シーケンスなどの制御接点信号の授受に使用している．

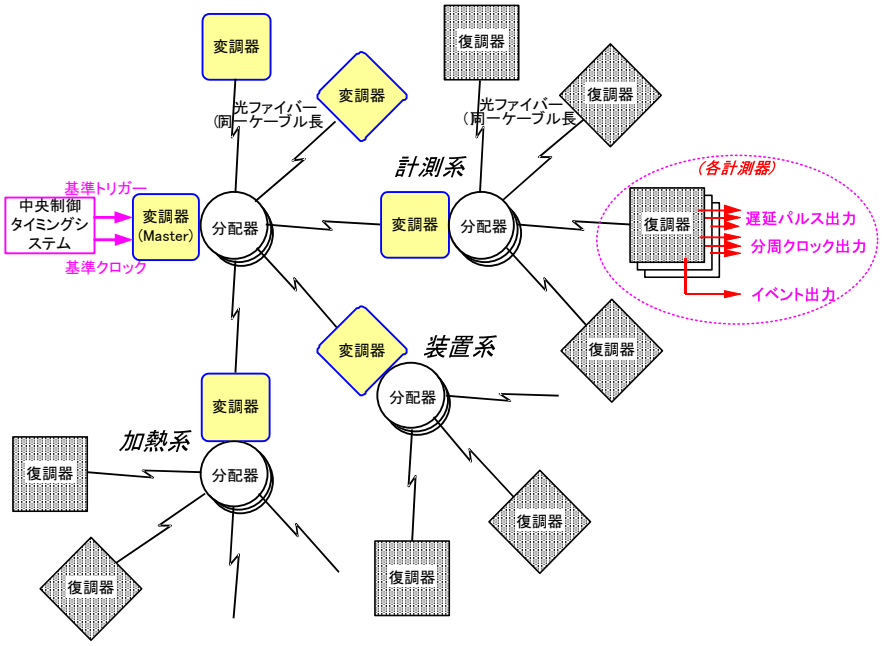


Figure 3.17: Schematic tree structure of LHD Diagnostics Timing System (DTS). There are three types in DTS modules; one master modulator, relay modulator(s), and end demodulator(s).

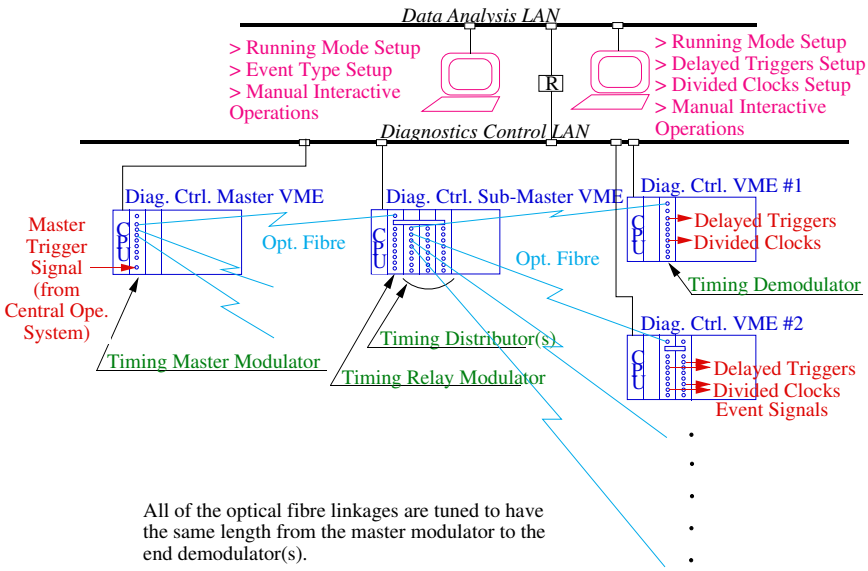


Figure 3.18: Connection view among DTS modules installed in VMEbus systems. Every connections between DTS modules are made by one optical fiber. Only the end demodulators can output the programmable triggers and divided clocks to run digitizers. As multiple fiber links can be branched off, the optical fan-out module is also applied as a VMEbus module.

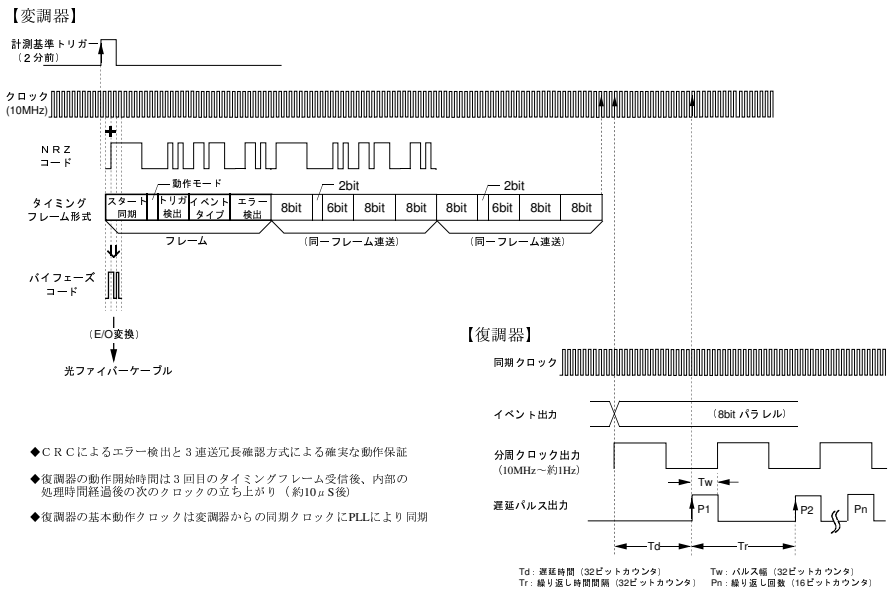


Figure 3.19: Timing message distributions in DTS: The DTS master modulator will encode input timing triggers into 32 bit message synchronized on 10 MHz clock by using NR coding rule, and then distributed three times to the demodulators.

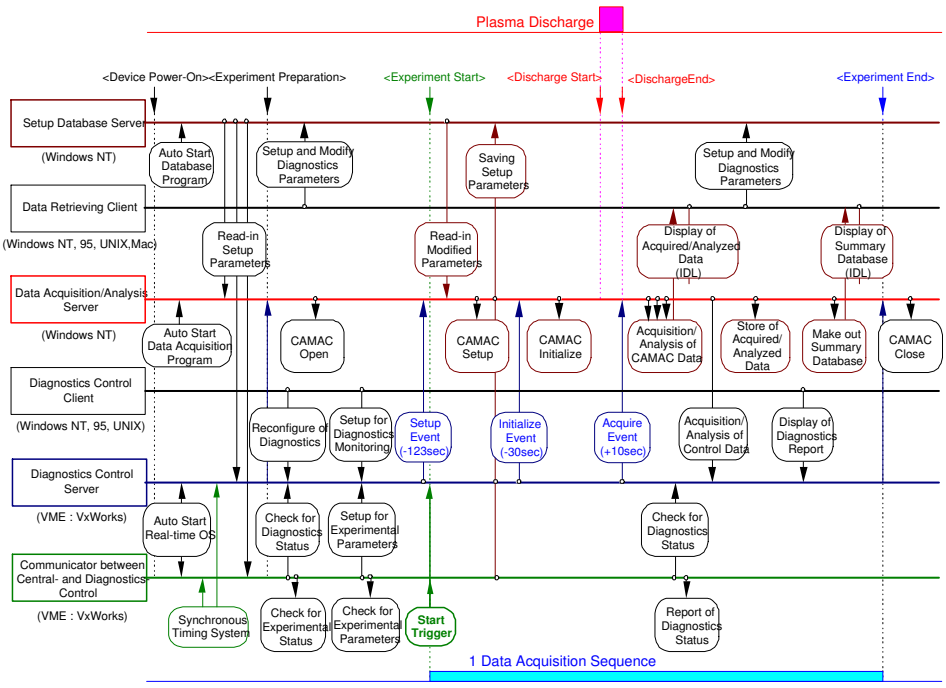


Figure 3.20: Sequence flow chart of each task computers:

3.5.4 計測制御チャネルのオブジェクト指向的データ伝送

計測制御サーバ計算機は，以上述べてきた通り基本的にネットワークを経由したクライアントからの要求を，自律的に解釈及び実行し，その結果を返すという動作を繰り返す．

LHD 計測制御系では，データ処理系と同様にオブジェクト指向方法論によったシステム開発を行っている．計測制御系では一般に，制御すべき機器や信号チャネルの種類が多くなるのが特徴であるが，各々の制御チャネルは RS-232C や GP-IB に代表されるような共通の通信インターフェースを使ったり，あるいは真空排気装置など自律制御ルールが共通であったりと，オブジェクトのクラス定義継承や処理の自律性などオブジェクト指向の特質によく合致した応用分野であるため，

- 制御機器など実際のハードウェアに則して，制御機能の主体をオブジェクト (制御オブジェクト) として自律的に記述できる

  ▼
- 開発分業の効率やコードの再利用性の向上，主体間相互関係の明瞭化
- データ変換などデータ種に依存する手続きを隠蔽することで，抽象化されたデータ取扱いと統一した制御 I/F が実現.
- ネットワーク層の隠蔽によりデータ通信をオブジェクト間のメッセージ授受として実装可能

  ▼

開発負荷の低減とシステムの見通しの向上が期待できる．このため LHD 計測制御系では，Figure 3.21 のように，VME の制御監視モジュール毎に制御主体 (マネージャ・オブジェクト) を，I/O チャネル毎に論理的データ主体 (チャネル・オブジェクト) を生成し，共通クライアントプログラムからの操作を可能にしている.

計測制御系の C/S 間通信は，一般的にデータ共有と機能呼出しとに大別できるが，この計測制御系ではデータ収集系が機能分散により独立して存在し，メッセージ授受による対話的機器制御が主となるため，通信形態はデータ量の少ない後者である．そのため実装としては，リアルタイム応答性が求められる計測制御システムでの利用であり，できる限りシステム負荷の軽いネットワーク処理が必要なことから，インターネット技術のうち遠隔手順呼出し (Remote Procedure Call) として実績のある ONC RPC 4.0 をセッション層プレゼンテーション層に，TCP/IP をトランスポート層ネットワーク層として利用した.

制御機器やチャネルの種類が多い計測制御では，多様な制御ハードウェアや通信形態の違いを吸収・隠蔽し，クライアント側が制御ハードウェアの違いを認識しなくて済む抽象

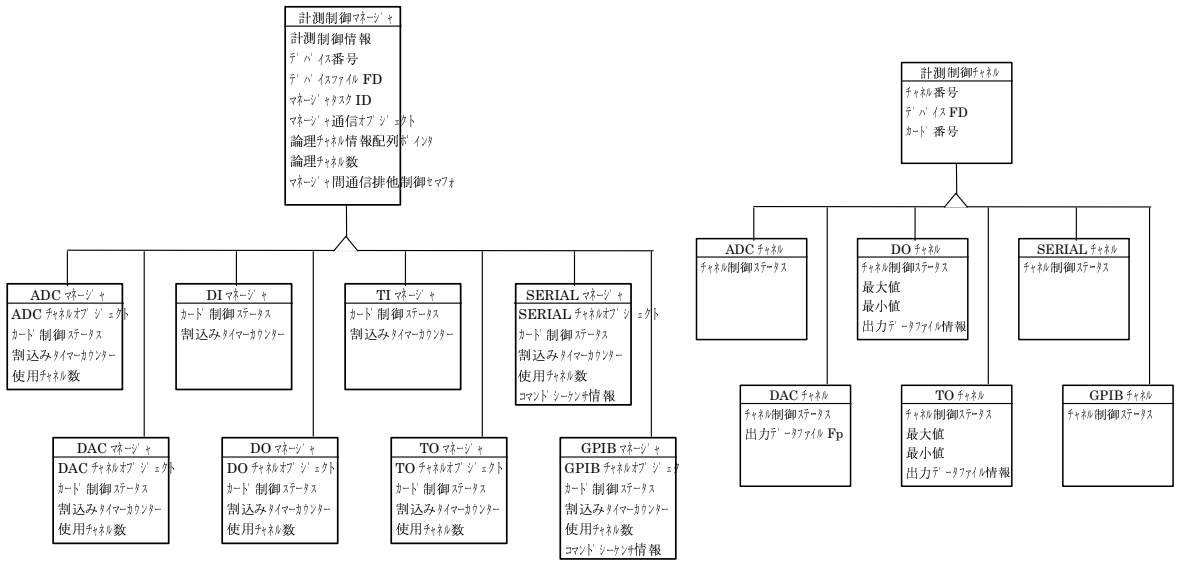


Figure 3.21: Objects' class definitions of I/O manager modules and signal channels Sequence flow chart of each task computers:

化機構が重要となるため，手順呼出しは各オブジェクトの内部メソッドとして隠蔽され，授受されるメッセージにより駆動される (Figure 3.22)．機器状態の監視モニタデータもまた，抽象化された論理チャネルとして統一され，リモートクライアントからのデータ要求メッセージに応じて内部的に RPC 経由で伝送される．

またユーザの計測器遠隔操作支援に必須となる計測制御 GUI 画面についても，Figure 3.16 のバルブ開閉状態アイコンの例などのように，Windows 環境で共通して使える GUI コンポーネントとして ActiveX (OCX) コントロールで構成し，これらが通信のための論理チャネルと

制御チャネルオブジェクト ⇔ 論理チャネルオブジェクト ⇔ ActiveX コントロール

のように一意に関連をもって接続している．オブジェクト間のメッセージ授受として，VME サーバとクライアント GUI 間の通信が抽象化されているため，クライアント GUI 側の開発に対しても大きな利益が得られる．

このように計測制御系においては，大規模な並行分散というよりもむしろ C/S モデルや I/O 別マネージャなど機能別の分散処理が前面に出てくるが，データ処理システムと呼応したスケーラブルな分散形態をとることは可能であり，その再結合手法には同じくインターネット技術とオブジェクト指向方法論が有効であることが，ここでも十分に検証されている．

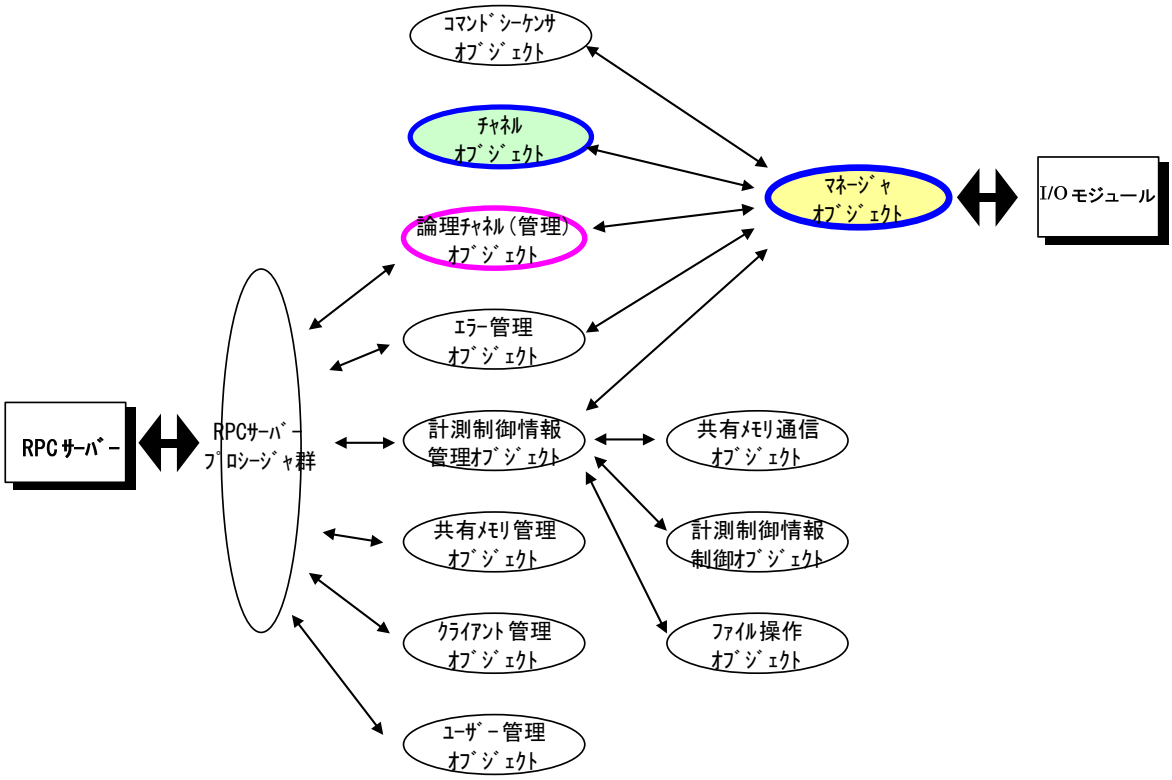


Figure 3.22: Objects' relationship between VMEbus I/O modules and TCP/IP RPC calls.

3.6　階層化された分散データ・ストレージ

2.1.5 節にも述べたとおり，実験データを長期的に格納保管する場所であるデータ・ストレージには，LHD のように特に大容量化が著しい場合であっても，常に収集あるいは作成されたデータが格納できると共に，一旦保存したデータの参照が円滑に行なえることが要求される．大容量の記憶媒体においては，大容量性とアクセス応答性とは一般的に背反しており，アクセス性能も，任意抽出 (ランダム) アクセスと順次 (シーケンシャル) アクセスとでは全く状況が変わってしまう．本研究では，こうした入出力性能からの諸条件・要請による分散システム形態を，OOM に基づいて再結合させ仮想的なマクロ・マシンとして振舞わせる手法の実現が重要な課題になっており，本節ではデータを保管・オンライン参照機能を提供するデータ・ストレージでの取組みについて述べる．

大規模な実験装置のデータ保存システムに求められる基本要件は，上記のデータ参照インターフェースの仮想的な統一化や，将来のデータ量の増加に対処できるシステム柔軟性などの機能面の他にやはり I/O 性能があり，具体的には，保存されている過去データを円滑かつ高速にユーザが取り出せる処理速度が最も重要な性能となる．しかし，データ・ストレージ装置において大容量性と高速性とは背反する性能であるため，データの取出しサイズ，保存容量共に大きくなる大規模実験のデータ処理システムでは，単一機種によるストレージ構築が事実上困難となっている．このため異機種のストレージ装置を複数組み合わせた分散ストレージ・システムを構成することで上記の両性能を確保すると同時に，それらを透過的に見せることで一つの仮想的なオブジェクト・ストレージとして振舞わせる必要がある．

また，Table 3.2 にも例示したとおり，核融合プラズマの計測データは取出し単位となる各計測・各ショット毎のサイズでおよそ数 MB〜数 10 MB 程度であり，素粒子物理実験などの計測データ ～数 10 kB/event から較べてかなり粒度が大きいデータとなっている．しかし，昨今の大容量データの典型例である音声・動画像を伴ったマルチメディア・データの典型的サイズの数 10 MB〜GB や，あるいはデータバックアップ記憶媒体の容量：数 GB〜数 100 GB に較べるとやや小さく中間的な範囲にあるともいえる．LHD 計測データ収集の諸元を簡単にあげると以下のようになる．

- 1 実験（ショット）毎に 30 計測からデータを収集（150 ショット/日）
- データの取扱いはチャネル単位（数 100 kB〜数 MB/チャネル）
- チャネル・データはソフトウェアにより圧縮（zlib 使用，データは 20% 程度に圧縮）

- 収集データ量は約 740 MB/ショット（平成 14 年度実績）
- 1 日の全収集データ量は約 13 GB（圧縮後，平成 14 年度実績）

データ数としては全ショット数×計測数となり，例えば年間 10,000 ショットの実験を行った場合，10 年で $10,000 \times 10 \times 30 = 3,000,000$ といったような多数にのぼる．全実験データを保管・提供するオンライン・データストレージには多数ユーザから任意にデータ参照要求が行われるため，こうした多数のデータ中から所定のものを取出す動作は，核融合プラズマ計測の実験ごとの独立性の高さとも相まって，ランダム・アクセス性が極めて高くなっている．以上の考察から，核融合プラズマ実験のデータ・ストレージでは，全てのデータ・ストレージをランダム・アクセス・メディアを利用して構築することが望ましいことが容易に理解される．

更に，昨今ほとんどの大規模物理実験はすべからく共同研究利用に供せられるため，過去に収集されたデータは基本的に常時全てネットワーク経由で参照可能になっていることも求められている．

以上述べたデータ・ストレージ・システムに対する諸要件の分析から，本システムのデータ・ストレージ設計概要を列記すると以下のようになる．

- 異機種装置を用いた 3 階層オンライン分散ストレージ．バックアップ装置は分離．
- 中・長期ストレージはディスクアレイとライブラリ装置の併用で，アクセス応答性・入出力速度と大容量性を相互補完．
- 分散オブジェクト・ストレージを一体の仮想的マクロ・マシンとして振舞わせるためのデータ所在情報の仲介サービス．
- 実験サイクル毎に必要な容量だけ増設することで投資コストを抑制．
- 低コストで有効な災害対策を講ずる．

続く各小節ではこれら各要件と対応するシステム設計について順次詳細な議論を加える．

3.6.1 保存データベースの分割・階層化

ODBMS が提供する仮想記憶ボリュームを用いて BLOB の実験データオブジェクトを永続化・保存するいわゆる保存データベースは，3.4 節でも述べたとおり LABCOM システムの場合，MPP 形態をとる分散データ収集サーバ上に同数だけ分散して存在する．これらは当然データ参照クライアントからの要求に応じて，保存されているデータオブジェクトを読出し・転送するのだが，

- 約 30 セットの分散収集サーバは小型計算機ベースであり，ローカルに持てるデータ・ストレージの容量に限界がある．
- データ収集中に過去データ参照要求が多く入ると 1 つの収集サーバ上でも I/O 過多となり収集動作に悪影響を及ぼす恐れがある．
- 実験シーケンスに同期して他の大容量ストレージ装置にデータを集めるのは，そこで I/O 集中が起こるため不可能．

という条件から，収集サーバがもつローカル・ストレージの容量はあまり大きく出来ないため，必要とされる大容量ストレージ装置を別のサーバに設ける必要がある．換言すると，容量の大きい計測データ・オブジェクトを格納するストレージは複数に分離されており，しかもその I/O を制御するホスト計算機も分散する必要があるため，ネットワーク分散ストレージ構成が求められる．そしてデータ移送 (Migration) は，実験シーケンスが止まりユーザからの参照要求も少ない夜間に実施するということになる．

核融合実験をはじめとする大型装置を用いる物理実験一般では，通常，その装置を運用する実験期間が比較的長い場合が多く，10〜20 年といった単位で実験プロジェクトが継続されることになる．当然，計測・蓄積される実験データの総量は相当なサイズになるが，これを全て保管するだけの大規模ストレージを一度に導入するのは，長期にわたる実験期間中の計算機技術の進展を無視して，10 年以上前のストレージ技術に依存することになり，極めて非効率でかつコスト/パフォーマンスが悪いことは自明である．LHD 実験プロジェクトでも全く同様のことが想定されるため，ここでも大規模ストレージ装置を採用することは止め，生成された実験データ量だけ逐次，保管ストレージ量を増設していく方策を採っている．

特に数年以上を経過した過去の実験データについては，一般に参照される頻度が下がっていくため，比較的新しく参照頻度の多い実験データと同じストレージ装置に格納して，常に参照要求に備えているのは，コスト的に非効率となる．しかし遠隔実験データ参照や共同研究利用の目的を通常併せもつ大型実験装置では，全ての過去データを常時参照可能にしておく必要があるため，古くなった過去データについては，保存 (メディア) 単価が安く，しかも保存メディアが自由に増設可能で，しかもある程度の時間内にデータが参照可能になる，メディア交換機能のあるライブラリー装置を使うのが合理的である．

以上の考察に基づいて，保存データベース群を構成する LHD データストレージは，保存期間とデータアクセス速度の違いによって，以下のような 3 階層に分類された分散ストレージ構成を採ることにした．

第 1 層　短期ストレージ

- 計測毎にローカル接続されたディスクアレイ（50GB，RAID-5）
- データは ObjectStore データベースへ格納
- チャネル・クラス単位での保存（メモリ・イメージをそのまま保存）
- 計測により保存可能な期間が異なる（短いものでは約 2 週間）

第 2 層　中期ストレージ

- ネットワーク上に用意された大容量ディスクアレイ (720GB+1600GB，RAID-5)
- データ保護のため 2 組のディスクアレイをミラーリング (RAID-1)
- データは専用ファイル形式に変換して格納.
  全チャネルを含むショット単位での保存（パラメータ＋バイナリデータ）
- 実験サイクル毎に増設

第 3 層　長期ストレージ

- 光メディアを使用したメディア交換型大容量ライブラリ装置
  MO ジュークボックス (1200GB × 3)，DVD チェンジャ (3100GB)
- データ保護のため 2 セットの MO ジュークボックスに同じデータを保存
- データは専用ファイル形式に変換して格納.
  全チャネルを含むショット単位での保存（パラメータ＋バイナリデータ）
- 1 ライブラリで複数実験サイクルのデータを保管

この 3 階層構成の分散ストレージの模式図を Figure 3.23 に示す．なお，短期および中期ストレージについては，磁気ディスク (HDD) アレイの標準規格である RAID (Redundant Array of Inexpensive/Independent Disks) を用いて，データ転送の高速化と冗長化による高信頼性を確保している．長期ストレージについては 3.6.4 節で改めて述べる．

3.6.2　ODBMS の限界と RDBMS

LHD データ収集系に於けるデータ保存，設定，検索用データベースは，実験データが全てオブジェクト単位で扱われることを考慮すると，オブジェクト指向データベース (OODB: Object-Oriented Database) で統一され，全てのデータが一個の OODB 空間中に格納されているかのように見える事が，管理・運用コストの面を考慮しても望ましい．しかし OODB あるいはオブジェクト指向データベース管理システム (ODBMS) 中でオブジェクト集合体から所定のオブジェクトを検索する速度は，RDBMS 中のテーブルからレコードを検索する場合に較べると明らかに遅い場合が多く，保存・登録されるデータ件数が将来増えた場

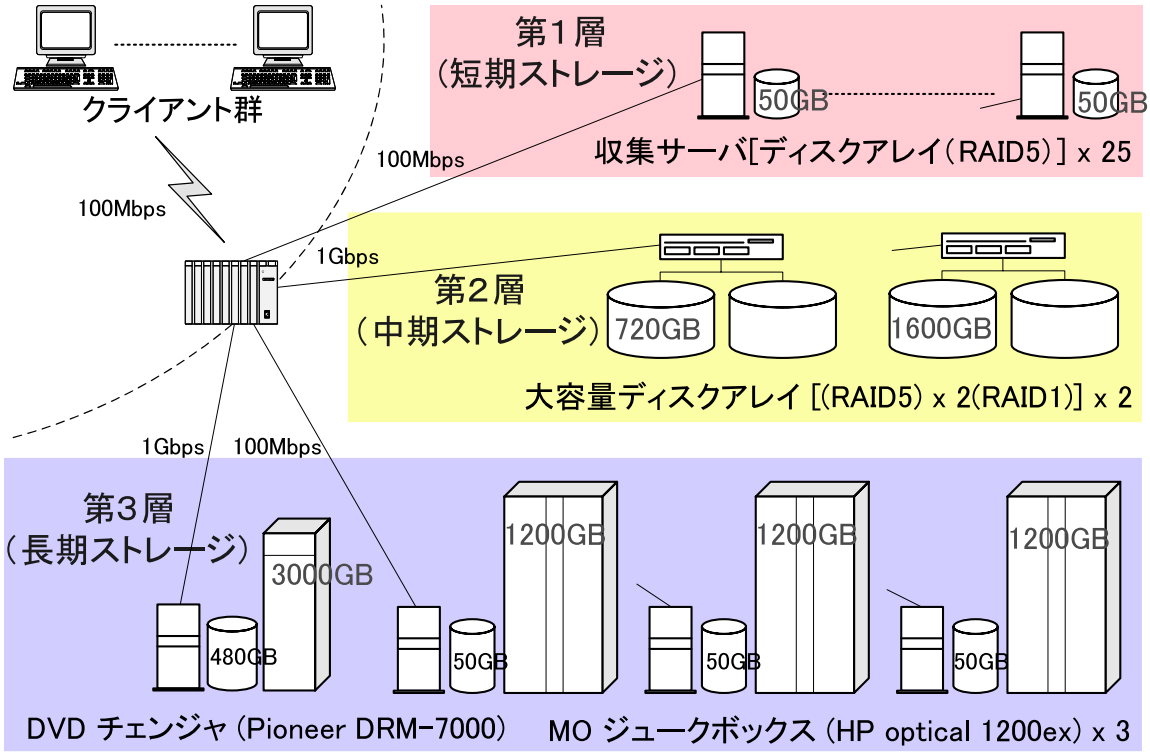


Figure 3.23: Three layers of LABCOM data storage system. Each layer has been distinguished by its I/O throughput, capacity, and number of distribution.

合に性能が大きく劣化する恐れがある．

ODBMS と RDBMS との検索処理を比較すると，単純な二次元テーブル構造のデータのみを取り扱う RDBMS に対して，任意のデータ構造を C++/Java 言語等で記述されたオブジェクト指向アプリケーションから透過的に取り扱える ODBMS は非常に高機能であり，当然サーバ側の DBMS もその分だけ重い処理となっている．それに加えてアプリケーション・スキーマとの透過性を実現する ODBMS クライアントでは，RDBMS の場合と比べ，

ODBMS　クライアント・セントリックな処理構造

アプリケーション・スキーマとの透過性をサーバ側で処理してすると負荷が集中しすぎるため，主たるアプリケーション処理やキャッシュ機能は，クライアント側のアプリケーション・プロセス (の API) で分散して行っている．このため，特に BLOB の集合体などが格納されている場合は，C/S 間の通信量が非常に多くなり，通信によるボトルネックが発生しうる．

RDBMS　サーバ・セントリックな処理構造

サーバ・プロセスでのクエリ処理が主で，キャッシュもサーバ側で行われるため，

C/S 間の通信量は SQL 文などというように少なくなる．

というような処理構造の違いがある．このため，ODBMS は分散形態が比較的容易に実現できるが RDBMS では中央集中型の形態しか取れないためサーバが巨大化する欠点がある．しかし，C/S 間データ授受が無くサーバプロセス中で処理が完結する RDBMS のほうがオーバーヘッドが少なく，また B-tree，hash テーブルといったインデックス機能を利用して検索速度を容易に向上できるため，検索速度に関しては一般に高速である．

Table 3.7 に具体的な LHD 計測のいくつかの計測データを用いて，ODBMS (O_2) と RDBMS (PostgreSQL) とを使用した際の検索速度を比較した結果を示す．この場合の ODBMS は，大きなバイナリデータオブジェクト (BLOB) [注6]そのものを保存する，いわゆる実験データ保存データベースではなく，RDBMS と同等の検索処理を行わせるために，ODBMS ホスト計算機上で検索クエリを処理する専用のサーバアプリケーション Indexd を動作させて，C/S 間のネットワーク通信オーバーヘッドを無くしたうえでの動作結果である．OODB 中で検索キーとなる主要メタ・データをインデックス化して 20〜30 倍程度の高速化を施しても，同様のインデックスを張った RDB とでは，検索速度がまだ 1 桁以上も異なるなど性能差は歴然としている．

Table 3.7: Elapsed time difference for data searching query (unit: micro-second). Here the total number of registered records is 559939 and the host computer is dual Pentium-III 500MHz machine with 512 MB memory.

Diagnostics	O_2 wo/ index	O_2 w/ index	PostgreSQL w/ index
SXmp	7771701.4	-	23099.4
FPellet	6356770.8	-	23276.0
SXfluc	-	289713.5	23112.0
FIG	-	277018.2	20182.3

RDBMS ベースで開発された市販 DMBS の中にはオブジェクト指向対応を銘打つ製品もあるが，純粋な ODBMS に較べてユーザによるクラス定義などオブジェクト指向モデリングに制約があり，リレーショナル・データベースに ODBMS 機能を多少拡張しただけの Object-Relational Database Management System といわれる物が多い［79, 73, 80］．こうした製品でのデータストレージ実現は今後の検討課題の一つではあるが，価格が非常に高

[注6] Binary Large Object の略．狭義のデータそのものである次元数値配列等を含んだ比較的サイズの大きいデータ塊のこと．

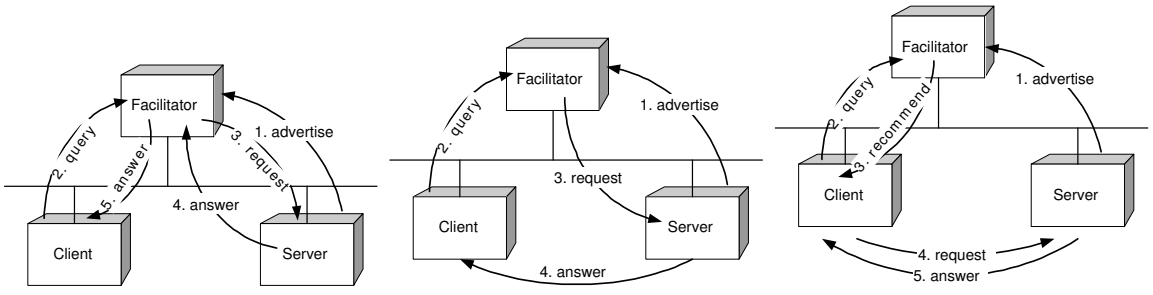
価であったり，小型計算機用の OS では動作しなかったりと，システム全体の可搬性を下げる点も多いため，本システム開発での導入は見送っている．

3.6.3 独立した遠隔問合せ機構と検索データベース

分散データベース/データ・ストレージを実現するためには，本来一個の OODB が全ての分散データ・オブジェクトを格納し，それらの参照サービスを提供できれば良い．しかし，3.6.1 節で要件分析した通り，大容量性とアクセス速度，増設性の要請から，OODB 以外のファイル/ファイルシステムを用いたストレージ装置を併用せざるを得ない．また，前節でも述べたとおり，ODBMS のみではデータベース中の登録・保存件数が増えた場合，その検索速度の遅さも問題になってくる．

このため本システムでは，データ参照クライアントに対してデータの所在情報を検索・仲介する検索データベース・サーバを，実際に計測データを格納し参照要求に応じて転送・提供する実験データ保存データベース・サーバとは別機構として独立して存在させる設計をとった．この検索データベースは，取り扱う情報が単純なデータ所在テーブルであることから，3.6.2 節の RDBMS/ODBMS 検索速度比較から考えても，RDBMS によって運用されるほうが 1 桁以上高速な処理が可能となる．

ここでの検索データベースの役割としては，(データ名，ショット番号) をキーとしたデータサービス要求に対して，保有ホスト名などの所在情報を検索・推薦するネーミング/トレーディング・サービスであり，このように要求メッセージをルーティングしてくれるサーバは特にファシリテータ (Facilitator) と呼ばれている［56］．Figure 3.24 に示すとおり，ファシリテータには要求メッセージのルーティング方法の違いによって (a) broker, (b) recruit, (c) recommend の各型がある．



(a) broker serbvice (b) recruit service (c) recommend service

Figure 3.24: Types of request message routing by the facilitator［56］.

こうしたファシリテータを介した遠隔 (問合せ) メッセージ・パッシングは，OOM モデルにおける分散オブジェクト間の動的結合の一つの形態であり，OOM コンピューティングでより自由度の大きいオープンシステムを実現するために重要な自律的適応と捉えることができる［56］．即ち，ファシリテータを利用することで，システム全体，特に複数データサーバへの動的な負荷分散や部分故障への対処機能も同時に実現している事になる．

LABCOM データストレージシステムでは，大容量データを取り扱う観点から，I/O 負荷がファシリテータに集中する (a) broker 型や C/S モデルでのデータ転送が組み難い (b) recruit 型でなく，ファシリテータのうちの (c) recommend 型をベースにしたアーキテクチャを適用している．Figure 3.25 に見られるとおり，本システムのファシリテータ Data Index Table サーバと，データサーバ Transd，データ参照クライアント retrieve の三者間の連携動作は Figure 3.24 に示したファシリテータの推薦 (recommend) 型とよく合致している．この場合のメッセージ通信言語は，ファシリテータが RDBMS であることから標準の SQL (Structured Query Language) となる．

ファシリテータを用いて分割・階層化されたストレージ・サーバ群を，3.6.1 節で述べた異機種ストレージの階層的な分散利用形態によって提供するためには，記憶媒体・機構の違いを吸収してクライアント側に統一的に見せる中継 (変換) 機構の介在が必須である．これらは一般的にアプリケーション・サーバと総称されている．

Figure 3.25 に示したとおり，本ストレージ・システムのデータ転送サーバ Transd は，OODB ボリューム，RAID 上のファイルシステム，MO ジュークボックス/DVD-ROM チェンジャ上の仮想ファイルシステム，という異種記憶装置で構成される 3 階層ストレージの全サーバ上で動作し，OODB 中の永続性オブジェクト，それをシリアライズしたファイル・インスタンスを，もとの C++アプリケーション・オブジェクトに復元してクライアント側に転送している．また収集直後の高速データ参照のため，データ収集サーバ上で稼動する Transd は，共有メモリ上のメモリーマップト・ファイル化されたインスタンスの読出しにも対応している．

前の 3.6.1，3.6.2 節で述べたとおり，ODBMS によって大容量ストレージ装置を直接管理するには，容量・速度性能・媒体管理機構など様々な面で制約・制限が多いが，このファシリテータと分散アプリケーション・サーバの組合せによって，ODBMS のストレージ面での制約から解放されつつ，しかも仮想的な大規模オブジェクト・ストレージ空間を形成するのに成功した本手法は，新たなシステム技術上の知見であるといえる．

なお，この分散ストレージ・サーバ群を仮想的な一個のオブジェクト・ストレージ空間としてシームレスに見せる機構の性能等評価については，データ参照クライアント側の視

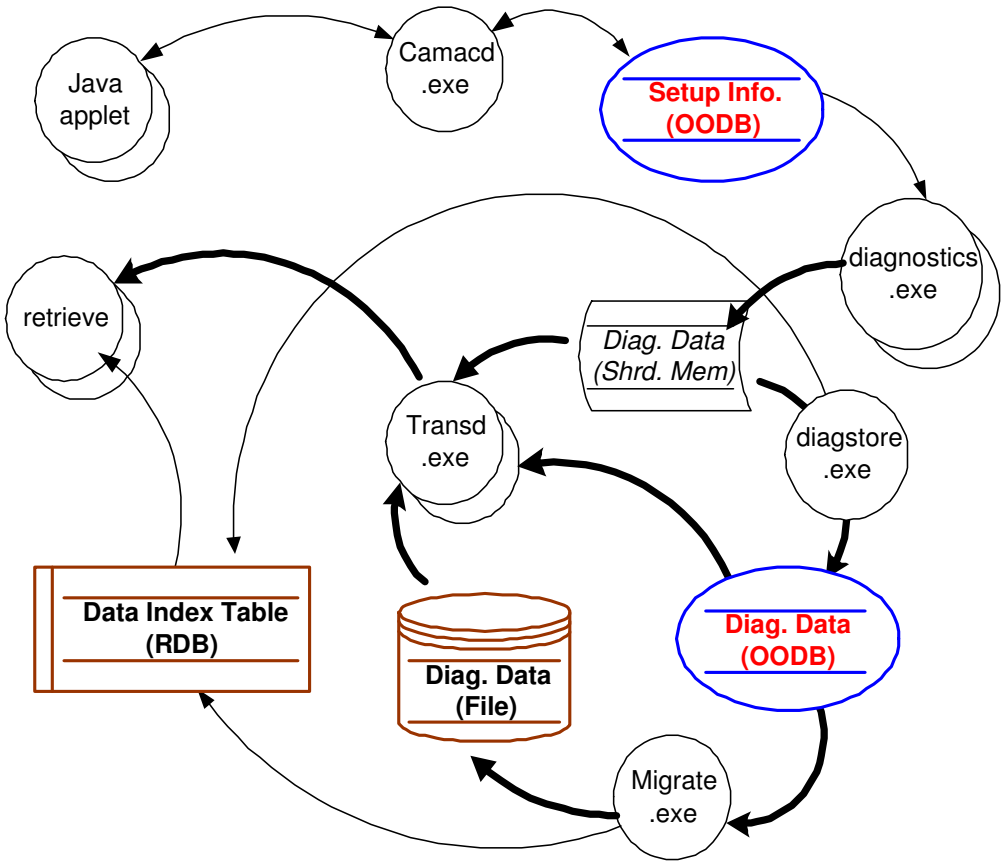


Figure 3.25: Flow diagram of 3-layer data transferring service: Just after acquiring from the CAMAC digitizers, the raw data will be stored once in the volatile memory area for the rapid retrieval without any disk access. Another data storing program “diagstore.exe” will make them persistent into the OODB virtual volume. For the consistency with the file/filesystem-based mass storage system, the persistent objects will be converted into files again. However, the data transferring service program can deal with all of three media types; the memory mapped files, OODB object instances, and files in filesystems.

点から，改めて後の 3.7.2 節で詳述している．

RDBMS でのインデックス付き検索の処理速度に関しては，登録レコード件数 N に対する検索時間が機構的に $\log_2 N$ に対応するため[注7]，Table 3.7 の登録件数 56 万件を 240 万件に増やしても検索時間がほとんど変わらないことを確認している．しかし性能的な上限は存在するため，登録件数の増加によって検索にかかる時間が無視できないほど長くな

[注7] インデックスとして一般的な B-tree を利用した場合．

り，システム性能に影響を及ぼす可能性も残っている．この性能上限については，改めて 3.9.3 節で考察を加える．なお，本システムでは，ファシリテータ・エンジンとしてフリー RDBMS で最も信頼性の高い PostgreSQL を用いている．

3.6.4 大容量ストレージ装置と仮想ボリューム管理の評価・導入

全ての実験データをオンラインで継続的に参照可能とすることを目的および定義としている大容量ストレージは，通常状態ではオンラインサービスを目的としないいわゆるデータバックアップ装置とは異なり，

1. 高速でかつアクセス性の良い短期保存装置
2. 比較的低速でアクセス性が劣るが大容量を実現できる長期 (半永久) 保存装置

の 2 つの機能の装置をうまく組み合わせてオンライン・ストレージを構成することが多い．つまり，高速な短期保存装置と大容量の長期保存装置との連携を，キャッシュ技術等を採り入れてうまく図る事が重要な課題である．これは，格納されたデータの参照時にユーザが短期及び長期保存装置の違いを認識する必要が無くなる利点があり，また同時に，実際のデータ格納作業を透過的に仮想化・自動化して大容量データ取扱いの管理者負担を大幅に軽減することにもつながる．

ここで，現在流通しているデータ保存媒体を挙げると，紙，磁気テープ (MT)，磁気ディスク (HDD)，光磁気ディスク (MO)，光ディスク (CD，DVD) 等がある．紙 (連続紙) に関しては，現在はペンレコーダ等の低速監視装置での用途のみで，データの保存に使われることは殆んどない．各種ストレージメディアの性能比較表を Table 3.8，3.9 に示す．

LHD データシステムの長期保存ストレージ装置には，記憶メディアの可搬性，格納データ容量とランダム・アクセス性を考慮して，設計当初から DVD (Digital Video/Versatile Disk) の導入を計画していた．DVD は普及している CD-ROM と同一サイズながら，記憶容量が片面約 4.7 GB に大幅に向上した光ディスクである．音楽 CD (CD-DA) 規格に対して CD-ROM 規格がデータ保存メディアとして普及したように，ビデオ映像メディアとしての用途を担って流通し始めた DVD-Video 規格に対するデータ保存規格 DVD-ROM も同様に一般に普及し，また，メディア単価のみならず CD-ROM チェンジャのようなメディアライブラリー装置も民生品の登場でデータ専用 MSS に較べてかなりの低価格で導入可能になることは十分に予見された．

しかしながら書込み可能型 DVD メディアの規格策定は，各メーカー間での規格調整が難航したため大幅に遅れ，結果としても DVD-RAM と，DVD-R/-RW，DVD+R/+RW

Table 3.8: Specification comparison among various kinds of popular mass storage systems.

	MO	MT (DTF2)		DVD-RAM	DVD-ROM
series	hp SureStore	Sony PetaSite		ASACA	Pioneer
product	1200ex	DMS-8400	DMS-B35	AM-1450DVD	DRM-7000
media cap.	4.8 GB	200 GB	200 GB	9.4 GB	4.7 GB
total cap. (compressed)	1.2 TB	26 TB (67.3 TB)	7 TB (18.1 TB)	10.3 TB w/ 24 drives	3.4 TB
write (MB/s)	2.3	24	24	2.76	2.76
read (MB/s)	4.6	24	24	2.76	2.76
drives	6	4	1	max. 24	max. 16
media exchg. (s)	10	60	20	5 (11-16)	<9
load/unload (s)	5.5/3.0	-	-	14	-
random seek (s)	0.035	71.4	71.4	0.1	0.12
power spent (W)	560	4090	860	450	400
weight (kg)	223	810	323	230	91.7
unit cap. (TB/m^2)	1.8	13	9.9	29	12.1

Table 3.9: MT performance comparison against the SCSI HDD drive: As HDD can take RAID formation, MT and DVD-RAM library also have the similar array called RAIT or RAIL. Capacity values are non-compression ones.

type	spec.	capacity	media rate	remarks
HDD	Ultra160 SCSI	73 GB	49 MB/s	(Seagate Cheetah X15/73)
MT	DTF2	200 GB	24 MB/s	(Sony DTF2 GY8240)
	LTO Ultrium 2	200 GB	35 MB/s	
	LTO Ultrium	100 GB	15 MB/s	
	SuperDLT	110 GB	11 MB/s	
	DLT8000	40 GB	6 MB/s	

陣営に分かれるなど製品化にはかなりの時間がかかった．このため平成 9 年度末に実験を開始した LHD 用のストレージ装置として書込み型 DVD を導入するに至らず，その時点で同じ容量が実現でき，かつランダムアクセスメディアである光磁気ディスク (MO: Magneto-Optical disk) ジュークボックスを暫定的に採用することになった．

階層型ストレージ管理システム (HSM)

LHD 長期保存ストレージ装置として最初に導入したのは，媒体自動交換ロボット機能を持つ光磁気ディスク・ジュークボックス HP 製 SureStore Optical 1200ex Jukebox で，8 倍密度の両面 MO メディアを 238 枚収容可能で 1 台あたり約 1.2 TB の総容量となる．物理的な記憶ボリュームの単位としては，メディア片面の 2.4 GB であり HDD の容量と較べても十分な大きさとはいえない．このため複数の物理ボリュームをつなぎ合わせて，一つの仮想ボリュームに見立てて利用する仮想ボリューム管理機能が，大容量ストレージとして使用される際には一緒に用いられる場合がほとんどである．

こうした仮想ボリューム管理機構の一つとして広く普及している技術が HSM と呼ばれる階層型ストレージ管理 (Hierarchical Storage Management) システムである．HSM とは，テープ／ディスクなどの外部ストレージ装置を階層的に考えて，参照頻度や使用頻度の高いデータはより高速な装置 (HDD など) に配置し，参照頻度が低いデータについては低速な装置 (テープ装置/光ディスク/MO) に移動する仕組みのことである［81］．一般的にテープなどは HDD に比べコスト (バイト単価) が低いため，長期間使用しないときはテープに保存し，処理するときだけディスクに移動するのが効率的となる．この HSM の機構は半導体メモリーなどによるキャッシュ機構とは異なり，全てのデータ・ファイルは基本的に同時に一箇所しか存在せず，高速媒体⇒低速媒体 (stage-out)，低速媒体⇒高速媒体 (stage-in) の移動をファイル・ベースで行うものである．

この HSM 機能により MO ジュークボックスを大容量のデータ記憶装置として管理・運用可能にするのが，東芝/日本テクノラボ製 MicroHSM である．これは，

- 専用 PCI コントローラボードによりホスト CPU の負荷を軽減
- RAM/SCSI HDD をキャッシュや高速 stage に使用して I/O を高速化
- 複数枚メディアをグループ化して仮想ドライブに割り当て，OS のファイル・システムとして使用可能
- GUI による運用管理ツールで運用設定や保守管理が容易
- Windows NT 以降のファイルシステム NTFS で利用可能

等の特徴を持っており，PC 用に開発されて比較的利用が容易であったため，小型計算機ベースの分散ストレージ・サーバにふさわしいと判断し導入した．運用は Windows NT4.0 上の NTFS5 ファイルシステムとして仮想ボリュームを利用している．しかしながら，MicroHSM による MO ジュークボックス運用上の問題点としては，

- 書込み処理中は読出し速度が大幅に遅くなる
- 毎日のデータ収集量増加に伴って書込みに長時間を要するようになった
- MicroHSM は OS 依存のデバイス・ドライバが必要なため，OS バージョンアップへの対応が遅れる
- 仮想ボリュームの内部構造がブラックボックスとなり，またメディア毎の可搬性が無いため，障害発生時に必ずメーカ解析が必要になり対応・保守コストが上昇
- データファイルは stage-on/out で複製コピーではなく移動されるため，移動中に障害が発生するとデータを消失する危険性が高い

等が判ってきたため，当初導入した 3 台のみを読出し専用で運用し，それ以降の容量増強などは MO ジュークボックスには行っておらず，新規保存データについては後述する DVD-ROM チェンジャーの利用へと移っている．

ダイレクト・アクセス型仮想ボリューム管理

複数記憶媒体をまとめて一つの大容量の仮想ボリュームとして提供する機構としては，前述の HSM のほかに半導体キャッシュ機構と同じ動作をするダイレクト・アクセス型仮想ボリューム管理方式がある．これを実現する代表的なソフトウェア製品が ADIC 社製 AMASS for UNIX [注8]である．

AMASS では HSM が持っていたファイルの stage-in/out 機構に付随するデータ損失の危険性と運用の難しさを解消するべく，狭義のキャッシュアルゴリズムをそのまま，仮想ボリューム上の物理セクターとキャッシュディスク上に応用することで，キャッシュ機構による仮想ボリュームへの実効的な I/O 高速化を目指している．これは利用側に対して HSM のように運用負担を発生させない優れた方式であり，MO ジュークボックスにつぐ長期保存ストレージとして LHD でも試験導入を行った．AMASS を DVD-RAM チェンジャにて動作させた場合の転送速度調査例を Table 3.10 に示す．

LHD での試験導入では，MO ジュークボックスより大容量が必要であることから 10TB

[注8] http://www.adic-japan.com/Product_AMASS.htm

Table 3.10: I/O rate example of AMASS filesystem with ASACA's DVD-RAM library AM-750DVD. In the 2nd condition, its slower CPU limits the /dev/null I/O speed and thus their throughputs becomes worse because the read outputs are thrown into the null device.

condition	elapsed time (/s)	total throughput (kB/s)
500MB single write	14'16"	598.13
500MB double write	21'50", 21'28"	781.68
500MB triple write	26'54", 26'54", 27'13"	940.60
500MB single read	8'37"	990.33
500MB double read	14'41", 14'40"	1162.32
500MB triple read	17'10", 17'14", 17'18"	1479.77
Note: HP9000 E35/HP-UX 10.20. Data and cache partitions are on the same disk.		
500MB single write	13'10"	648.10
500MB double write	12'58", 12'52"	1316.19
500MB triple write	13'22", 13'14", 13'00"	1915.21
500MB single read	8'48"	969.69
500MB double read	23'36", 23'35"	723.16
500MB triple read	33'24", 34'07", 34'06"	750.36
Note: Sun SS20/Solaris2.6. Separated data and cache disks.		

の容量を持つ SONY 製 DTF2 テープライブラリ装置 PetaSite を用い，ホスト計算機に Sun Enterprise 4500 とキャッシュディスクとして 100MB/s FibreChannel 接続 1.3TB RAID 装置を使用した．I/O 性能実測結果を Table 3.11 に掲載するが，MO 書込みと同程度の 55kB/s 程度の非常に遅い速度しか出ていない．この主原因は AMASS と PetaSite の取扱いブロック長の違い等によるものであることが判明しているが，結果として AMASS 等の各種調整では十分な速度改善に至らなかった[注9]ため，結局，AMASS による大容量オンライン・ストレージとしての利用をあきらめ，ネットワーク対応バックアップ・ソフトウェアである BakBone 社製 NetVault によるバックアップ装置として PetaSite を転用している．

[注9] この原因究明と対策については主に NIFS 江本雅彦氏により作業がなされた．

Table 3.11: Writing speed comparison among various kinds of storage medias. Applied data size is totaly 4.57 GB with 903 files.

storage	medias/exchange time	elapsed time	throughput	
MO jukebox	4.8 GB×238 = 1.2 TB	11 h 4 min	112 kB/s	write
(w/ MicroHSM)	(ave.) 10 s	21 h 34 min	57 kB/s	write,stage-out
DVD changer	4.7 GB×670 = 3.1 TB	27 min	2.8 MB/s	write
(w/ jukeman)	(max.) 9 s	50 min	1.5 MB/s	write,verify,compare
SCSI Disk	-	14 min 25 s	5.3 MB/s	NTFS5 write
DTF2 PetaSite	200 GB×50 = 10 TB			
(w/ AMASS)	(max.) 90 s	–	55 kB/s	400 MB write
(w/ NetVault)		–	11∼12 MB/s	556 GB write

DVD-ROM チェンジャーと UDF フォーマット

DVD 記憶媒体の大容量ストレージへの採用については，民生用ライブラリ製品等の発売・普及が LHD 実験開始 (1998.3) に間に合わなかったことにより一旦断念されたことは前述した．しかしその後，民生用の DVD-ROM メディア自動交換機構つきライブラリ装置が 1999 年に Pioneer 社から発売された．光ディスクチェンジャー DRM-7000 がそれで，Table 3.8 にも示したとおり最大 720 枚 (3.38TB) の CD/DVD 媒体を収容可能で標準価格 120 万円である．この価格は同程度の容量の MO ジュークボックスや DVD-RAM チェンジャーに較べるとほぼ 1/10 であり価格対容量比が非常によいばかりか，小型 (PC) 計算機での利用を前提にした製品であるため，管理用のソフトウェアも一般流通製品で信頼性・実績があり価格的にもこなれている．多くの点で PC 計算機をホストにした分散ストレージの構築に非常に適した装置であるといえる．

DVD-ROM チェンジャ用の管理ソフトウェアとしては，CDROM ジュークボックス装置用としても実績のある iXOS 社製ライブラリ管理ソフトウェア Jukeman [注10]を選定した．これによりチェンジャ内の複数ドライブやメディアを一元管理すると共に，キャッシュ機能の有効利用により高応答性を実現している．関連する主な特徴は以下の通り．

1. 各メディアは仮想ディレクトリ内のサブフォルダとして一元管理
2. ディレクトリ・キャッシュとデータ・キャッシュによる高速アクセス

[注10] iXOS 社の iXOS-Jukeman から Smart Storage 社 SmartStor Jukeman へ，更に OTG Software 社を経由して LEGATO 社 ArchiveXtender Jukeman として現在に至っている．

3. 仮想ボリューム管理は未サポート
4. 対応ファイルシステムは ISO9660/UDF ブリッジ形式のみ

Jukeman による管理が前述の HSM や AMASS と大きく異なるのは，仮想ボリューム管理機能が無い点である．このことにより，ストレージに格納される大量のデータに見合った大容量サイズのファイルシステムを仮想ボリューム上に構築することが出来ず，DVD-ROM メディアの標準ファイルシステム UDF (Universal Disk Format) [注11] による物理的ボリューム単位での利用に制限されることになる．しかし逆に，仮想ボリュームの欠点であった各記録メディア単位での可搬性が無い点は解消され，DVD-ROM メディアをジュークボックスから 1 枚取り出してそれを UDF ファイルシステムとして単体で利用することも可能になっている．これは 3.8 にも改めて記述するとおり，障害復旧やデータ単位の保守には極めて好便であり，こうした保守性の良さが欠如している HSM 利用への反省からも Jukeman 採用に踏み切っている．なお，その他の DVD チェンジャおよび Jukeman の利用上の問題点としては，

- DVD の実ボリューム容量 4.7GB に制約されるため，ショット毎になる保存データ区分との関連で各メディア上に未使用領域が発生し利用効率が下がる
- 実ボリュームが多数にのぼるため，ファシリテータ (検索データベース) に登録するデータ所在情報の種類が増える．
- Jukeman は DVD-R による書き込みもサポートしているが，専用 DVD-R ドライブは高価で，かつメディア単位での書き込みしかできない
- 外部ドライブで書き込みを行った後，チェンジャに入れる操作を人手を介して行っている

等があるが運用に際してどれも致命的な問題とはなっていない．以上をまとめて LHD データシステムで稼動しているライブラリ装置の比較総合評価を Table 3.12 に示す．

最後に，Figure 3.26 に LHD 実験第 6 サイクル終了時までの全保存データ量の推移を示すとともに，データ収集系，データ保存系，データ取り出しクライアントを含めた LABCOM 計測データシステムの全景を Figure 3.14 に掲げる．

階層化された分散データ・ストレージについてまとめると，

[注11] UDF は国際標準化機構 (ISO) の制定した ISO-13346 規格および Optical Storage Technology Association (OSTA) の UDF インプリメンテーションガイドラインにより規定されている．第一世代 CD-ROM の異なる OS 間における互換性を保証するために開発・規定された ISO-9660 ファイルシステムと同じく，UDF は DVD-RAM，DVD-ROM，CD-RW，CD-ROM 等の光ディスクの開発，マスタリング，再生における幅広い互換性を実現するための標準的な国際規格となっている．

Table 3.12: Advantages and disadvantages between types of LHD mass storage equipments.

device	read	write	occupation	management	recovery	cost	future
MicroHSM+MO	○	×	○	○	△	×	×
AMASS+MT(PetaSite)	○	×	△	○	○	×	△
NetVault+MT(PetaSite)	○	○	○	○	◎	×	○
Jukeman+DVD	○	(○)[a]	△	○	○	◎	◎

[a] with external writing drive

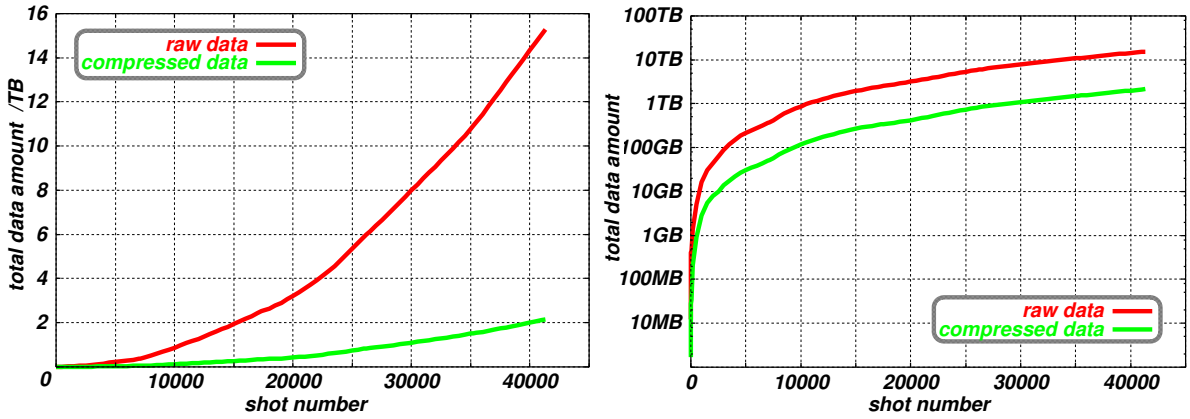


Figure 3.26: Total data amount in archives:

1. ディスクアレイと光メディアを使用したライブラリにより 3 階層のオンライン・ストレージを構築し，全実験データの半永久的な参照を可能にした.
2. 第 1 層では，信頼性を備えた RAID 上に ODBMS 永続オブジェクト領域を分散収集サーバ毎で保持・利用して，システム構築の負荷を大幅に低減した.
3. 第 2 層として，RAID 冗長構成の大容量ディスクアレイを用いて全実験データの高速入出力バッファを実現，データ参照速度を向上させた.
4. 第 3 層に従来の MO ジュークボックスに加えて，新たに DVD チェンジャの導入・検証を行い，保守・管理を含むトータルなストレージ・コストの低減を実現した.
5. DVD メディアを複製し別棟に保管する災害対策を実施した.

といえる．最終項目については改めて 3.8 節にて述べる.

3.7　データの仮想的取扱いとネットワーク・アプリケーション

システム概念設計を行った 3.3.5 節でも述べたとおり，ユーザからの対話的なデータ操作や取り出し・可視化などの処理には，ネットワークを意識すること無くデータ呼び出しが出来る単純で統一されたユーザ・インターフェースが提供される必要がある．このデータ参照 I/F が異機種/異 OS 上で同様に動作すれば，ユーザはデータ処理システムを完全なブラックボックスとしてのみ用い，自分の好みのデータ解析環境上に転送・再現されたデータの後処理，つまり解析や表示処理などに専念出来る．

LHD データ処理システムでは，データを再現・提供するシステム側処理とデータを解析・表示するユーザ側処理を明確に分離し，解析・表示処理タスクをユーザ側開発として解放することにした．これは大規模なデータ収集系構築にかかるシステム開発負荷を開発者とユーザとが分業・分担して開発案件の積み残し解消を目指す，エンド・ユーザ・コンピューティング (EUC) あるいはエンド・ユーザ開発 (EUD) と呼ばれるオープンシステム特有の手法である．

3.7.1　データ参照クライアントのオープン化

3.3.5 節でも述べたとおり核融合実験のデータ参照要求は，ショット番号と呼ばれる実験の通し番号と，各計測データにつけられるデータ名 (計測名) との組で一意に同定される．そのため

```
> retrieve, "Bolometer", 12345, "1:16", dataArray
```

というユーザからの参照要求で，ショット番号#12345 の Bolometer 計測の 1〜16ch に相当するデータ配列が dataArray に転送されるデータ取出しインターフェースが最も受け入れられやすい．更に，これで得られたデータ配列の第 1 チャネルが

```
> dataArray = fullScale * (dataArray - offset)
> plot, dataArray(1,*)
```

のように変換処理されたり x-t プロットされる対話的なデータ解析・可視化アプリケーションがあれば，データ処理システムとしてのユーザ・インターフェースは基本的に完備されることになる (Figure 3.27 参照).

こうした対話的データ取扱い環境を提供するツール・アプリケーションは，商用製品で存在し IDL (Interactive Data Language) アプリケーションと一般に呼ばれている．LHD デー

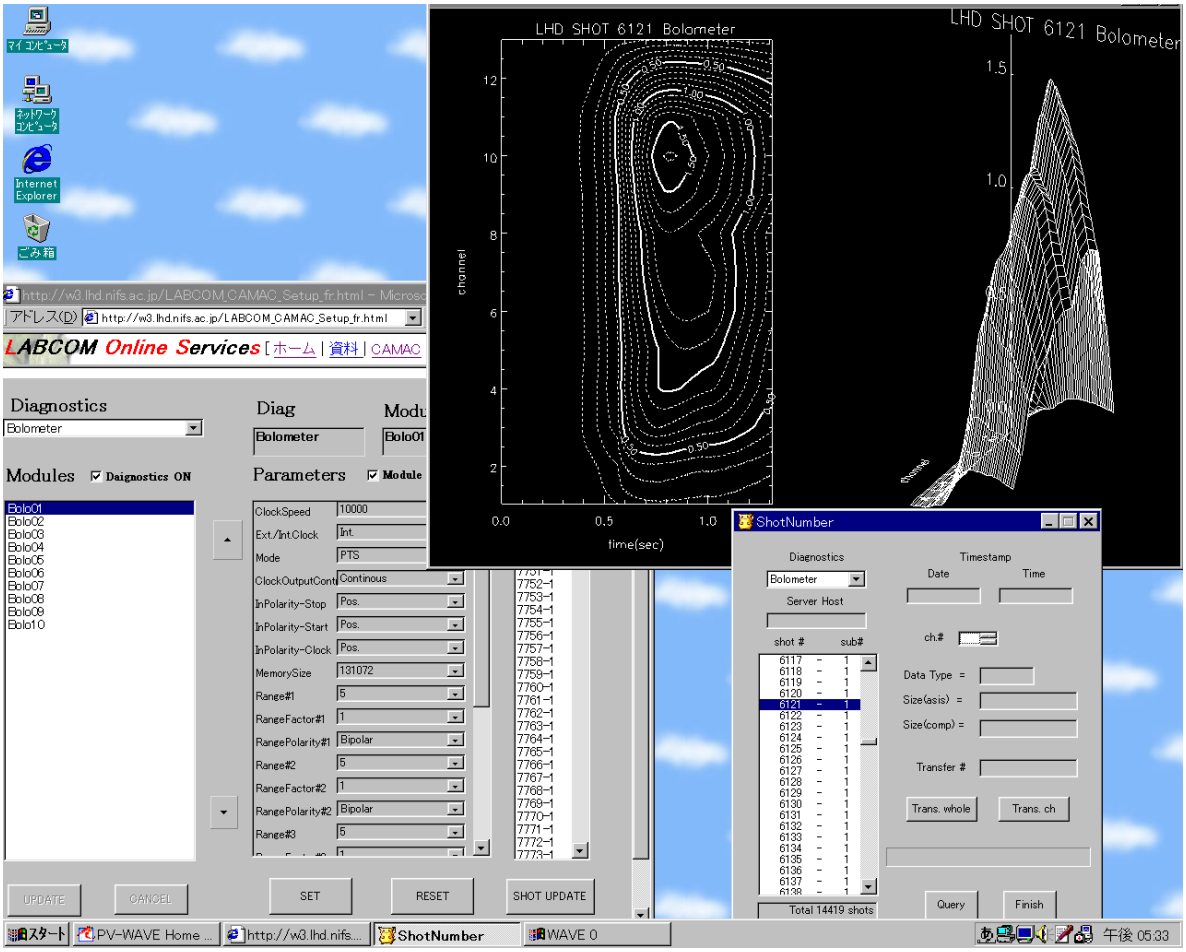


Figure 3.27: Typical GUI programs to access the CAMAC data acquisition and retrieval system in LHD. The PV-WAVE plotting output, CAMAC setup Java Applet, and the acquired shot number viewer can be seen on Windows graphical user environment.

タ処理では，データ・システムを利用するユーザのコンピュータ利用スキルを考慮して，一般的な Windows OS 上と普及しつつある PC UNIX Linux 上で動作する Visual Numerics 社の PV-WAVE/IMSL［64］および Research Systems Inc. 社の IDL［65, 66］をサイトライセンスを取得して標準ツールとしている．これらの利用環境中では，ユーザは自由にプロシージャと呼ばれるスクリプト言語を用いて一連の処理・表示ルーチンを作成・利用が可能なため，対話的データ処理環境であると同時に，データ解析・可視化の自動処理を容易に開発できる EUD 環境にもなっている．

また，自身で API を利用して直接データを取出し計算処理などをプログラムしたいというユーザの要望にも応えて，データ・インターフェースは Win32 DLL および Linux

(UNIX) Shared Library 形式で提供・公開し，C，C++，FORTRAN などプログラミング言語で書かれた個別プログラムからの直接呼出しにも対応した．

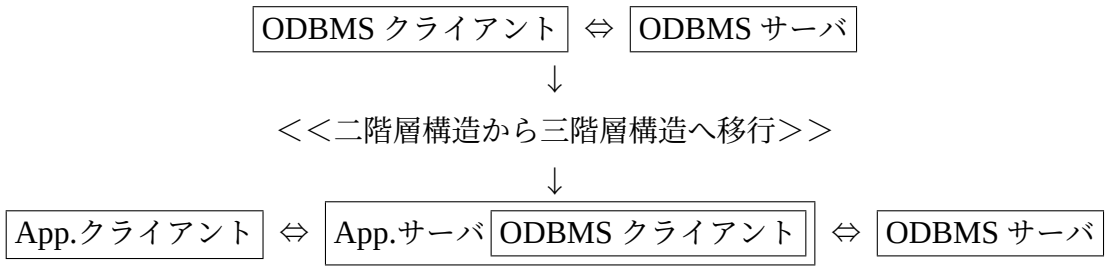
このように統一されたデータ参照インターフェース (Retrieve パッケージ) と EUD 環境ツールの組み合わせにより，エンドユーザでも容易に解析・可視化プログラムの開発を可能にした事は，システムの可搬性をあげる効果もあり，その意義は大きい．同時に，ユーザのデータ可視化プログラムに対する要求・要望条件の多さ・多様さに対応することを考えると，IDL 製品を EUD 開発ツールとして採用したことによる開発側の負荷低減の効果も無視できない．

3.7.2 三階層化されたデータ取出し機構

ユーザからのデータ参照とは，基本的に OODB 中に保存されているデータオブジェクトをクライアント計算機で動作するアプリケーションのメモリー中に複製することである．この場合のアプリケーションは，ODBMS の言語 API を用いて ODBMS サービスを呼び出す ODBMS クライアント・プログラムとして，ODBMS のサーバプロセスと直接通信するクライアント/サーバ (C/S) の動作形態をとることになる．

しかし 3.6.2 節でも説明したとおり，この ODBMS の C/S という二層構成でデータ参照を行うと，RDBMS とは異なりクライアント・セントリックで重い処理を行う ODBMS の特性から，データベース空間中に存在するオブジェクトを検索する際に目的外オブジェクトの一部情報までクライアント側に転送されるなど，C/S 間で必要以上に多くの情報の授受が発生する．特にクライアント・プログラムが遠隔からサーバ・プロセスを呼び出す場合は，この情報の授受はネットワーク通信経由，通例では TCP/IP ベースのインターネット通信となるため，通信帯域を消費するとともに処理そのものにも不要なオーバーヘッドを負わせることになる．これは結果としてデータ参照処理の速度低下として現れる．

こうした DBMS の処理メカニズムに起因する速度低下を回避するためには，情報授受量が大きくなる ODBMS の C/S 間通信を同一ホスト内のプロセス間通信 (Inter Process Communication: IPC) にとどめ，転送オーバーヘッドが大きくなる遠隔通信は行わない形態へ移行することである．そしてクライアント側からのデータ転送要求に対して，C/S 間のデータ授受を必要最低限にしてオーバーヘッドを回避する専用の C/S 関係を新たに加えることで，遠隔 C/S 間の通信速度を改善することができる．この手法は三階層構造 (型) のネットワーク・アプリケーション，あるいはアプリケーション・サーバ利用などと呼ばれ，中間に介在するアプリケーション・サーバ・プログラムが同時に ODBMS 言語 API を呼び出す ODBMS クライアント・プログラムにもなっている．模式的に書くと，



となる．実際の本システムの構成は Figure 3.28 に示すとおりである．

DBMS のクライアント/サーバ二階層構造からの脱却は，C/S 間の通信帯域幅が見込めない一般のインターネット WWW アプリケーションなどで多く見られ，検索エンジンとなる DBMS と WWW サーバ，および WWW クライアントの三階層形態などで広く普及している．Figure 3.25 にも示したとおり，本システム中でも遠隔のユーザ・インターフェース (UI) から要求を受ける下記の 2 つの要求処理サーバでもこの三階層構造をとっており，

1. CAMAC 運転設定 GUI
   Java Applet ⇔ App.サーバ CAMACd ⇔ ODBMS 設定サーバ

2. データ参照 UI
   参照クライアント Retrieve ⇔ 転送サーバ Transd ⇔ ODBMS 保存サーバ

C/S 間ネットワーク通信が処理ボトルネックを発生させることを回避しつつ，しかももとのサーバ側オブジェクトをクライアント側でもオブジェクトとして扱えるようにするのに成功している．Table 3.13 に Retrieve パッケージの典型的なデータ取出し速度を示す．実効的なデータ I/O 速度が 1.71 MB/s と，ODBMS の最大読出し速度 2.1 MB/s (Table 3.6 参照) にほとんど一致していることから，三階層構造をとったデータ取出し UI がシステム最大 I/O 性能を引き出すのにほぼ成功していることが確かめられる．

Table 3.13: Typical results of ’Retrieve’ performance. In case of the excellent compression ratio (PHARD), its elapsed time is spent almost for searching and the decompression calculation.

diagnostics	data size (B)	compressed (B)	ch#	time (s)	I/O (MB/s)	restore (MB/s)
Langmuir2	62914560	21456008	30	11.98	1.71	5.01
PHARD	85458944	249528	16	3.26	0.073	25.0
MMimg	67623152	29466439	72	20.47	1.37	3.15

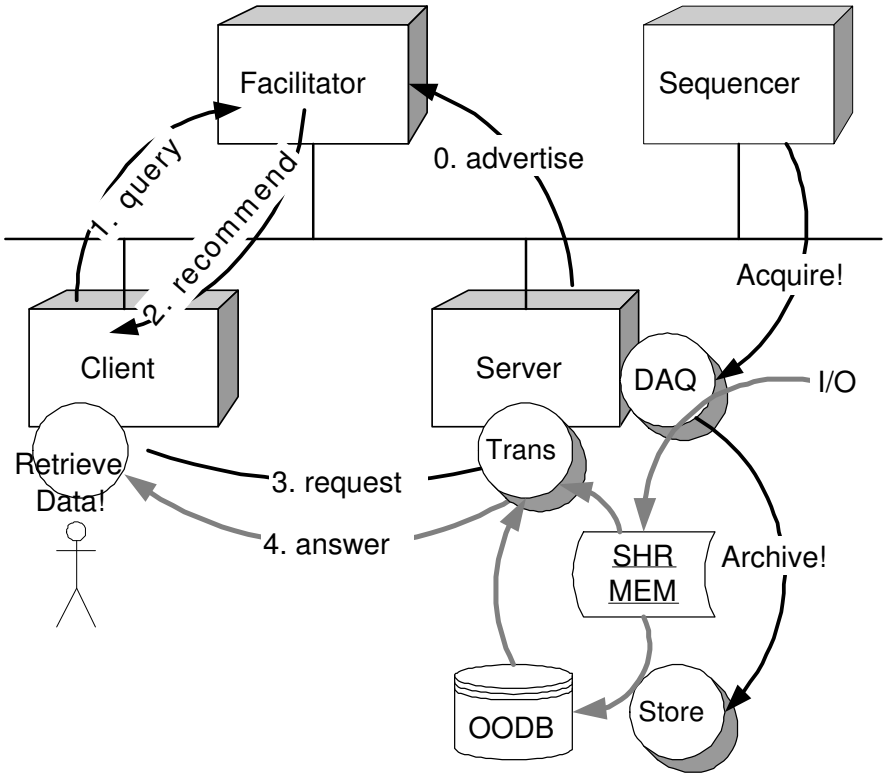


Figure 3.28: Cooperation between the recommending facilitator and the data transferring application server. “Trans” server can read not only the OODB volume but also the shared memory objects; the latter is for the rapid read-out just after every plasma discharge ends.

3.7.3 ユーザ作成データの再保存機構

前節までに述べたとおり，LABCOM データシステムのデータ取出しインターフェースは統一された Retrieve パッケージとしてユーザ一般に公開しており，これは LABCOM システムで制御・運転されているデジタイザから収集された生データのユーザ・インターフェースである．これに対して，ユーザが生データを解析して生じた一次処理データや LABCOM システム以外の個別集録装置で収集された生データを，同じく Retrieve で取出し可能な LABCOM データシステムに格納したいという要望がある．これに応じて公開しているのが dbStore とよぶ任意 (BLOB) のデータ登録・保存が可能なユーザ・インターフェース・パッケージである．

dbStore パッケージは次の 4 つのプロシージャで構成されており，

1. dbsOpen, 'MailAddress', 'DiagName', ShotNo, SubShotNo, DataType [, result]
   ショットデータの登録を開始する．
2. setParam, pStrArr, 'KeyName', Value, ValType [, result]
   パラメータ格納用の文字列配列 (パラメータ数 ×3) を作成する．省略も可能．
   事前に文字列配列の定義が必要．(e.g.) pStrArr = STRARR(2*3)
3. dbsWrite, pStrArr, pDataArray [, result]
   1 チャネル分のデータを格納する．チャネル数だけ 2.〜3.を繰返し実行する．
4. dbsClose [, result]
   データ転送と登録を完了する．

これらを順次呼出してデータの格納を行う．最後の dbsClose 呼出しによって，全チャネルの前処理データから登録保存用の圧縮済データファイルが内部的に作成され，匿名 FTP ベースでの転送とインデックス DB への登録 (SQL 発行) がなされる．

一旦登録されたデータの更新や削除は，全て Web ホームページにより実行可能になっている．削除申請フォームからの入力で申請された削除内容を消去するための専用暗号キーを自動作成し，それに対応する削除フラッグをインデックス DB 中にセットする．そして dbsOpen で登録したユーザアドレス宛にその暗号キーを用いて実際に削除を実行する Web ページ URL を発行し，以下のように送付する．

To:(宛先)　　ユーザが登録したメールアドレス
Subject:(件名)　【dbStore データの削除申請を受け付けました】
○○様

XXXX 年 XX 月 XX 日　XX:XX:XX
LABCOM グループ

【dbStore データの削除について】
申請のありました dbStore データの削除の準備が整いましたのでご連絡いたします。

記

データ登録名：XXXXXX
削除範囲：XXXXXX〜XXXXXX
データの削除を完了するか削除申請の取り消しを行うには、以下のアドレスに
アクセスしてください。
（アクセスする前にご使用のブラウザの Java Script を有効にしてください）
http://w3.lhd.nifs.ac.jp/w3/dbStore/del.asp?key=XXXXXXXXXXXXX

以上

お問い合わせは LABCOM@LHD.nifs.ac.jp または内線 XXXX までお願いします。
NIFS LABCOM

---

dbStore の登録処理は Retrieve と同様に，BLOB 転送を行うネットワーク通信帯域を最大限有効活用する必要があり，専らクライアント・セントリックな処理方法を採って，クライアント側で圧縮済データの作成や SQL 発行を行っている．削除処理についても同様に，上記の通り WWW 機構をうまく用いることでサーバ側のみの処理にして，不要なネットワーク通信を無くすことに成功している．

3.8　ディザスタ・リカバリ（災害/事故復旧）対策

LHD のある NIFS は平成 15 年現在，東海地震および東南海地震の地震防災対策強化指定地域に隣接しており，予想震度も震度 5 弱〜5 強と相当の災害が予想される強度になっている．こうした災害あるいはその他の事故発生の場合には，その後，相当の期間を設けて復旧・再確認をおこなうことは十分に想定されるべきことであるが，それまでの実験研究の成果が損なわれることだけは回避しなければならない．

実験研究の成果は基本的に計測データに帰着するため，実験成果の災害事故に対する保護とは，計測データの安全な保管を行うことに正に該当する．LHD 計画は特に核融合実験プロジェクトの中でも大規模なものであり，また共同研究利用施設という位置づけからも，計測データの安全保管は極めて重要になってくる．LABCOM システムでは以下のように災害事故対策を考察している．

1. 災害・事故事象の同定 … 地震・火災
2. 情報システムの防災対策 … 耐震・防振・免震および耐火
3. 災害発生後に情報システムの速やかな復旧を可能にする対策
4. レプリケーション (複製) によるデータ保護 → データの複製を遠隔地に保持

⇒　DVD メディアの複製を別の建物に保管

非常の事態に対しては，やはり実験成果の安全な保護が最優先事項となるため，実験当日の計測データよりも過去に収集された膨大な蓄積量の長期保存ストレージに重点が置かれる．常時オンライン参照サービスを行っている階層ストレージは，ネットワーク接続上の問題からも設置場所を限定されているため，大規模ストレージ装置自体をより安全な場所に置くことは難しい．そのため複製されたデータをオンライン・ストレージとは別の離れた安全な場所に保管するのが現実的である．

LABCOM ストレージでは大容量データの複製媒体にかかる費用も考慮し，広く一般に普及しており入手容易で可搬性のある DVD-ROM メディアを用いて，過去データを別棟にオフライン保管している．3.6.4 節でも述べたとおり，DVD チェンジャ中に格納される DVD-ROM メディアは各々 UDF フォーマットになっており，それぞれの単体での読出しが可能なので，DVD-ROM (チェンジャ) で UDF フォーマットを用いている場合の災害対策性は，仮想ボリュームを用いて各々メディア上の記憶内容がブラックボックスになる HSM などの大規模ストレージに較べて非常に好適だといえる．

また 3 階層の分散ストレージ構成をとっている点を活かして，各階層ストレージ中でも

最大 3 つの複製データを保持していることで，いずれかのデータが損失されても層間で相互に再複製して復旧にかかる時間を最低限に抑えることを可能にしている．更に，データの参照場所を推薦するファシリテータ機能により，いずれか一部のストレージ装置が破損・停止しても，ユーザへのデータ参照サービスを続行可能な体制になっている．これは平時の装置障害発生時にもサービス停止を回避できる機構として非常に有効に機能している．

次に，災害当日を含めたより多くの計測データを安全に保管するためのデータ移送・複製頻度については，計測データ I/O 量の多さから実験シーケンスに同期したデータ転送は困難であることから，Figure 3.29 の通り，日々の実験終了後の夜間に階層ストレージへの移送と複製の処理を実施している．

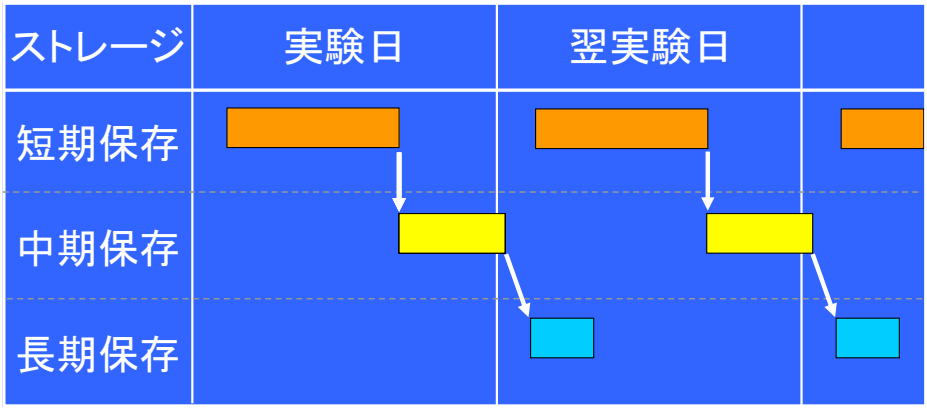


Figure 3.29: Data migration schedule in LABCOM 3-layer storage system.

上述の通り，最重要な研究資産である計測データの安全保持については，階層化分散ストレージ中で同時保持・多重化されており，更にオフラインで DVD-ROM バックアップメディアにも保管されて万一に備えられている．しかしデータ容量が巨大であることから，ストレージ装置へのデータ再移送が必要になるとそれに要する読出し・再書込み時間はかなり長くなるため，復旧が遅れることになる．このため，実験データの安全保持とは別に復旧の長時間化を回避する手段として，特に常時記憶媒体が動作している第 2 階層の RAID 装置などに対して免震装置などを導入することも併せて検討されるべきである．[注12]

[注12] 予算・費用の都合で H14 年現在，まだ LHD では実現していない．

3.9 大規模分散データ処理システム開発のまとめと評価

今回新たに開発した核融合プラズマ計測のためのデータ処理システムで，最も重要な基本設計は，オブジェクト指向に基づいた大規模な並行分散処理系の開発であった．本システムの稼動は，核融合実験の分野で今までなされてこなかった OO 開発へのパラダイム・シフトを実証した世界初のケースであり，ここで得られた新たな知見が当該分野に与える意義は大きい．

オブジェクト指向の分散コンピューティングにおいては，個々の機能の自律的な処理主体であるオブジェクトが他のオブジェクト群と連携して外部からの要求を処理することで，より大きな機能主体を形成することになる．これは即ち，大規模な分散システムであっても，要素間のオブジェクト指向的連携によって，外部の利用者からはあたかも一個の巨視的な仮想マシンが動作しているように見えるのであり，これこそが本研究が実現を目指したシステム形態である．しかし，これには克服すべき以下のような問題点があった．

核融合実験は予め規定されたプリ・プログラムドで時系列的な手順の進行であり，大量の計測データが同時タイミングで一斉に出現するため，その I/O 処理には高い最大瞬間性能が求められる．このため長い間，核融合実験の計測データ処理システムではそうした利用形態に合致する I/O オーバーヘッドの少ない構造化システムの利用が続けられてきた．即ち，オブジェクト指向方法論の最大の利点である開発負荷の増加抑制の目的で，構造化システムより処理オーバーヘッドが大きいオブジェクト指向システムへの移行が可能であるか否かという命題に，LHD 以前の核融合実験は直面してこなかった．

これに対して，従来のデータ処理量とは桁違いの I/O 性能を要求された LHD データ処理システムでは，中央集中型構成でのハードウェア資源の集約化による性能の高度化や，人的開発能力の集中投入などによる開発対応が不可能であった．要求性能を将来のスケーラブルな増強を含めて実現するためには，必然的に大規模な分散処理のシステム形態への移行と，分散システムの開発負荷増を補償する，より高効率な開発パラダイムへのシフトが不可避となった．

これら二つの大きな課題に対して，どのような工夫を用いて解決することができたかが本研究で得られた新たな知見として重要であり，続く二小節において改めて列挙している．

本研究で開発したデータ処理システムは，Figure 3.10 に示される基本構成エレメントが非常にコンパクトであり，PC，SCSI，Ethernet といった広く一般に普及している小型計算機用ハードウェア/周辺機器が活用できるため，今までよりも遥かに柔軟かつ容易にシス

テム構築や増設・増強等が行えるのが特徴である．最新の大型実験装置である LHD の実働システムでは，多種多量の計測データ収集を扱うために基本構成に加えて検索用データベースや大容量ストレージ，それにギガビット級の高速ネットワーク等を利用している．しかし基本的には，このコンパクトなデータ収集エレメントを一計測毎に独立に設置し，並列分散度を約 30 セットに増やして全体を稼動させている事に変わりはない．

Figure 3.12 に示したとおり，LHD 計測データ処理システムの実稼動後の順調な処理量増加によって，本研究で開発されたデータ処理システムのもつスケーラビリティ，拡張性，自由度の大きさが十分実証されたといえる．

なお遠隔実験への積極的な参加を支援するオープン・データ処理システムの観点から，内外ユーザが関連情報を自由に入手できる情報ポータルサイトを立ち上げており，LHD データ処理システムで開発した公開可能なソフトウェアや開発関連文書などは Web サーバ **http://w3.LHD.nifs.ac.jp/** にて全て公開している．

3.9.1 開発効率化への取組みについて

本章で述べたとおり，大規模な分散処理システムには OOD が適しており，核融合プラズマ計測のデータ収集保存システムにこれを応用する研究を行ってきた．大規模な分散処理システムは，プログラム開発の集約化を妨げその負荷を上げてしまうため，OOM やその他の手法を用いて具体的にどのように開発効率化に取組み，LHD のような大規模システム開発に成功したかが，他分野に対しても重要な知見となる．そのため，最後にもう一度，本研究で行った具体的な取組み方法をまとめて挙げてみると以下の通りである．

UML による仕様分析・設計

統一モデル化言語 (UML) を仕様分析・設計に全面的に用いることで，仕様の明確化作業が容易になると共に，開発者間の意思統一の助けとなった．

ODBMS による実験データの格納

主記憶上のアプリケーション・オブジェクトをそのまま永続化できるため，データ保存・参照のための I/O 手順プログラミングの開発負荷を大幅に削減できた．

EUD に基づく解析・表示クライアント開発の分離

対話的データ言語 (IDL) ツールとデータ取出しインターフェースとを併せてエンドユーザ開発 (EUD) 環境とし，ユーザ側にデータ解析・表示開発を解放した．負担の大きかったユーザインターフェース開発を削減しシステム開発を大きく効率化できた．専らデータを利用するのみだったユーザが，自身で容易にデータ解析・表示機

構を作成できるようもなった.

3.9.2 I/O 性能の改善について

元来，データ収集保存システムは巨大なデータ I/O 系であり，その最重要なシステム性能もやはり I/O 能力である．本研究では，最近の大容量化の要請に対応するため，大規模な並行分散形態をとり並列数によってスケーラブルに I/O 能力を加減できるデータ収集システムを開発したわけだが，ここで収集エレメントの素性能やシステム全体の I/O 性能を最大限に引き出すため，様々なレベルで工夫を施した諸点を今一度，列記してみる.

データ収集直後の圧縮組込みと以降の取扱いデータサイズ低減

データ収集の直後に圧縮処理を組込んで，以降の保存・取出し・転送などのデータ取扱いを全て圧縮データ・オブジェクトで行うことで，サイズ低減による見かけ上の I/O 速度を改善した.

データ収集・圧縮・保存プロセスのマルチスレッド化

デジタイザとの接続 I/O ポートと保存用 HDD への I/O ポート，およびデータ圧縮で消費する CPU 計算能力と利用する計算機資源が異なるデータ処理プロセスをマルチスレッド化して並行に進行させることで，I/O 利用効率を改善した.

直前収集データの共有メモリー上での保持

実験は連続して行われるため直前の収集データを主記憶領域の共有メモリー上に保持しておくことで，参照頻度の高い最新実験データの取出し応答性を向上させた.

I/O 能力の高い ODBMS の選択

処理の遅い O_2 から ObjectStore にデータ保存用 ODBMS を置き換えることで，データ保存・取出し速度を改善した.

収集動作同期メッセージ授受の低水準 I/O 化

分散データ収集サーバ群を実験シーケンスに同期動作させるための同期メッセージ授受には，当初利用を計画した仮想オブジェクト共有空間 HARNESS がノード数増加 (〜10) と共に負荷が急上昇するため，より軽負荷の TCP/IP ソケットをベースにネットワーク API を専用のクラス定義に隠蔽実装して利用する手法に変えた.

3 階層分散ストレージによる大容量性と高速性の両立

オンライン・ストレージを新しいデータ順に短期・中期・長期の 3 階層に分化し，各々相応しい速度性能のランダム・アクセス記憶媒体を用いた．これにより I/O 速度で 2MB/s 弱，クライアント計算機上への計測データ復元速度で数〜25MB/s と，

非常に高速でストレスのない常時データ取出し環境を実現した．

RDBMS によるデータ所在情報検索ファシリテータ・サーバ

実験・計測毎のデータの所在情報は表形式でレコード数が多いため，RDBMS で表情報の高速検索結果を提供する分散コンピューティング・モデルにより，高速データ検索と分散ストレージ構成を可能にした．

データベース参照・データ取出しのネットワーク三階層モデル化

収集データを保存する ODBMS からのデータ取出しを C/S の二階層から，通信量の少ない三階層モデルに変えて専用ネットワーク API を実装することで，通信負荷を下げると共に取出し速度を改善した．

上記には，開発効率化を目的とした OOM あるいは OOD システムの性能低下を補償するためのものも含まれる．

しかしながら，ビッグ・サイエンスの一分野である核融合研究の大規模データ処理システムで，実用に耐えるべく施した具体的な各種の解決手法は，今後，他分野のシステムの性能向上などにも応用の可能性があり，計算機科学の面から見ても普遍性をもつ新たな知見であるといえる．

3.9.3 性能上限に関する考察

本章で研究開発した分散型データ収集保存システムは，収集エレメントの独立性が高いため並列数を増やすことで容易にシステム全体の収集 I/O 率をスケーラブルに能力増強できるのが最大の特徴の一つである．このため近い将来を含めて核融合実験のデータ収集には基本的にすべからく対応が可能なのであるが，基本エレメントの素性能とシステム全体性能の二つの面から性能上限値は存在する．

エレメントの I/O 素性能の上限を決めているのは基本的にデジタイザ接続ポートのデータ伝送速度である．CAMAC の場合では CAMAC データウェイの実効転送速度〜1 MB/s であり，これが CAMAC 収集系の性能上限を決めている．CAMAC よりも高速な接続ポートを持ったデジタイザ規格の場合は，収集サーバ PC の拡張バス速度までは特に上限がない．PC の拡張バス性能は，一般的に PCI バス規格の理論性能 133 MB/s，実効的には約 100 MB/s がその典型的な値になる．

データ保存を担う ODBMS および HDD への書込み速度は，本章で述べたとおり PCI バスに較べると低速である．しかし，次章で述べる定常運転時を除き，通常の短パルス実験の際のバッチ処理運転時には，本開発システムは，収集データを一旦，主記憶領域の共有

メモリー上に書込んで保存処理を完了し，ODBMS，HDD へは別プロセスによる遅延書込みで行っている．このため ODBMS，HDD への書込み速度はデータ収集のスループットには影響を与えず，性能上限を上げることに成功している．但し，収集データ量に対して主記憶領域の大きさが十分あり，データ圧縮計算を十分こなせる CPU 計算能力があるという条件になるが，現在の PC 水準でこれは十分容易に実現可能となっている．

システム全体の性能上限はもう少し複雑である．本システムの収集エレメントは基本的に短パルス実験時のバッチ収集では完全に独立して並行動作するため，LHD の実績では CAMAC の約 1 MB/s を 30 台並行収集であったが，100 MB/s のデジタイザ収集を 100 台あるいは 1000 台，それ以上でも同時に動作させることが可能である．

このためシステム性能上限は収集面ではなくデータ参照の応答速度からとなる．LHD では約 4 万実験を経過した時点で，ファシリテータに登録されているデータ・エントリ数が約 3〜4 百万レコードになっている．この登録レコード件数が 1 桁上がった程度では検索速度はそれほど低下しないが，これが 2 桁以上多くなると RDBMS に格納できる件数の問題や検索にかかる時間が無視できなくなり実質的な上限になる可能性はある．

しかし本研究を行っている時点で，核融合プラズマ計測でのこうした事態の発生は極めて想定しにくく，他分野へ本研究のシステム技術を応用する際などには状況に応じて再検討が必要となる可能性はある．

第 4 章

超広帯域実時間データ収集系の開発

LHD 装置は超伝導磁場コイルを採用したことにより定常的なプラズマ保持実験が可能になっている．この特性を活かした 10000 秒程度の準定常プラズマ保持を行う実験も，重要な実験テーマの一つとして計画されている．このため計測データ処理システムも他装置と同様，定常運転化が求められており，今後の ITER 等の次期装置計画への知見を得るという目的でも不可欠な研究課題の一つとなっている．

前章までに述べてきたとおり，計測データ収集システムの定常運転化への根本対応は，CAMAC に替わる新たなデジタイザ系統を開発利用することである．このため本研究では，以下に挙げる核融合プラズマ計測用のデジタイザ要件を満たしたデジタイザ・フロントエンド (DFE) の研究開発と検証を行った．

1. 0.5〜1 MS/s・ch, 12〜16 bit （プラズマ揺動計測に利用可能）
2. 一筐体で 100 チャネル以上収容可能
3. 無停止連続データ転送機能．転送レート >100 MB/s
4. DFE–ホスト間リンクが光絶縁され〜500 m 以上延長可能
5. PC と親和性が良い（低コスト化のため）

現在の PC 技術環境においては PCI バスが実質的標準となっており，PCI bus rev.2.2 (PCI-X) 以降で >100 MB/s の転送帯域も充分実現可能になっている．この PCI bus のフィールド版が CompactPCI 規格であり，論理的特性は PCI と全く等価であるため PC との親和性が非常に良い．それに伴って低コスト化の点でも非常に進んでおり，2.1.2 節でも述べたとおり，同程度のバス性能を備えた VME64x などの他規格の追随を許さない状態で，市販モジュール製品も順調に増えてきている．このため本研究では，この CompactPCI を採用し従来にない高速広帯域で無停止のデータ収集系を新たに設計することにした．

4.1　新システムへの要件

本システムは，高速サンプリング速度で多チャネルのプラズマ物理計測データの生成，収集，転送，格納，演算処理，表示 (可視化) を連続無停止の実時間で実施するデータ処理 (計測制御) 装置である．本システムは，大別して以下に挙げる各機能を持つものとする．

- 多チャネルデジタイザによるデータ生成.
- コンピュータからのデジタイザおよび同集合体の (遠隔) 制御と光絶縁.
- 生成データのコンピュータへの実時間転送.
- データの実時間格納.
- データの実時間解析演算.
- データの実時間可視化．遠隔での表示.
- データの高度解析処理による実時間機器制御信号の出力.

以下に各機能の要求仕様の詳細を述べるが，特定分野での利用に限定されないよう，基本的にオープンスタンダードに可能な限り準拠した仕様で実現するものとする．

4.1.1　ストリーム出力デジタイザによる A/D 変換仕様

核融合プラズマ実験では，通常，アナログ信号をデジタル化するのに最も一般的に使用されるのが，トランジェントレコーダ型 ADC といわれる時系列波形を配列データ化するデジタイザである．ここでもこの方式の高速 ADC モジュールを開発する．デジタイザに要求する仕様は以下の通り．

概略仕様

ADC モジュールは，最大 1 MHz のサンプリングレートで 8 チャネル同時サンプリング可能な，A/D 変換モジュールとする．各チャネルはそれぞれに独立したプリアンプ，16 ビット ADC，およびワンショット動作のための 4 M ワードの波形バッファメモリを持ち，プリアンプ増幅率は 0.1〜100 倍で，入力レンジは± 2.5V，± 5V，± 10V，+2.5V，+5V，+10V に設定可能とする．サンプリングクロックは 8 ch 共通で，ローカルクロック (の分周) か外部クロック入力を選択できる．

動作モードには大別して，ワンショット (バッチ処理) モードと，リアルタイム・ストリーム処理モードとを持ち，前者では CAMAC ADC 互換の動作を行う．後者では一筐体

内に収納される 100 ch のアナログ信号を同時に無停止連続でデータ収集・転送できる能力を持たせる．Table 4.1 に仕様の詳細を示す．

ADC ボードの設計

上述の概略仕様をもとに実時間動作可能なトランジェント・レコーダー型 CompactPCI ADC モジュールの概略設計を行った．ADC モジュール内部および連結する PCI バスのブロック図を Figure 4.1 に示す．ADC の経路設計を決定する上で特に重要な仕様としては,

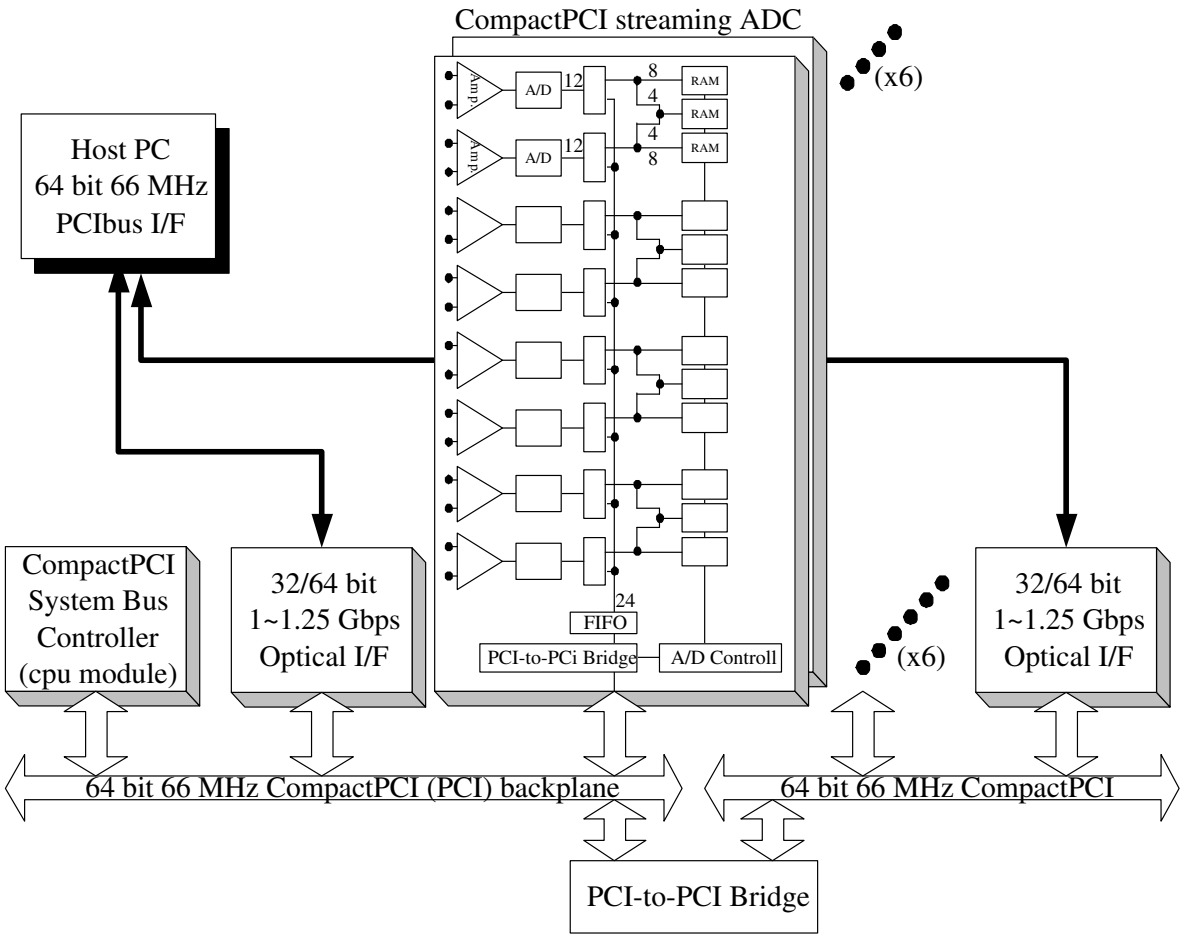


Figure 4.1: Schematic view of CompactPCI streaming ADC: To realize the effective 100 MB/s data streaming, the basic CompactPCI/PCI specification whose bandwidth is max. 133 MB/s on 32 bit 33 MHz bus transfer is not enough. The new revision PICMG rev.3.0 which defines 64 bit 66 MHz broader one will be expected.

Table 4.1: Required conditions toward the new streaming ADC.

論理的仕様		
	ADC 分解能	16 ビット.
	サンプリングレート	最大 1 MHz.
	クロック	1 kHz, 2 kHz, 5 kHz, 10 kHz, ..., 500 kHz, 1 MHz, および DC〜1 MHz の任意外部クロックが選択可能. デューティ 50 %.
	デジタル入力	外部クロック，スタート/ストップトリガーが デイジーチェーンで同期可能．TTL レベル動作.
	アナログ入力レンジ	± 2.5V，± 5V，± 10V，+2.5V，+5V，+10V ch 毎に設定可能.
	入力インピーダンス	100 kΩ 以上.
	デジタル化精度	0.1 %以下 (フルスケール比).
	プリアンプ	0.1〜100 でゲイン値が ch 毎に設定可能. DC〜1 MHz のアナログ信号帯域で歪率が 0.1 %以下. ch 間の入力絶縁のため平衡入力を推奨.
	割込み処理	デジタイザから収集制御 (ホスト) コンピュータに 割込み出力が出来ること.
物理的仕様		
	コネクタ形状	プッシュ・プル式かプッシュロック (バヨネット) 式. ねじ込み式は不可. アース絶縁のため 2 極 LEMO コネクタを推奨.
	コネクタ数	1 筐体に約 100ch のアナログ入力信号が収容できる.
	モジューラー構造	デジタイザはモジュール単位で着脱増減可能なこと.
機械的仕様		
	筐体形状	CompactPCI 6U あるいは CAMAC クレート程度. 19 inch ラックに収容可能であること.
電気的仕様		
	アナログ絶縁性	チャネル間でアナログ部は絶縁されていること.
	耐環境性	高い耐静電 (ESD)・電磁ノイズ (EMS) 特性をもち 強電磁ノイズの実験室環境で安定動作すること.
	CMRR	コモンモード除去比 70 dB 以上.
	ch 間クロストーク	-60 dB 以下.

プラズマ放電実験の 2 つの運転モード:短パルス運転と長時間定常運転，の両方の運転モードで同様に動作する事である．しかし，定常運転モードでの計測サンプリング速度は DFE からホスト計算機へのデータ転送レートで制限されてしまうため，実際には放電終了後の一括処理でよい短パルス運転時の方が，サンプリング速度の制約がなくなり，より高速サンプリングが可能になる．

本研究を行っている時点では，まだ 64 bit 66 MHz PCI を提供するホスト計算機や CompactPCI ボード製品も出回り始めたばかりでありコスト・パフォーマンスが良くないため，将来的には必要なくなる可能性が高いが，ここの ADC ボード設計では，短パルス一括処理と定常モードの二つの運転モードを実現できるよう，実時間転送のための FIFO バッファのほかに大容量のローカルバッファを備えることにした．

4.1.2 デジタイザ・フロントエンドへの要求機能

以下，デジタイザモジュールの集合体とそれを格納し動作させるシャーシ，バックプレーン，コントローラなど全体一式をデジタイザ・フロントエンド (DFE) と考え，これへの機能的要求を検討する．

ホストから DFE の制御は，CAMAC や SCSI 機器で行われているような予め規定されたコマンドベースの命令動作であることが，既存の CAMAC システムとの親和性も良い．しかし PCI/CompactPCI には，CAMAC, SCSI とように規格上でそうした制御コマンドが規定されているわけでなく，入出力ポートアドレスと割込みレベルで制御される拡張バスである．このためコマンド駆動型の API が提供される可能性はほとんど無く，CompactPCI のドライバを API 呼出しにより排他的に制御する専従サービスプログラムを立てることによって，はじめて複数アプリケーションからの制御依頼の管理が可能になる．当然，DFE 側から割込み処理要求を受けアプリケーションに返す機能も必要である．その他の要求仕様は以下の通り．

- DFE はホストの周辺機器として DFE 1 台毎にドライバで正しく認識される事.
- 各モジュールはスロット順に静的にアプリケーションから認識できる事.
- DFE-ホスト間の接続メディア・I/F にはファイバーチャネル (FibreChannel) 規格等 1〜1.25 Gbps の通信帯域をもつ規格を使用し，DFE–ホスト間は一対一あるいは多対一で接続可能である事.
  但し，転送レート上問題があると判った場合は，一対多接続も検討する．
- DFE-ホスト間の電気的絶縁のため，光ファイバーによるデジタル接続である事.

- DFE-ホスト間の光通信は，500 kHz 100 ch (1 MHz 50 ch) の全 ADC 出力である約 100 MB/s 以上を，有限の遅れ (30ms) 内でホストの主メモリー上まで無停止連続転送できる事.

また，DFE からホスト計算機への実時間転送に関しては，生データ転送前後での実時間圧縮・展開技法や，イベント駆動型データ収集法などによる生データの実時間削減手法の導入が，将来の改良時の課題とあげられる.

4.1.3 実時間データ格納と転送・解析・可視化

ホスト計算機の主メモリ上に刻々と収集されてくる生データに対して，その後必要となるのは格納，クライアントへの転送，解析処理，可視化表示などである．収集データ帯域は 100 MB/s を取り扱うことから，この 100 MB/s の実時間生データストリームをそのまま格納できる能力があることは必須条件となる．また，複数台の PC で解析演算と格納の並行分散処理が進められる PC クラスタ構成をとることで，計算機への負荷状況によってその分散度合 (台数，構成) を増減・調整できるようにして，分散形態の利点を効率よく活用することも重要である．その他，サーバ側仕様としては以下の通りである.

- 解析演算の結果得られたリアルタイム処理後データも実時間格納できること.
- 実時間データ転送では，不必要に詳細な生データをそのままクライアント側に転送することが無いようデータの間引き転送ができること.
- 動作 OS は Linux，Solaris あるいは MS Windows とする.

また可視化表示クライアント側としては,

- Linux/Solaris 上で X window System (X11) ベースの表示プログラムを提供し，設定やユーザの簡単なコンフィグ・スクリプト記述により，多様なデータ表示ができるよう可能な限り汎用性を持たせる.
- Windows 系統の場合，遠隔表示はターミナルサーバ機能を利用する.
- 処理後音声波形化されたデータの音声出力もおこなえるようにする.
- リアルタイム表示画面は対話的にユーザが切替えられるようにする.

などの要件が挙げられる.

実時間計測系の解析演算や可視化表示機構の開発では，汎用のデータ取出しインターフェースの提供と EUD 開発の手法だと，データ I/O への処理時間要件が厳しく技術的に困難と判断されるため，計測器ごとに異なったプログラムを作成する必要があると考えら

れる．そのため，実時間解析の演算ルーチン・出力情報，クライアント可視化プログラムとサーバ側との演算処理分担，実時間データ可視化・表示機構のグラフ画像・データ更新など表示情報等について改めて開発ターゲット毎に計測担当者と定める作業がそれぞれに必要である．LHD 実時間計測系への応用については，開発ターゲットして先ず磁場揺動計測を取り上げて付録 B.1 節にその調査結果を示した．

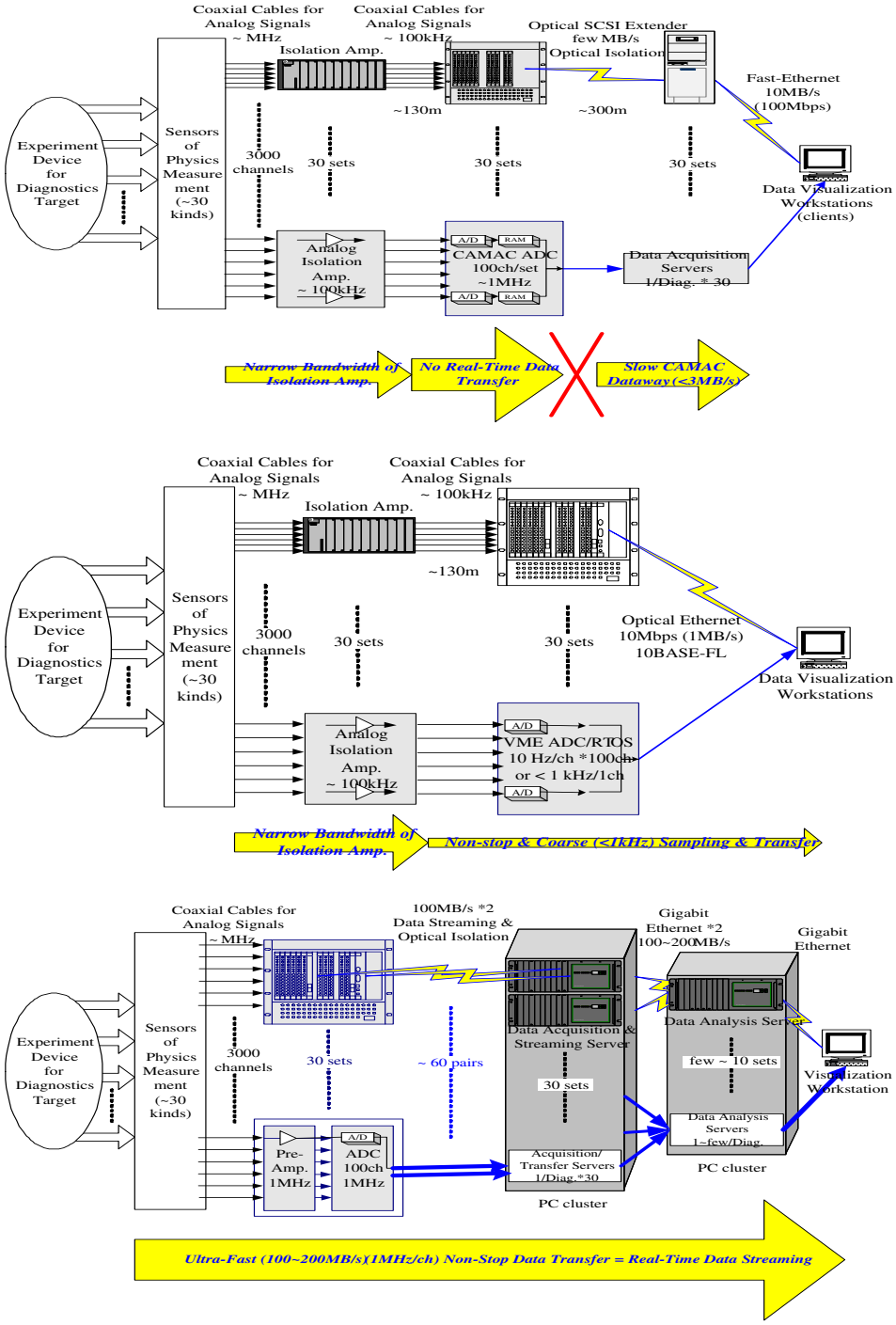


Figure 4.2: Structural comparison for non-stop streaming acquisition: First one is the conventsional CAMAC batch-processing acquisition, and the second is the real time monitoring system running on VMEbus. 100 MB/s continuous data acquisition and simultaneous streaming are the required conditions for the next generation data acquisition, however, both of the first ones cannot have enough capabilities to achieve them.

4.2　プロトタイプ・システムの構築と性能評価

前節で述べた CompactPCI 規格の高速 ADC の設計開発と並行して新開発システムの全体性能を試験・評価するため，平成 14 年に National Instruments 社の協力で以下の仕様の CompactPCI 評価試験 (プロトタイプ) 装置を導入した.

- 4 ch 10 MS/s ＋ 64 MB buf. トランジェントレコーダ型 ADC PXI-6115 … 7 台
- 3-U ハイト，8 スロット PXI シャーシ (32-bit, PCI rev.2.1) PXI-1000B … 1 台
- MXI-3 1.25 Gbps C-PCI/PCI 光リンクインターフェース PXI-PCI8335 … 1 セット

同プロトタイプ・システムでは，ADC モジュールへの要求仕様 8 ch 1 MS/s に対して，4 ch 10 MS/s と少しずれており最大 28 チャネル動作しかできない．しかしリアルタイムデータ生成・伝送試験には，各チャネルのサンプリング速度を増すことで DFE 全体としてのデータ生成率をあげられるため，～100 MB/s の連続データ生成を十分模擬できる．このため以下では，本プロトタイプによるシステム構築を行った結果を示している.

4.2.1　データの実時間収集・・・DFE ＝主メモリ間

C++の専用アプリケーションプログラムで実施したデジタイザ–主メモリ間の転送処理試験を以下に示す．ここでは PXI-6115 モジュールの諸条件を変化させ，C++アプリケーションが確保したバッファメモリまでデータ転送を持続できるかどうかの試験をした．変化させた条件は，サンプリングレート，使用モジュール数、1 モジュールあたりの使用チャネル数，である．設定上，各モジュールの使用チャンネル数を個別に設定できるが，今回は全て同一にした.

【試験環境】

```
CPU: Intel Xeon 2GHz Dual,  MEMORY: 2GB
OS: Windows2000 Server SP3
    SystemLargeCache=0  SecondLevelDataCache=512
Compiler: MS-VC++ 6.0
Driver: NI-DAQ 6.9.3
```

アプリケーション上のダブルバッファ量：約 16MS/モジュール

アプリケーションプログラムは、各モジュール単位にユーザメモリ（約 8 MS）に転送することを 2 分間連続して正常にできることで合否を判定した。NI-DAQ ドライバーからの

Table 4.2: Result of continuous data transfer between DFE and computer's main memory:

rate (MS/s)	modules	ch/module	channels	bandwidth (MB/s)	success
10	7	1	7	140	–
	6	1	6	120	–
	5	1	5	100	–
	4	1	4	80	OK
	4	2	8	160	–
	3	2	6	120	–
	2	2	4	80	OK
	1	3	3	60	OK
	1	4	4	80	OK
5	7	1	7	70	OK
	7	2	14	140	–
	6	2	12	120	–
	5	2	10	100	–
	4	2	8	80	OK
	4	3	12	120	–
	3	3	9	90	–
	2	3	6	60	OK
	2	4	8	80	OK
2	7	1	7	28	OK
	7	2	14	56	OK
	7	3	21	84	–
	6	3	18	72	OK
	6	4	24	96	–
	5	4	20	80	OK
1.67	7	3	21	70	OK
	7	4	28	93	–
	6	4	24	80	OK
1.43	7	4	28	80	OK
1	7	4	28	56	OK

アクセスはモジュール単位であり，マルチスレッドはサポートしていないので，ループ内で各モジュールをチェックしながらデータを取得するシングルスレッド・タスクとした．

Table 4.2 に試験結果をしめすが，DFE–ホスト主メモリ間の伝送帯域は，DFE 内のデジタイザの運転パラメータにかかわらず，一律に約 80 MB/s となっていることが確認される．また同時に CPU 使用率を観測したところ，他のアプリケーションは動作していない状態で，アプリケーション動作中はほぼ 50%前後であり，入手できるサーバ用としてかなり高速な CPU をもってしても試験した連続データ転送の処理負荷がかなり大きくなるこ

とが判る．

4.2.2 データの実時間格納・・・主メモリ＝ HDD 間

核融合実験に限らず物理計測一般の実験成果とは計測データそのものである．このため計測器センサー類から収集された計測データは基本的に全て保管される必要があり，データ処理系はその能力を備えていなければならない．この条件は実験が長時間・定常化しても変わることはないため，超広帯域化した実時間データ収集系で，実際には長期保存しておくデータの取捨選択を行うことになるとしても，一旦収集されたデータを全て保存するできるシステム性能を実現していることは重要である．

Table 3.9 にも示したとおり，通常半永久記憶装置としては I/O が高速である HDD でも内部転送レートとしては〜50 MB/s 内外であり，前節で確認された実時間収集の試験結果 (〜80 MB/s) にははるかに及ばない．このため仕様条件でもある 80〜100 MB/s を達成するには，HDD を複数台並列に利用して I/O 速度を上げる必要がある．こうした複数台の HDD を 1 セットとして同時に利用する手法をストライプ・セットとよび，その動作をストライピング動作という．ディスクアレイ装置の構成規格である RAID では RAID0 というレベルに相当する．

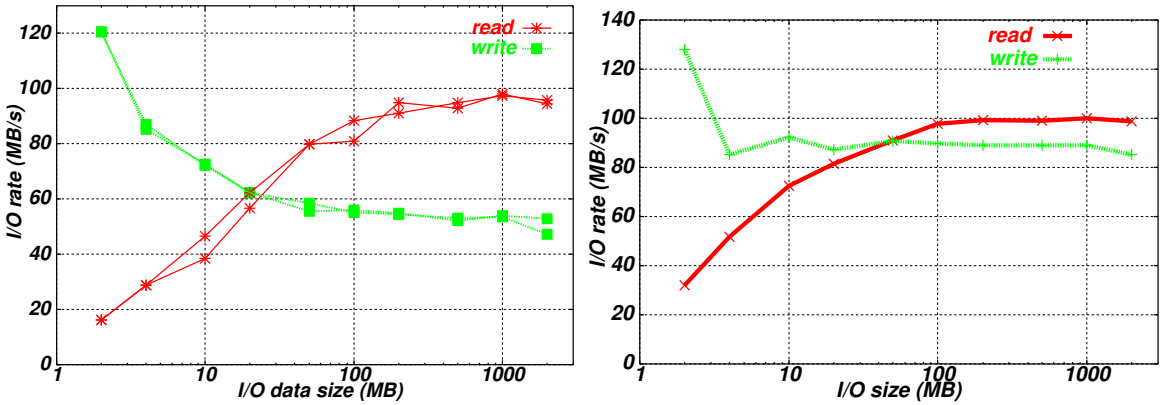


Figure 4.3: Achieved I/O rate of PCI RAID controller with RAID0 striping set: Their filesystem types are (top) NTFS ver.5, and (bottom) FAT32 filesystem. This benchmark has been obtained by using the HDBENCH ver.3.40 beta.6.

Figure 4.3 には，最近急速に普及し，大容量化とともに I/O 性能も上がっている Ultra-ATA/100 HDD (IBM DTLA 307075; 75 GB, 7200 RPM, 8.5 ms, 2 MB buf.) を 4 台と，PCI RAID カード (Promise FastTrak100/TX2000) にて RAID0 ストライピング構成を組んだ際

の read/write 性能を試験した結果を示している．効率的なある程度以上のブロックサイズの入出力では，read 性能はほぼ 100 MB/s に近く，またジャーナリング機能をもっていることで書込みオーバヘッドが大きい NTFS では性能が下がるものの，よりオーバヘッドの少ない FAT32 ファイルシステムでは write でも〜90 MB/s の性能が出ており，簡便な PCI RAID0 カードと HDD の組み合わせでも実時間収集データの保存を十分可能にする性能が出ることが確認された．

以上述べたとおり，プロトタイプシステム環境での動作検証では，連続 AD 変換時に〜80 MB/s でホスト PC 主メモリ上への無停止データ転送動作を確認することができた．これは NI 社が公表している MXI-3 光インターフェースユニットの最大伝送性能〜84 MB/s (連続時) とよく一致しており，実効性能としてプロトタイプシステムの現状でも十分満足のいく連続動作性能を出すことができたといえる．今後は LHD 短パルス実験での導入，CAMAC からの交替と長時間実験での実稼動インサイト・デモンストレーションを通して，普及と利用拡大を図っていくことになる．

またオープンな標準規格である CompactPCI の利点を生かしたより低コストな ADC モジュールの開発・利用と，目標とした性能仕様を満たすべく高速転送 PCI(64-bit, 66MHz, 100MHz, 133 MHz) への対応，それに 2-G FibreChannel, Myrinet2000, BEX-3 など 2 Gbps 級の広帯域光インターフェース導入の検討などを継続して行う予定である．

4.3　既存デジタイザとの共存

前節で開発検証を行った CompactPCI デジタイザ系は，機能性能的には従来の CAMAC 系統やその他のフロントエンドを置換できる能力があることを確認することができた．しかしコスト面では比較的低価格になると想定される CompactPCI/PCI 規格を用いても，やはり 2000 チャネルにもおよぶ全計測信号を一斉に置換するのは現実的に困難である．

そこで重要になるのが，CAMAC など既存デジタイザ資産の有効活用法の開発である．特に LHD など大型装置で比較的多くのデジタイザ資産を保有する場合には，新規デジタイザの開発・移行と既存資産の有効活用法の開発の両方が不可欠になっている．

4.3.1　CAMAC バッチ収集系の長時間運転対応

プラズマ計測で良く使用されている CAMAC デジタイザは，規格的に連続運転できないため，新デジタイザの導入を含めて新たな対応策を講ずる必要があるが，大規模装置では既存デジタイザリソースの有効活用が重要となる．このため CAMAC 系統をどのように定常実験に対応させるかが大きな課題である．LHD データ収集では以下のような運転方法を想定し検討を行った．

1. 1-way 低速サンプリング
2. イベントトリガー実装
3. 2-way 交互運転
4. 繰返し運転

厳密には上記のうち 3. のみが連続運転に対応しているが，2000 チャネルに及ぶ LHD の計測信号を 2 系統化する実現性を考慮して，4. を標準対処法と位置付け計測に応じて 1.または 2.を選択可にする形態とした．Figure 4.4 に長パルス持続中に短パルス連続運転と類似した周期トリガータイミングを生成して，CAMAC 系を繰返し運転させる模式図を示す．

ロングショット中で計測データ収集系のみ繰返し運転を実施するため，1 実験シーケンス中に複数のサブ（短）ショット・シーケンスを生成する局所ループ回路を計測タイミングシステム系統（VME ＋ PLC）に今回新たに付加した．Figure 4.5 を参照．

ここで生成された 2〜3 分間隔で繰返すサブショット・シーケンス (s3〜s9) は，元のロングショット・シーケンス (S1〜S10) に重畳されて下位に頒布されるため，末端の計測データ収集システムは特に通常実験時と変わることなく長パルス運転に対応できる長所がある．クライアント側の同期のためには，IP マルチキャストにより上記重畳シーケンス (S1

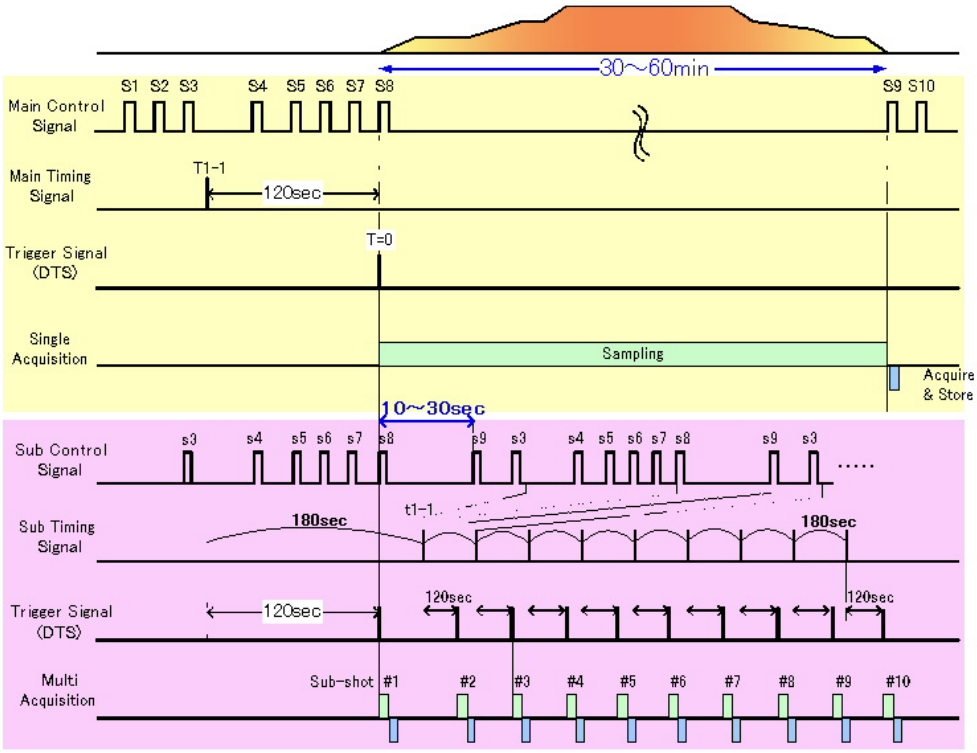


Figure 4.4: Sequence timings and triggers distributed by DTS: Normally the loop interval of the subshot sequences will be the same 180 s as the usual short pulse experiment in LHD.

～S10 ＋ s3～s9) をパケットメッセージとして頒布している．

4.3.2　計測制御系/制御データ処理システムとの住分け

3.5 節で述べた LHD 計測制御系でも VME モジュール型の ADC によりリアルタイム信号モニターが可能である．しかし機器制御が本来の目的である計測制御系では，ADC での信号データサンプリングは一つ一つが VMEbus 計算機 CPU への割込み＆ソフトウェア処理要求になるため，サンプリング・レートは基本的にあまり上げられない．

通常，CPU の種類を問わず CPU の割込み処理性能はほとんど上限 1～10 kHz 程度となっている．複数チャネルを取り扱ったりその他の I/O からの割込みも処理しなければならないことから，LHD の計測制御 VME 計算機の ADC では 10 Hz/ch のサンプリング・レートを上限にしている．

計測制御系と同様の機器監視を目的とした LHD 制御データ処理システム［82］は，同じく定常運転時には 1 Hz/ch のサンプリングを 24 時間連続で運転可能で，LHD の超伝導コ

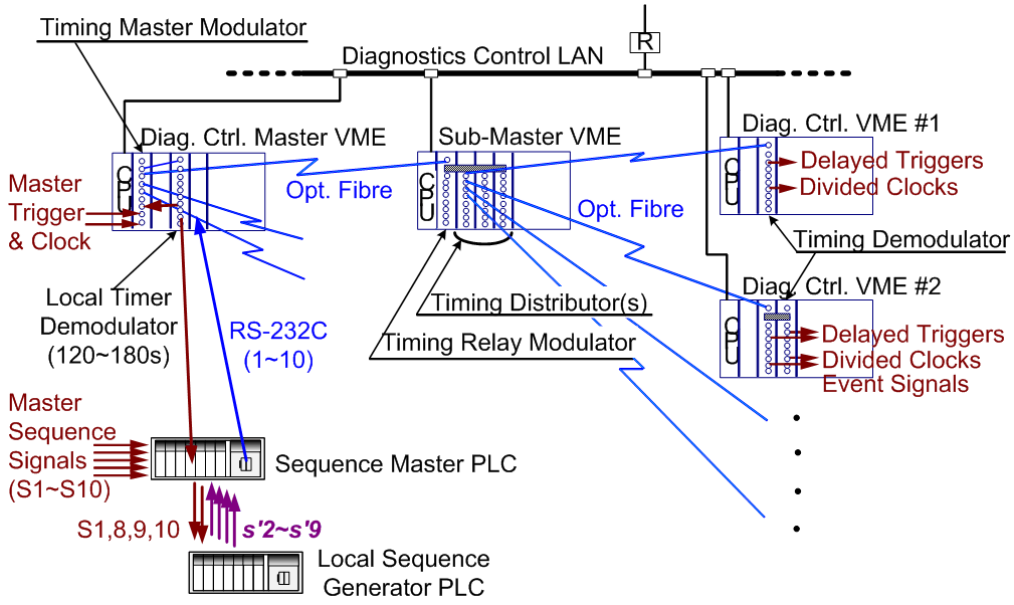


Figure 4.5: Upgraded structure of the LHD Diagnostics Timing System(DTS): For the multiple subshot timing generation, a combination of a DTS demodurator and a PLC work together as the local timing and sequence loop oscillator. Sub-sequence messages will be distributed on different channel than the original master sequence, which only distributes the one pulse sequence even in the long-pulse experiment.

イル監視システムとして～512 ch 以上稼動しており，同時サンプリング・チャネル数が多くできるのが特徴である．もう一つの特徴が，予め定義されたイベント発生時に 1 Hz モニターと並行動作しながらその前後だけサンプリング速度を 1 kHz に変えて収集するバッチ処理できる機能を併せ持つ点である．

定常化プラズマ実験においては長時間にわたって各種プラズマパラメータが安定しており，その運転状態の監視モニターとしてはこれら 1～10 Hz サンプリング程度の間隔で十分であることが多い．しかしプラズマ・パラメータの変化傾向を観測するのではなく，その局所的な揺動を解析するなど物理計測の目的では，こうしたサンプリング周期では不十分であり，そうした意味でも物理計測デジタイザ類とは棲み分けができているといえる．

4.3.3 中速度 (< 1 kS/s) モニター計測系の改良

2.1.2 節で述べた横河電機 (株) 製 PC ベース計測器 WE7000 シリーズでは，WE7272 などトランジェント・レコーダー型 ADC モジュールと複数のフロントエンド筐体，専用光

インターフェースの組み合わせ (デイジーチェーン接続) により，数 100 チャネル以上におよぶ多チャネル実時間収集系を構築，専用の WE7000 コントロールソフトウェアで運転することができる．この収集システムでは 1 kHz/ch で ADC をリアルタイム動作させることができるので，前節の実時間機器監視系の 1〜10 Hz サンプリングと，いわゆるプラズマ物理計測用の 10 kHz〜1 MHz サンプリングとの間隙を埋めるサンプリング周期であり，多チャネル構成が容易な点からも有効なシステムであるといえる．

しかしプラズマ計測用に良く用いられる WE7272 トランジェント・レコーダー型 ADC で 100 kHz サンプリングまで可能であるにも関わらず，専用コントロールソフトウェアでは 1 kHz/ch の動作制限がかかっている．これは，専用コントロールソフトウェアが数 100ch の利用を前提としているためであるが，通常の多チャネルプラズマ計測でも高々 100 内外のチャネル数であり，プラズマ計測の要請には必ずしもそぐわない状況になっている．

そこで本研究では，CAMAC ADC の 1/2〜1/3 程度と比較的安価な WE7000/WE7272 を中速度 (10〜20 kHz/ch) の実時間収集デジタイザ系として活用すべく，Figure 4.6 のように LABCOM データ処理系に組み込むシステム開発を行った．これは WE コントロール API

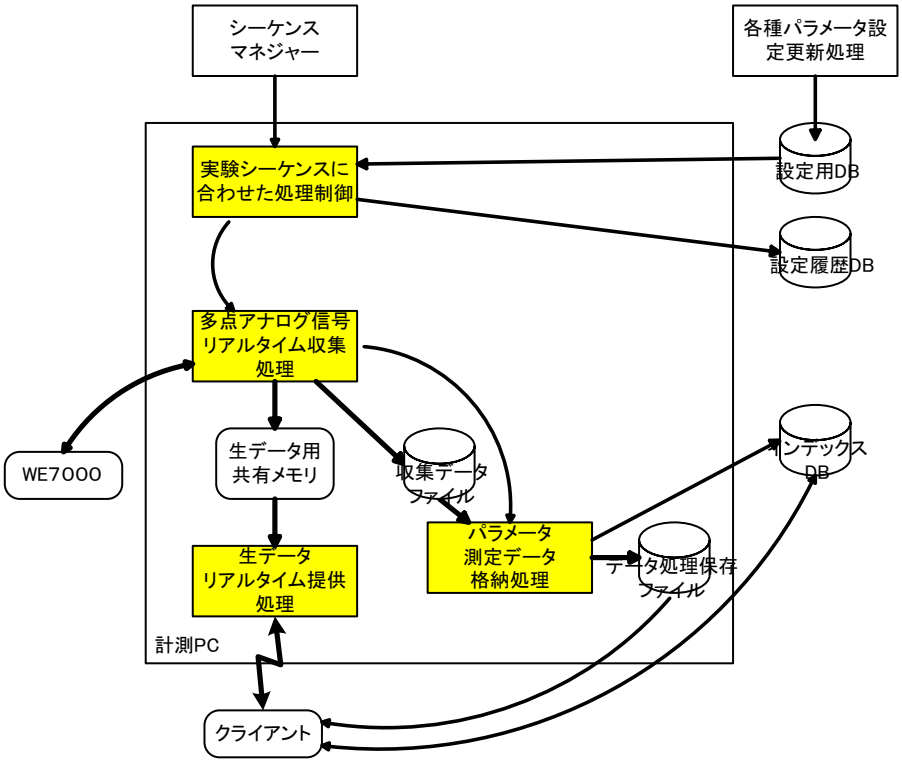


Figure 4.6: Block diagram of real-time streaming processes for Yokogawa WE7000.

を直接よぶプログラムを作成することで，1 kHz/ch 制限問題を回避している．このプログラムで最も注意すべき点は生データ用共有メモリのアクセス制御であり，収集プロセスからの書込みとデータ提供プロセスからの読出しに同時に対応できるように，ダブル・バッファあるいはリング・バッファの機構を用いて，アクセス待ちを回避する必要がある．

Table 4.3 が同収集プログラムにより 1 時間程度の連続収集試験を行った結果である．測定環境は以下の通り．

【測定条件】

WE7000 ： WE800 × 1, WE7272(MAX.100kHz) × 8 = 32 ch,
WE7038 光通信モジュール (MAX.250Mbps)
計測 PC ： Xeon 2GHz × 2, Mem.2GB, Windows2000 Server SP3
ソフトウエア ： WE コントロール API Ver.4.0.5.0
ファームウエア： WE800:4.01, WE7272：3.10.0001 × 8, WE7038：3.02

Table 4.3: 1-hour durability examination on WE7272×8 digitizers: In fault cases of 1-h running, the block sizes are chosen as the best one which gives the longest sustained time. This result shows that 32 ch 100 kHz sampling are even possible in short pulse (∼10 s) experiments.

sampling rate	ch.#	data rate	block length	sustained time	1-h run
10 kHz	32	625 kB/s	4 kB	3600 s	◯
10 kHz	32	625 kB/s	8 kB	3600 s	◯
20 kHz	32	1250 kB/s	8 kB	1065.7 s	×
20 kHz	32	1250 kB/s	16 kB	3600 s	◯
50 kHz	16	1562 kB/s	32 kB	3600 s	◯
50 kHz	32	3125 kB/s	32 kB	41.9 s	×
100 kHz	7	1367 kB/s	32 kB	3600 s	◯
100 kHz	32	6250 kB/s	32 kB	13.4 s	×

上記の試験結果をみると，リアルタイム収集時のシステム全体性能がほぼ WE7000 転送帯域の性能上限である 1.5 MB/s になっていることから，本開発システムが 20 kHz×32 ch の連続リアルタイム収集能力を十分持っていることが確認できた．また 10 秒程度の短パルス運転では，WE7272 の最大サンプリング速度 100 kS/s での動作も可能であることから，本収集系が短パルス運転時にも利用可能である事も同時に確認できた．

非常に高性能であることが本研究で実証された CompactPCI ベースの超広帯域ストリーム収集系であるが，既存のデジタイザリソースやより簡便な製品を適材適所に組み合わせることで，コスト・パフォーマンスも最適化することができる．本研究で開発および検証した３つの実時間速度帯域のデジタイザをチャネル単価と共に比較すると以下のようになる．

デジタイザ	運転形態	サンプリング速度 (ch 数)	ch 単価
CAMAC	繰返し運転	∼1 MHz (any)	150k∼250k 円/ch
計測制御 VME	実時間	∼10 Hz (8)	20k∼50k 円/ch
WE7000	実時間	∼20 kHz (32)	55k∼80k 円/ch
CompactPCI	実時間	∼1 MHz (28)	100k∼200k 円/ch

この通り，実時間サンプリングの速度範囲とチャネル単価との対応は明からである．LHD のような大規模で多チャネルの収集系を構築する場合には，要求性能に見合ったデジタイザ系を選定することが重要だといえる．

4.4 LHD 定常実験への適用と今後への展望

計測データ収集系の定常運転対応の問題を根本から解決するために，本研究では CompactPCI 規格をベースにした新データ収集系の開発検証を行った．その結果，プロトタイプにて～80 MB/s の無停止連続ストリーム収集・保存に成功し，プロトタイプ・システムでも一筐体に収容できるチャネル数を下げれば十分そのままで実用化可能であることが示された．

残る問題としては，2 Gbps FibreChannel, Myrinet2000, BEX-3 といった 2 Gbps 級の光ファイバー伝送 I/F の導入と，CompactPCI rev.3.0 対応による 64 bit 66 MHz (=533 MB/s) バックプレーン転送帯域の向上により設計目標性能である 100 MB/s の達成を目指すことになる．今後は CAMAC デジタイザの代替のため低コストな同 AD モジュールを開発・導入する検討も必要になるだろう．

長時間放電実験に対応するための既存デジタイザ，特に CAMAC ADC の有効活用法の開発については，LHD 計測データ収集・処理システムにおいて既存 CAMAC デジタイザ系統の 2～3 分間隔の断続繰り返し運転モードを平成 13 年度までに策定し，実験第 5 第 6 サイクルで動作を確認，実運用に供した．これは必ずしも完全な実時間データ収集を実現したわけではないが，既存デジタイザリソースと 2,000 チャネルに及ぶ計測チャネル数を取り扱うための現実的な選択肢である．

システムとしては，既存の計測タイミングシステム (DTS) と同ロジック制御 PLC のシーケンス系統に対して，2～3 分間隔のローカルタイマーを持った追加の DTS+PLC の組合せを重畳することで，透過的に重畳された計測タイミングシーケンスの送信を可能にした．そのためそれを受信して動作するデータ収集サーバは，全くシステム変更の必要なく繰返し運転モードに移行できるという優れた長所を持っている．

また実時間収集機能を持っている横河電機 (株) 製 PC ベース計測器 WE7000 用の専用実時間データ収集システムを開発し，1 kHz に制限されていたサンプリング速度を 20 kHz (32 ch.) 連続まで伸ばすことに成功した．これにより，再掲，

デジタイザ	運転形態	サンプリング速度 (ch 数)	ch 単価
CAMAC	繰返し運転	～1 MHz (any)	150k～250k 円/ch
計測制御 VME	実時間	～10 Hz (8)	20k～50k 円/ch
WE7000	実時間	～20 kHz (32)	55k～80k 円/ch
CompactPCI	実時間	～1 MHz (28)	100k～200k 円/ch

のように，ほぼプラズマ計測に必要な周波数帯域全域を網羅することが可能になり，実時間計測に必要なサンプリング速度・チャネル数・予算などに応じて具体的なデジタイザ準備の戦略が策定できるようになった.

本研究での LHD リアルタイム収集系開発の網羅的かつ具体的な成果は，LHD に続いて今後世界で運転が開始されてくる定常化プラズマ実験装置や大型の次期装置として世界から注目を集めている ITER などに対して，核融合実験の分野で世界最初の事例として非常に大きな知見を与えるものである．現時点ではまだ，ITER データ収集保存系に要求される仕様の詳細は不明であるが，ITER 建設時までの計算機，ネットワークなど要素技術性能の向上を考慮すると，本研究で実証したシステム技術により十分実現可能であるといえる.

第５章

結論と展望

本論文は，核融合科学研究所が現在注力して実験を行っている大型ヘリカル実験装置 (LHD) の計測データ処理システムを構築するために行った開発研究の集大成である．この開発研究には，データ処理システムの守備する範囲の広さゆえに，アナログ信号を伝送・変換するハードウェアから収集データを可視化するユーザエンドのソフトウェアに至るまでが含まれており，電子回路，計測制御，ネットワーク，計算機/システム・アーキテクチャ，ソフトウェアの各工学の応用のほかに，プログラミング方法論やプロジェクト・マネジメント論など開発パラダイムにも検討や工夫が及んだ．

元来，核融合プラズマ実験では他分野の大量データを収集する実験・観測装置とは異なり，長い時系列を伴った比較的粒度が大きいプラズマ計測データを，各々データの種類が異なる多種多様な計測器から同時に収集処理することが求められてきた．また，多数回の実験データを全参照して統計処理を行い，経験論的法則性いわゆるスケーリング則を抽出するデータ解析もしばしば行われ，この際の極めて特異なデータ参照パターンにも対応してきた．このため核融合実験の計測データ処理システムは，他分野にはない独自のデータ収集および取扱いの能力を実現してユーザに提供してきている．

しかし最新鋭の核融合プラズマ実験装置 LHD のデータ処理システムに対する新たな要件は，従来装置での実績を多くの点ではるかに凌駕するものであった．ここに再度，課題となった点を挙げてみると以下のようになる．

1. 1 回の放電実験あたり総量 0.6〜1 GB の生データ収集
2. 30 種に及ぶプラズマ物理の計測器
3. 短いデータ処理時間 (3 分に 1 回繰り返す実験周期)
4. 計測の多チャネル化・マルチメディア化
5. 遠隔共同研究に不可欠なオープン・システム・アーキテクチャ/インターネット活用

6. データ処理系の定常運転化 (〜10000 秒の連続プラズマ放電など)
7. 現有デジタイザ資産の活用

そのため，LHD 計測データ処理システムの開発研究においては，全く新しい設計思想を適用し，各々の課題に合わせてそれぞれ工夫を凝らすことで，多数の要件を満たすことに成功した.

従来の核融合プラズマ計測のデータ収集系では，高速な処理を実現するために計算機資源を集約した中央集中型のシステムアーキテクチャを採用する場合がほとんどであり，現在もまだ中〜大型の核融合実験装置ではこの型のデータ収集システムが依然主流として稼動している．しかし世界最新鋭の LHD 装置では，こうした既存装置をはるかに凌駕するデータ収集量と計測数とに対応を迫られ，中央集中型システムではデータ入出力の集中によるボトルネック発生が実験装置運転に深刻な遅延をもたらしかねなくなった.

この問題を解決するための新たなシステム設計が，データ収集伝送系の並行分散化である．従来の中大型計算機に１〜数系統でデータを伝送していた形態を廃し，分散配備した複数の小型計算機に個々に伝送系統を備えることで，同時並行したデータ収集処理が可能になりシステム全体のスループットを大幅に向上させることができた.

このシステム形態では，多数の計測機器に同時並行かつ独立して収集処理を進行させることができるため，LHD の約 30 種のプラズマ計測にそれぞれ並行して対応させることで，１ユニット性能 (〜1 MB/s) の約 30 倍のシステム・スループット (〜30 MB/s) が得られ，核融合実験で使われる CAMAC データ収集システムとして世界最速を達成している．大型のプラズマ実験装置では極めて稀な 3 分間隔の LHD 短パルス繰返し運転で，世界最大のデータ量 (〜1 GB/shot) を各実験終了後の約 1 分以内に収集処理できるのも，このシステムの高速性ゆえである．また，システム全体に影響を与えずに計測機器の増設・削除等に容易に対処できる点も，独立性の高い並行分散システムの長所の一つである.

このように小規模な 1 計測 (〜1 MB/s) から大規模な 30 計測 (〜30 MB/s) までスケーラブルに拡張でき，それ以上の規模へも自在に性能を伸長できる計測データ収集システムを開発・実証できたことで，同分野に対して大きなブレイクスルーをもたらすことに成功したといえる．また，それを世界に先駆けて LHD システムに適用，実用化したことは非常に意義が大きい.

しかしその反面，機能分離・並列分散を徹底することで，例えば，今まで一つのプログラム開発で済んでいたタスクは二つ以上に増やされるため，システム開発負荷が大きく増すという問題も生じてきた.

こうした大規模分散システムの開発を従来と同じ構造化手法によって進めると，上述の通り，開発負荷の増加は深刻な問題となる．これを回避するには，より高い開発効率が得られる新たなパラダイムが必要であり，本研究では分散コンピューティング・アーキテクチャと親和性の高いオブジェクト指向方法論を全面的に適用した．

オブジェクト指向に則った C++/Java 等のプログラムコード開発は，開発時のプログラミングの見通しを良くするとともに，複数メンバーによるプログラミング分担の際にもその取り合いを判りやすくした．類似のコードを多数作成することが多い CAMAC デジタイザや VME の I/O モジュールの制御には，オブジェクト指向開発言語のクラス継承によって，コード開発負荷が大幅に低減されている．

オブジェクト指向化は LABCOM システムのほとんど全てのシステムプログラムに及んでいるが，特にシステム・スループットの面で高性能を要求されるデータ収集プログラムでは，処理速度改善のためのシステム再検討に UML を用いるなど，システムの要求分析や仕様策定にもオブジェクト指向開発技法を有効に活用した．このようにオブジェクト指向という明確な開発パラダイムを適用することにより，新しい構成のシステム構築に際しても，有限の開発リソースで着実に大規模システムを開発することができたといえる．

計測データ処理システムの遠隔利用については，計測器に付随してこれを制御しその機能を提供するサーバ・プロセスと，提供された機能を利用するクライアント・プロセスとを完全に分離し，その間を遠隔通信のデファクト・スタンダードであるインターネット技術で再結合する構成が要求される．このクライアント/サーバ間の分離は，それぞれの処理の実体を主体的にかつ非常に独立性よく記述できるオブジェクト指向言語を用いることで，極めて容易に実現することができた．以上は，データ収集や計測制御といった機器寄りの分散システム開発において，オブジェクト指向パラダイムの適用が非常に有効であることをよく実証している．

通常，オブジェクト指向プログラミングがもつ生来的な冗長性によって，例えば構造化言語 C で記述されたものよりも C++プログラムでは処理オーバーヘッドが大きくなる．このため，今回のように大容量の I/O を取り扱う系に適用した場合，データ伝送速度の低下により全体のスループットが上がらないことが憂慮された．しかし，システム各部でマルチスレッド・プログラミングやデータ圧縮技法などをそれぞれ工夫して利用することで十分にそれを補償し，よりスケーラブルな収集システムへと進化させることが出来た．オブジェクト指向開発は LHD のような巨大な I/O 系にはそのままではそぐわない面があったが，本研究での実証により新たな知見が加えられたといえる．

計測データ収集系の定常運転化は，核融合実験のデータ処理分野では非常に新しい課題

であり，しかも従来の一括 (バッチ) 処理形態を一変させるものである．核融合計測で一般的に使用されてきた CAMAC 規格デジタイザは高速定常運転が機能上不可能なため，kHz〜MHz 帯のサンプリングレートを下げられない MHD 揺動計測等では，CAMAC に替わる新たな高速リアルタイム動作のデジタイザを必要としている．本研究では性能とコスト，周辺技術との親和性の点から CompactPCI 規格に注目し，これを核融合実験の計測デジタイザ・フロントエンドとして実用化することに成功した．

実際の利用環境に近い AD モジュール 7 台を 1 フロントエンド内で動作させ計 28 チャネルの収集を行った際に 80 MB/s まで連続転送が可能であることが確認された．この実効帯域は稼動時に想定される 1 MS/s サンプリング (×28ch) 動作の全 AD モジュールが生成する 56 MB/s のデータストリームを十分に転送できる性能であり，本システムの実用性をよく示している．

また中速 (〜20 kHz) デジタイザとの組み合わせで，既存のデジタイザ資産を活用するデータ収集系の定常運転戦略は，同様の問題を抱える中〜大型実験装置に対して，実証された具体的解決技法を提示できた点も重要である．

以上に述べた研究成果は，C++/Java コードライブラリあるいは複数の研究報告文書として，LABCOM ホームページ http://w3.lhd.nifs.ac.jp/ から公開している．LHD 計測データ処理システムは平成 13 年度の実験第 5 サイクル末で，平成 7〜9 年度に作成された計測計画が想定した 600 MB/shot を超えて 620 MB/shot の総データ収集量に達した．その後も翌年の実験第 6 サイクル末には 740 MB/shot にまで収集量を増やすなど，そのスケーラブルな性能を遺憾なく発揮しており，本開発研究の目的は十分に達成されているといえるであろう．

なお一実験あたりの計測データ収集量でも，世界の核融合実験装置の収集量と較べて十分抜きん出た値となっているが，実験の繰り返し間隔を考慮した一日あたりの収集量に換算すると，3 分間隔という非常に短い運転周期に対応できている LABCOM システムの収集実績は，一桁以上も他システムの処理量を上まわっており，圧倒的な差で世界一である点を再度強調しておきたい．

今後に向けた研究課題と展望

本研究では，コンパクトなデータ収集エレメントを多数並行に動作させることによって，高性能な大規模システムを構築することが実証できた．しかし本研究で検討したデータ収集エレメントはデジタイザから収集コンピュータおよびそれ以降の部分であり，計測セン

サーからデジタイザへの接続は，既存のアナログ信号伝送技法にて行っている．

計測センサが生成する生信号は通常微弱であり，他の計測器や強電力装置などから信号線路を伝ってまわりこんでくるさまざまな電磁的ノイズの影響をうけて汚染されやすいため，従来から絶縁増幅器 (アイソレーション・アンプ) を各計測チャネルごとに用いてアナログ計測信号 (源) を保護・伝送する手法が確立している．しかしアナログ信号の絶縁方法は，誘導トランス型にせよフォトカプラ型にせよ一般に周波数特性が悪く，プラズマ計測に要求される〜MHz 帯のものは非常に高価で，計測エレメントの可用性を大きく下げてしまっているのが現状である．今後よりいっそう計測データ収集エレメントの可搬性・可用性を向上させるためには，このセンサ＝デジタイザ間の技術革新が必要であり，計測センサの近くに設置できるより可搬性のあるデジタイザの検討・開発が課題となっている．

今回の開発研究では，データ収集エレメントの独立性を重視し，1 ユニットから多ユニット構成まで同一アプローチを採ることでスケーラブルな MPP システムを構築した．普及した小型コンピュータをエレメントに用いて，従来に比べてシステム全体の大幅なコストダウンも可能になっている．その反面，この構成では各エレメントで利用できるコンピュータ性能が基本的に同じであり，計測ごとにデータ収集量や計算処理の差が大きい場合に処理時間が大きく食い違い，システム全体として処理完了タイミングが高負荷のエレメントに引っ張られてしまうという問題を抱えている．実際 LHD 計測の CAMAC データ収集系でも，一計測あたりのデータ収集量が最小は 1.5 MB から最大は 85 MB までと大きく隔たっており，処理の完了時間も数秒〜145 秒と違いが大きいのが実情である．

この問題を解決するには，データ収集コンピュータ間で負荷分散 (ロード・バランシング) を実現する必要がある．

今までのデータ収集系の基本思想では，デジタイザあるいは DFE は一意にその I/O サーバ・コンピュータに制御される，多対一もしくは一対一の接続関係であった．LHD 計測でも CAMAC–PC 間は SCSI による一対一の独立した伝送経路である．これに対して，デジタイザの入出力チャネル/モジュールを抽象化・仮想化し，なおかつデータ I/O の伝送路に回線交換機技術を適用することができれば，各サーバの I/O 負荷状況によって I/O 先をふりわける多対多のデータ収集・転送系が実現されうる．この場合，分散されている各コンピュータ資源・I/O 性能がより有効活用できるようになるため，システム全体の処理完了までにかかる時間が大幅に改善・短縮されることが予想される．

本研究で得られた分散処理系の成果をそのまま延長することにより，ITER 等の将来の超大型核融合実験への対応が可能になったことはこれまでに述べてきたとおりである．将来展望に述べた課題への対応は，これを更に効率的なものにする可能性を含んでいる．

謝辞

本論文をまとめるに至るまでの長い道のりで，非常に多くの方々にご支援を頂き，またご心配をおかけしてまいりました．この場にて深く感謝の意を表します．

本論文作成をつぶさにご指導いただいた核融合科学研究所の岡村昇一教授，須藤滋教授，上村鉄雄教授，長山好夫教授，および北陸先端科学技術大学院大学の日比野靖教授，松阪大学の奥村晴彦教授に深く感謝いたします．また，出身研究室の指導教官であった東京大学の井上信幸，小川雄一，吉田善章，二瓶仁，森川惇二の諸先生方には多年にわたりご心配をおかけ致しました．改めてお詫びと御礼を申し上げます．

本研究の主眼となった新システム開発はまさしく LABCOM グループメンバーの協力によってはじめて成りえたものです．小生の着任以来ずっとシステム開発の辛苦を共にしてきた小嶋護，大砂真樹の両氏なくしては本研究は成立しなかったと思われます．また，秀熊茂氏，江本雅彦氏，杉崎秀樹氏，野々村美貴氏，飯代清司氏，吉田正信氏，今津節夫氏，木股あやみ氏には多面にわたりご協力をいただきました．加藤丈雄，駒田誠司の両氏には LHD 実験 LAN 運用グループの活動を通して何かとご助力いただきました．プラズマ計測研究系，技術部計測技術課，制御技術課の皆様にも感謝いたします．特に同期の伊藤康彦氏には公私にわたって様々な相談に乗っていただきました．

また核融合データ処理研究会の共同研究の活動を通して多くの先生方にご指導と激励を頂戴しました．中部大学の山口作太郎先生，職業能力開発総合大学校の寺町康昌先生，日本サンマイクロシステムズの安光正則さん他皆様，日本原子力研究所の松田俊明，米川出，坂田信也，及川聡洋の各先生方，大阪府立大学の宝珍輝尚先生には様々なご指導を頂きました．九州大学の上瀧恵里子先生，京都大学の岡田浩之先生，筑波大学の板倉昭慶先生にもいろいろお世話になりました．同じ大学院研究室で学び修了後も変わらぬ親交をいただいている同窓の諸兄にも事毎に励ましをいただきました．

本当にありがとうございました．

平成 15 年 8 月

中西秀哉

付録 A

共同研究支援ネットワークの構築

A.1 核融合実験共同研究 ISDN 網 (FECnet) の構築

プラズマ核融合分野の共同研究ネットワークとして，平成 7 年度に研究高度化推進経費を受け核融合実験共同研究ネットワーク (Fusion Experiments Collaboration Network: FECnet) が開設・運用を開始した［27］．このネットワークを全面的に後援する立場から，LHD 計測データ処理系においては，予めデータ処理系に対する実験操作/処理要求クライアント計算機を，所内外を問わず全て一律にネットワーク端末化することで，共同研究先からのネットワーク経由での遠隔実験参加が，所内の実験用端末からの実験操作と全く同一の利用/操作環境となるように当初よりデータ処理システムの設計を行なってきた．

具体的実現方法としては，遠隔地間通信を容易にした昨今のインタネット技術を最大限に採り入れ，

1. データ収集/計測制御用の計算機を，遠隔自動処理機能を提供するサーバと，ユーザ・インタフェースを提供するクライアントに各計測器毎に完全に分離
2. サーバ–クライアント間通信をインタネット化することで，ネットワーク上の距離的制約から解放

する事で，LHD を実際に擁する核融合科学研究所々内で LHD 実験に参加する場合と全く同じ実験環境を，遠隔地の共同研究先にも提供する．Figure A.1 に FECnet と LHD データ処理系との関連を示すが，所内実験者が利用するクライアント計算機と遠隔地から ISDN 回線によって接続されるクライアント計算機とが，RPC などの通信プロトコルを用いて，データ収集/計測制御/実験データベースの各種サーバに全く同等に接続されることが理解できる．

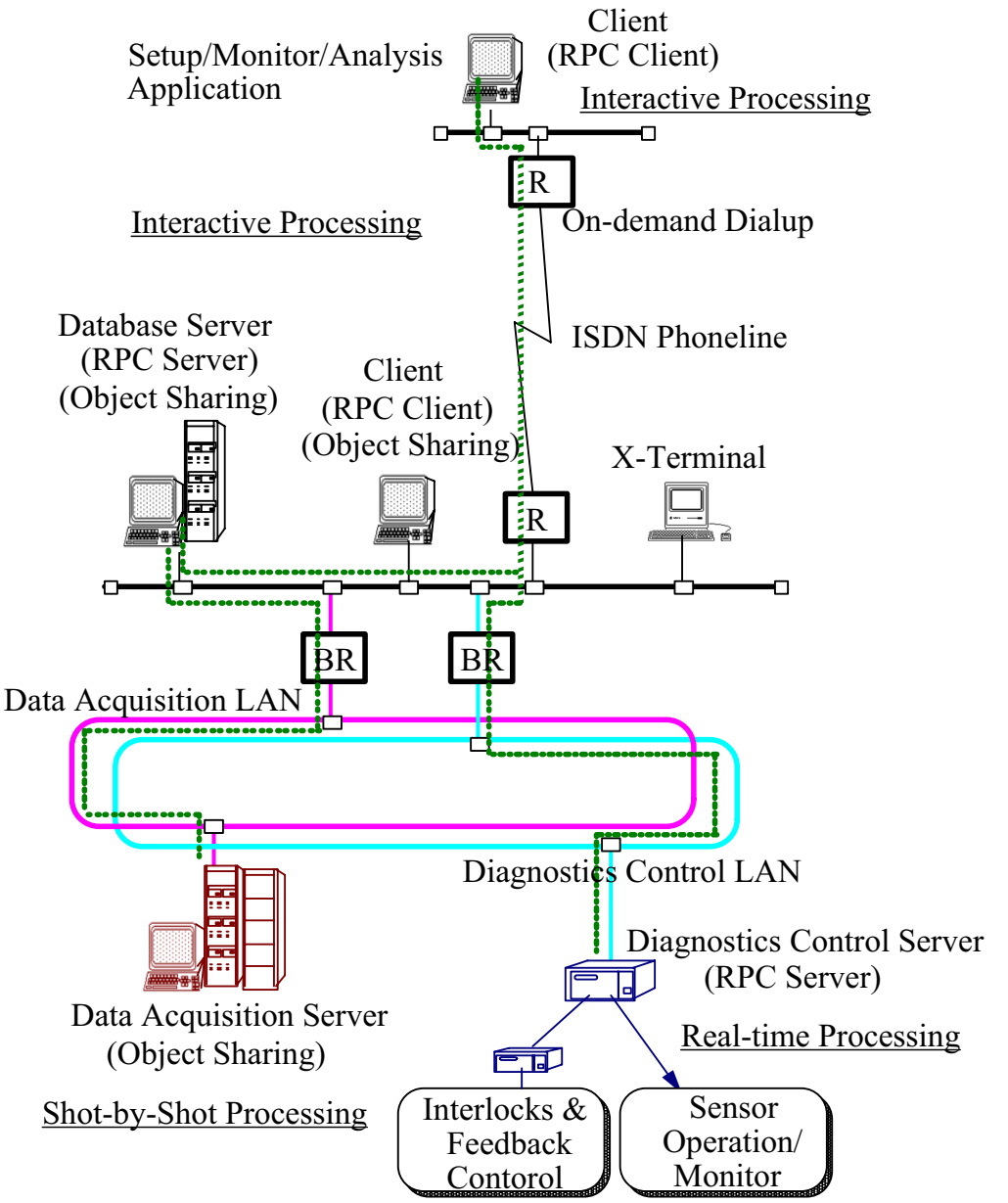


Figure A.1: Schematic view of the FECnet remote communication to access data acquisition and diagnostics control computers: Remote clients can communicate interactively with every server computers by using the same protocol as local clients. RPC is a kind of the higher-level protocol of TCP/IP.

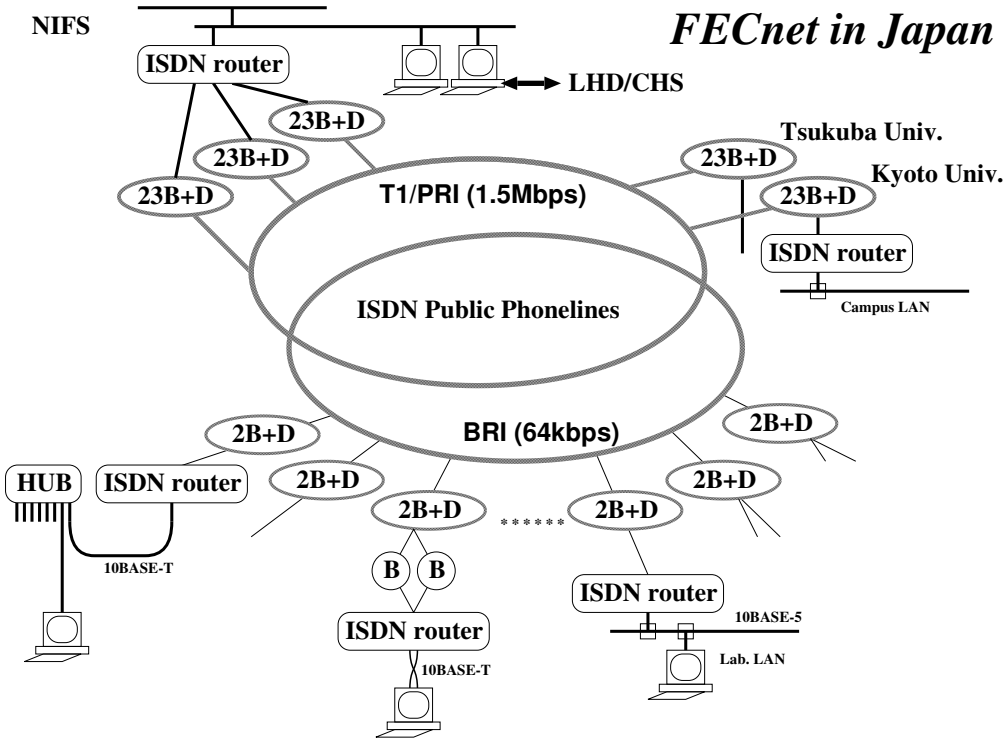


Figure A.2: Diagram of inter-connections between LHD local area network and FECnet ISDN networks.

この形態は，計測制御サーバに直接接続できる事から，遠隔地においても計測器の制御機器をリアルタイムで遠隔操作可能となっており，従来の実験データの転送/共有方式の共同研究支援には無い，実時間遠隔操作/監視という新たな共同実験形態を提供する特徴を持っている．

FECnet で利用した ISDN 技術は，より高速な ADSL 網などの普及によって徐々に利用が縮小しており，今後の LHD 実験のための共同研究ネットワークとしては次節に述べる SuperSINET 利用へと移行してきている．

A.2　SuperSINET の導入と活用

大学間を結ぶ幹線ネットワークとして ATM 技術をベースに構築された学術情報網 SINET にかわり，DWDM(Dense Wavelength Division Multiplexing：高密度波長分割多重) 伝送や広域 Gigabit Ethernet 技術の基盤にしたスーパー SINET が平成 13 年に新たに構築され平成 14 年 1 月 4 日から運用が開始された［83］.

大容量の計測データを収集し，また共同研究利用サイトとして遠隔実験環境の開発を求めていた LHD データ処理系では，SuperSINET 導入当初から積極的に参画することで，高エネルギー・核融合科学研究部会として 10 Gbps IP ×1 および 1 Gbps Ethernet ×3 の SuperSINET 主要ノード獲得に成功，Figure A.3 のように構内実験ネットワーク環境もそれに併せてギガビット級 LAN に増強できた.

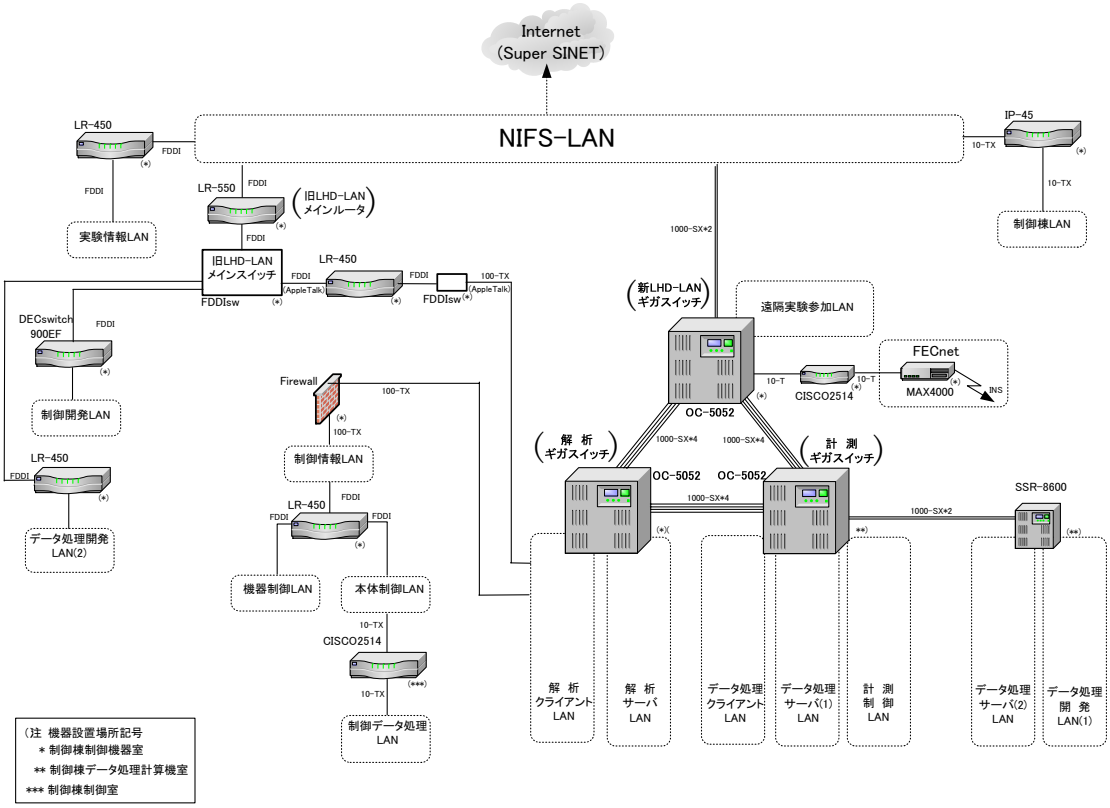


Figure A.3: Overview of the recent LHD network based on the Gigabit Ethernet: These network fabrics had been introduced for the SuperSINET in 2001. (Reprinted from http://lhdnet.lhd.nifs.ac.jp/Giganet/Giga020822.html)

これにより他大学等の共同研究先への超広帯域なネットワーク環境を実現され，本研究で

開発したような～100 MB/s 級の超広帯域実時間収集データのリアルタイム遠隔伝送も視野に入るようになってきている．

平成 14 年度以降，共同研究先である京都大学や名古屋大学，東京大学，九州大学との間でそれぞれ 1 Gbps IP 専用回線を設置し，大容量の計測データ遠隔参照と遠隔実験策定のための高速ビデオ会議システムの導入を実現してきている．

付録 B

LHD 実時間計測系への調査と応用

B.1 リアルタイム磁場揺動計測のための解析演算と可視化仕様

以下に LHD 磁場計測担当者との面談調査および意見交換により，本研究で開発した超広帯域実時間データ収集系の適用とその際に必要な開発諸元について得られた結論を示す．特に LHD 磁場揺動計測は必要とするサンプリング速度，チャネル数からして超広帯域実時間データ収集系の格好の開発ターゲットと考えられる．

Magnetics 計測の現況

Magnetics 計測には，大別してプラズマの平衡をみる磁場平衡計測と，MHD に関連して揺動磁場の変化を見る磁場揺動計測とがあり，いずれもデジタイザのサンプリング速度は 100kHz だが，前者は～10ch，後者は 196ch をもっている．

長時間のプラズマ放電時の計測に対応するため，平衡計測用には，VME をベースにした新計測システムを計画しており，現在，CHS でのプロトタイプ試験機が設計・開発・製作発注済みである．このシステム開発の最大の目的と特徴は，従来までのアナログ方積分器では数 10 秒以上の長パルス運転時に，オフセットドリフトが無視できなくなるため，リアルタイムのドリフト補正機能をシステム側に持たせた新設計の積分器を利用している．

このシステムはピックアップコイル全体 196ch をカバーするものではなく，あくまでも Wp，Ip，Mp，Saddle ループなど数点の信号 (6ch+13ch) に限定して利用する．その他の約 200ch の信号のデータ収集については，未だ収集計画が確定されていない．

新磁場平衡計測システムの仕様条件など

基本的に，新磁場平衡計測システムで処理される Wp，Ip，Mp，サドルループ信号など約 20ch(入力 10ch）に関しては，ほぼ同等以上の機能を超広帯域実時間データ収集系システムでも実現可能である．諸値は時間変化の一次元データなので，リアルタイムで横スクロールする x-t プロットを信号の数だけ実装する．

- アナログ入力信号チャネル … 10 ch 微分信号 (磁場揺動用ピックアップコイルは全 196ch ある)
- 入力信号のサンプリングレートは 100 kHz
- 1ch アナログ入力をマルチプレクサで 1/3 ずつ時系列分割して，それぞれを A/D 変換後に数値積分．積分時間 … 2ms
- 1/3 ずつの積分波形を足し合わせて積分信号を復元.
- 基本ソフトは CHS で組込み用途に使っている CINOS で言語レスで実装

磁場揺動計測

現在，100kHz で通常の CAMAC ADC にてバッチ・サンプリング＆処理している．チャネル数は 196ch．これらのデータをリアルタイムで処理する具体的計画は今のところ無い．データ解析・表示方法としては，現在以下の通り．

- 時系列 2048 サンプルで FFT をかけている
- 空間的なモード数 (m,n) を算出するのに最低空間点 22ch が必要だが，全 196ch を使用できればより良い
- モード数計算のルーチンは，各 ch の相互相関を取って位相差を出すもの
- 磁場 (揺動) の内部構造を再構成して等高線 (contour) 図などに表現するには，モデルを仮定する必要があり，実験データのリアルタイム表示としては，曖昧性が入るので余り積極的には行いたく無い．

リアルタイム処理に望む性能としては，画面出力のリフレッシュは必ずしも連続的である必要は無く，1 秒〜5 秒程度までの更新レートでもよい．仕様詳細は以下の通り．

リアルタイムに表示を行いたい画面のうち，Wp，Ip など平衡に関する諸値は 1 次元データなので通常の x-t プロットでよい．表示として面白いのはやはり揺動計測の結果で，磁場揺動計測の約 200 チャネルのピックアップコイルアレイの実時間処理では以下のようなプロットがあれば，望ましいと考えられる．

Power Spectrum Density(PSD) の時間変化をカラー contour プロット　196ch 全てを表示する必要は無いが，100kHz サンプリングの時系列データを 2048 サンプル毎に区切って各チャネル FFT 計算し周波数成分に変換．横軸：時間，縦軸：周波数のグラフで，周波数成分が強いところは赤く，弱いところは青くカラー表示し，どの周波数の揺動波が立っているか，またその時間変化を一目で判るように表示する．

Coherence contour plot　PSD contour plot が周波数成分だったのに対し，空間のプラズマモード数 (m,n) の各成分強度を全 196 チャネル間の相互相関解析によって算出，縦軸：(m,n)，横軸：時間で同じく強度の強いところを赤，弱いところを青くするなどでモード数成分の時間変化を表示する．

モード数毎の成分強度変化　縦軸：モード成分の強度, 横軸：時間で (m,n)=(1,1),(2,1),(2,2),... のそれぞれの成分強度変化グラフを縦に幾つか並べて x-t プロット＆横スクロール．

磁気面表示　縦長断面位置，横長断面位置での内部磁場構造を contour プロット．磁気島が生成しているなどがリアルタイムで見られると面白い．内部構造の演算モデルを埋め込む必要がある点は留意する必要あり．オプションとして検討．

以上の表示画面が同時に複数表示できることが望ましい．一部は切替えで可．

将来に実現してみたいシステムとしては以下が想定される．

- 磁場揺動により磁気島が生成しているような場合，揺動成分の波形位相に同期させて CD 波を入射することで，磁気島の抑制が可能となるので，磁気島制御に利用してみたい
- サドルディテクタ・ループコイルと制御用ループコイルとを設置し，サドルディテクタで検出した揺動成分に反応させて制御ループにフィードバック信号を送る能動プラズマ制御実験も試みたい．

磁場揺動により磁気島が生成するような場合，揺動成分の波形位相に同期させて CD 波を入射することで磁気島の抑制が可能となる等アクチュエータの準備により制御も可能なので，磁気島 (成長) の検出と制御信号の出力が，現在のところ LHD 磁場揺動計測では，想定しやすいフィードバック信号出力の一つと考えられる．また，磁気島注意報・警報システムなどを簡単なコンピュータポートで出力する (光 or 音) ことも検討課題である．

参考文献

［1］ 藤井啓文. 新しい計算機環境 –高エネルギー物理実験から–. プラズマ・核融合学会誌, Vol. 73, No. 4, pp. 439–450, Apr. 1997.

［2］ 南茂夫. 科学計測のための波形データ処理. CQ 出版社, 東京, 1986.

［3］ 山口勝美, 森敏彦. 計測工学. 機械システム入門シリーズ, No. 2. 共立出版, 東京, 1993.

［4］ 高見豊. 作りながら学ぶ高速/マルチポート通信. インタフェース, No. 8 in ’95, p. 51. CQ 出版社, 東京, Aug. 1995.

［5］ 畔津明仁ほか. バス・インタフェース設計法. インタフェース, No. 1 in ’92, p. 119. CQ 出版社, 東京, Jan. 1992.

［6］ 鈴木忠, 佐藤健一, 宇都公博. 標準入出力インターフェースの利用法. インタフェース, No. 5 in ’92, p. 116. CQ 出版社, 東京, May 1992.

［7］ 徳田雅史ほか. 標準バスの定義と仕様 最適な VME システムの構築. インタフェース, No. 4 in ’93, p. 85. CQ 出版社, 東京, Apr. 1993.

［8］ 元吉信一. フィールドバス入門. 日刊工業新聞社, 東京, 2000.

［9］ 岡村廸夫. 標準ディジタル・バス (IEEE-488) とその応用. CQ 出版社, 東京, 1981.

［10］ トランジスタ技術編集部（編）. IBM PC と ISA バスの活用法. 別冊 トランジスタ技術 増刊. CQ 出版社, 東京, 1996.

［11］ KineticSystems Corporation, Illinois. *CAMAC Product Catalog*, 1993.

［12］ LeCroy, New York. *1994 Research Instrumentation Catalog*, 1994.

［13］ 熊原忠士ほか. CAMAC: データ処理用モジュール型計測装置の規格. JAERI-M 6003, JAERI, 1973.

［14］ IEEE Standards Board. *IEEE Standard FASTBUS Modular High-Speed Data Acquisition and Control System*. John Wiley & Sons, New York, 1984.

［15］ 岡村周善. Vme バス・システム完全マスタ. インタフェース, No. 2 in ’87, p. 191. CQ 出版社, 東京, Feb. 1987.

［16］ 下山智明, 城谷洋司. SUN システム管理. アスキー, 東京, 1991.
［17］ KineticSystems Corporation, Illinois. *VXIbus Product Catalog*, 1995.
［18］ 日本ヒューレット・パッカード, 東京. *Test & Measurement Catalog*, 1996.
［19］ National Instruments, Texas. *Instrumentation Reference and Catalogue*, 1995.
［20］ ソニー・テクトロニクス, 東京. 総合カタログ, 1996.
［21］ Tektronix, Colorado. *VXI Product Catalog*, 1994.
［22］ 横河電機 (株) Test&Measurement 事業部. PC ベース計測器 WE7000. http://www.yokogawa.co.jp/Measurement/Bu/WE7000/, 2003.
［23］ Interface 社, http://www.interface.co.jp/catalog/ctp/glossary.asp. CompactPCI 用語集, 2003.
［24］ R. Craig Klem (Fairchild Semiconductor). Signal Integrity, Bandwidth and Backplane Termination. In *Systems Chip Conference 99*, http://www.fairchildsemi.com/products/ backplane/syschip/flexbp.pdf, 1999.
［25］ 泉谷建司. Ethernet と FDDI. ソフト・リサーチ・センター, 東京, 1993.
［26］ 清水洋, 鈴木洋. *ATM-LAN*. ソフト・リサーチ・センター, 東京, 1995.
［27］ http://www-dg.LHD.nifs.ac.jp/FECnet/.
［28］ 西坂真人. “シリアル”の勝者は誰？ ZDNet/JAPAN, http://www.zdnet.co.jp/news/0207/17/nj00_cypress.html, 2002.
［29］ 長島章, 次田友宜ほか. JT-60 のデータ処理設備. 核融合研究, Vol. 59 Supplement, pp. 303–318, 1988.
［30］ REPUTE Group. Low Aspect Ratio Tokamak Experiments. REPUTE-1 Annual Report 1993, Univ. of Tokyo, 1994.
［31］ 浅川誠. 実験データ処理システムの例 — WT-3 のデータ処理 —. プラズマ・核融合学会誌, Vol. 73, No. 2, pp. 162–167, Feb. 1997.
［32］ FORCE COMPUTERS. *SPARC CPU-2CE Technical Reference Manual*, 1982.
［33］ 中西秀哉ほか. 光 ethernet を用いたネットワーク化データ収集システム. プラズマ・核融合学会 第 10 回秋期講演会予稿集, p. 43, 1993.
［34］ Yasuo Takeuchi, et al. Development of Data Acquisition System using RISC/UNIX Workstation. KEK preprint 92-80, TIT-HEP-92-05, KEK, TITECH, 1992.
［35］ ftp://onlnews.kek.jp/pub/kek/camac/.
［36］ 森川惇二. 実験データ処理システムの例 — REPUTE-1 のデータ処理 —. プラズマ・核融合学会誌, Vol. 73, No. 2, pp. 168–173, Feb. 1997.
［37］ REPUTE Group. Data Acquisition Networking System with Optical Ethernet for

REPUTE-1. REPUTE-1 Annual Report 1992, Univ. of Tokyo, 1993.

［38］ Toshiaki Matsuda, Shigeru Hidekuma, Mamoru Kojima, and M. Suzuki. New data acquisition system for the JFT-2M tokamak. *Rev. Sci. Instrum.*, Vol. 66, pp. 515–517, 1995.

［39］ T.W. Fredian, J.A. Stillerman, and M. Greenwald. Data acquisition system for Alcator C-Mod. *Rev. Sci. Inst.*, Vol. 68, No. 1, pp. 935–938, Jan. 1997.

［40］ J.A. Stillerman, T.W. Fredian, K.A. K.A. Klare, and G. Manduchi. MDSplus data acquisition system. *Rev. Sci. Inst.*, Vol. 68, No. 1, pp. 939–942, Jan. 1997.

［41］ Y. Neyatani, et al. Feedback control of neutron emission rate in JT-60U. *Fusion Eng. Design*, Vol. 36, pp. 429–433, 1997.

［42］ K. Kurihara, et al. Plasma Real-Time Control System for Advanced Tokamak Operation Scenarios in JT-60. *IEEE Trans. Nucl. Sci.*, Vol. 47, pp. 205–209, 2000.

［43］ I. Yonekawa, Y. Kawamata, T. Totsuka, H. Akasaka, M. Sueoka, K. Kurihara, T. Kimura, and the JT-60U Team. JT-60U CONTROL SYSTEM. *Fusion Sci. Tech.*, Vol. 42, No. 9-11, pp. 525–529, 2002.

［44］ 坂田信也ほか. 新データ処理設備実時間処理計算機 (RTP) の開発、および高性能化. JAERI-Tech 2000-043, JAERI, 2000.

［45］ 窪田敏之. 組み込み OS の今を探る. In *Software Design*, No. 7, 組み込み OS 新時代突入, pp. 128–132. 技術評論社, 東京, July 2003.

［46］ 上瀧恵里子, 伊藤智之. 実験データ処理システムの例 — TRIAM-1M のデータ処理—. プラズマ・核融合学会誌, Vol. 73, No. 3, pp. 330–334, Mar. 1997.

［47］ *MPI*. http://directory.google.com/Top/Computers/Parallel_Computing/Programming/Libraries/MPI/, 2003.

［48］ PC クラスタコンソーシアム, http://www.pccluster.org/score/dist/score/. SCore Cluster System Software 5.4 ドキュメント, 2003.

［49］ Ian Foster and Carl Kesselman. *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann Pub., 1998.

［50］ TOP500 Supercomputer Sites. http://www.top500.org/list/2002/11/, 2002.

［51］ Atsuo Iiyoshi and Kozo Yamazaki. The next large helical devices. *Phys. Plasmas*, Vol. 2, No. 6, pp. 2349–2356, Jun. 1995.

［52］ T. Matsuda, T. Totsuka, T. Tsugita, T. Oshima, S. Sakata, M. Sato, and K. Iwasaki. DATA PROCESSING AND ANALYSIS SYSTEMS FOR JT-60U. *Fusion Sci. Tech.*, Vol. 42, No. 9-11, pp. 512–520, 2002.

［53］ Jean Bacon. 並行分散システム. トッパン, 東京, 1996.
［54］ Edward Yourdon. オブジェクト指向システム設計. トッパン, 東京, 1995.
［55］ Phil Sully. オブジェクト指向モデリング. 日経 BP 出版センター, 東京, 1995.
［56］ 河込和宏, 中村秀男, 大野邦夫, 飯島正. 分散オブジェクトコンピューティング. 共立出版, 東京, 1999.
［57］ Dynamics Research Corp./日本エスキューブ, 東京. Parallel Processing/Network Communications with HARNESS, ユーザーズ・ガイド, 1993.
［58］ 小野沢博文. CORBA 完全解説 基礎編/応用編. ソフト・リサーチ・センター, 東京, 1999.
［59］ 日本サン・マイクロシステムズ（編）. Java プログラミング Java RMI. サイエンス, 1998.
［60］ Randy Otte, Paul Patrick, and Mark Roy. 分散オブジェクト指向 CORBA 分散プログラミングから大規模分散システム構築まで. プレンティス・ホール出版, 東京, 1996.
［61］ JUAS オープンシステム研究部会・基幹システムの CSS 化の課題 WG（編）. 基幹業務のクライアント／サーバーシステム化への移行方法 (USC データベース). (社) 日本情報システム・ユーザー協会, http://www.juas.or.jp/usc/report/511-i.htm, 1994.
［62］ 木暮仁. EUC/CSS を成功させるには －新パラダイム時代での情報システム運営－. 日科技連出版社, 東京, 1996.
［63］ 小嶋護, 秀熊茂, 居田克己, 佐藤浩之助. プラズマ実験におけるデータ処理 — 実験データ処理システムの例 — JIPP T-IIU のデータ処理. プラズマ・核融合学会誌, Vol. 73, No. 1, pp. 93–99, Jan. 1997.
［64］ 日本ビジュアルニューメリックス, http://www.vnij.com/products/wave/. PV-WAVE バージョン 7.5, 2003.
［65］ Research Systems, Inc., http://www.rsinc.com/idl/. *IDL Software*, 2003.
［66］ アダムネット, http://www.adamnet.co.jp/scs/products/idl/index.html. データ解析ビジュアライゼーション ソフトウェア IDL, 2003.
［67］ 日本ラショナルソフトウェア. 日本語版 UML (Unified Modeling Language) ドキュメント ver1.1. http://www.rational.co.jp/uml/.
［68］ 菅谷誠一. SCSI-2 詳細解説. CQ 出版社, 東京, 1994.
［69］ 鈴木博. C 言語と SCSI 制御. 工学図書, 東京, 1994.
［70］ Helen Custer. *Inside Windows NT*. アスキー出版局, 東京, 1993.
［71］ Helen Custer, Jeffrey Richter, Kraig Brockschmidt, et al. Win32 プログラミングワークショップ. アスキー出版局, 東京, 1994.

［72］ Robert L. Kruse, Bruce P. Leung, and Clovis L. Tondo. C によるデータ構造とプログラム設計. トッパン, 東京, 1994.
［73］ Setrag Khoshafian. オブジェクト指向データベース. 共立出版, 東京, 1995.
［74］ オージス総研/三井物産. Objectivity テクニカルオーバービュー Ver.4, 1996.
［75］ ウインドリバーシステムズ, 東京. VxWorks チュートリアル, 1991.
［76］ ウインドリバーシステムズ, 横浜. VxWorks プログラマーズガイド, 1994.
［77］ John Bloomer. RPC プログラミング. アスキー出版局, 東京, 1995.
［78］ G. Raupp and H. Richter. The timing system for the ASDEX Upgrade experiment control. *IEEE Trans. on Nucl. Sci.*, Vol. NS-39, p. 198, 1992.
［79］ 宇田川佳久. オブジェクト指向データベース入門. ソフト・リサーチ・センター, 東京, 1992.
［80］ Ian Graham. オブジェクト指向概論. トッパン, 東京, 第 2 版, 1996.
［81］ 日本 IBM, http://www-6.ibm.com/jp/storage/basic/word/hsm.htm. 階層型ストレージ管理, ストレージ講座 – ストレージ略語/用語, 2003.
［82］ 山口作太郎, 庄司主, 三戸利行, 山崎耕造, 刈谷丈治, 奥村晴彦, 江本雅彦, 寺町康昌, 大須賀闥雄. 実験データ処理システムの例 — 超伝導コイル実験監視システム —. プラズマ核融合学会誌, Vol. 73, No. 3, pp. 335–342, Mar. 1997.
［83］ 国立情報学研究所. スーパー SINET とは. http://www.sinet.ad.jp/s_sinet/index.html.