

JAERI-M
94-050

磁気流体平衡高速処理装置 FAME の概要

1994年3月

坂田 信也・萩野谷裕文・鶴岡 貞哉・渡辺 秀人^{*}
青柳 哲雄・齊藤 直之・原田 裕夫・谷 啓二

JAERI-M レポートは、日本原子力研究所が不定期に公刊している研究報告書です。
入手の問合せは、日本原子力研究所技術情報部情報資料課（〒319-11茨城県那珂郡東海村）
あて、お申しこしください。なお、このほかに財団法人原子力弘済会資料センター
(〒319-11 茨城県那珂郡東海村日本原子力研究所内) で複写による実費頒布をおこなって
おります。

JAERI-M reports are issued irregularly.

Inquiries about availability of the reports should be addressed to Information Division,
Department of Technical Information, Japan Atomic Energy Research Institute, Tokai-mura,
Naka-gun, Ibaraki-ken 319-11, Japan.

© Japan Atomic Energy Research Institute, 1994

編集兼発行 日本原子力研究所
印 刷 株原子力資料サービス

磁気流体平衡高速処理装置 FAME の概要

日本原子力研究所那珂研究所炉心プラズマ研究部

坂田 信也・萩野谷裕文・鶴岡 貞哉・渡辺 秀人*

青柳 哲雄・斎藤 直之・原田 裕夫・谷 啓二

(1994年2月14日受理)

JT-60 の非定常プラズマ輸送解析に必要な 100 点を越える時系列磁気流体平衡の計算を約 20 分のショット間に処理するため、磁気流体平衡高速処理装置 FAME が開発された。

FAME は、多段スイッチで接続された 20 個のマイクロプロセッサーから成る MIMD 型の最大理論性能 250 MFLOPS の小規模並列計算機である。そのソフトウェアとして磁気流体平衡解析コード SELENE 及びその入力データ作成コード FBI が並列処理を念頭にチューニングされた。その結果、既存の大型汎用計算機 FACOM M780-10S に比べ 7 倍の計算速度が得られた。

本報告書は、この FAME システムについて、そのハードウェア、ソフトウェア及び周辺装置の概要をまとめたものである。

Outline of Fast Analyzer for MHD Equilibrium "FAME"

Shinya SAKATA, Hirofumi HAGINOYA, Takuya TSURUOKA
Hideto WATANABE*, Tetsuo AOYAGI, Naoyuki SAITO
Hiroo HARADA and Keiji TANI

Department of Fusion Plasma Research
Naka Fusion Research Establishment
Japan Atomic Energy Research Institute
Naka-machi, Naka-gun, Ibaraki-ken

(Received February 14, 1994)

The FAME (Fast Analyzer for Magnetohydrodynamic (MHD) Equilibrium) system has been developed in order to provide more than 100 MHD equilibria in time series which are enough for the non-stationary analysis of the experimental data of JT-60 within about 20 minutes shot interval.

The FAME is an MIMD type small scale parallel computer with 20 microprocessors which are connected by a multi-stage switching system. The maximum theoretical speed is 250 MFLOPS.

For the software system of FAME, MHD equilibrium analysis code SELENE and its input data production code FBI are tuned up taking the parallel processing into consideration. Consequently, the computational performance of the FAME system becomes more than 7 times faster than the existing general purpose computer FACOM M780-10s.

This report summarizes the outline of the FAME system including hardware, soft-ware and peripheral equipments.

Keywords: Parallel Computer, MHD Equilibrium

* Nuclear Energy Data Center

目 次

1. FAME システムの概要	1
2. ハードウェアの概要	3
2.1 並列計算機本体	3
2.1.1 ハードウェアの構成	3
2.1.2 CPU ボードの構成	4
2.1.3 スイッチングユニットの構成	6
2.2 並列計算機用フロントエンドワークステーション (EWS)	6
2.3 ファイルサーバ	6
2.4 動画処理装置	7
2.5 コマ録り装置	7
3. ソフトウェアの概要	9
3.1 MHD 平衡解析ソフトウェア	9
3.1.1 FBI コード	9
3.1.2 MHD 平衡コード	11
3.1.3 MHD 平衡コードの並列化と最適化	11
3.1.3.1 ジョブの制御方法	12
3.1.3.2 入出力の C 言語化	13
3.1.3.3 出力データの圧縮	14
3.1.3.4 入力データ読み込みの改善	14
3.1.3.5 WRITE 文の制御	14
3.1.3.6 画像データ処理	14
3.1.4 並列化の結果	15
3.2 動画処理ソフトウェア	17
3.2.1 画像データ	17
3.2.2 動画処理	17
3.2.3 コマ録り処理	20
4. あとがき	21
謝 辞	21
付 錄 A	23
付 錄 B	28

Contents

1.	Outline of FAME System	1
2.	Outline of Hardware	3
2.1	Parallel Processor	3
2.1.1	Hardware Composition	3
2.1.2	CPU Board Composition	4
2.1.3	Switching Unit Composition	6
2.2	Frontend Workstation	6
2.3	File Server	6
2.4	Moving Image Processor	7
2.5	Video Processor	7
3.	Outline of Software	9
3.1	Software for MHD Equilibrium Analysis	9
3.1.1	FBI Code	9
3.1.2	MHD Equilibrium Code	11
3.1.3	Parallel Programming and Optimization of MHD Equilibrium Code	11
3.1.3.1	Job Control	12
3.1.3.2	C Programming of I/O Routines	13
3.1.3.3	Output Data Compression	14
3.1.3.4	Improvement of Input-data Read	14
3.1.3.5	Control of WRITE Statement	14
3.1.3.6	Image Data Processing	14
3.1.4	Results of Parallel Programming	15
3.2	Software for Moving Image	17
3.2.1	Image Data	17
3.2.2	Moving Image Processing	17
3.2.3	Video Processing	20
4.	Conclusions	21
	Acknowledgements	21
	Appendix A	23
	Appendix B	28

1. FAME システムの概要

トカマク実験における計測データの解析においては、磁気流体平衡計算の結果をベースに、線積分量として得られている計測情報局所物理量に写像変換することがまず行われる。J T - 6 0においては、電磁気測定データを用いて磁気流体平衡計算を行い、それをもとに写像変換情報を出力する最外殻磁気面同定コード（以下では、F B I (First Boundary Identification) コードと称す）が整備され利用されている。

計測データの解析は、トカマクコードを中心とする。プラズマ中の熱、粒子の輸送解析を中心とするが、実験の進展、それに伴うコード整備の進展にあわせてデータ解析も高度化しつつあり、従来の定常計算に加え今後非定常解析にその比重が移りつつある。F B I コードは、1 タイムスライス毎に1回実行する必要があるが、非定常計算の場合1つのショットのデータ解析に数十回から百数十回のF B I コードをショット間に実行させる必要がある。その計算処理量は、現有のデータ処理大型計算機（スカラー計算機F A C O M M 7 8 0 - 1 0）の能力をはるかに超えるものである。また、ベクトル計算機（F A C O M V P 2 6 0 0）を用いるには、計算機の運用形態が異なるため実験に即したデータ処理は不可能である。

前述の、数十回、百数十回に及ぶタイムスライス毎のF B I コードの計算は、各々独立に実行可能である。これは、マイクロプロセッサをネットワークで接続し、多数のプロセッサを同時に並列実行させる並列演算技術の応用により、極めて高速に処理することが可能であることを意味している。並列計算機F A M E は、この点に着目し、F B I コード及びM H D 平衡コードを高速に処理するために開発されたものである。F A M E の計算結果は、M H D 平衡データベースとしてディスク装置に格納され、ネットワークを通じてユーザに提供される。同時に、計算結果は動画処理され、光ディスク上に保存される。F A M E は、図1.1に示す様に、これ等の機能を実現するための2 4 G B (ギガバイト) の専用ハードディスク装置、ビデオコマ録り装置、光ディスク装置などの動画処理装置がシステムとして一体化されている。

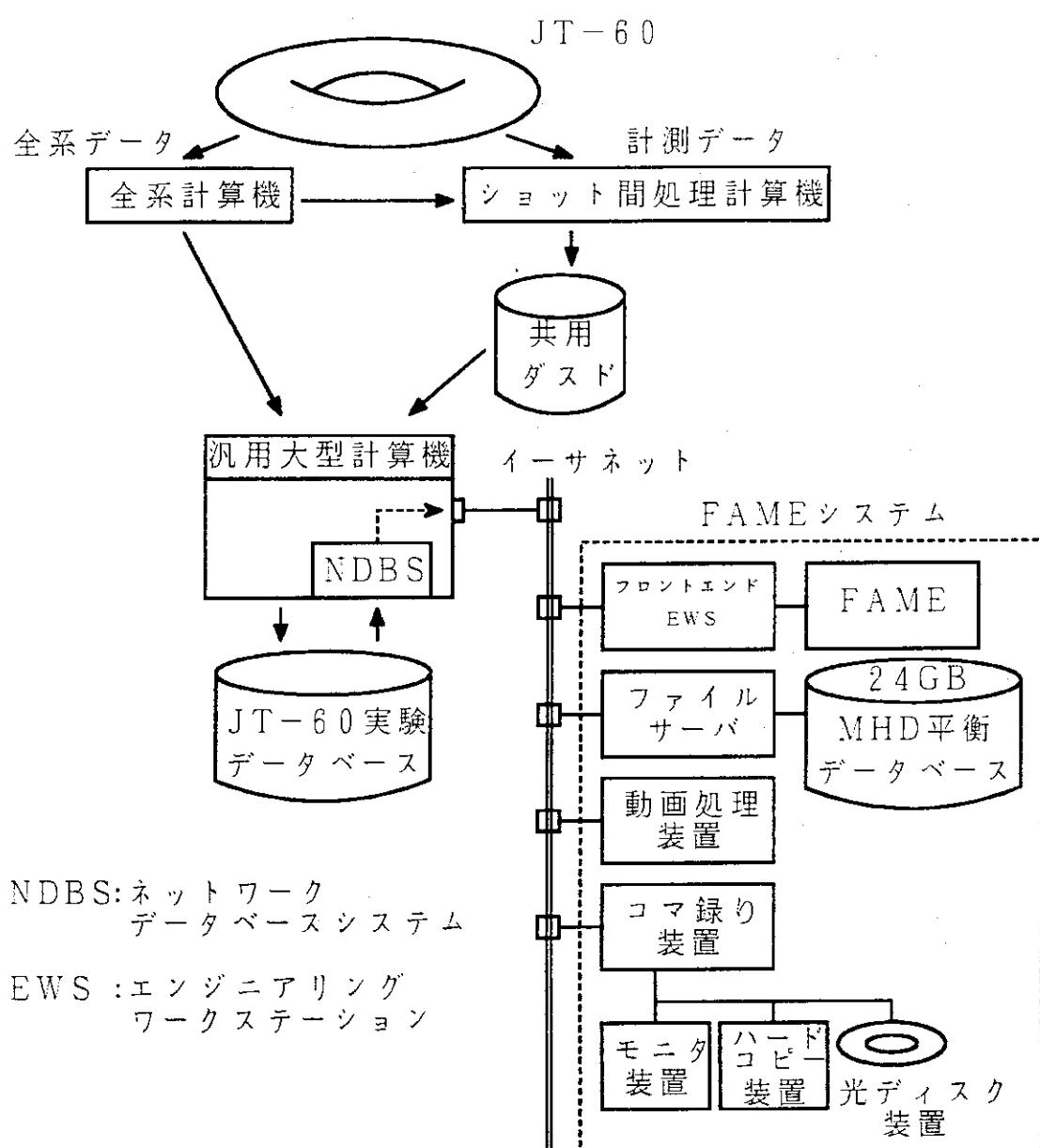


図1.1 FAMEシステムに関連するフロー

2. ハードウェアの概要

FAMEシステムは第1章に述べた様に、並列計算機本体、並列計算機用フロントエンド計算機、ファイルサーバ、動画処理装置及び、コマ録り装置により構成されている。以下に各々についてのハードウェア概要を記す。

2.1 並列計算機本体

以下の基本機能を満足する小規模並列計算機である。

(1) アーキテクチャタイプ

MIMD (Multi Instruction stream / Multi Data stream)型

(2) 理論性能

250MFLOPS (100万浮動少数点演算/秒)

(3) プロセッサエレメント数(並列数)

20

(4) ネットワーク形態

多段接続方式

(5) 1プロセッサエレメントあたりのメモリ容量

64MB (メガバイト)

以下にハードウェア構成、CPUボード、スイッチングユニット構成の概要をまとめます。

2.1.1 ハードウェアの構成

並列計算機本体では、 4×4 のゲートアレイ（スイッチングユニット）を構成要素とする多段接続網を用いてプロセッサを結合する。図2.1に並列計算機本体の外観を示す。

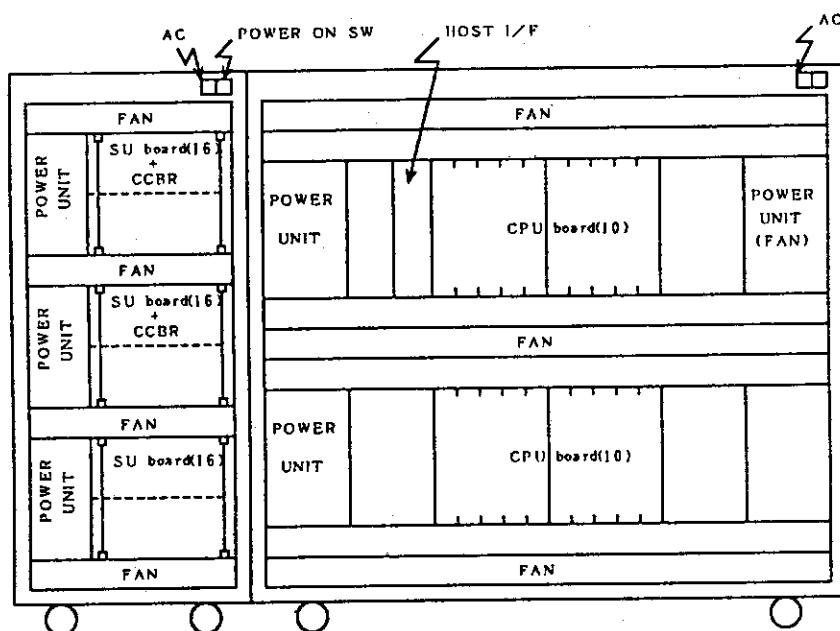


図2.1 並列計算機本体の外観

並列計算機本体は、ネットワーク部とプロセッサ部から構成される。ネットワーク部には、スイッチングユニットボードが48枚(16枚×3段)装備され、さらにプロセッサ部には、C P U ボードが20枚(10枚×2段)装備されている。図2.1に示した様に上段の向かって左から、プロセッサの物理番号0, 1, 2, 8, 9、下段の向かって左から、プロセッサの物理番号10, 11, 18, 19となっている。プログラムの実行時には、0から連続の値をつけた番号が論理番号として与えられるために、プログラマへの影響はさほどない。しかし、ハードウェアエラーが生じた場合などは、この番号が用いられる。並列計算機用フロントエンド計算機との通信は、任意のプロセッサと通信可能であるが、通常は論理番号0のプロセッサがフロントエンド計算機と通信することが多い。また、プロセッサ間の同期なども全てプロセッサ0が主導権を握って行うため、場合によっては、プロセッサ0をマスタプロセッサ、他のプロセッサをスレーブプロセッサと呼ぶ。

2.1.2 C P U ボードの構成

並列計算機の基本部分となるC P U ボードは、プロセッサ部、メモリ部からなる。図2.2にプロセッサ部の構成を示す。

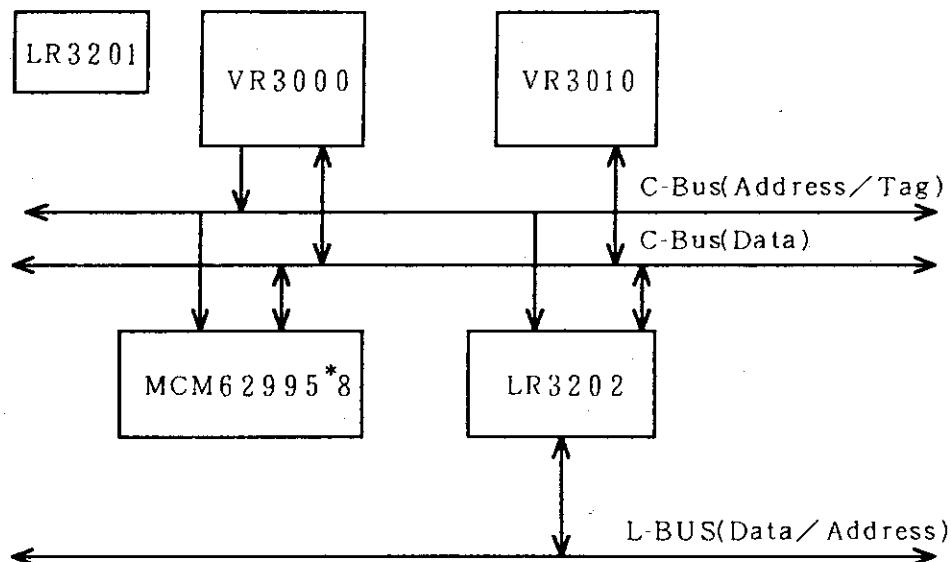


図2.2 プロセッサ部の構成

プロセッサ部では、以下のハードウェアがC-BUSで結合されている。

(1) VR3000 (C P U)

32 bit RISC (Reduced Instruction Set Computer)型マイクロプロセッサであり、並列計算機におけるプログラムの実行を行う中枢部である。

システムクロック 25MHz (メガヘルツ)

理論性能 25 MIPS (100万命令/秒)

(2) V R 3 0 1 0 (F P U)

実行プログラム中の浮動少數点命令を実行するコ・プロセッサであり、V R 3 0 0 0 マイクロプロセッサからの指示に基づき、浮動少數点の演算を高速に行う。

システムクロック 2 5 M H z (メガヘルツ)

理論性能 1 2 . 5 M F L O P S (1 0 0 万浮動少數点演算／秒)
单精度、倍精度共

(3) M C M 6 2 9 9 5 (キャッシュメモリ)

高速プロセッサ V R 3 0 0 0 の性能を發揮するため、プログラムのメモリアクセスの局所性を利用して、高速 S R A M をキャッシュメモリとして 8 個配置している。

命令用 6 4 K B (キロバイト)

データ用 6 4 K B (キロバイト)

(4) L R 3 2 0 1

電源投入時、及びリセット時にプロセッサを初期化して、プロセッサ動作条件を決定する。

(5) L R 3 2 0 2

プロセッサバスである C-B U S と、メインメモリ、ネットワーク、I/O デバイスなどが接続される L-B U S 間でデータ転送を行う場合に、両者のバスを制御し、データを変換する。

また、メモリ部は、以下のハードウェアより構成されている。

(1) L R 3 2 0 3

L-B U S と D R A M を接続するための D R A M コントローラであり、L-B U S のアクセス要求が D R A M であった場合に、D R A M に対して読み出し、書き込み要求を行う。

また、D R A M のリフレッシュも行う。

(2) L R 3 2 0 4

D R A M と L R 3 2 0 3 を接続するためのバッファであり、複数の D R A M を接続するために用いる。

(3) D R A M μ P D 4 2 4 0 0 0

並列計算機に搭載されている主記憶（メインメモリ）であり、各プロセッシングエレメント当たり、6 4 M B (メガバイト) 使用している。

高速 D y n a m i c D R A M のモジュール (S I M M) を使用している。

2.1.3 スイッチングユニットの構成

並列計算機のプロセッサ間ネットワークは、 4×4 のスイッチングユニットを構成要素とする多段接続網である。

スイッチングユニットは、4入力4出力のクロスバススイッチを実現するゲートアレイであり、並列計算機のネットワークボードには、信号線数の関係から1枚につき1個のスイッチングユニットが搭載されている。基本的には、入力側に FIFO がありデータの行き先となるポートが空いていれば、そのポートにデータを供給する。

スイッチングユニットのハードウェアブロック図を図2.3に示す。

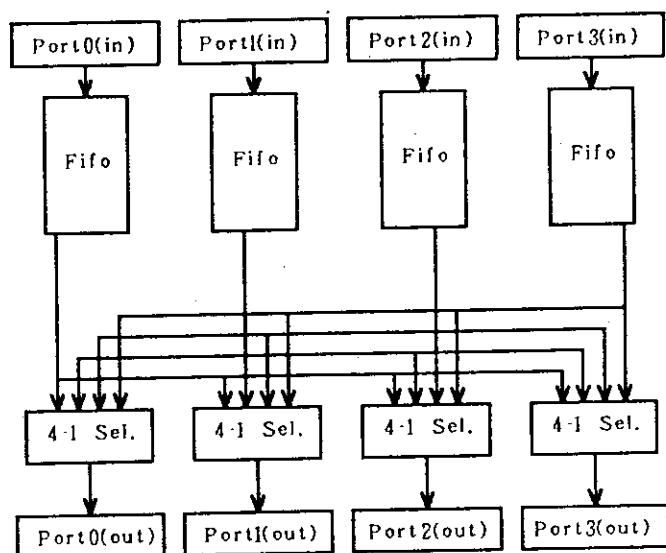


図2.3 スイッチングユニットハードウェアブロック図

2.2 並列計算機用フロントエンドワークステーション（EWS）

EWS 4800/230（以下では、フロントエンドEWSと称す）は、FAMEシステムにおいて並列計算機の実行管理、並列プログラムの作成、並列計算機へのファイル環境の提供の役割を果たす。フロントエンドEWS本体には、中央処理装置、64MB（メガバイト）の主記憶部、5インチ662MB（メガバイト）ハードディスク、3.5インチ1MB（メガバイト）フロッピディスク、RS-232Cインターフェイス、プリンタインターフェイス、BRANCH4680/4780インターフェイス、SCSIインターフェイスが実装されている。並列計算機本体とは、DRII/W準拠のDMAチャネルで結合されている。

2.3 ファイルサーバ

JT-60の大型計算機上のMHD平衡データベースは、約1000ショット分のデータがディスク上に存在しユーザから自由にアクセスできる形になっている。FAMEシステムとしては、実験データ解析上、これ等とほぼ同等のショット分のMHD平衡データがデータベースとして利用可能な状態にある必要がある。そのためには、

20 MB (メガバイト) / ショット × 1200 ショット = 24 GB (ギガバイト) の容量の磁気ディスクが必要となる。また、実験期間中には多数のクライアントからの大量のデータアクセスがあるため、サーバには極めて高い性能が求められる。そこで、UP4800 (以下では、ファイルサーバと称す) は、以下の機能を有する磁気ディスクシステムを採用している。

プロセッサ処理装置	110 MIPS (100万命令/秒)
主記憶容量	64 MB (メガバイト)
VMEバス	10 スロット
基本インターフェイス	
(1) LAN ネットワーク	(TCP/IP プロトコル)
(2) SCSI インターフェイス	
(3) パラレルインターフェイス (セントロニクス準拠)	
(4) シリアルインターフェイス (RS232C)	
入力装置	
(1) 3.5インチフロッピーディスク装置	
(2) カートリッジテープ装置 (1/4インチ)	
(3) 磁気ディスク装置	24 GB (ギガバイト)

中央処理装置は、米国MIPS社の高性能RISC (Reduced Instruction Set Computer) チップであるR4000を採用するとともに二次キャッシュを実装することにより高速度の演算処理能力を実現している。主記憶部は16 Mbit (メガビット) DRAMを採用し、128 MB (メガバイト) の増設単位で主記憶を大容量に拡張できる。

2.4 動画処理装置

実験計測装置の整備の進展に伴い、あるいは、それ等実験データの解析ツールの進展に伴い、より大量のデータをモニタする必要性が高まりつつある。大量のデータをモニタする最も効率的な方法は、可視化である。SUN SPARK Station 10 (以下では、動画処理装置と称す) は、計測データあるいは、それ等の高レベル処理データを高速に動画処理することを目的とした装置である。動画処理装置のハードウェア構成は、プロセッサ性能がクロックレート35 MHz (メガヘルツ) 、メインメモリ64 MB (メガバイト) 、また、Ethernet、SCSI、シリアル/パラレル、ISDN、SBUSのそれぞれのインターフェイスを装備している。記憶装置は、3.5インチフロッピーディスク、400 MB (メガバイト) の内蔵ハードディスク、1.3 GB (ギガバイト) の外部記憶装置ハードディスク、8mm磁気テープ装置及び、CD-ROMドライブ装置により構成されている。

2.5 コマ録り装置

NVS2000 (以下では、コマ録り装置と称す) は、動画処理装置で作成したCG画像

データをネットワークを介してF T P (File Transfer Protocol)で受け、ビデオ信号をビデオテープレコーダに録画するための制御をする機能と、ビデオディスクレコーダを用いて光ディスクに保存する機能を有している。コマ録り装置は、ビデオディスクレコーダ、カラー モニタ、及び、カラービデオプリンタから構成される。ビデオディスクレコーダは、録画／再生性能が動画で24分／面、静止画で43500フレーム／面、アクセス時間が0.5秒である。また、ビデオ記号方式は、N T S C カラー方式でR S - 4 2 2 、R S - 2 3 2 C のインターフェイスを装備している。

3. ソフトウェアの概要

J T - 6 0においては、「実験データベース」と称されるものは、

1. 全系データベース
2. 計測データベース

の2つのデータベースから成っている。前者は、主として、J T - 6 0のシステム制御に係わるデータを収納し、後者は、プラズマ計測装置から出力される信号を変換し、物理データとしてデータベース化したものである。全系データは、図1.1に示す様に全系計算機でデータ収集された後に大型計算機F E Pに送られデータベース化される。一方、計測データは、計測データショット間処理装置I S Pで収集され、同時にI S P / F E P共用ダスド上にデータベース化される。この共用ダスド上のデータベースは、一定の期間後にF E Pにおいて実験データベース専用ダスドに全系データベースと共に格納される。

通常、J T - 6 0における1ショットのプラズマ放電時間は10数秒であり、これが約20分間隔でくり返される。ショット放電後直ちにデータ収集、データ処理が開始されるが、図3.1に示す様に、全系、計測の各々のデータベース処理が完了するまでに約9分の時間が必要である。一方、全系システムにおける、次ショットの準備は、放電終了後約6分から既に始まっている。F A M Eは、実験データベースのデータ取得からその処理が開始される。従って、F A M Eの結果を次ショットに反映させることは、もともと不可能である。そこで、F A M Eでは、その処理結果を次々回の放電に反映させられる様にソフトシステムの整備を進めている。

現在、F A M Eシステムにインストールされているアプリケーション・ソフトウェアは、

- 1) M H D 平衡解析ソフトウェア
- 2) 動画処理ソフトウェア

に大別される。前者はF A M E本体で動作し、後者は、周辺装置の1つである動画処理専用E W Sで動作する。これ等2つのソフトは、実験データベース作成完了の信号をI S Pよりソケット通信で受けることにより起動し、オンライン的に全ての計算が逐次に処理される。

以下、上記2つのソフトについてその概要をまとめる。

3.1 M H D 平衡解析ソフトウェア

F A M Eシステムの計算処理ソフトウェアには、フロントエンドE W Sで動作するF B Iコードと、フロントエンドE W Sでジョブ管理し、並列計算機で動作するM H D平衡コード「S E L E N E」がある。

以下では、二つの計算コードの概要と並列化、並びにそれに関連した最適化について説明する。

3.1.1 F B I コード

F B Iコードでは、実験放電で測定したポロイダル磁場、磁束、コイル電流を基に、プラズマ電流を6本のフィラメント電流で仮定し、プラズマの最外殻磁気面を最小自乗法を用いて求めている。また、その結果からプラズマの形状を示す諸パラメタ($R_p, \alpha_p, \kappa, etc.$)や

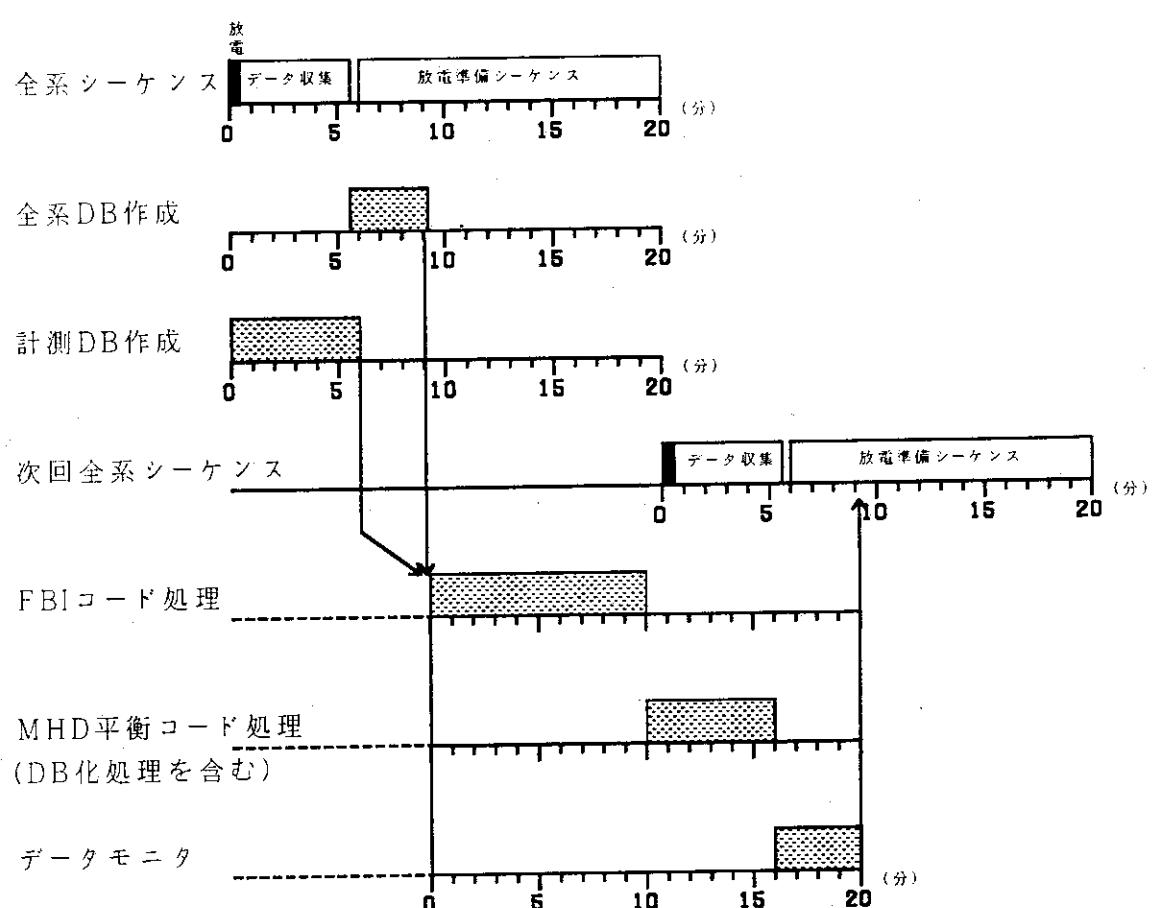


図3.1 JT-60運転シーケンスとMHD平衡データベース作成タイミング

安定性を示すパラメータ(n_{index} , etc.)、さらに、反磁性測定を基にプラズマの蓄積エネルギー(β_p , I_i)を求め、最終的には、MHD平衡コードSELENE用の入力データを出力するものである。

FBIには次にあげる2つのバージョンが存在する。

- (1) 準定常プラズマ電流解析バージョン
- (2) 非定常プラズマ電流解析バージョン

トカマク放電においては、何らかの磁気流体的不安定性が原因ではないかと考えられる“ディスラプション”と呼ばれる急激なプラズマ電流の消滅が観測されることがある。このディスラプション時や、急激なプラズマ電流のランプアップ、ランプダウン時には、真空容器導体上に渦電流が流れる。上記(2)は、FBIのSELENE入力データ生成処理において、磁気プローブ信号に対して、この渦電流の影響を考慮したバージョンであり、(1)は考慮しないバージョンである。本FAMEシステムには、上記いずれのバージョンも用意されフロントエンドEWSで簡単にスイッチできる様になっている。

FAME版のFBIコードでは、後述するMHD平衡コードの計算を並列計算機によって処理することから、大型計算機版のFBIコードのように一時刻のみの計算だけでなく、数十から数百の時刻について、まとめて計算しておく必要がある。そのため、各時刻について、毎回データベースを初期設定及び終了設定したり、少量の取得データを転送していくは効率が悪いので、必要となるデータ全部を一回にまとめて検索・転送できる様になっている。

3.1.2 MHD平衡コード

MHD平衡コード「SELENE」では、測定したポロイダル磁束を境界条件として、Grad-Shafranov方程式を直接差分法であるBuneman法を用いて磁気流体平衡配位を求めていく。

入力データは、プラズマ電流、コイル電流の実測値と、FBIコードで計算したポロイダル磁場で規格したプラズマ平均圧力 β_p と内部インダクタンス I_i としている。ただし、プラズマ圧力の非等方性などにより $I_i < 0.8$ となる場合には、 β_p を $\beta_p + (I_i - 0.8)/2$, I_i を0.8と置き換えて計算し、得られた磁気面関数から、磁気軸位置や正確な垂直磁場の曲率 n_{index} 等を求め、実座標から磁気面座標へ変換を行う。

3.1.3 MHD平衡コードの並列化と最適化

一般に、JT-60に代表される様な大型トカマク装置においては、ある磁気流体平衡から別な平衡に遷移する時間は放電時間に比べ十分短いため、磁気流体平衡は、各々の時刻において独立な事象と考えることができる。すなわち、MHD平衡の時間発展を解く場合各時刻の平衡計算は完全に並列に処理することが可能である。

この特徴を利用し、FAMEではSPMD (Single Program Multiple Data stream) モデルを採用した。このモデルでは、全プロセッサに同じプログラムがロードされ、各プロセッサは、自分の論理IDを基に、自分の担当範囲を決定し計算を行う。

並列数の大きいシステムの場合、上述の各時間の処理に関する粗い粒度の並列化に加え、さ

らに粒度の細かいプログラミングを行うことにより、1時刻の平衡計算を複数のプロセッサで並列処理することも考えられる。しかし、この場合、プログラムの大幅な修正が不可避である。JT-60においては、MHD平衡計算コードSELENEは汎用大型計算機で開発されて来ており、今後もこの大型計算機版SELENEを標準バージョンとしてソフトが改良され管理していく。このことから、FAMEにおいても大型計算機版SELENEとの互換性を確保する必要があり、上述の細粒度並列化は採用しない。但し、3.1.3.2以下で述べる様な最適化を行うことにより、処理時間の短縮を図った。

3.1.3.1 ジョブの制御方法

FAMEで動作するMHD平衡コードの構成と処理の流れを以下の図3.2に示す。

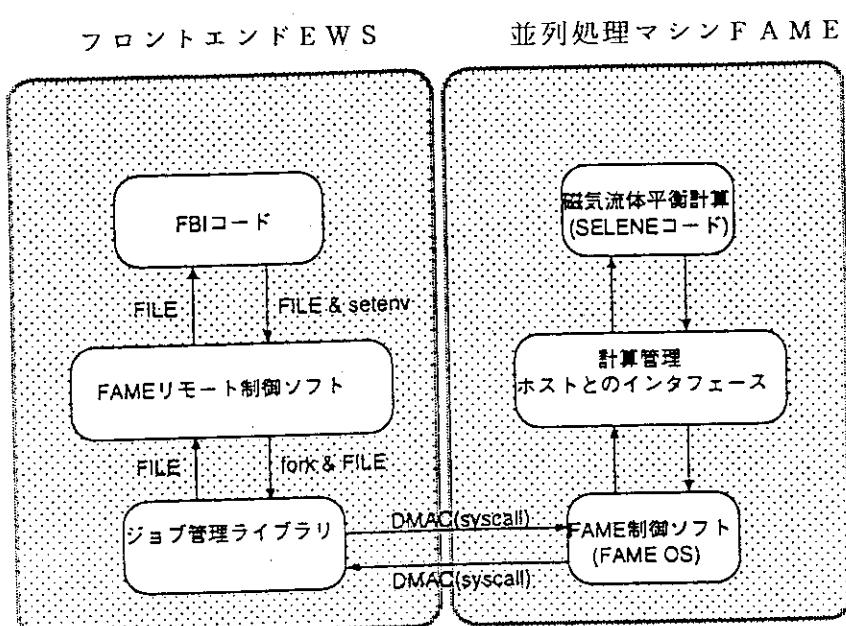


図 3.2 並列計算用MHD平衡コードの構成と処理の流れ

このように、FAMEシステムで動作するMHD平衡コードでは、フロントエンドEWSでジョブ管理を行い、その指示に基づき、バックエンドで並列計算機が動作する仕組みになっている。フロントエンドEWSと並列計算機との間の制御・交信は、基本的にフロントエンドEWSのディスク上に作成されるファイルを介して行われる。そして、フロントエンドEWS上で、並列計算機リモート制御ソフトを実行すると、自動的に並列計算機上で、MHD平衡コードが並列に動作するので、並列計算機の使用を意識する必要がない。

図3.1 の制御シーケンスを以下に示す。

(1) FAMEの初期化

並列計算機リモート制御ソフトを起動すると、まずFAMEの各プロセッサに対して専用のオペレーティングシステムの常駐が行われる。その後、初期化ルーチンが実行さ

れ、MHD 平衡コードのロードモジュールの常駐が行われる。

(2) ジョブの待機

フロントエンド EWS からジョブが投入されない限り、並列計算機の各プロセッサは待ち状態を保つ。

(3) FAMEへのジョブの投入

フロントエンド EWS では、並列計算機の各プロセッサに対して、入力データファイルの名前が書き込まれたジョブ制御用のファイルをディスク上に作成する。そして、この時点では、既に並列計算機の各プロセッサに常駐している MHD 平衡コードは、このファイルの存在を 5 秒毎に確認しているので、制御ファイルの存在が判明した時点でその内容を読み取って、指示された入力データファイルによる MHD 平衡計算を開始する。

(4) 計算時間の計測

フロントエンド EWS では、並列計算機の各プロセッサ上で動作している MHD 平衡コードの計算時間を計測している。

(5) ジョブの終了通知

並列計算機の各プロセッサは計算が終了すると、その結果をフロントエンド EWS のディスクに書き込み、ジョブの終了をフロントエンド EWS に通知するための制御ファイルを作成する。フロントエンド EWS 側で動作している並列計算機リモート制御ソフトは、MHD 平衡コードが動作している間、このファイルの存在を定期的に確認している。そして、その存在が判明した時点で、引き続いて計算するための入力データが存在する場合には、(3) の処理を実行し、入力データが存在しない場合には、他のプロセッサの終了を待って、(1) の状態に戻す。

(6) 計算時間の制限について

並列計算機リモート制御ソフトでは、指定された時間内に計算が終了しないプロセッサが見つかった場合には、そのプロセッサだけでなく、全てのプロセッサの計算を止め、メッセージ出を行い、ジョブが停止される。

3.1.3.2 入出力の C 言語化

並列計算機における入出力は、オペレーティングシステムによって、1 回に 32KB (キロバイト) までに制限されている。そこで、この制限を越えるデータの入出力を扱う部分の処理を、C 言語で書き換えることによって対応した。また、この修正では、入出力処理が FORTRAN で行う場合よりも簡略化されているため、データの入出力にかかる時間が短縮されている。

3.1.3.3 出力データの圧縮

FAMEシステムにおけるファイルサーバの容量は24GB（ギガバイト）であるが、一回のショットについて、約20MB（メガバイト）の計算結果が出力されるので、1200ショット分のデータしか格納出来ない。そこで、出力の大半を占める平衡データを、MHD平衡コード内で圧縮して出力されている。

圧縮方式のアルゴリズムについて簡単に説明すると、まずデータの最大値と最小値を求めて、その間を1024区間に等分し、量子化する。そして、量子化したデータの差分をとり、エントロピを減少させる。最後に、Huffman法により符号化する。

その結果、ファイル容量は10分の1に圧縮されると共に、データ転送時間の短縮にも有効である。また、この時の有効数字は三桁まで保証されているが、計測器の精度を基に考慮すると、この精度は十分妥当と言える。

3.1.3.4 入力データの読み込みの改善

MHD平衡コードにおいて論理機番4から読み込まれるポロイダル・コイルの磁場データは、600KB（キロバイト）程度の大きさで、時刻に依らず一定であるため、このデータを毎回フロントエンドEWSから並列計算機に転送するのは無駄である。そこで、並列計算機の各プロセッサにおいて、一回目の計算時のみデータを読み込み、かつ、それと同時に内容を作業領域に格納し、二回目以降の計算では、作業領域の参照を行うだけで済むように工夫されている。

3.1.3.5 WRITE文の制御

大型機で動作しているMHD平衡コードでは、論理機番6から計算経過をリスト出力していたが、並列計算機で動作するMHD平衡コードにおいては、計算時間の短縮のために省略している。これは、以下のようにソースコードを修正し、Cのプリプロセッサを利用することで実現している。

```
#ifdef WRITED
WRITE(6,*)
#endif
```

この修正により、コンパイルオプションを追加して、ソースコードを再コンパイルするだけで、フロントエンドEWSの画面上に出力リストを表示させることが出来る。但し、この場合、二個以上のプロセッサを用いた計算を行うと、各プロセッサの出力のタイミングがまちまちであるため、表示される計算経過の判読は不可能となる。

3.1.3.6 画像データ処理

FAMEシステムにおけるMHD平衡コードの計算結果は、数値データとしてだけでなく、動画処理装置上で、静止画や動画として確認することが出来る。

まず、F A M E で画像データを作成する。画像データ処理は、M H D 平衡コードの中に組み込まれており、ロジックとしては、1 時刻の平衡計算が終了した直後に、その結果を图形処理（カルコンプー G K S インターフェイス）プログラムを介してベクターデータ（G K S メタファイル）に変換しフロントエンド E W S のファイルへ出力するというものである。

次に、動画処理装置でフロントエンド E W S のファイル中のベクターデータをメモリへ読みませ、それをダブルバッファ機能を用いて表示させる。これについての詳細は、3.2 動画処理ソフトウェアを参照されたい。

3.1.4 並列化の結果

以下では、今回の最適化を含めた並列化によって向上が見られた計算時間、及び計算結果に対する考察を述べる。

F A M E システムにおける、F B I コードとM H D 平衡コード、及び動画処理ソフトの処理時間と、大型計算機において二つの計算コードを動作させ、結果を图形端末上に画像表示させた時の処理時間を表3.1 に示す。

表3.1 F A M E と大型計算機との処理時間の比較（単位：秒）

使用計算機 コード	F B I	M H D 平衡	作図及び表示	合 計
F A M E	578	310 (画像データ作成を含む)	191 (作図181 表示10)	1079
汎用大型計算機	2575	4635		7210

使用データ

ショット番号 : E 0 1 9 4 7 0

メッシュサイズ : 128

ポイント数 : 1. 1 秒 ~ 11. 3 秒 103 ポイント

注) 大型計算機のデータは、1. 1 秒での計測値を103倍したものである。

また、その時間は経過時間である。

F B I コード、及びM H D 平衡コードの大型計算機における計算結果とF A M E における計算結果は、入力データが全く同じものであっても完全には一致しない。

これには二つの理由が挙げられる。まず一つは、数値の内部表現の違いにより、実数にお

ける仮数部の有効桁数に違いがあることである。もう一つは、FORTRANの仕様において、実数の精度を上げる時に、大型計算機では桁上げされる数値が保証されていないのに対して、フロントエンドEWSでは保証されていることが挙げられる。

しかし、大型計算機での計算結果とFAMEでの計算結果の違いの大きさは、有効数字三桁未満であり、測定器の測定精度を考えると十分妥当なものといえる。

3.2 動画処理ソフトウェア

動画処理ソフトウェアは、並列計算機による計算結果として出力される画像データの入力をを行い、動画処理装置のウィンドウに動画として再生したり、再生した内容について保存する必要のある画像については、ネットワークによって接続されているコマ録り装置に転送及び保存することを目的として全体システムに位置づいている。

以下に画像データ、動画処理及びコマ録り処理について説明する。

3.2.1 画像データ

画像データは、動画処理がモニタに出力する動画画面の各画面の情報を構成するデータ群であり、画面の開始、画面内の作画データ列及び画面の終了を意味する内容で構成される。F A M E システムでの画像データは、F B I コード及びM H D 平衡コードに組み込んだ画像データ生成用のインターフェイスライブラリを経由することによって並列計算機からフロントエンドE W S のハードディスクに出力される。画像データの容量は全データをバイナリとして保存することによって極力最小となるよう工夫し、1画面あたり約68 K B (キロバイト)の容量で保存される。その結果コマ録り装置の保存単位である 648×480 ピクセルのデータ容量で比較した場合、ラスターデータ(画面内容をR G B画素で表現したデータ)で保存する場合と比較して1画面あたり約250 K B縮小された。

インターフェイスライブラリのF A M E による並列処理への対応に伴い、出力処理部分をF O R T R A N 言語からC言語に書き換えることにより、画像データの出力に要する時間を短縮した。

3.2.2 動画処理

動画処理は、起動コマンドを入力することによって図3.3の画面を表示する。

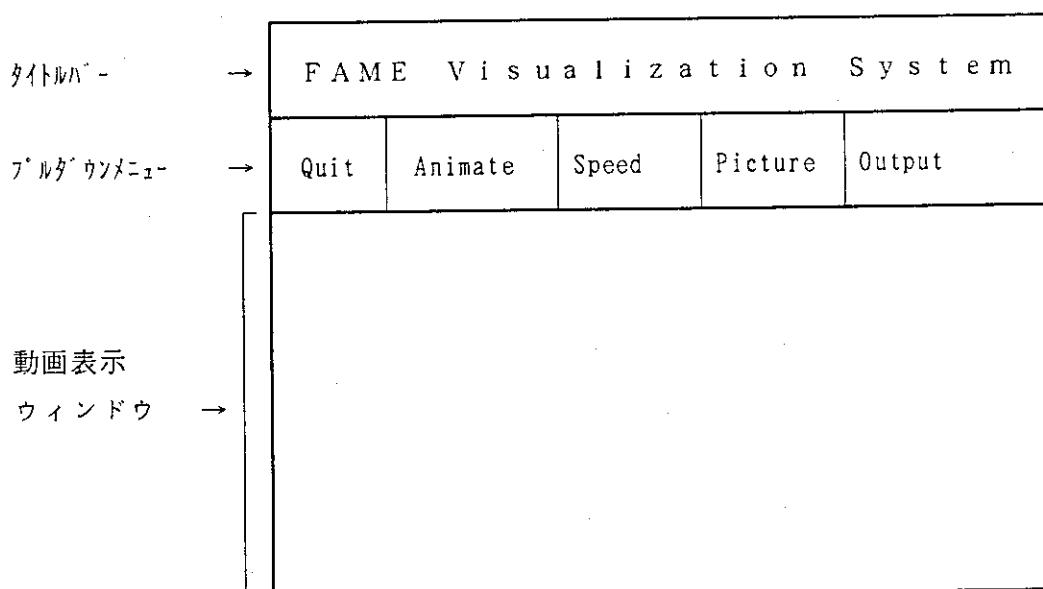


図3.3 動画処理画面

次にF B Iコード及びM H D平衡コードによって事前に作成された画像データを入力し、各画面の情報を連続的にウィンドウに表示することによって全画像データを動画として再生する。再生された動画は、ウィンドウに用意されたボタン内のオプションを選択することにより、表3.2に示す機能が選択可能である。各ボタンは、図3.4に示すようなプルダウンメニューとなっており、マウスにより選択を行う。

表3.2 動画処理オプションボタン機能

ボタン名称	オプション	機能概要
Q u i t		プログラムの終了
A n i m a t e	F o r w e r d	前方向への再生
	B a c k	後方向への再生
S p e e d	F a r s t	通常速度の再生
	S l o w	低速度の再生
	S t r o b o	画面単位の再生
P i c t u r e	S t a r t - P	開始画面へのジャンプ
	S t o p - P	終了画面へのジャンプ
O u t p u t	S u n - V i s i o n	市販動画処理プログラムのデータ作成
	N v s - 2 0 0 0	コマ録り装置のデータ作成
	P o s t - S c r i p t	ポストスクリプトのデータ作成

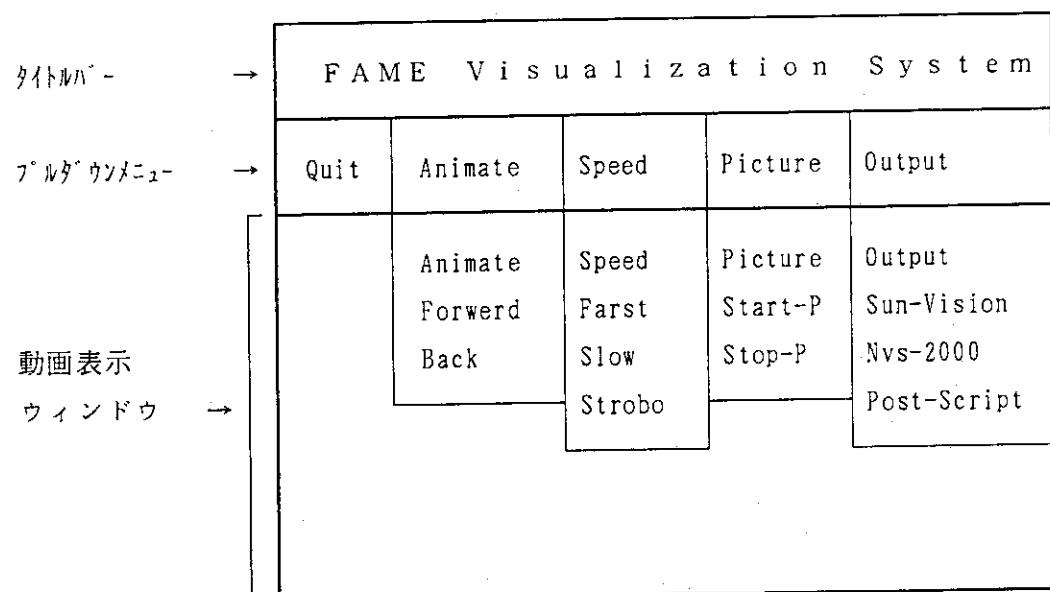


図3.4 プルダウンメニュー

コマ録り処理を行う場合は、動画処理に於けるプルダウンメニュー内のN v s - 2 0 0 0を選択することによって出力されるラスターデータファイルをコマ録り装置用のファイル形式に変換し、ネットワークを介してコマ録り装置に転送及び光ディスクに保存する。

プルダウンメニューOutput内のPost-Scriptを選択することによってポストスクリプト用データ作成が可能であり、当該メニューが選択された時点で図3.5に示すようなポストスクリプト出力対象ファイル選択用の画面が表示される。出力結果を付録Aに示す。

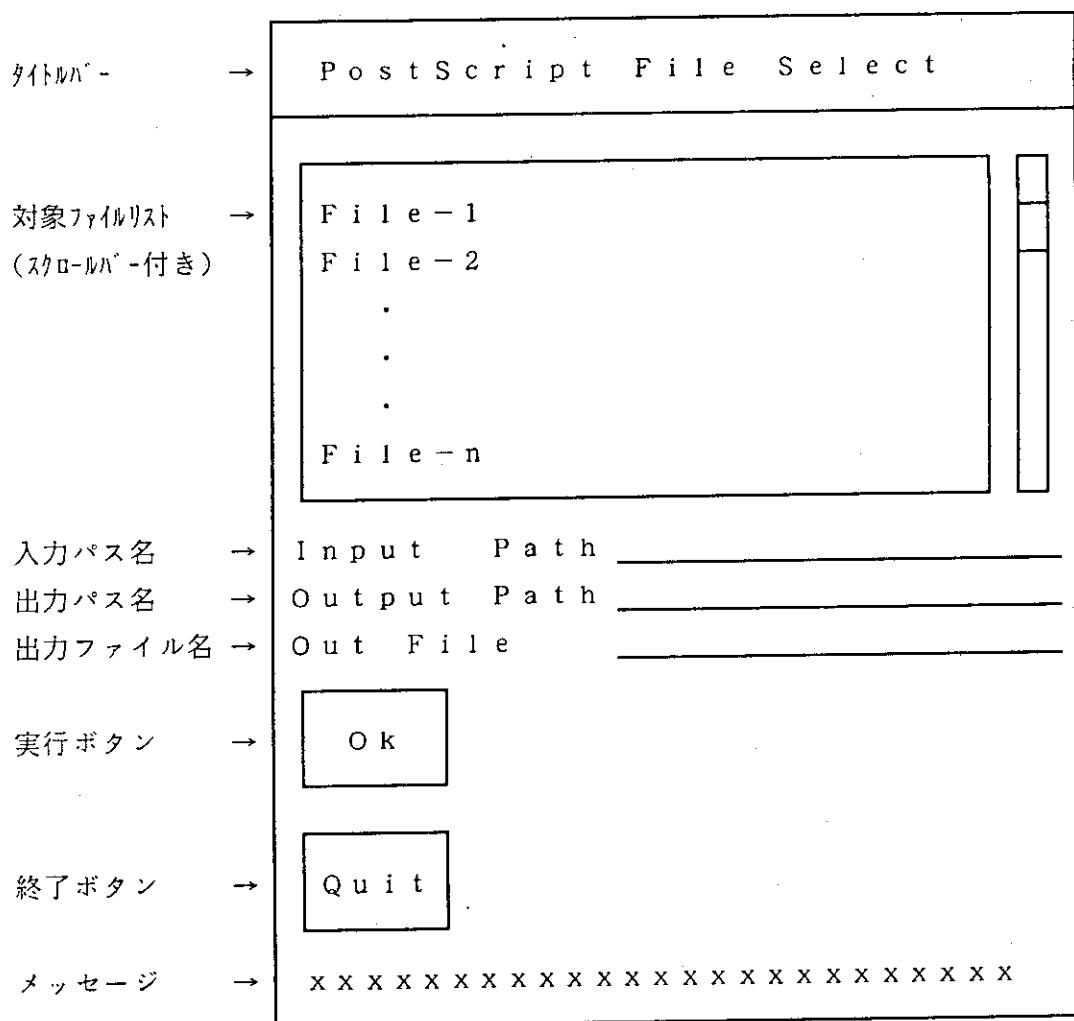


図3.5 ポストスクリプト出力操作画面

特に動画のような大量の画像を取り扱う場合は、実装メモリ容量を追加し、さらに高速なフレームバッファ（画像を表示するための転送バッファ）処理が必要になるため、当該動画処理装置では、実装メモリに64MB（メガバイト）（約140画面の許容量）を用意し、フレームバッファには、ダブルバッファ装置を実装している。これにより画像データのメモリ蓄積を可能とし、かつダブルバッファによる高速な画像転送処理を実現した。またネットワークを経由して動画処理を利用する場合でも高速な動画処理が可能となるよう考慮し、画

像データのメモリ蓄積は利用者側のワークステーションにすることとした。これにより一度利用者側に転送された画像データは利用者側のワークステーションに蓄積されるため、十分なメモリ容量を実装しているワークステーションであれば、ネットワークを経由した動画処理であっても動画処理装置を利用する場合と遜色のない利用を可能とした。

3.2.3 コマ録り処理

コマ録り処理は、前述の手順で作成した動画データの保存を目的として行うものであり、ネットワークを介してデータを転送することによって、付属の光ディスクに画像データの保存が可能となる。光ディスクの容量は、片面43500フレームの保存量であり、画面で87000フレームの画像データが保存可能である。

光ディスクに保存された画像データは、並列計算機及び動画処理装置とは無関係に任意の時点で再生が可能であり、再生した画像はビデオプリンタ及び市販のビデオテープレコーダに出力することが可能である。出力結果を付録Bに示す。以上のコマ録り処理は、ネットワークに接続されたワークステーションからのコマンド入力によって利用できる機能であり、この他に画像データの検索及びモニタ表示等に関連する光ディスクの管理機能も有している。

4. あとがき

J T - 6 0 の非定常プラズマ輸送解析に必要な 1 0 0 点を越える時系列磁気流体平衡の計算を約 2 0 分のショット間に処理するため、磁気流体平衡高速処理装置 F A M E が開発された。

F A M E は、C P U の性能が 1 2 . 5 M F L O P S (100 万浮動小数点／秒) のマイクロプロセッサが、 4×4 のゲートアレイを構成要素とした多段接続網で 2 0 個結合されている並列計算機で、最大性能が 2 5 0 M F L O P S (100 万浮動小数点／秒) である。また、周辺装置には F A M E のジョブ管理を行うためのフロントエンド E W S 、平衡データを保存するための 2 4 G B (ギガバイト) の容量を持つファイルサーバ、計算結果を動画として表示するための動画処理装置、そして、その画像を記録するためのコマ録り装置がある。

ソフトウェアでは、F B I コードがフロントエンド E W S で、並列化されたM H D 平衡コードが F A M E で動作するようにインストールされている。また、両計算コードの結果を動画表示するためのプログラムが開発された。さらに、各コードの処理時間の短縮を図るために種々の最適化が施された。その結果、従来の大型計算機による処理に比べて、飛躍的な処理時間の短縮が行えるようになった。

現 F A M E における磁気流体平衡解析に要する経過時間は表3.1に示す様に約 1 8 分である。ショット間の時間は約 2 0 分であるが、前述の様に、F A M E による解析結果を次々ショットに反映させるためには、全ての計算処理をこのショット間に終了させる必要がある。このためには、データベース作成に要する時間 9 分を差引けば、F A M E の処理時間としては 1 2 分しか残されていない。今後、現バージョンの処理でオーバヘッドとなっている F B I 計算をフロント・エンド E W S から並列計算機に移し、時間短縮を図っていく予定である。

謝 辞

F A M E システムは、1 9 8 8 年より約 5 年間プラズマ理論研究室において実施されたプラズマシミュレータ M E T I S の開発研究の成果を基本的に継承したものです。そのシステム開発は、ハードウェアソフトウェア両面にわたり日本電気株式会社 C & C システム研究所コンピュータシステム研究部小池部長、中田氏、松下氏、鳥居氏その他多数の方々の努力に負うところが大きく、ここに、感謝の意を表します。また、M E T I S で蓄積されたノウハウの利用を快く了承していただいたプラズマ理論研究室の竹田室長、さらには、本システム開発を終始支援していただいた炉心プラズマ解析室の安積室長、情報システムセンターの秋元室長に心より感謝いたします。

4. あとがき

J T - 6 0 の非定常プラズマ輸送解析に必要な 1 0 0 点を越える時系列磁気流体平衡の計算を約 2 0 分のショット間に処理するため、磁気流体平衡高速処理装置 F A M E が開発された。

F A M E は、C P U の性能が 1 2 . 5 M F L O P S (100 万浮動小数点／秒) のマイクロプロセッサが、 4×4 のゲートアレイを構成要素とした多段接続網で 2 0 個結合されている並列計算機で、最大性能が 2 5 0 M F L O P S (100 万浮動小数点／秒) である。また、周辺装置には F A M E のジョブ管理を行うためのフロントエンド E W S 、平衡データを保存するための 2 4 G B (ギガバイト) の容量を持つファイルサーバ、計算結果を動画として表示するための動画処理装置、そして、その画像を記録するためのコマ録り装置がある。

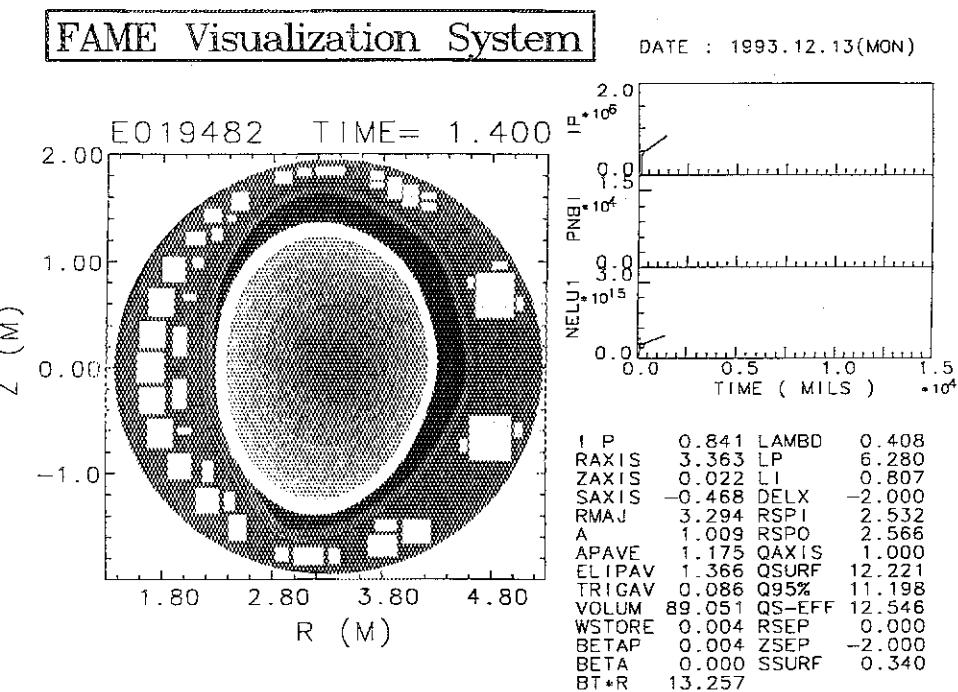
ソフトウェアでは、F B I コードがフロントエンド E W S で、並列化されたM H D 平衡コードが F A M E で動作するようにインストールされている。また、両計算コードの結果を動画表示するためのプログラムが開発された。さらに、各コードの処理時間の短縮を図るために種々の最適化が施された。その結果、従来の大型計算機による処理に比べて、飛躍的な処理時間の短縮が行えるようになった。

現 F A M E における磁気流体平衡解析に要する経過時間は表3.1に示す様に約 1 8 分である。ショット間の時間は約 2 0 分であるが、前述の様に、F A M E による解析結果を次々ショットに反映させるためには、全ての計算処理をこのショット間に終了させる必要がある。このためには、データベース作成に要する時間 9 分を差引けば、F A M E の処理時間としては 1 2 分しか残されていない。今後、現バージョンの処理でオーバヘッドとなっている F B I 計算をフロント・エンド E W S から並列計算機に移し、時間短縮を図っていく予定である。

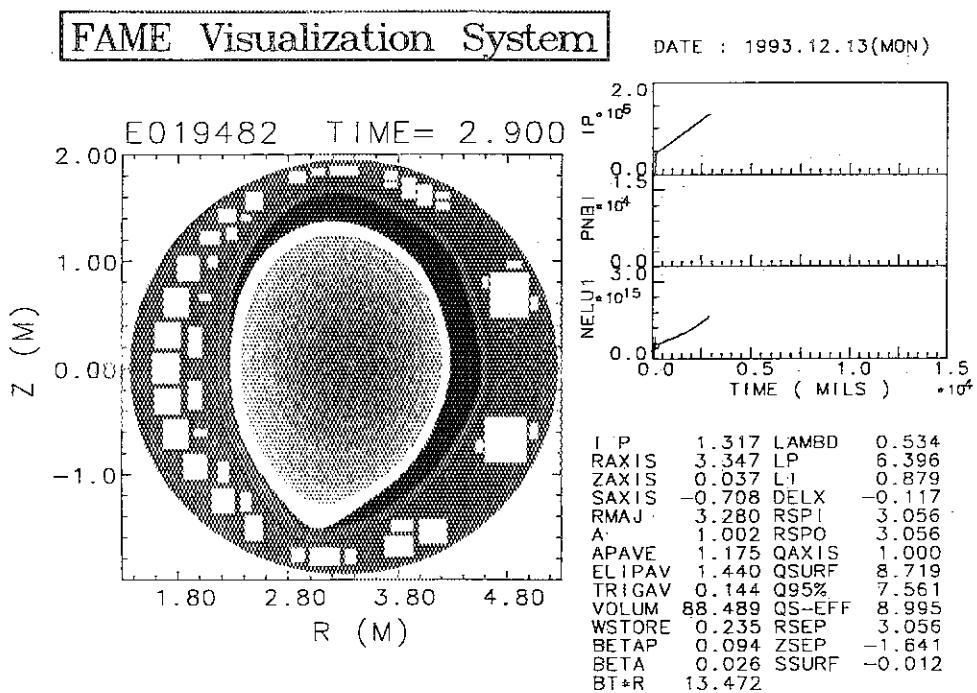
謝 辞

F A M E システムは、1 9 8 8 年より約 5 年間プラズマ理論研究室において実施されたプラズマシミュレータ M E T I S の開発研究の成果を基本的に継承したものです。そのシステム開発は、ハードウェアソフトウェア両面にわたり日本電気株式会社 C & C システム研究所コンピュータシステム研究部小池部長、中田氏、松下氏、鳥居氏その他多数の方々の努力に負うところが大きく、ここに、感謝の意を表します。また、M E T I S で蓄積されたノウハウの利用を快く了承していただいたプラズマ理論研究室の竹田室長、さらには、本システム開発を終始支援していただいた炉心プラズマ解析室の安積室長、情報システムセンターの秋元室長に心より感謝いたします。

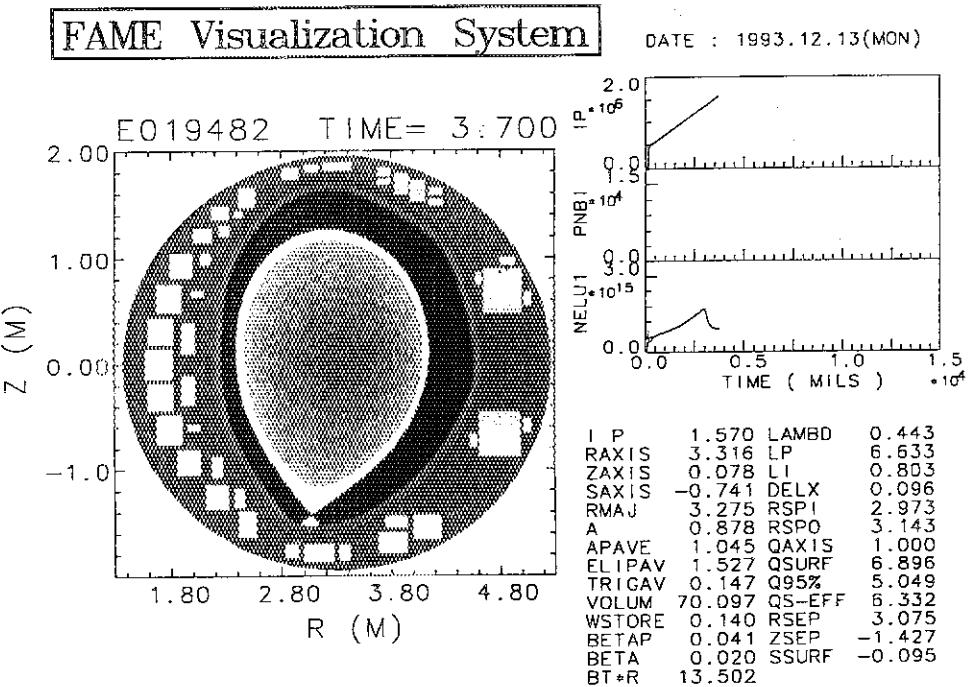
付 錄 A



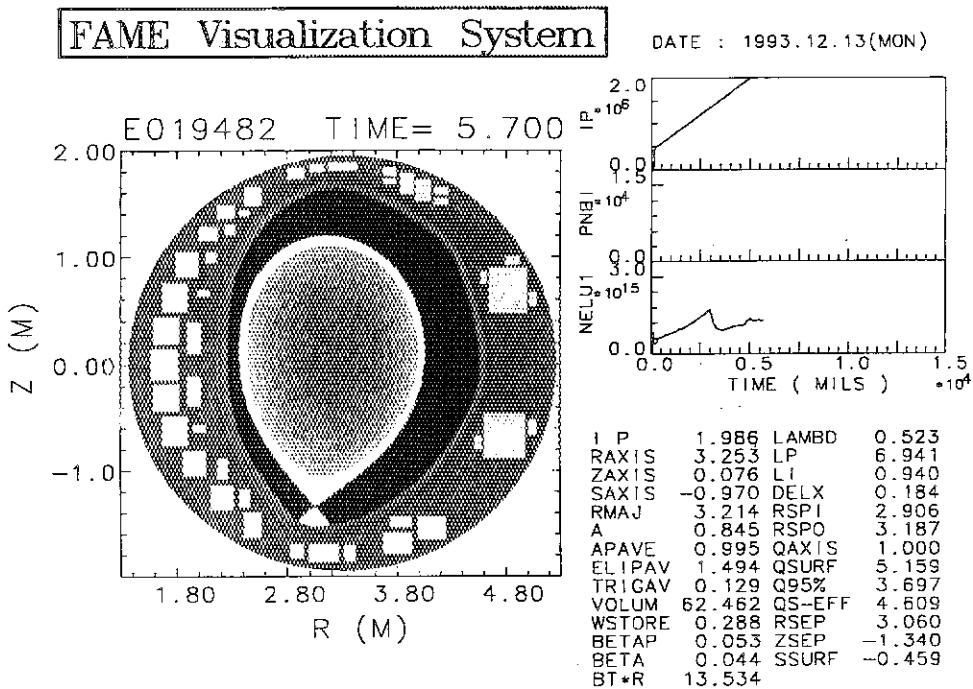
図A-1 ポストスクリプト出力による放電1. 4秒後の計算結果



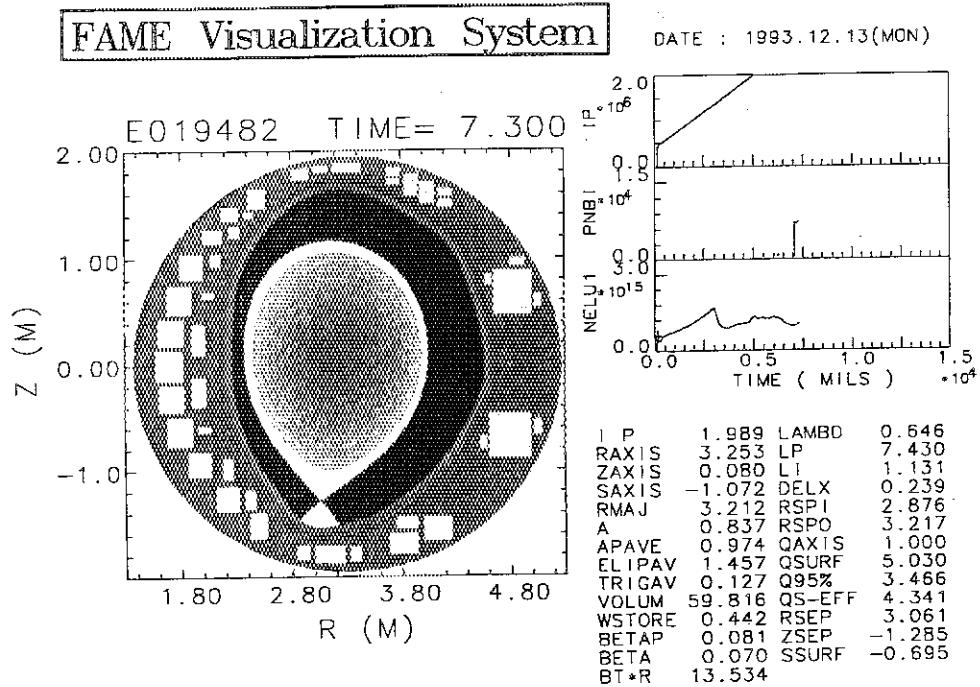
図A-2 ポストスクリプト出力による放電2. 9秒後の計算結果



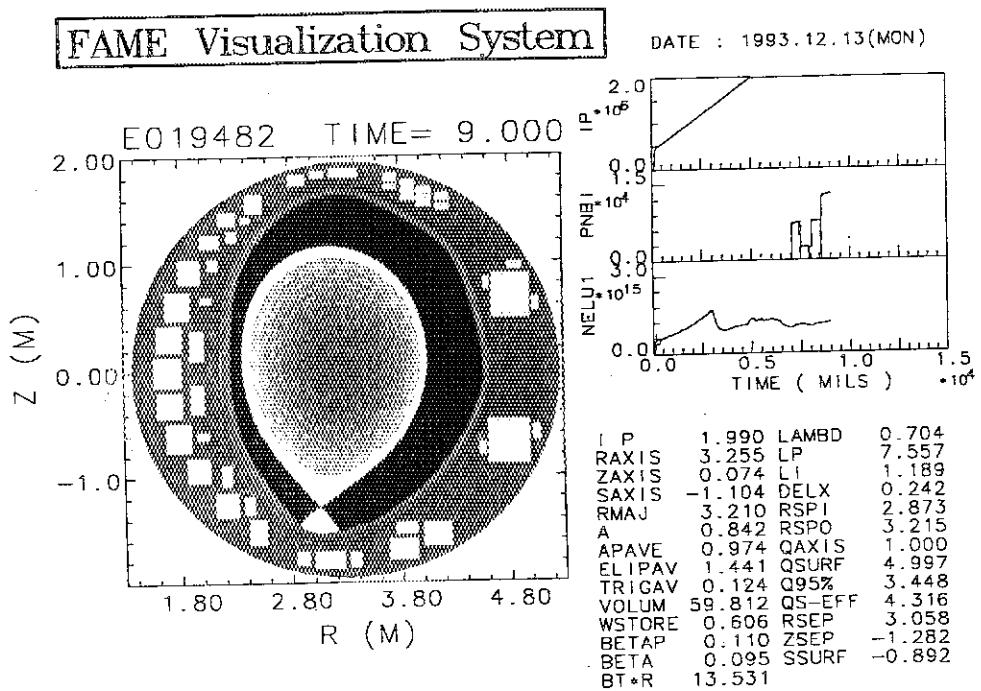
図A-3 ポストスクリプト出力による放電3. 7秒後の計算結果



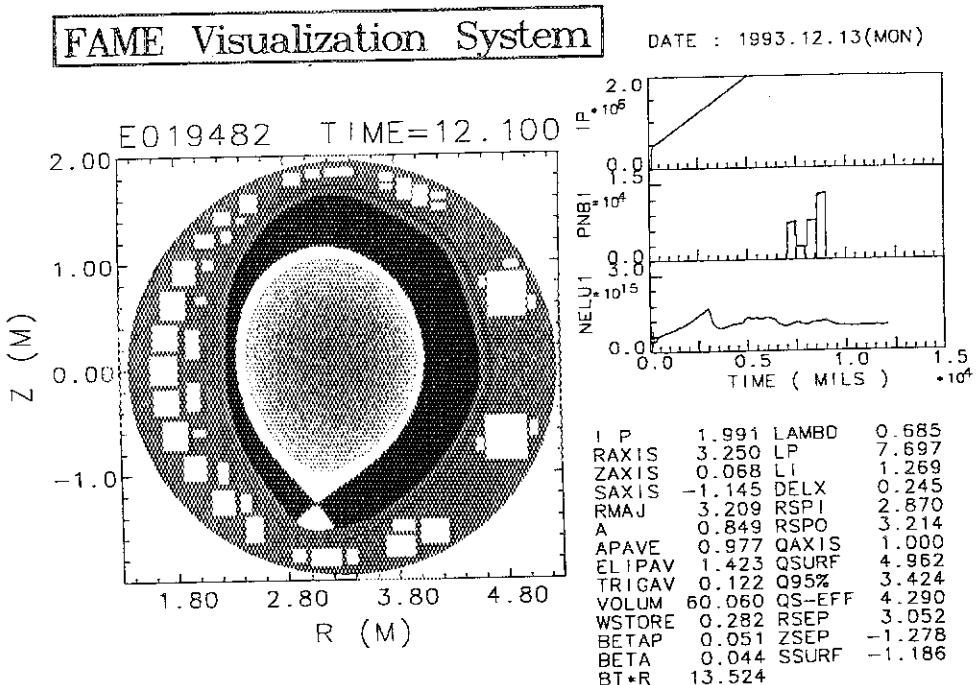
図A-4 ポストスクリプト出力による放電5. 7秒後の計算結果



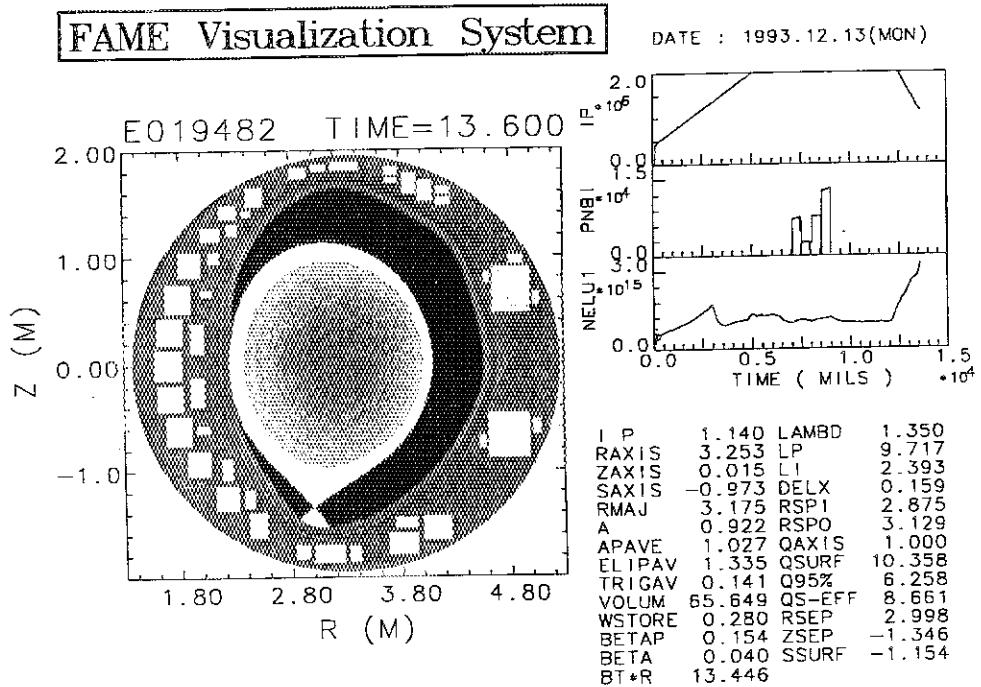
図A-5 ポストスクリプト出力による放電7.3秒後の計算結果



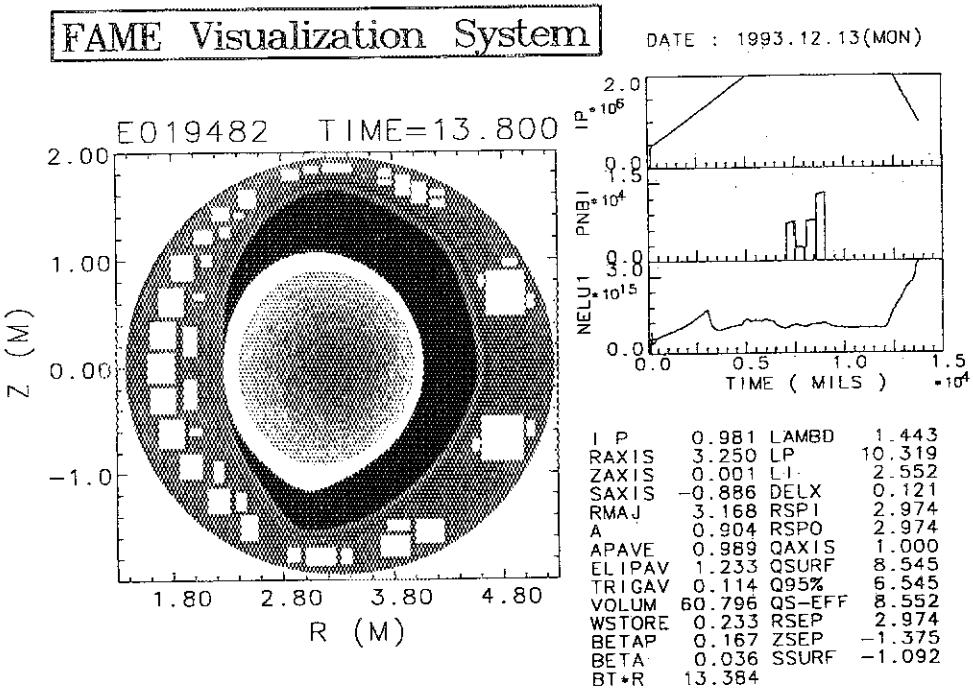
図A-6 ポストスクリプト出力による放電9.0秒後の計算結果



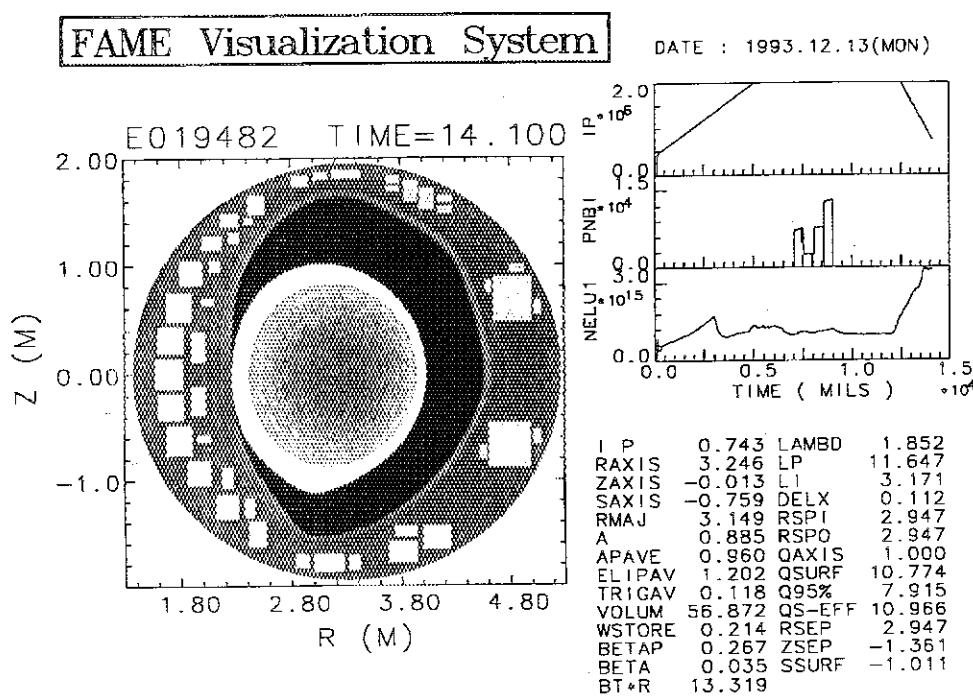
図A-7 ポストスクリプト出力による放電12. 1秒後の計算結果



図A-8 ポストスクリプト出力による放電13. 6秒後の計算結果

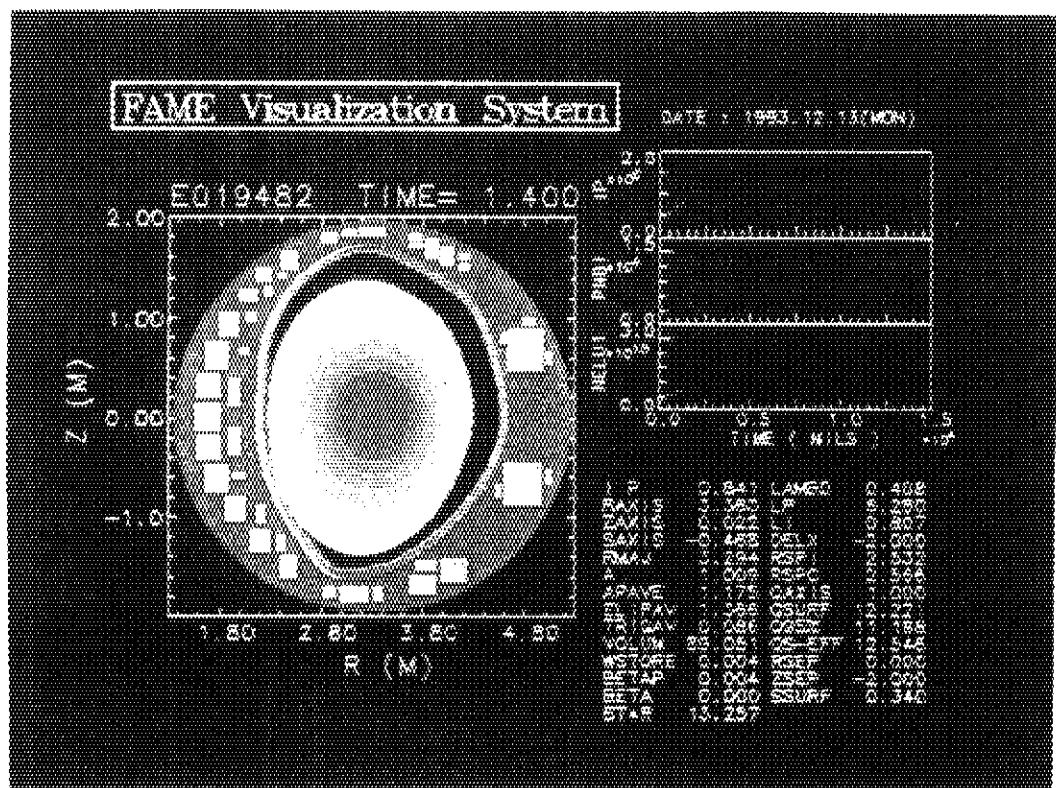


図A-9 ポストスクリプト出力による放電13.8秒後の計算結果

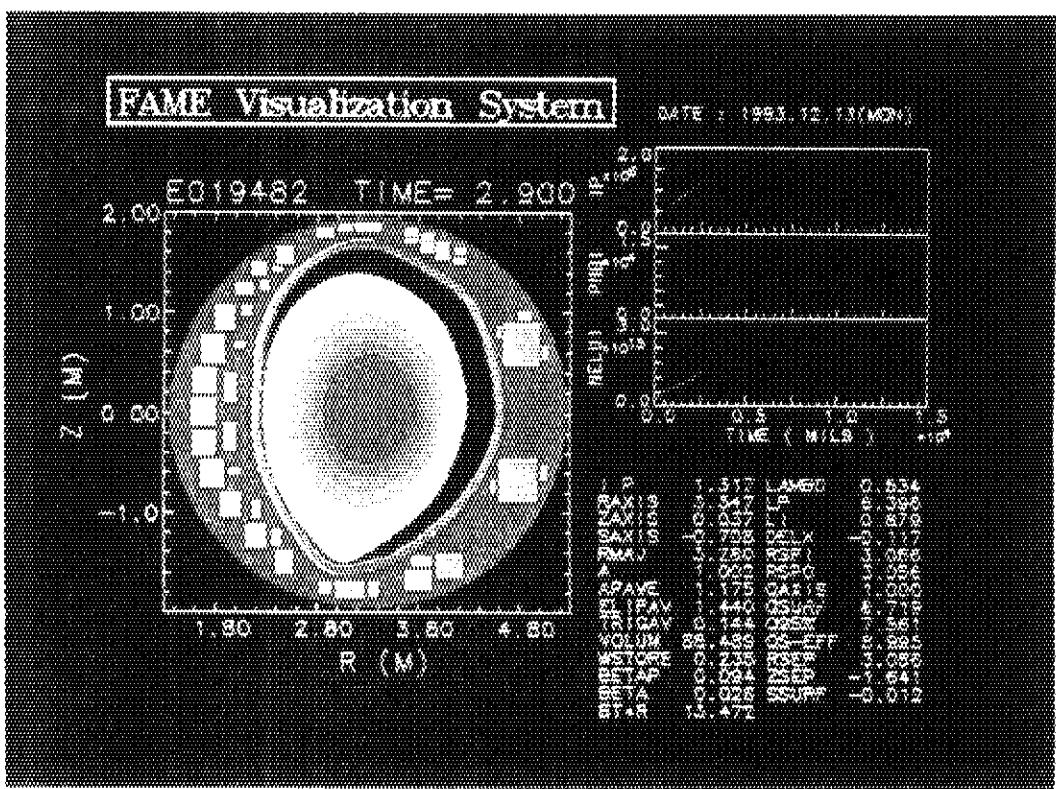


図A-10 ポストスクリプト出力による放電14.1秒後の計算結果

付 錄 B



図B-1 コマ録り装置に出力した放電1. 4秒後の計算結果



図B-2 コマ録り装置に出力した放電2. 9秒後の計算結果

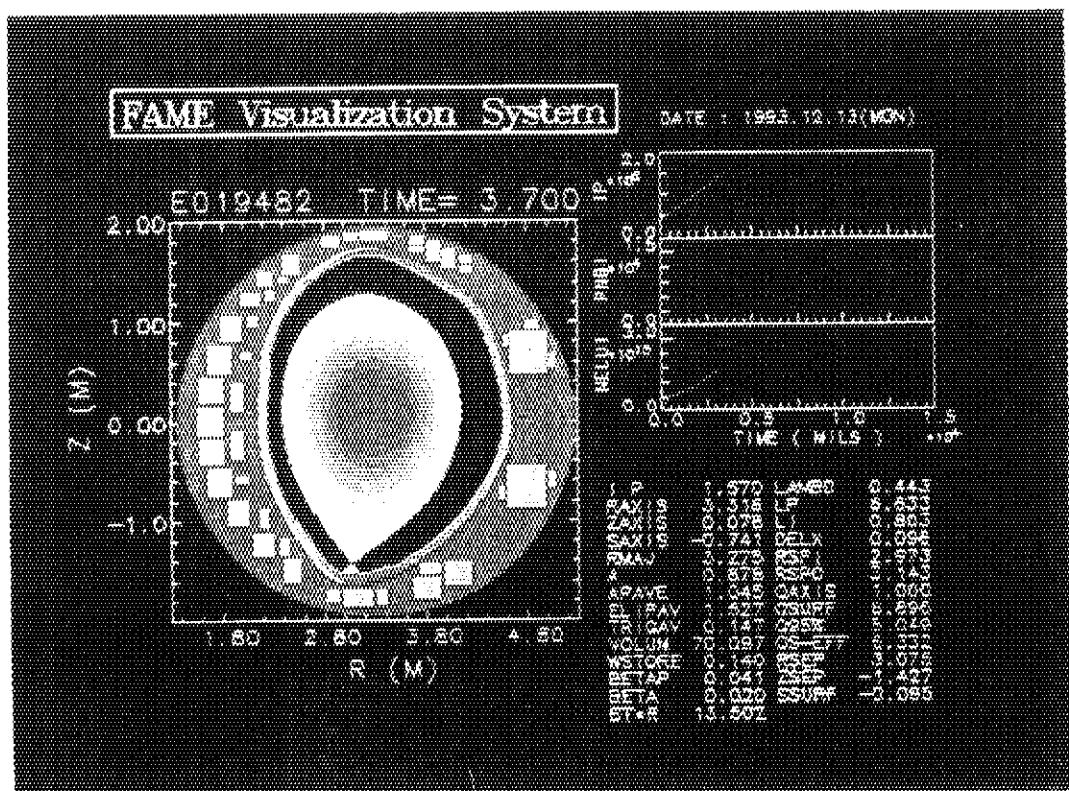


図 B - 3 コマ録り装置に出力した放電3. 7秒後の計算結果

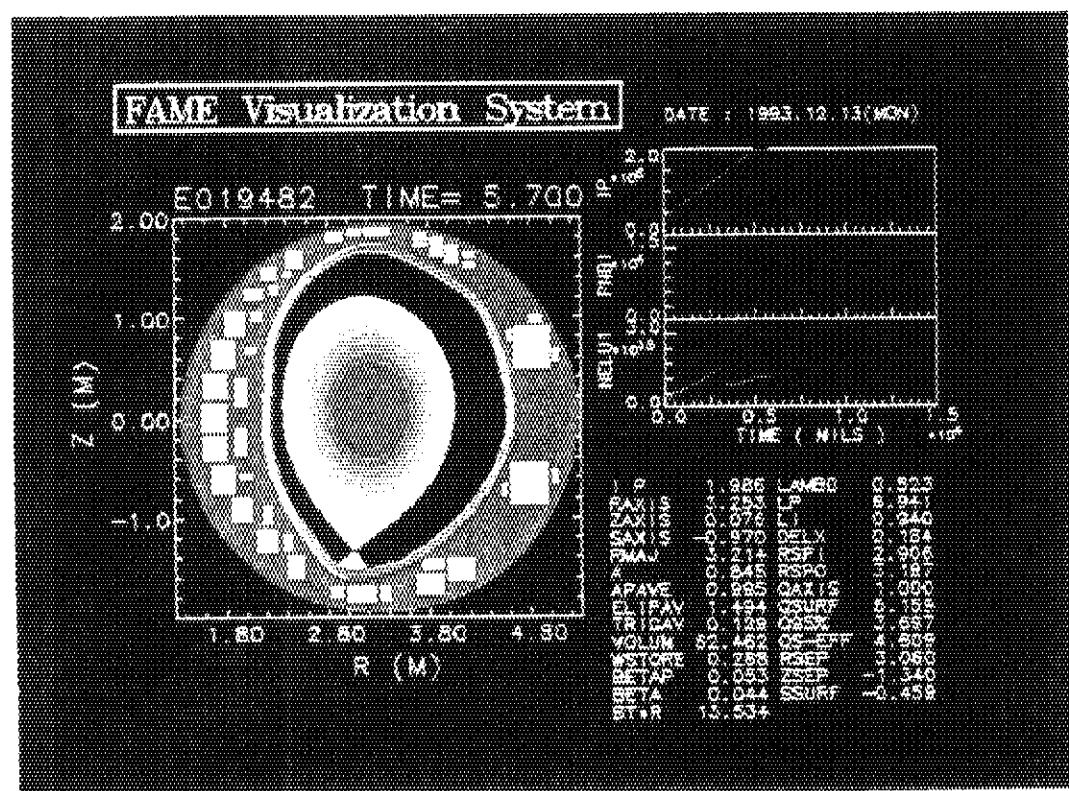
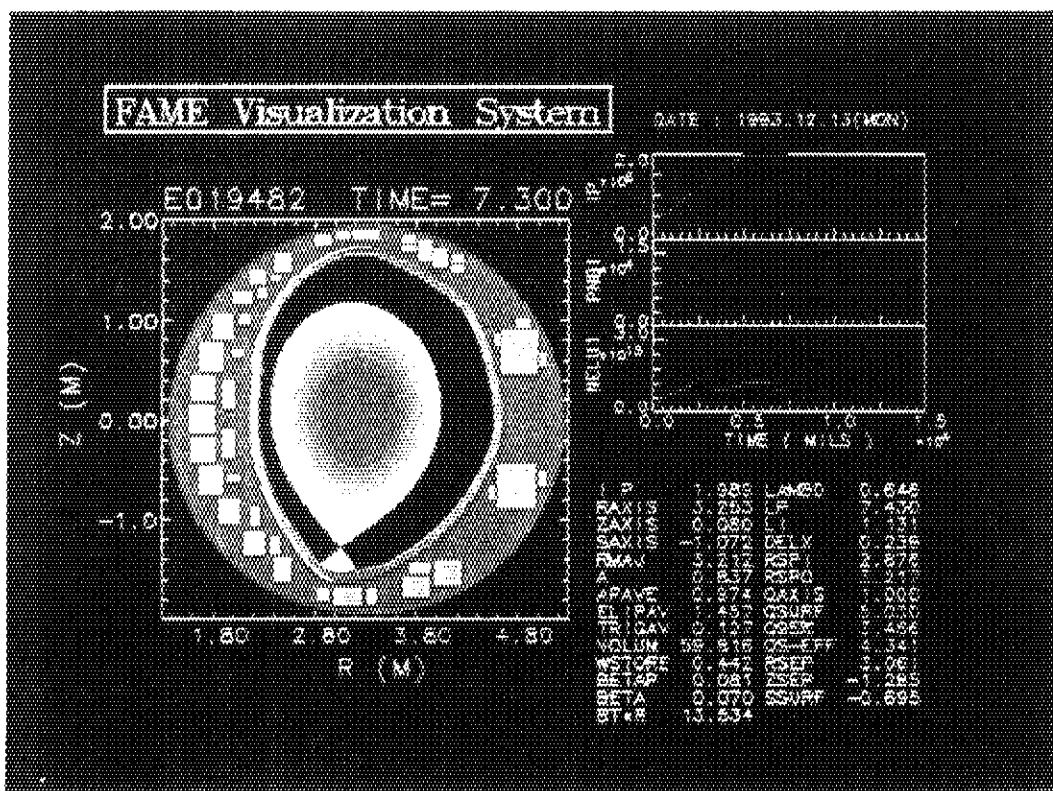
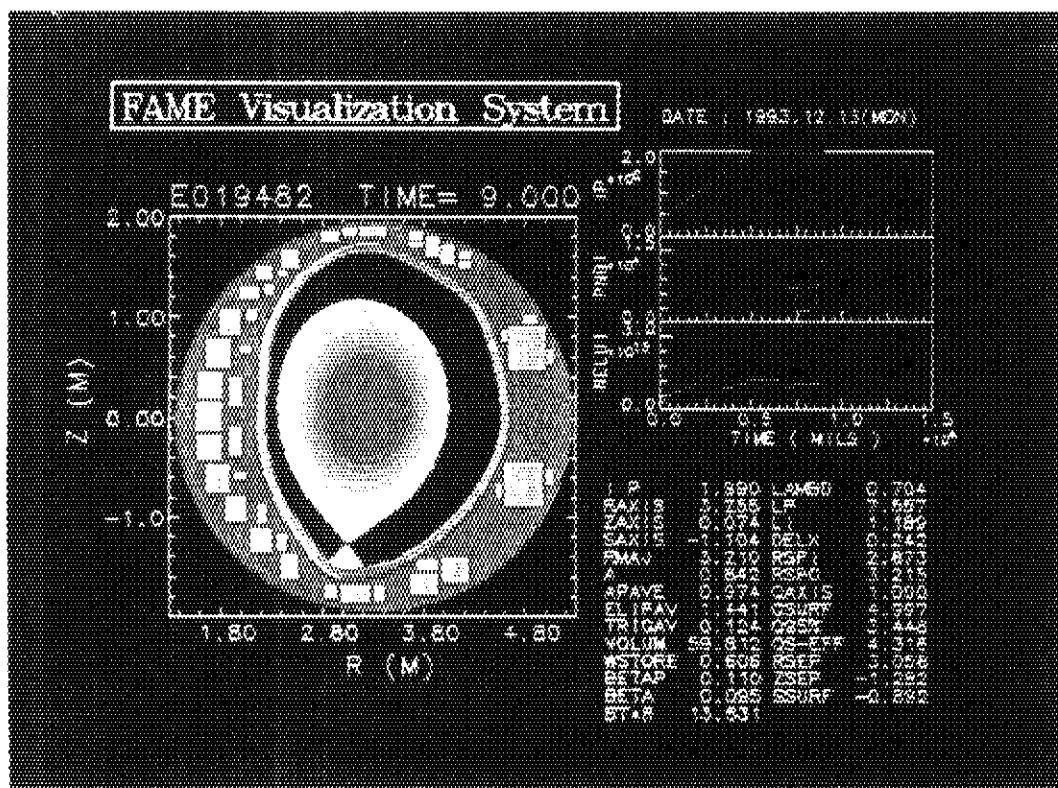


図 B - 4 コマ録り装置に出力した放電5. 7秒後の計算結果



図B-5 コマ録り装置に出力した放電7.3秒後の計算結果



図B-6 コマ録り装置に出力した放電9.0秒後の計算結果

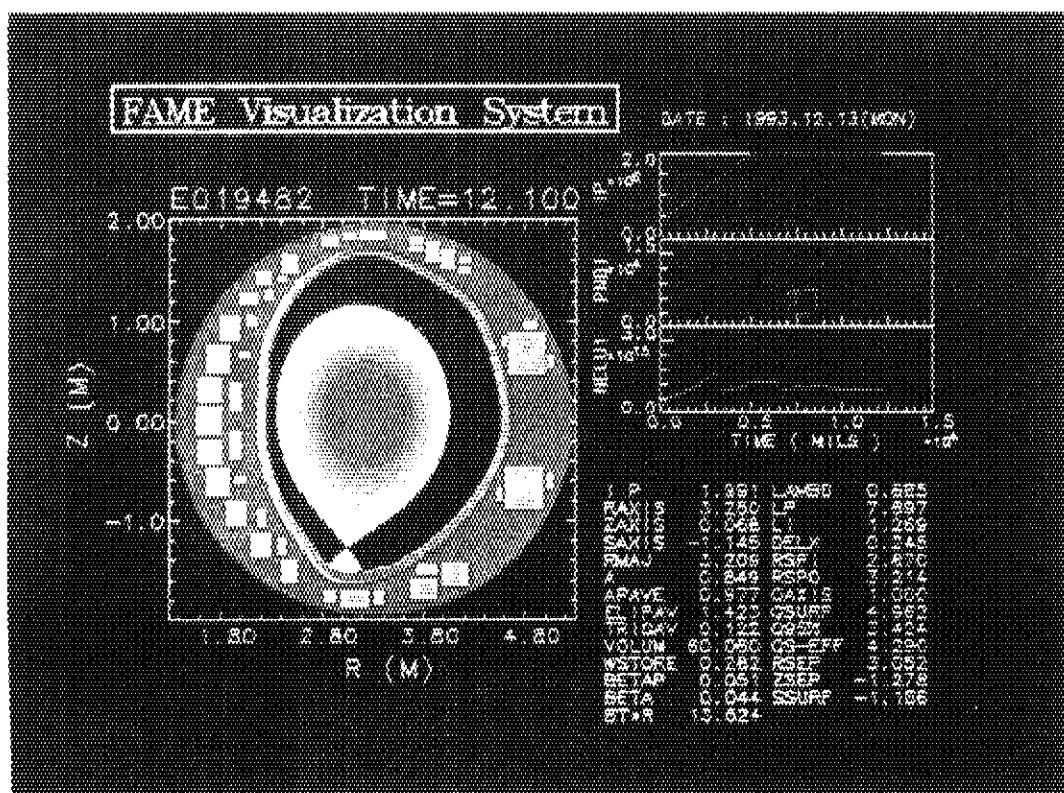


図 B-7 コマ録り装置に出力した放電 12. 1秒後の計算結果

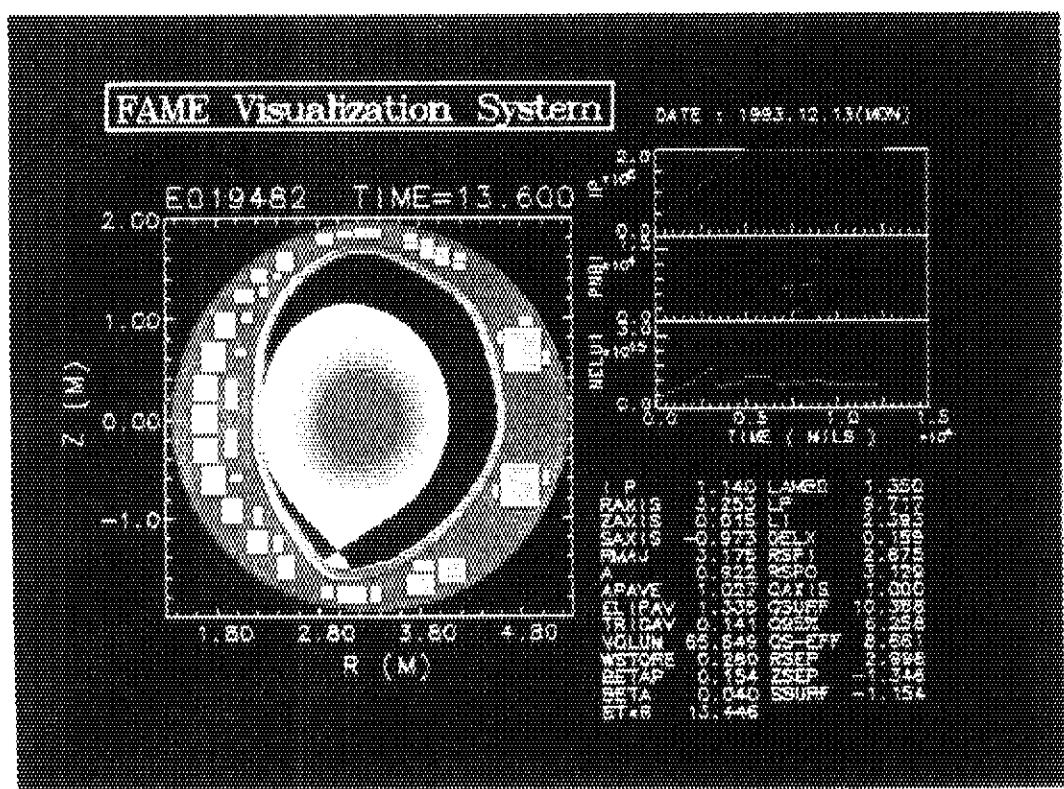


図 B-8 コマ録り装置に出力した放電 13. 6秒後の計算結果

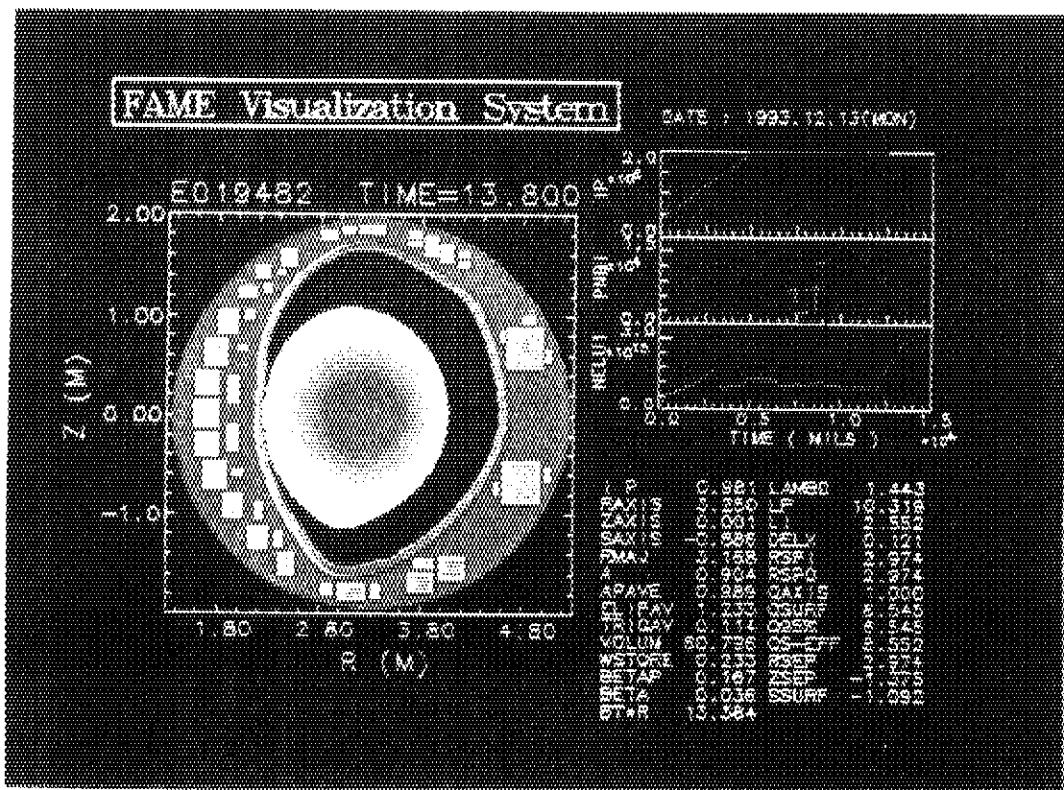


図 B-9 コマ録り装置に出力した放電 13.8 秒後の計算結果

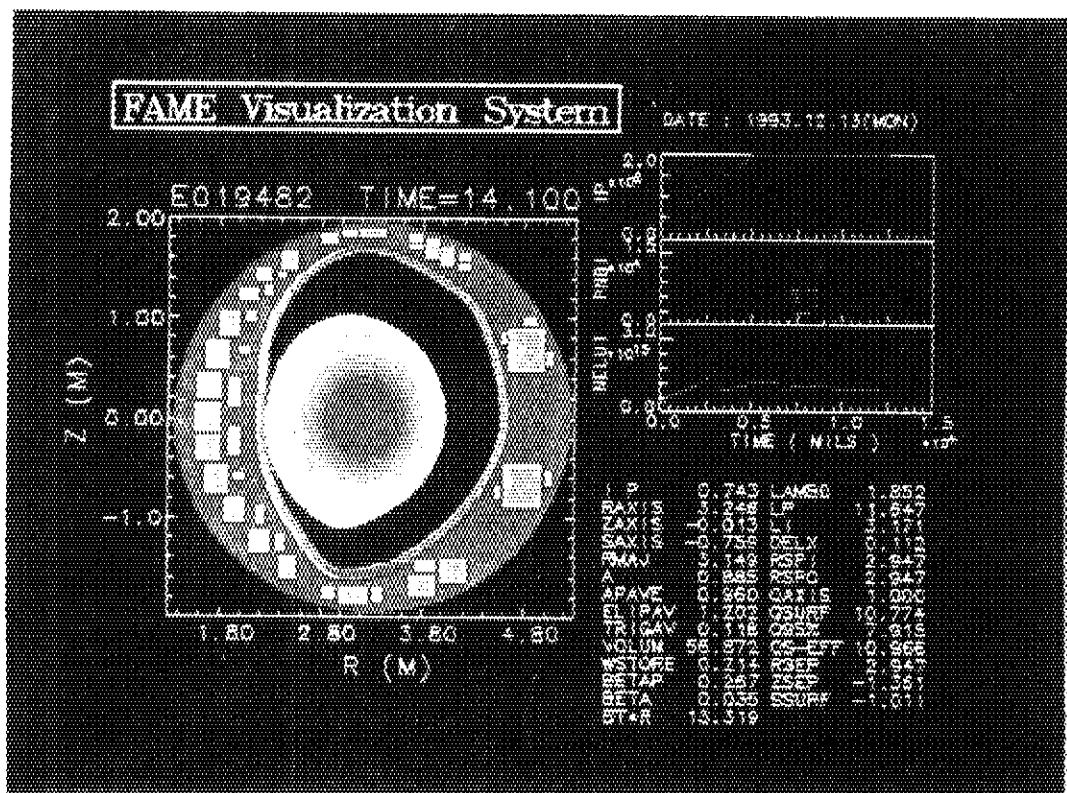


図 B-10 コマ録り装置に出力した放電 14.1 秒後の計算結果