



バイオグリッドプロジェクト

文部科学省RR2002

「スパコンネットの構築」

におけるデータグリッド

大阪大学サイバーメディアセンター

下條真司

<http://www.ais.cmc.osaka-u.ac.jp/~shimojo>

www.biogrid.jp



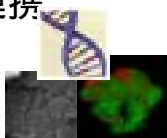
Technical challenge of Grid

- 大規模計算
 - 無限のPC
- 大規模データ管理・処理



HPC/Distributed Computing
Meets Internet

– データベース連携





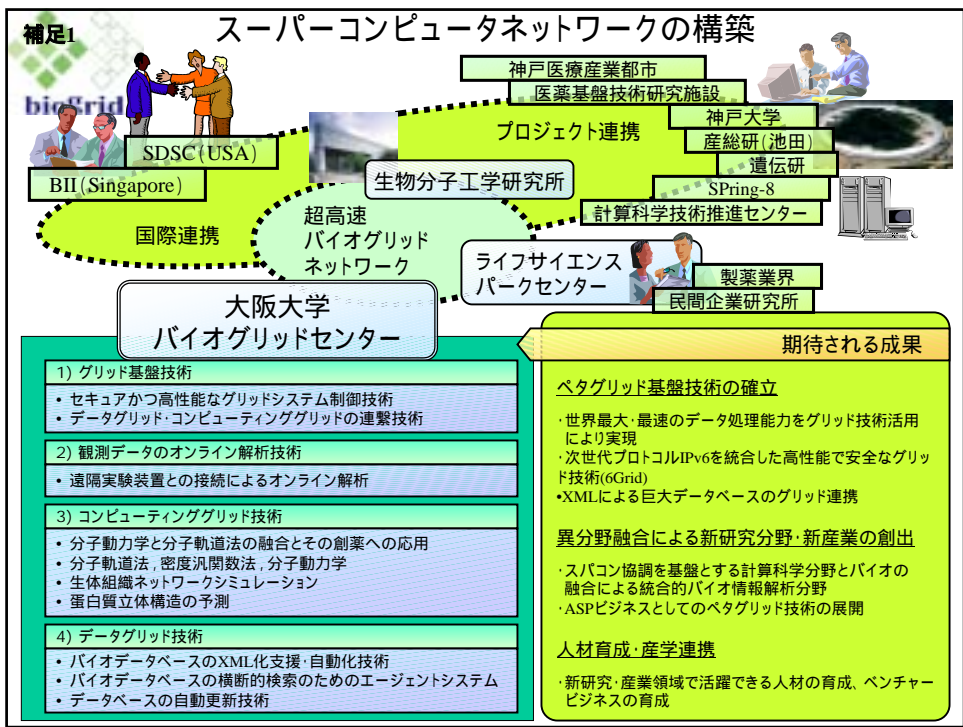
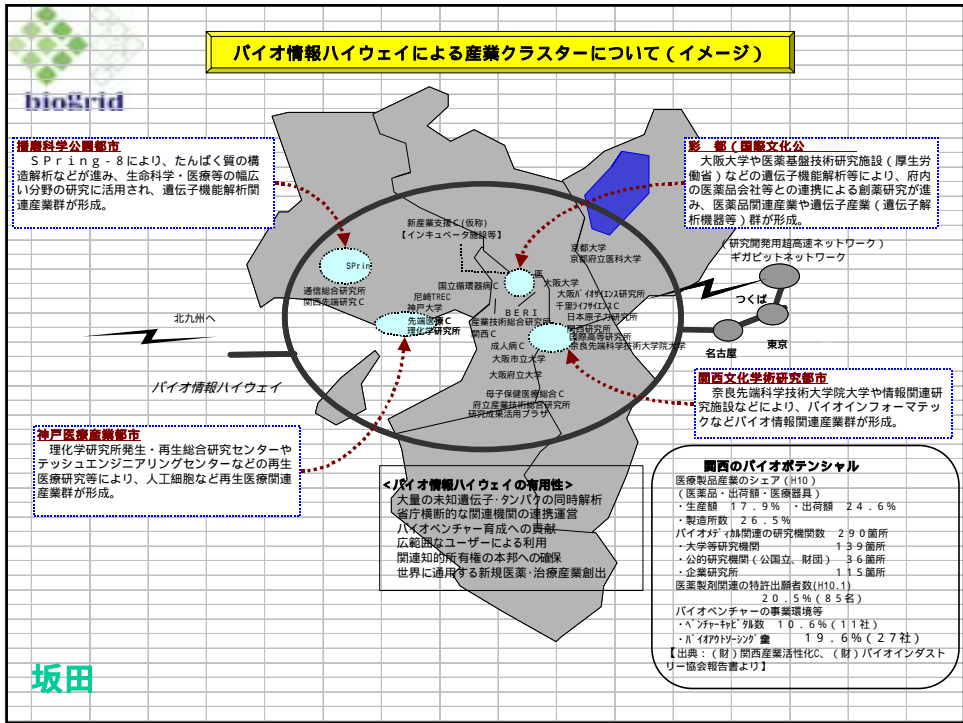
バイオグリッド・センター構想

- 超高速ネットワーク、超並列計算機、大規模データベースを核とするリエゾンセンター
- 情報通信技術(IT)の蓄積である大型計算機センターを利用
- 情報通信技術の本格的な取り込みと人材の育成,関連機関のネットワーク化、ビジネス化を行う
- Gridによる国際協調



バイオ・グリッド・センター構想





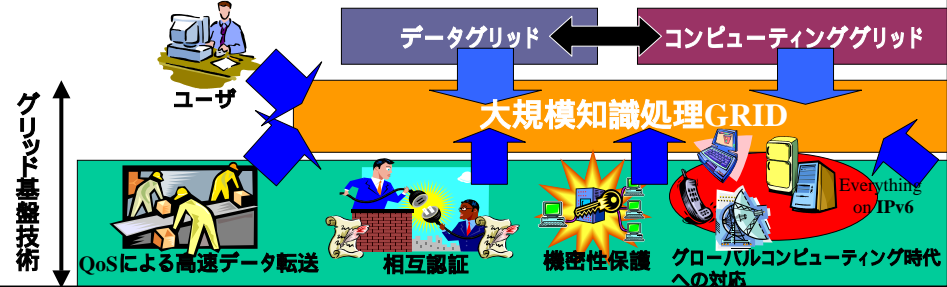
1) グリッド基盤技術

データグリッド・コンピューティンググリッド連携技術に関する研究

- データグリッド・コンピューティンググリッドをメタデータによって融合する
 - 異なるデータベース、グリッドを融合する効率的なデータ表現

セキュアで高性能なグリッドシステム制御技術の開発

- QoS (Quality of Service)技術の実証実験
 - データベース検索、シミュレーション結果などの大量データ移送を効率的に実現したグリッドミドルウェアとしてのスケジューリングシステム
- IPv6/IPSecによる新しいグリッドアーキテクチャと機密データの保護
 - 相互認証、データ暗号化、権限委譲のセキュリティメカニズムの統合
 - E-Commerce, E-Businessにおけるわが国独自の基盤技術へ(キラーアプリケーション)
 - ITBLとの相互接続実験



2) コンピューティング・グリッド技術

高速シミュレーションの異種大規模計算機群による分散処理とデータグリッド、Intelligent Gridとの融合

オープンソース開発環境の構築と並列化技法の開発

- 分子軌道法、密度汎関数法、分子動力学法、ネットワークシミュレーションなどの計算ソフトウェア(大阪大学等で開発)をオープンソース・プログラムとして整備。

グリッド上での並列化の推進、コンピューティング・グリッドを活用する新たなソフトウェアの開発とそのデータグリッドとの融合

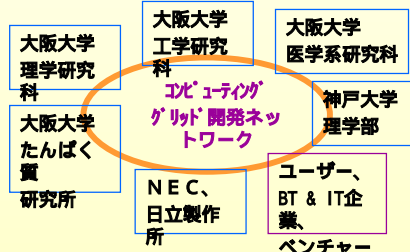
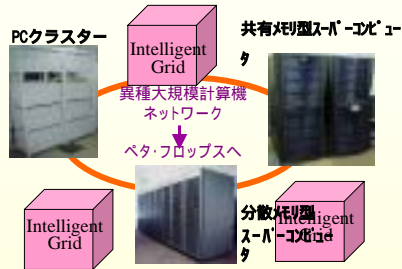
- in silico 薬物スクリーニング
- 薬物・受容体の電子状態および反応経路解析
- 生体分子立体構造予測
- 細胞内・生体組織のネットワークシミュレーション

データグリッド・コンピューティンググリッドの協調

- データグリッドとの協調によるバイオデータ解析・データマイニング (バイオ情報の新たな属性や関連性の探索)

コンピューティンググリッド・ユーザーのためのプラットフォーム構築

- 具体的な新産業創生の場をベンチャー企業等へ提供





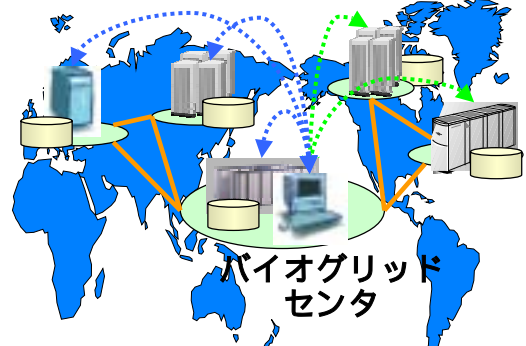
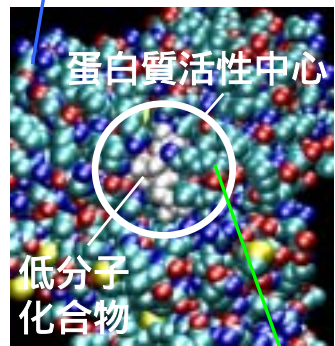
グリッドによる分子動力学と分子軌道法の融合と その創薬への応用に関する研究

候補新薬と蛋白質受容体との相互作用エネルギーを高速且つ高精度に解析するため、異なるレベルのシミュレーション計算を複合化した*In silico*創薬ソフトウェアの開発と実証

分子動力学の解析領域

分子動力学計算

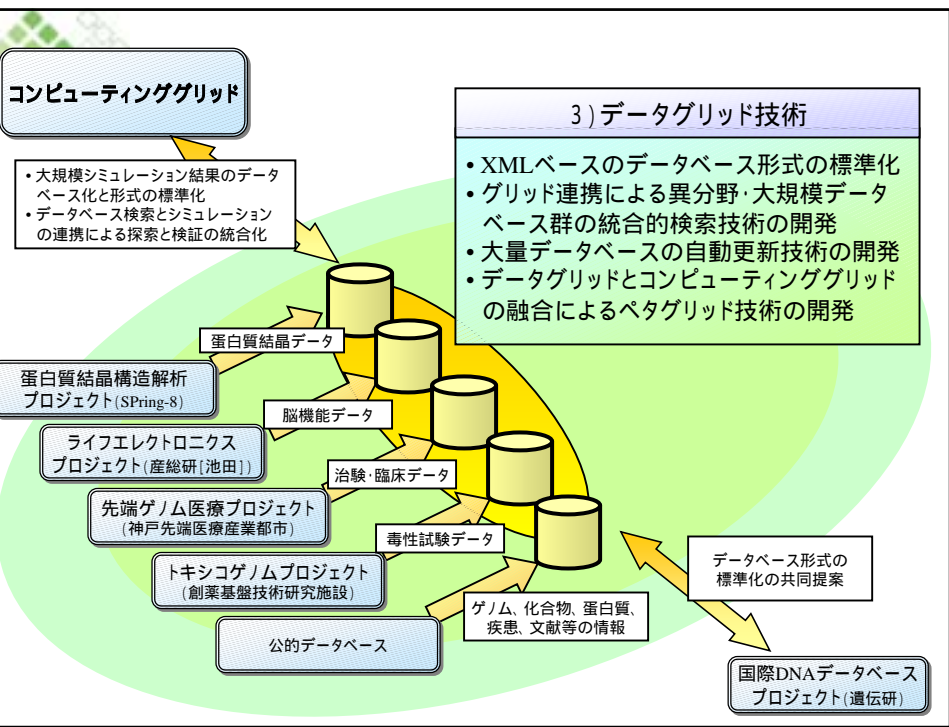
分子軌道計算



分子軌道法の解析領域

グリッド環境での連成解析

何(日立)





「統合データベース」と「データベース連携」

統合データベース

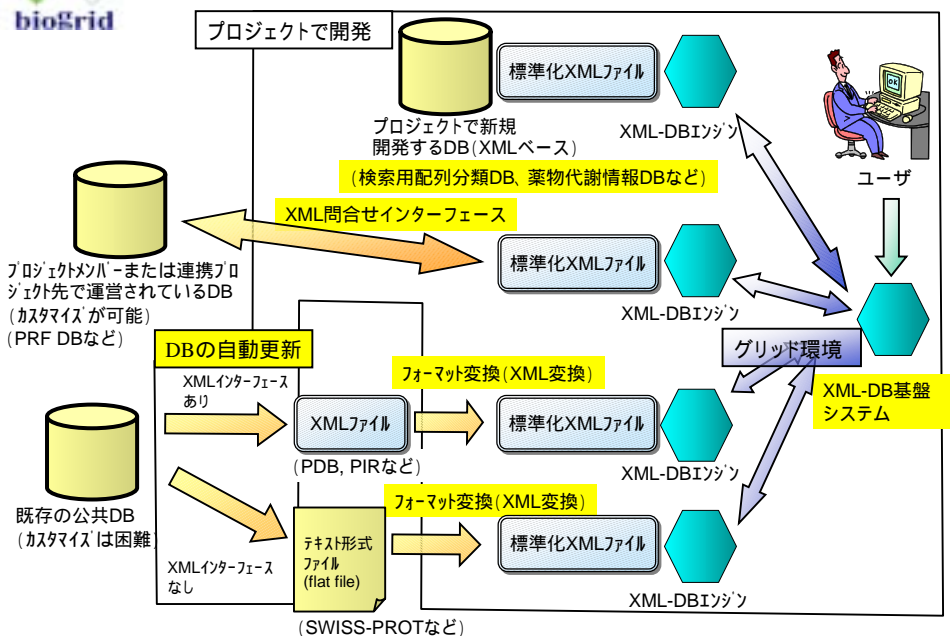
- 検索対象となる複数のデータベースを静的に結合して、物理的に単一のデータベースを構成。

データベース連携

- 検索対象のデータベースはそのまま保持し、検索時に各データベースに問合せを出して、その結果を動的に結合する。

	システム構築コスト	検索コスト	DB更新コスト
統合DB	× (統合DB構築に膨大なコスト)	(検索は通常のDBと同じ)	× (検索対象DB更新のたびに大きな更新コストが発生)
DB連携	(検索対象DBをそのまま保持)	(検索時に各DBの検索結果の結合が必要)	(検索対象DBの更新をそのまま反映させるだけ)

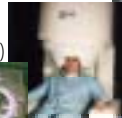
バイオデータグリッド



4) データ遠隔収集技術

bio 世界最高性能を誇る観測機器を超高速ネットワークで共有

- 超高圧電子顕微鏡 (UHVEM)
 - 大阪大学超高圧電子顕微鏡センター (大阪府茨木市)
- 脳磁計 (MEG)
 - 産業技術総合研究所ライフエレクトロニクスラボ (大阪府池田市)
- Spring-8
 - 高輝度光科学研究センター (西播磨)

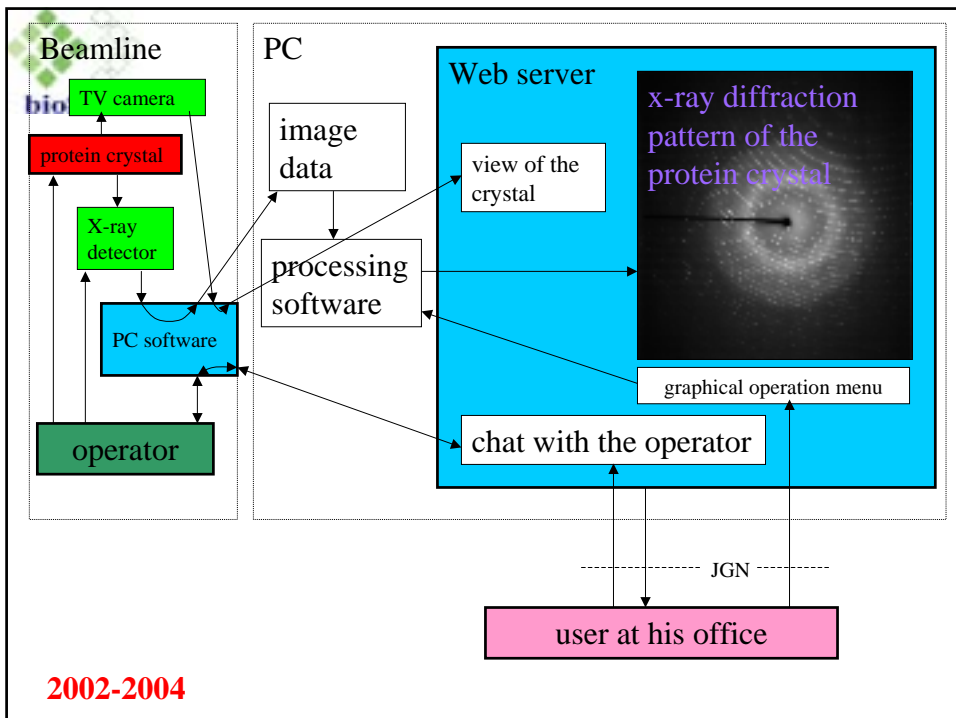


データ観測

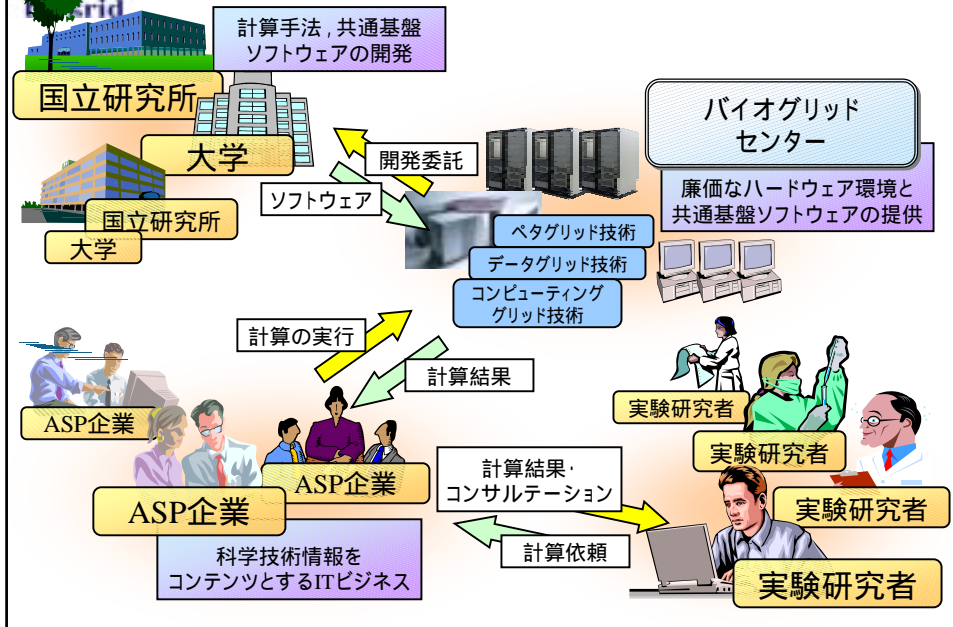
解析、シミュレーション

データグリッド、Intelligent Gridへの融合

- 仮想研究空間の実現 (E-Science)
 - データ解析手法、シミュレーション手法の共有 (知識の共有)
 - シームレスなオンライン解析の実現
 - 観測機器の遠隔操作
 - 観測データ、可視化データ等の可視化実現へ



バイオグリッドとASPビジネスモデル



bioGrid

MEGrid Project Overview

- Gridの特性を活かして、データや情報の移動 / 蓄積 / 共有 / 解析 / 可視化等のフローを統合

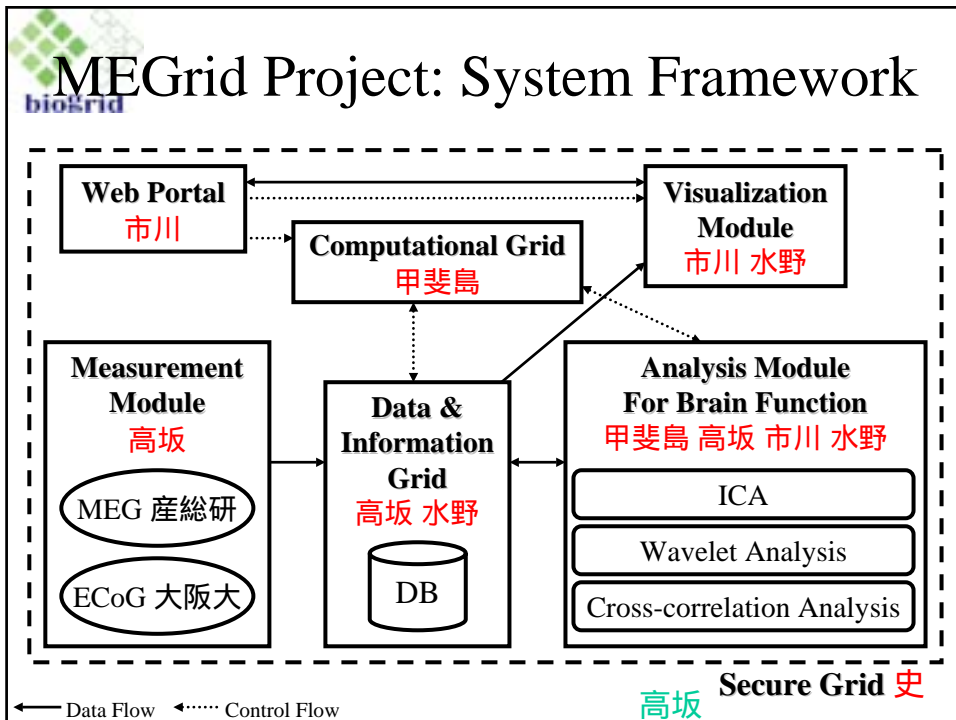
Computing

MEG

Data Generation

高坂 診断

研究



産総研MEG-阪大クラスタ間の接続

bioGRID

- 5月～9月にかけて作業・調整を行い、産総研@池田のMEGをネットワークに接続
- 計測システムのメモリから直接データを読み出している
- 計測中のデータをリアルタイムで伝送可

MEG@産総研

1GB線

データ量: 数十MB/分

阪大クラスタ

高坂



産総研MEG - 阪大クラスタ間の 接続(続き)

- 人脈確保 / 事務的手続きが大変
 - 産総研のアカウント取得
 - MEGのメーカやその代理店との連携
- 作業が大変
 - 施設予約が一杯: 夜中しか作業できない
- セキュリティポリシーの壁
 - 外部ネットワークに接続できない(物理的に接続を切り替えることで対処)
 - 外部から測定機器の操作はできない(ファイアウォール問題など)
- 技術的困難は比較的小

高坂



Computational Grid

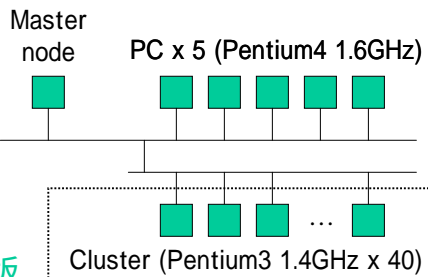
- 現在MEGridで用いている解析手法
 - Wavelet解析
 - Cross-correlation解析
 - ICA
- 目標: 解析時間の短縮
- アプローチ: ジョブに適したGrid用スケジューリングシステムの設計 / 実装

高坂

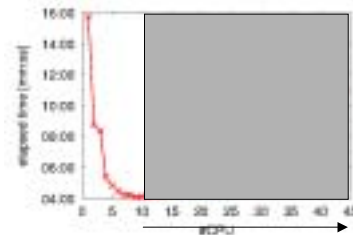


スケジューラの必要性

- 並列化の実験 (ICA)
 - PC x 5台と40ノードのClusterからなるグリッド環境
 - 1CPUで16分の解析時間が10CPUで4分に減少
 - 10CPUより大きくなると解析時間が増大: 同期取りの手間
 - 必ずしも計算機を増やせば増やすほど速くなるとは限らない 適切なスケジューラが必要



高坂



User Interface

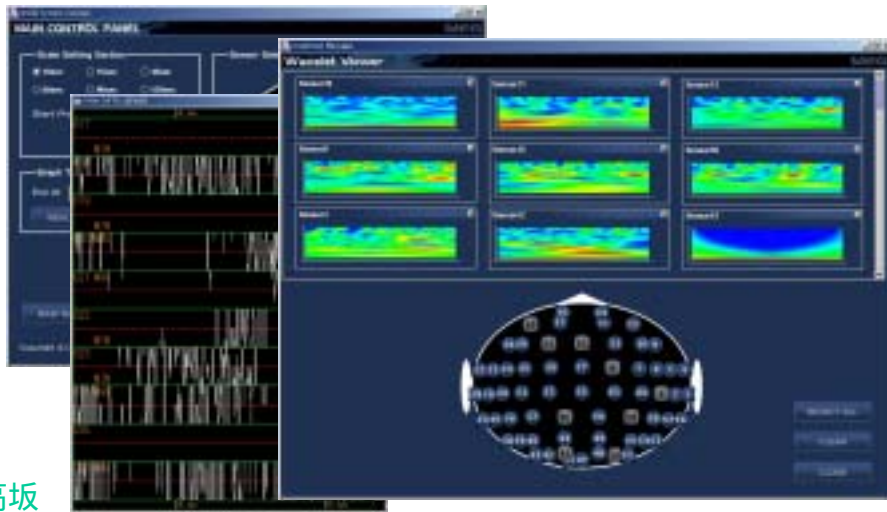
- Gridを「普通の」ユーザにも使えるものに
 - コマンドライン GUI
 - バッチ処理 インタラクティブ処理
- Webインタフェースを用いたGridポータル環境の構築
 - 認証
 - 計算要求の投入
 - データや解析結果の可視化

高坂



データ / 解析結果の可視化

- 現在構築中の可視化ソフトウェア



高坂

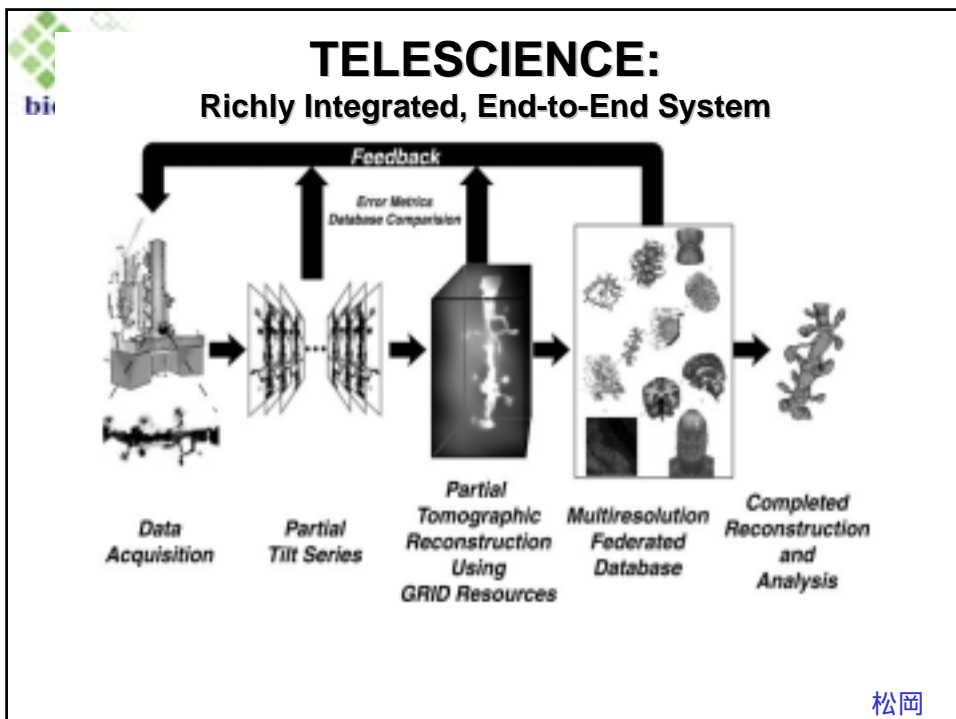


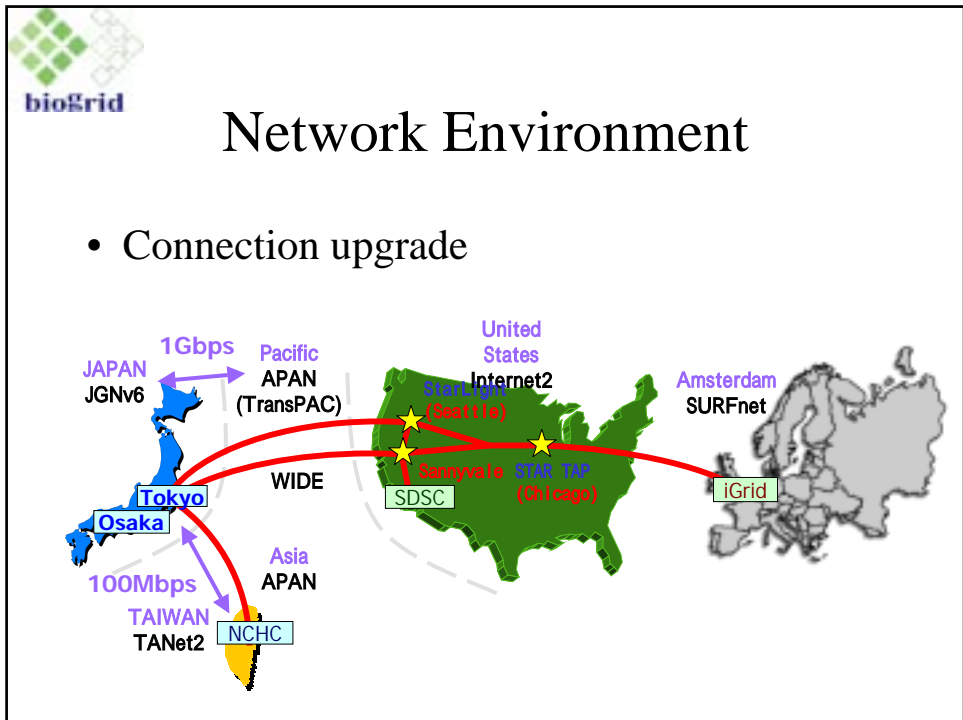
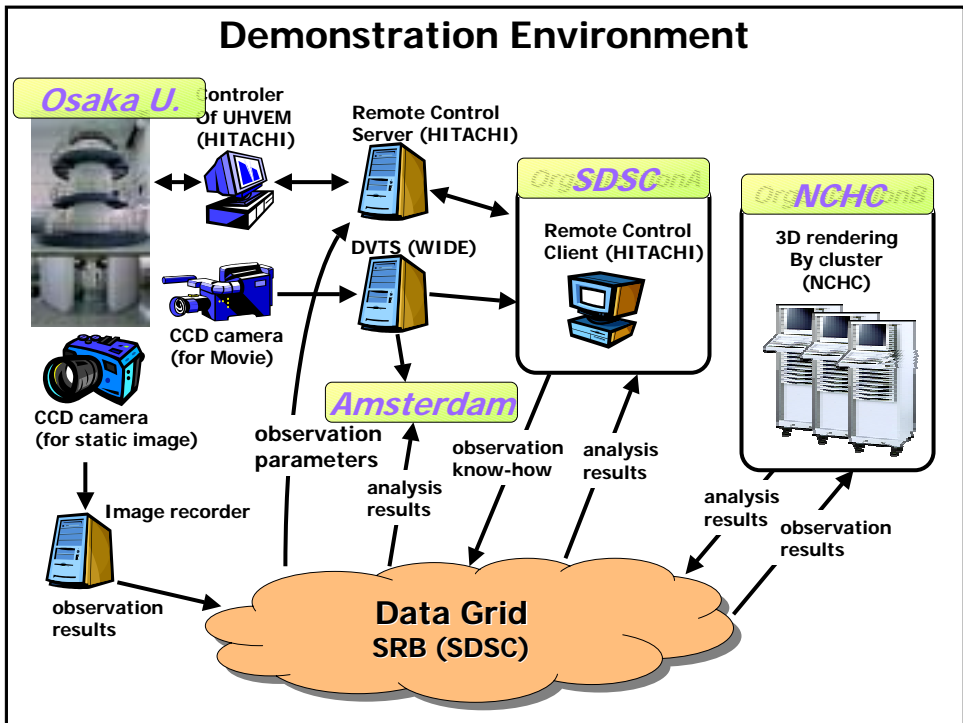




Telescience with Ultra-high Voltage Electron Microscopy

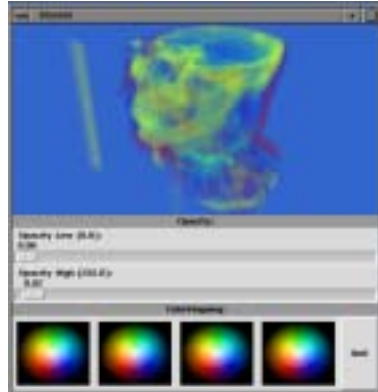
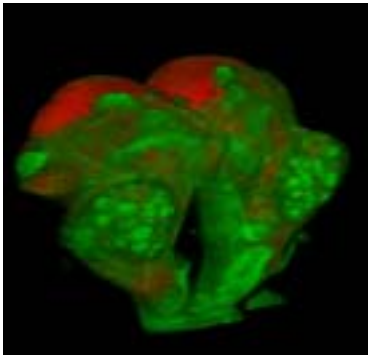
- 世界最大 (3MV) の超高圧電子顕微鏡
- telescience portalとして統合
 - SDSC (米国), NCHC (台湾)との国際連携
- ネットワークの課題
 - IPv6での広帯域伝送







Advanced Visualization Technologies



Model-based/Image-based VR

Slide courtesy of National Center for High Performance Computing
(<http://www.nchc.gov.tw>)





ネットワークへの課題

- 帯域は十分だったはずだが...
- アプリケーションごとに時間、性質、相手の違う要求
 - Gfarm, data resavoir
 - Telescience
- ミドルウェアとネットワークの協調
 - On demand overlay network
- IPv6による挑戦
 - 常に新しい課題が



SC2002 Bandwidth Challenge Results.htm

