

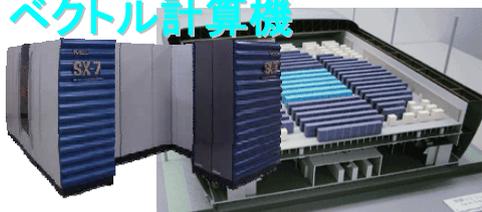
# 次世代スーパーコンピュータ 開発と応用研究

理化学研究所  
情報基盤センター  
次世代スーパーコンピュータ開発実施本部  
中央研究所生体力学シミュレーション特別研究ユニット  
姫野龍太郎  
himeno@riken.jp

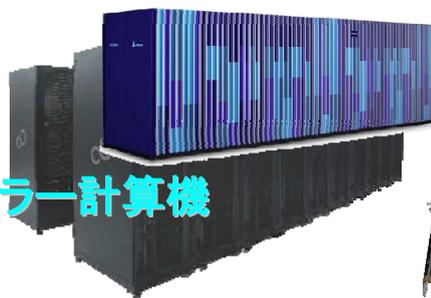
## 従来技術

- スーパーコンピュータは大別して3種類
  - ベクトル計算機: 地球シミュレータ、NEC SX8等
  - スカラー計算機: 富士通HPC2500、日立SR1100等
  - 専用計算機: 分子動力学専用機(MDGRAPE)等
- それぞれに特質があり、計算により向き不向き
- 従来は一種類のスーパーコンピュータだけを備えて全ての用途にサービス

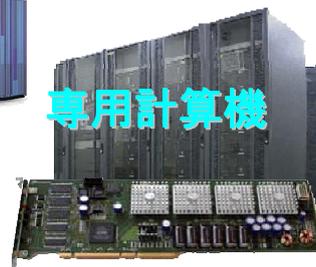
ベクトル計算機



スカラー計算機



専用計算機





# 2004年導入したRSCC

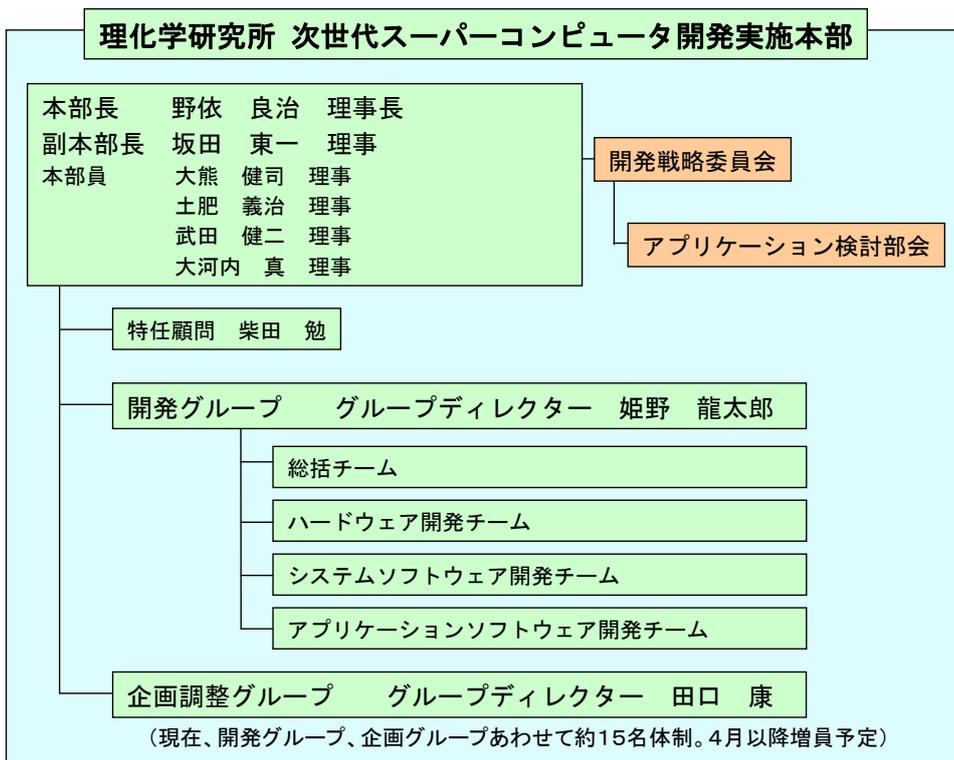


## 【チャレンジ】

- ・世界初のスカラ+ベクトル+専用機複合システム
  - 一つの計算機に異なる機種のコパイラ
  - 機種で異なるバイナリーファイルを統一
- ・日本で初めてグリッド技術を全面的に採用した計算機センター
  - 利用者に利用計算機を意識させない
- ・世界最大規模、日本で最速のPCクラスタ
  - Top500リスト(2004年6月)第7位
  - 高性能で低コスト



# 平成18年1月開発実施本部設置





# Schedule (draft)

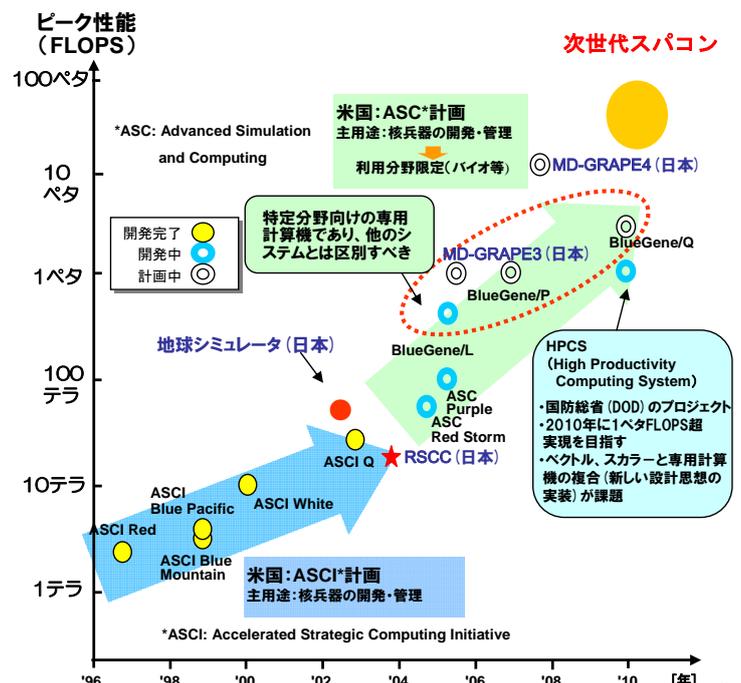
FY	2006	2007	2008	2009	2010	2011	2012
Event & Evaluation	★CSTP follow-up		★CSTP follow-up		★CSTP follow-up		★CSTP follow-up
R&D	★CSTP follow-up		★CSTP follow-up		★CSTP follow-up		★CSTP follow-up
System Software	GRID middleware etc., Design&Production			Evaluation			
	Next Generation NanoTechnology Simulation, Design&Production			Evaluation			
	Next Generation Life Science Simulation, Design&Production			Evaluation			
	Next Generation Life Science Simulation, Design&Production			Evaluation			
Grand Challenge Application Software	Next Generation NanoTechnology Simulation, Design&Production			Evaluation			
	Next Generation Life Science Simulation, Design&Production			Evaluation			
Hardware	Design		Implementation Technology Design & Evaluation		Production	Enhancement	
File System, etc.			Design	Production		Enhancement	
Geographical investigation, Construction	Investigation	Design	Construction				

CSTP: Council for Science & Technology Policy (CSTP) is one of the four councils of important policies of Cabinet Office.

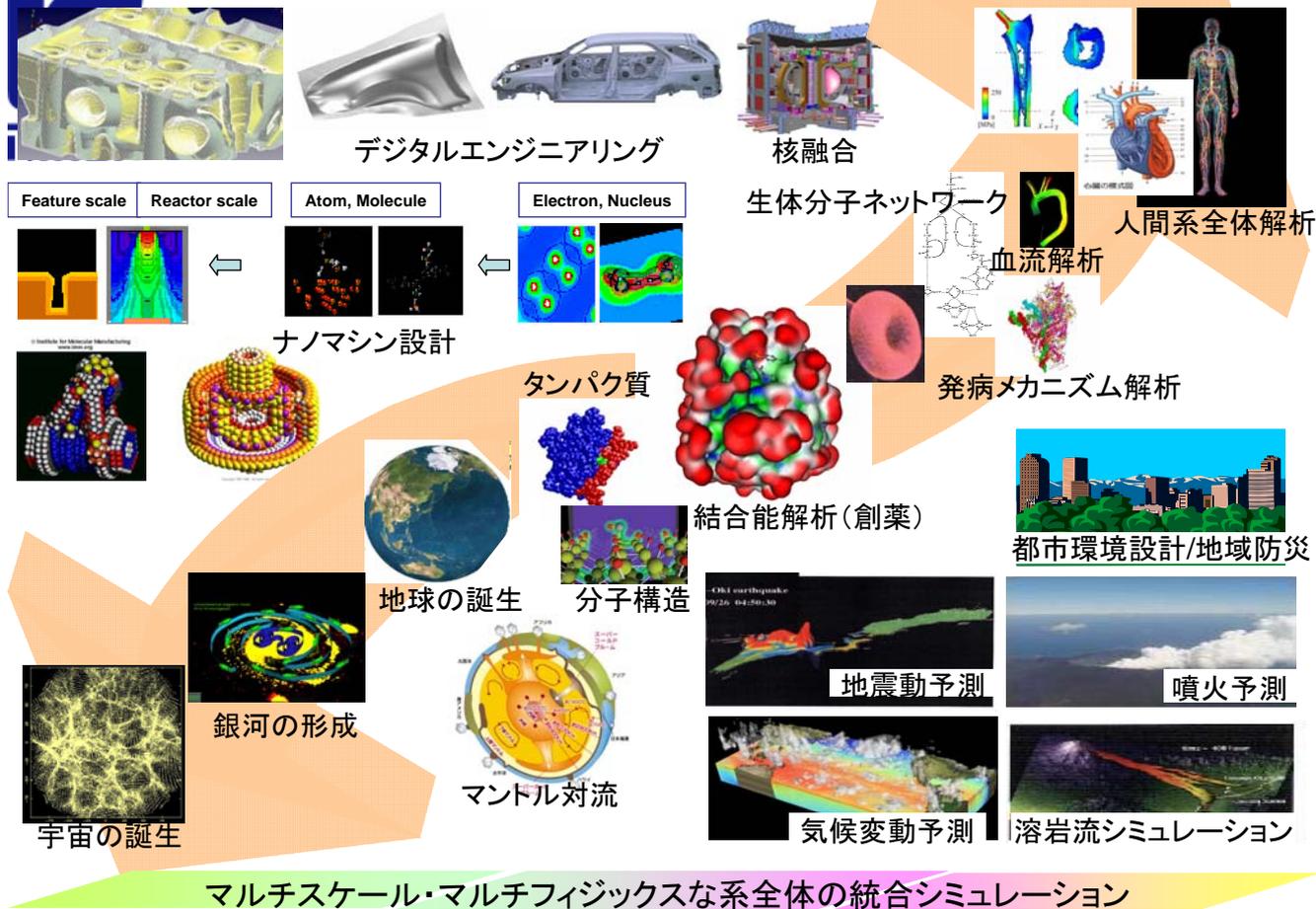


## 日米のスパコン競争

- 例えば、2010年 10PetaFLOPS: 地球シミュレータの250倍
- 従来の延長上では、電力150MW、スペース 75,000m<sup>2</sup>



# ペタフロップス超級SuperComputingが待たれている世界



## 性能向上は何に使う？

- 例えば流体の計算
  - 各方向2倍のメッシュ→8倍の計算点→時間方向も倍の計算量があるので、16倍の計算量 $n^4$
  - 100倍の性能向上→各方向3.16倍の密度向上
- 普通の研究では
  - 普通の大さの計算でパラメータをいくつも変えて計算、変化を調べる
  - たった一つの計算だけができれば良いのではない



## 世界最速は必要か？

- 世界最速を10年に1回とるより、数年ごとに性能向上を図る、あるいはより広い使われ方をする方が得策
  - 研究者の育成
  - 利用技術の向上
  - 産業界への普及
- ただし、世界でここだけしかできない計算を可能にすることは価値あり



## 性能向上は何に使うか？

- ある分野に特化した世界最速の計算機：新しい研究ツール
  - 世界でここだけしかできない計算
  - まだ、誰も手がけていない計算
  - これまで不可能だと思われてきた計算
- 従来規模の計算は、従来の大学計算機センター、研究所の計算機を利用
- 教育センターを併設
- 産業利用は促進



## 従来の計算機では不十分な計算

- バイオ・ナノの第一原理計算、量子化学計算
- うまくいけば
  - 新薬の開発、新材料・触媒の開発
  - 生命現象そのもののコンピュータシミュレーション



## 10PetaFLOPSで可能になること

- タンパク質の機能解析
  - 100万原子でも1日に20ナノ秒計算可能 (Direct)
- フォールディング
  - $N=10^4$
  - 合計シミュレーション時間 ~1ミリ秒/12週間
  - 比較的速くフォールドするタンパク質であれば、その過程を調べられる
- 高精度 *in Silico* スクリーニング
  - 5000 サンプル/日



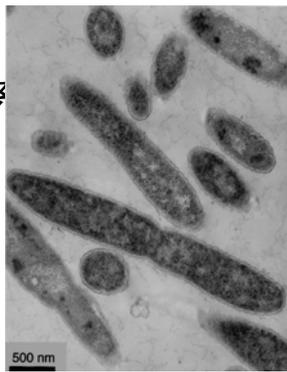
# 何を想定しているか(1)

実験データ解析から計算機シミュレーションへ

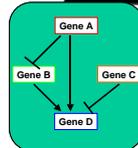
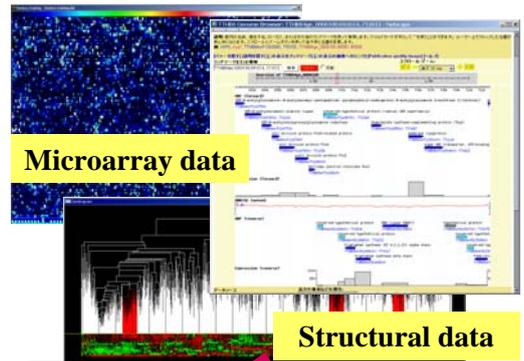
## 1) 生体分子の構造と機能

- DNA複製系・転写系・翻訳系
- DNA修復系
- アミノ酸・糖質・脂質代謝系
- ヌクレオチド代謝系
- エネルギー代謝系
- 輸送系・細胞膜系
- シャペロン
- その他

- 核酸・アミノ酸・糖・脂質
- 補酵素・金属イオン
- その他

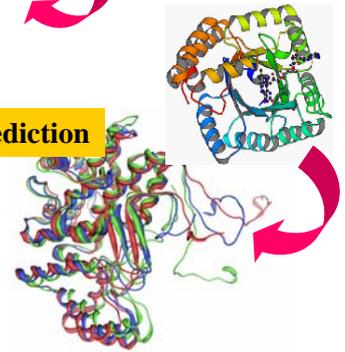
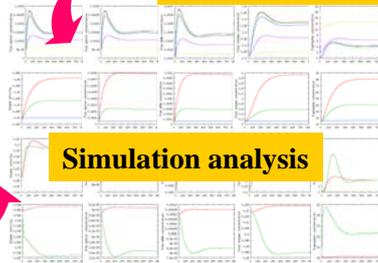
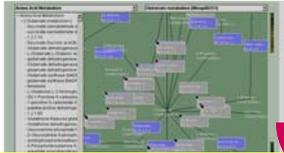


高度好熱菌



## 2) 各生体分子の存在場所・濃度, さらに, それらの時間依存性

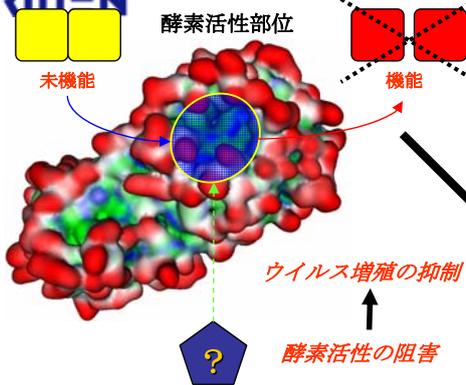
## 3) 細胞の全生命現象のシミュレーション



# 何を想定しているか(2)

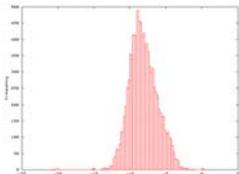
In silico スクリーニングによるヒット化合物の探索

標的タンパク質: SARSウイルス タンパク質分解酵素 (ウイルスの増殖に関与)



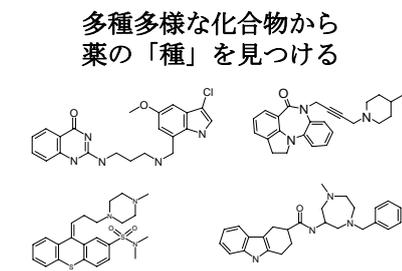
コンピューター上で実際の相互作用の様子をシミュレーションする

現在はスクリーニングの精度が悪い



スクリーニング結果を解析しヒット化合物を選択する

高精度化が必要



実験によるリード化合物の探索へ



# 何を想定しているか(3)



GSC林崎PD

ゲノムネットワーク研究と機能性RNA研究における計算量とソフトの問題

現状の成果

- ゲノム解析
- 完全長cDNAの網羅的収集とアノテーション
- トランスクリプトーム解析

ゲノムネットワーク研究における課題

[ゲノム機能情報の解析]

- ヒトゲノムの発現調整領域に関する情報の網羅的収集と解析
- 生体分子(タンパク質など)間相互作用に関する情報の網羅的収集と解析
- 全ゲノムTiling Arrayによる1塩基精度の発現解析
- 全ゲノムTiling Arrayによるクロマチン構造の解析
- [個別生命機能解析]
- 疾患解析 (がん、動脈硬化、高血圧、糖尿病など)
- 生命現象解析(発生、分化など)
- 薬の標的分子などの解析 等々

機能性RNA研究の課題

- RNAの生体高分子の構造と機能の解析
- 体系的な分子計算の理論モデルによる解析
- 分子動力学シミュレーション
- 高次構造モチーフ検索

数年後のトピックス

- \$1000ゲノム解析技術の商用化
- 全ゲノムTiling Arrayによる個人個人のゲノム全長配列解析、及び全SNP解析

計算量とソフトの問題の例

[超大規模なデータ量と解析計算量]

- 例)相互作用や制御関係の組み合わせ:  $-10$ の $16$ 乗( $10$ ペタ)の組み合わせ
- 例)1塩基Tiling Array:  $-5.6$ TB/サンプル  $\rightarrow$   $10$ 万サンプルで $560$ ペタB
- 例)発現調整領域解析のクラスタリング計算:  $-15$ 万個の計算に $256$ GBメモリの計算機で $1$ 週間
- 例)機能性RNAの高次構造モチーフ検索:  $-1$ モチーフ検索に $10$ 日間(現在 $379$ モチーフ)
- [並列化アルゴリズム開発の必要性]
- 例)前記解析のクラスタリング計算:  $-$ 現在 $1$ 週間の計算時間を数時間に

期待される成果

- 病気の原因から発症までのメカニズムの解明
- 健康な生活の実現
- 新たな治療法や創薬の開発
- 経済活性化の実現



# 何を想定してるか(4)

## マルチスケール人体シミュレーション

マイクロからのアプローチ

マクロからのアプローチ

MD・第一原理・量子化学シミュレーション

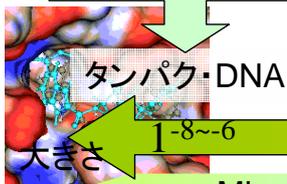
連続体シミュレーション

タンパク構造解析  
薬剤反応性解析

分子ネットワーク  
解析

流体・熱・構造  
化学反応の連成

血管系モデリング  
骨格モデル



大きさ 1-8~-6

1-5~-4

1-3~-2

1-1

10

Micro

Meso

Macro

薬剤開発

DDS  
再生医療

HIFU

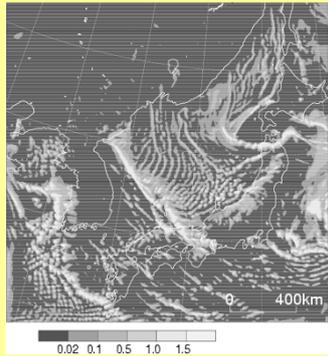
結石破碎  
外科手術  
カテーテル  
マイクロマシン

治療技術



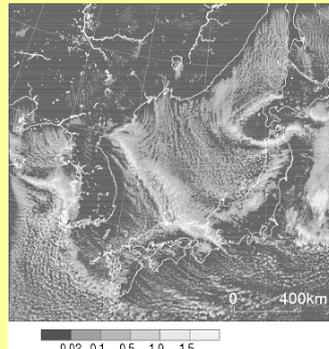
# 何を想定しているか(5)

## 現在の雲解像モデル解析



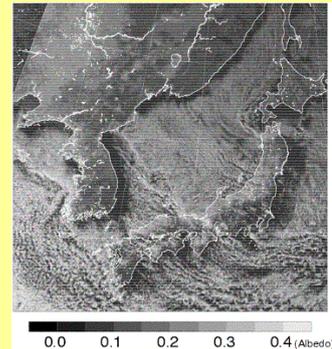
5 kmメッシュアジア域  
現時点での雲解像モデル解析

ベクトル：約1テラFLOPSで24時間分予測、2時間計算



1 kmメッシュアジア域  
地球シミュレータによる雲解像モデル解析

ベクトル：約10テラFLOPSで24時間分予測、12時間計算



実画像(ひまわり5号)

地球シミュレータの利用によって、雲解像モデル解析精度が向上

## 将来の雲解像モデル解析

10ペタFLOPSベクトル機では

- 1 kmメッシュ全球 24時間分予測 2時間計算
- 1 kmメッシュアジア域 24時間分予測 1分計算

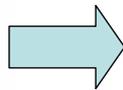
提供：気象研究所



# 何を想定しているか(6)

部品単位 → 全体

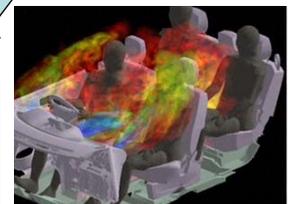
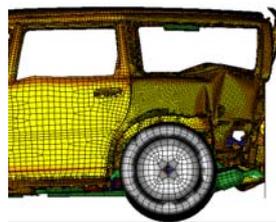
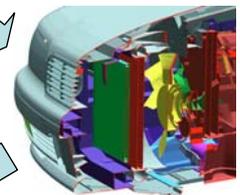
一現象 → 複合



金型・プレス機も含めたシミュレーション

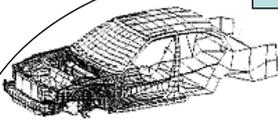


全ての部品を入れてモデル化

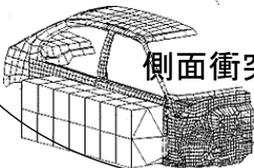


### モデルの統合

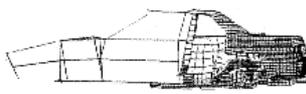
前面衝突用



側面衝突用



後面衝突用

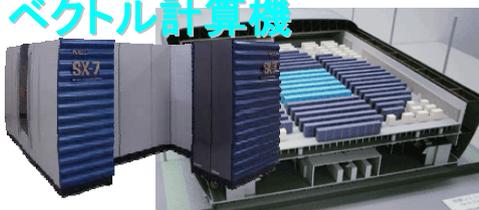


# 今後の応用の方向性と 計算機アーキテクチャ

## スーパーコンピュータの従来技術

- スーパーコンピュータは大別して3種類
  - ベクトル計算機: 地球シミュレータ、NEC SX8等
  - スカラー計算機: 富士通HPC2500、日立SR1100等
  - 専用計算機: 分子動力学専用機(MDGRAPE)等
- それぞれに特質があり、計算により向き不向き
- 従来は一種類のスーパーコンピュータだけを備えて全ての用途にサービス

ベクトル計算機



スカラー計算機



専用計算機





## 今後の応用分野の方向

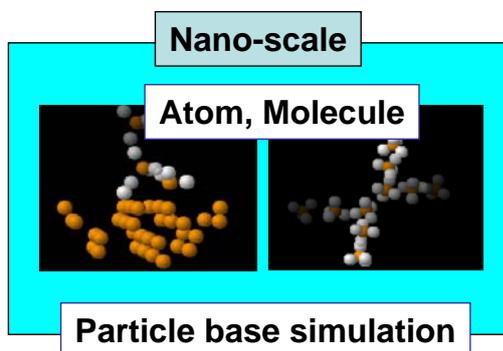
- Multi-physics
  - 流れ、構造、熱、音、電磁場、化学反応など
  - 複数の支配方程式
- Multi-scale
  - 原子レベルから人間のスケール
  - 複数の支配方程式



単一のアーキテクチャーではカバーできるだろうか

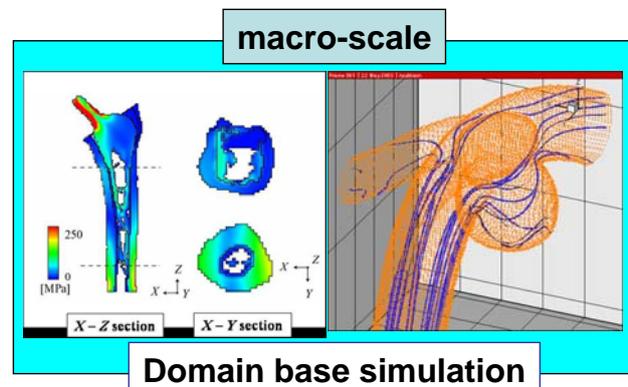


## Multi-scale simulation



Small data, Large computation

SPC:GRAPE, MD:GRAPE



Moderate amount of data and computation

Vector / RISC



# 2004年導入したRSCC

【システム構成】  
スカラ+ベクトル+専用機  
の複合システム

入出力機器

【スカラ部】  
12.4テラフロップス

【ベクトル部】  
0.28テラフロップス

【システム間接続】  
1ギガビット毎秒のネット  
ワークを使いグリッド接続

【専用部】  
1.7テラフロップス

## 【チャレンジ】

- ・世界初のスカラ+ベクトル+専用機複合システム
  - 一つの計算機に異なる機種のコパイラ
  - 機種で異なるバイナリーファイルを統一
- ・日本で初めてグリッド技術を全面的に採用した計算機セ
  - 利用者に利用計算機を意識させない
- ・世界最大規模、日本で最速のPCクラスタ
  - Top500リスト(2004年6月)第7位
  - 高性能で低コスト

産業技術大賞  
文部科学大臣賞・受賞  
2005年4月

京速計算機のフィージビリティ  
スタディーに利用する予定



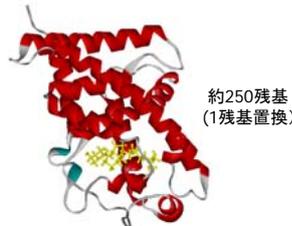
# 現状のハードウェアの問題

マルチスケール・マ  
ルチフィジックスの  
シミュレーションの  
実現

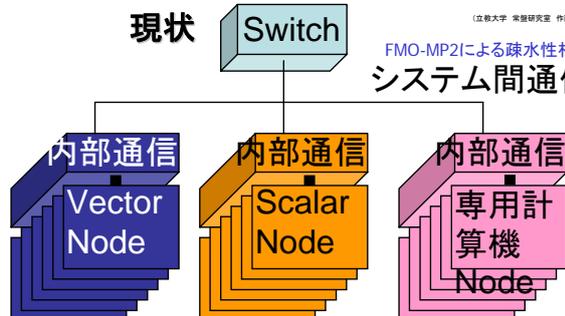
システム間  
通信の能力  
不足

それぞれの計  
算機システム  
の結合

例えば、QM/MM連成、亀  
裂から破壊にいたるシミュ  
レーション、燃焼シミュレ  
ーションなど



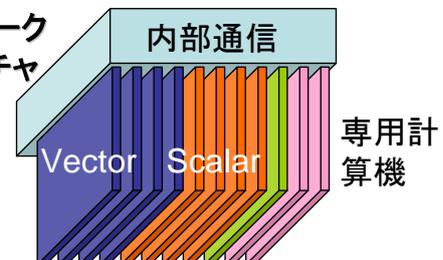
現状



FMO-MP2による疎水性相互作用の扱い  
システム間通信

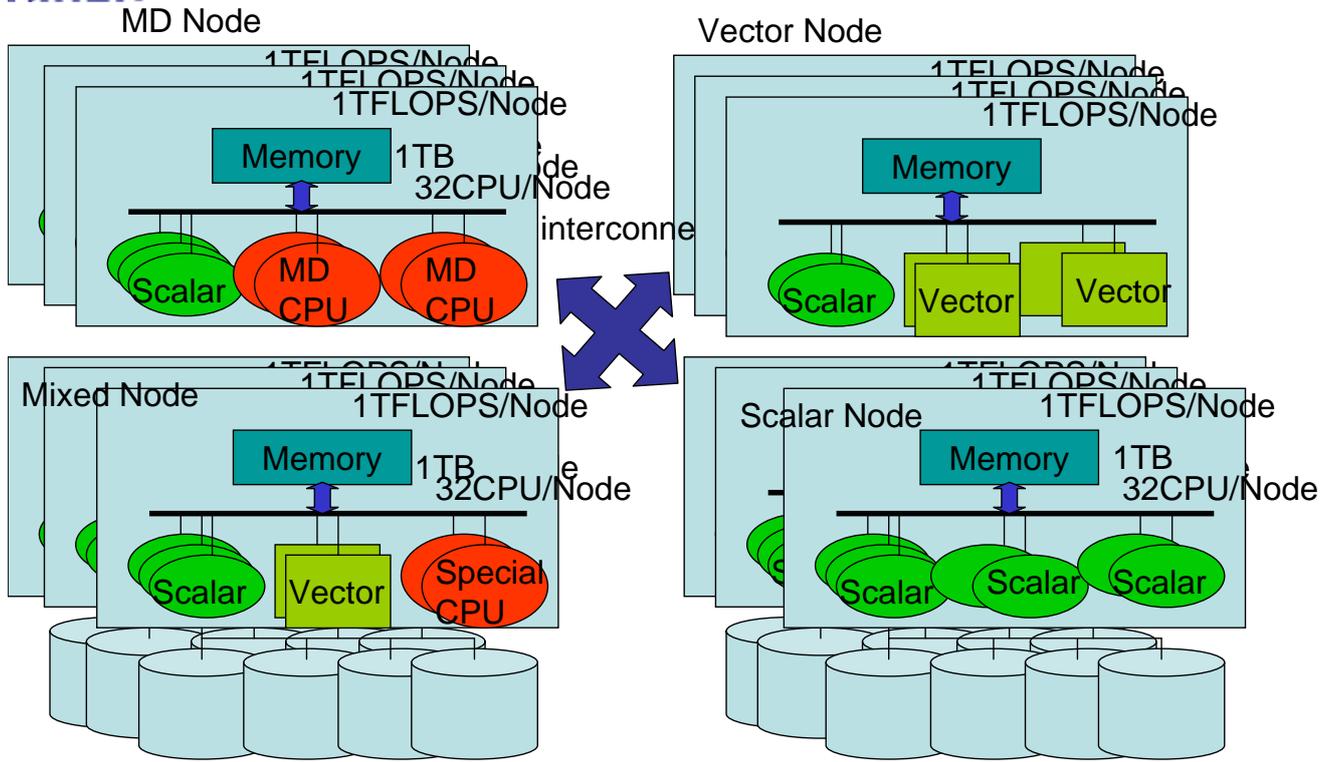
統合された複合計算  
機システムの必要性

新アーキ  
テクチャ

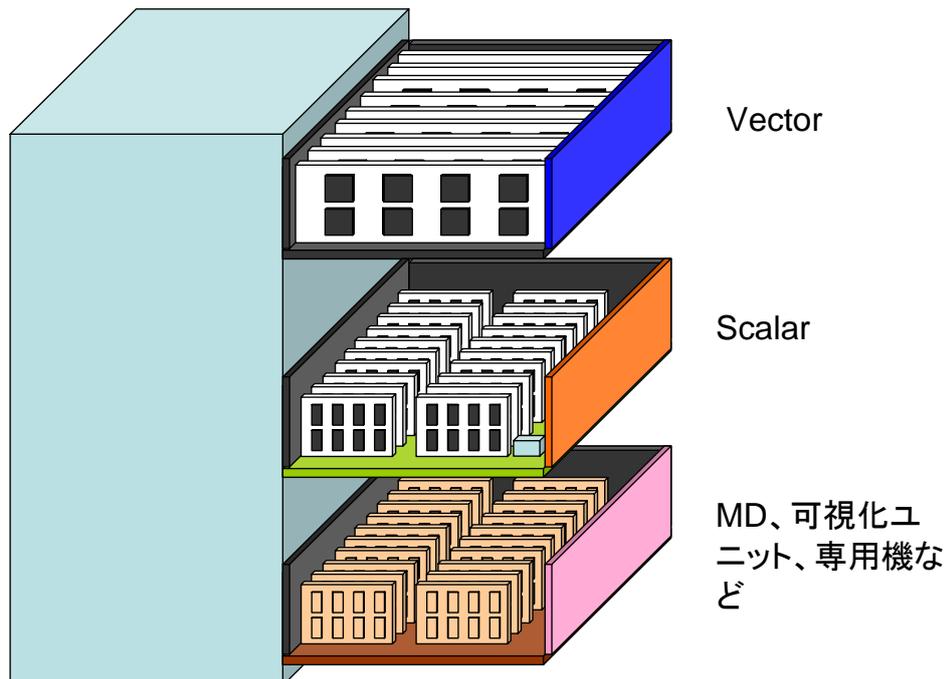




# システムのイメージ



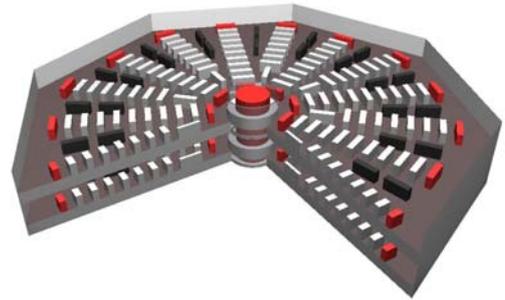
# 筐体イメージ





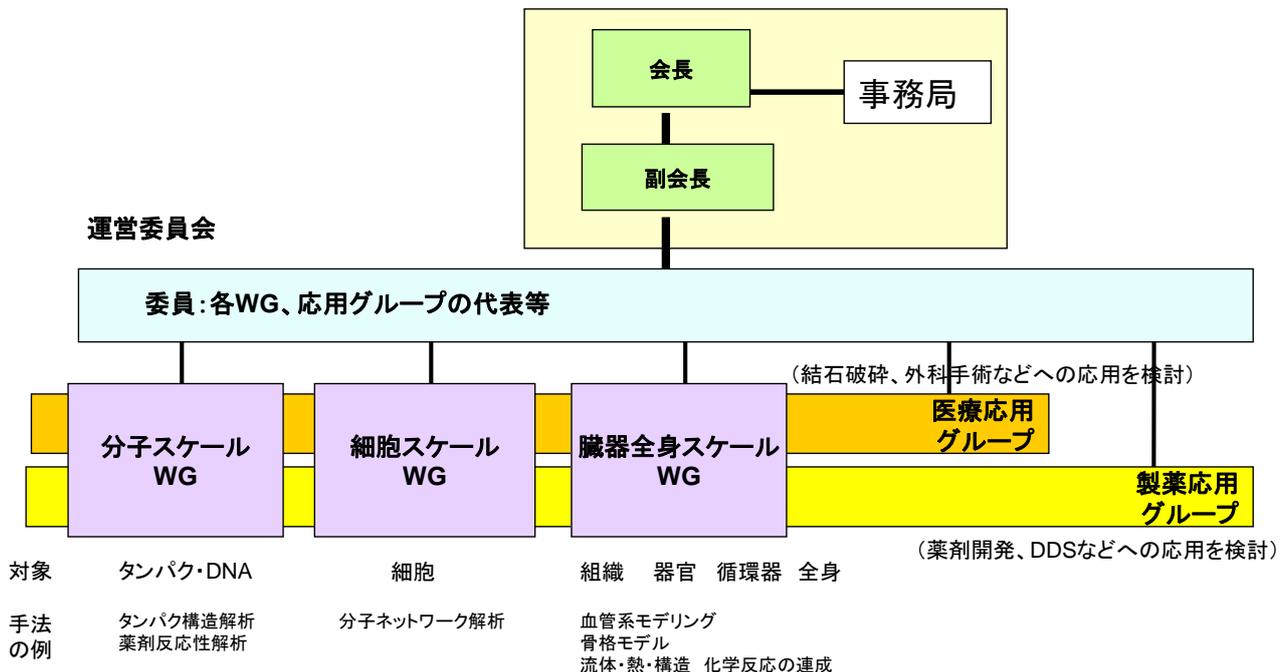
# 現時点での目標

- 理論ピーク性能
  - 10 Peta FLOPS
- 実効性能
  - 各種のアプリケーションで1.0 PetaFLOPS以上
- Memory 0.75PB
- 45 nm processing technology
- 光通信、光スイッチ
- 消費電力
  - 35MW
  - コージェネによる発電: 14MW
  - 蒸気炊き吸収式冷凍機の利用
- 設置延べ面積: 10,000平方m



## 次世代生命体シミュレーション研究会(素案)

- 運営委員会で、全体の計画、方針付けを行う
- 分子スケール、細胞スケール、臓器全身スケールの各WGで研究者が具体的な検討を行う
- それらのWGを横串で結びつける、医療応用グループ、製薬応用グループを設置する



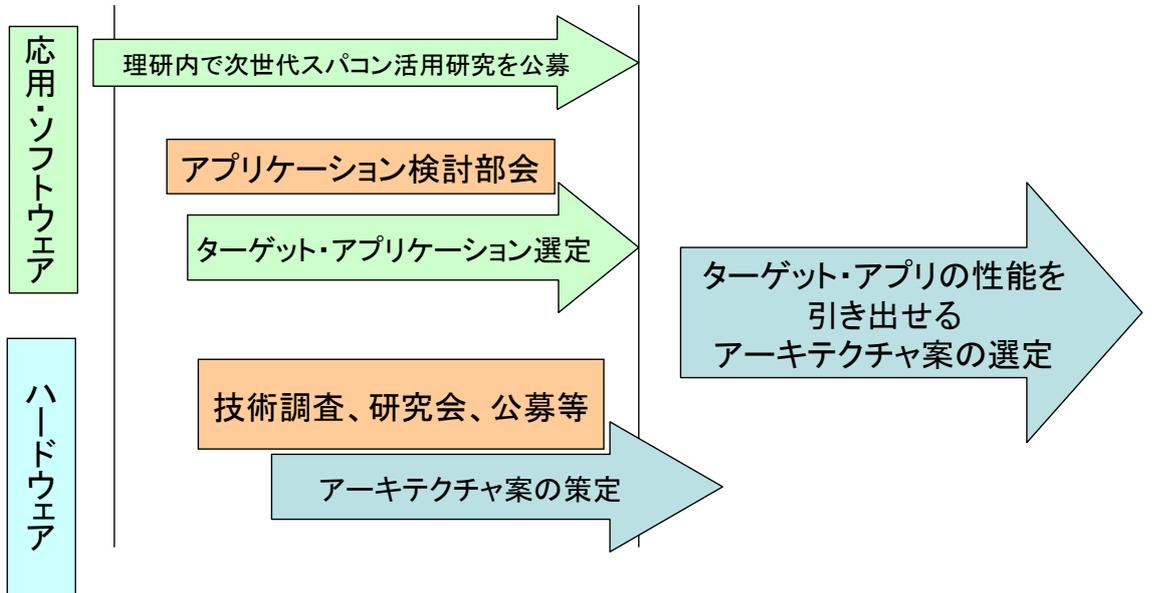


# アーキテクチャ策定に向けて作業中

平成18年1月

4月

7月



## まとめ

- 理化学研究所でライフ・ナノ・サイエンスの研究進展と、それ以外の領域でも十分な能力を発揮できるような複合計算機システム構想をまとめた。
- 2006年1/1、理研に次世代スーパーコンピュータ開発実施本部設置。現在、この構想実現に取り組んでいる。
- グランドチャレンジ問題としてナノとライフサイエンスが設定されている