

# e-サイエンスに向けた革新的アルゴリズム基盤

加藤直樹

京都大学大学院工学研究科

2011年11月22日

## 背景

- 平成22年3月の学術会議による大型研究計画の策定
- 「大型施設計画」(素粒子・原子核物理学、天文学、宇宙空間科学、核融合科学、地球科学など)は**日本は世界の第一線**
- 生命科学、地球環境科学など広範な学術の諸分野においても「**大規模研究計画**」の実施が、国際的視点も加えて緊急の課題
- 大型施設計画および大規模研究計画の遂行には、科研費等の枠では賄いきれない多額の予算が必要
- 大型計画は、新たな科学と技術の限界への挑戦であり、フロンティアを切り開き新たな知を創造する先端研究である。

## 背景

- ・ 大型計画により生み出される成果は、我が国の国際的地位を高め、広い関連分野の研究・教育を育て加速させるとともに、萌芽的研究を育成する研究基盤を広く強化することができる。
- ・ 学術の大型計画の継続的推進は緊急の課題
- ・ 国際的な共同協調に関する迅速で強力な対応が必要
- ・ 国民の、そして科学者コミュニティの理解が得られるような、科学に基づく透明なアセスメントの必要性
- ・ 大型施設計画を長期的かつ組織的に推進する仕組みの明確化
- ・ 長期間にわたって多くの研究者を組織し、通常の競争的経費では実施が困難であるような予算を要する「大規模研究計画」の必要性

## 背景

- ・ 科学的で透明性の高い評価および、所要経費、計画期間、期待される成果などについてのマスタープランの策定と、確実な推進の体制の必要性
- ・ **日本学術会議： 科学者コミュニティの専門的意見を集約して、大型施設計画および大規模研究計画の検討を行い、わが国として初めての全分野にわたる大型計画のマスタープランを策定**
- ・ **43計画を策定**

## 背景

- ・ 昨年12月から改訂作業 107の新規提案
- ・ 46計画採択(新規計画 10提案)
- ・ <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-21-h135-1.pdf>
- ・ <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-21-h135-1-2.pdf>
- ・ <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-21-h135-1-4.pdf>
- ・ 情報学分野では3計画が採用
  1. 国民生活を支える実空間型情報学基盤の研究(坂村健)
  2. 大規模計算・科学データ共有のためのアカデミッククラウド基盤(坂内正夫)
  3. e-サイエンスに向けた革新的アルゴリズム基盤(加藤直樹)

- ・ Jim Gray 博士：E-サイエンスの重要性を提唱した
- ・ **第1パラダイム数千年前：経験科学**  
ニュートンの法則、マクスウェルの方程式など、理論中心の科学
- ・ **第2パラダイム数百年前：理論科学**  
複雑な現象をシミュレーションにより予測・再現する科学
- ・ **第3パラダイム：数十年前：計算科学**  
複雑な現象をシミュレーションにより予測・再現する科学

## e-サイエンスとは

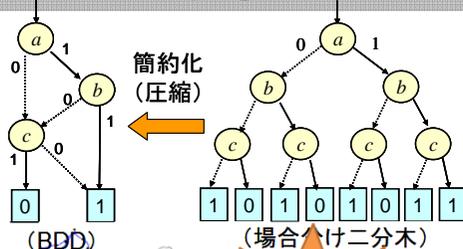
- **第4パラダイム: E-サイエンスまたはデータセントリックサイエンス (データ中心科学)**
- **理論、実験、およびシミュレーションを統合**
- **データ検索、データマイニング技術を用いる。**
  - **実験器具からのデータを取得**
  - **シミュレーションからデータを生成**
  - **ソフトウェアによって処理**
  - **科学者は、データベースやファイルを分析**

- **3月5日 学術会議 情報学シンポジウム**
- **Prof. Hopcroft 特別講演**  
**今後重要な研究課題：**
  1. Large graphs (数十億点)の性質
  2. clustering
  3. extract signal from noise
  4. extract information from large data
  5. ranking
  6. collaborative filtering
  7. dimension reduction

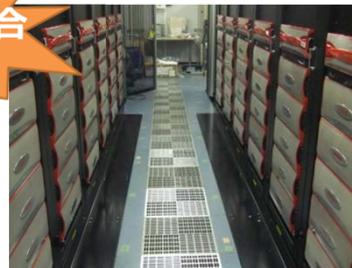
計画の目的、概要と科学的な意義

- 従来手法では解決不可能な大規模かつ複雑な諸問題を数理解析に基づく**革新的なアルゴリズムによって解決**することを目指す。
- アルゴリズム科学と実問題の数理的モデル化と解決のため**共同研究拠点**を作る
  - アルゴリズムの設計と応用の科学的体系を構築し、HPC技術、ソフトウェア実装技術等の関連分野を融合した問題解決手法の開発と標準化を行い、地球規模のネットワークや高密度な計算環境での高度活用を実現する情報技術開発基盤センターとして整備
- 科学の個別分野で開発された**アルゴリズム手法を整備し、標準化して諸分野に提供する体制を確立**することにより、広範な学術向上に寄与すると共に、国民の生活の質の向上、新規産業創出による数兆円規模の経済発展が期待できる

最先端理論(Algorithm Theory)



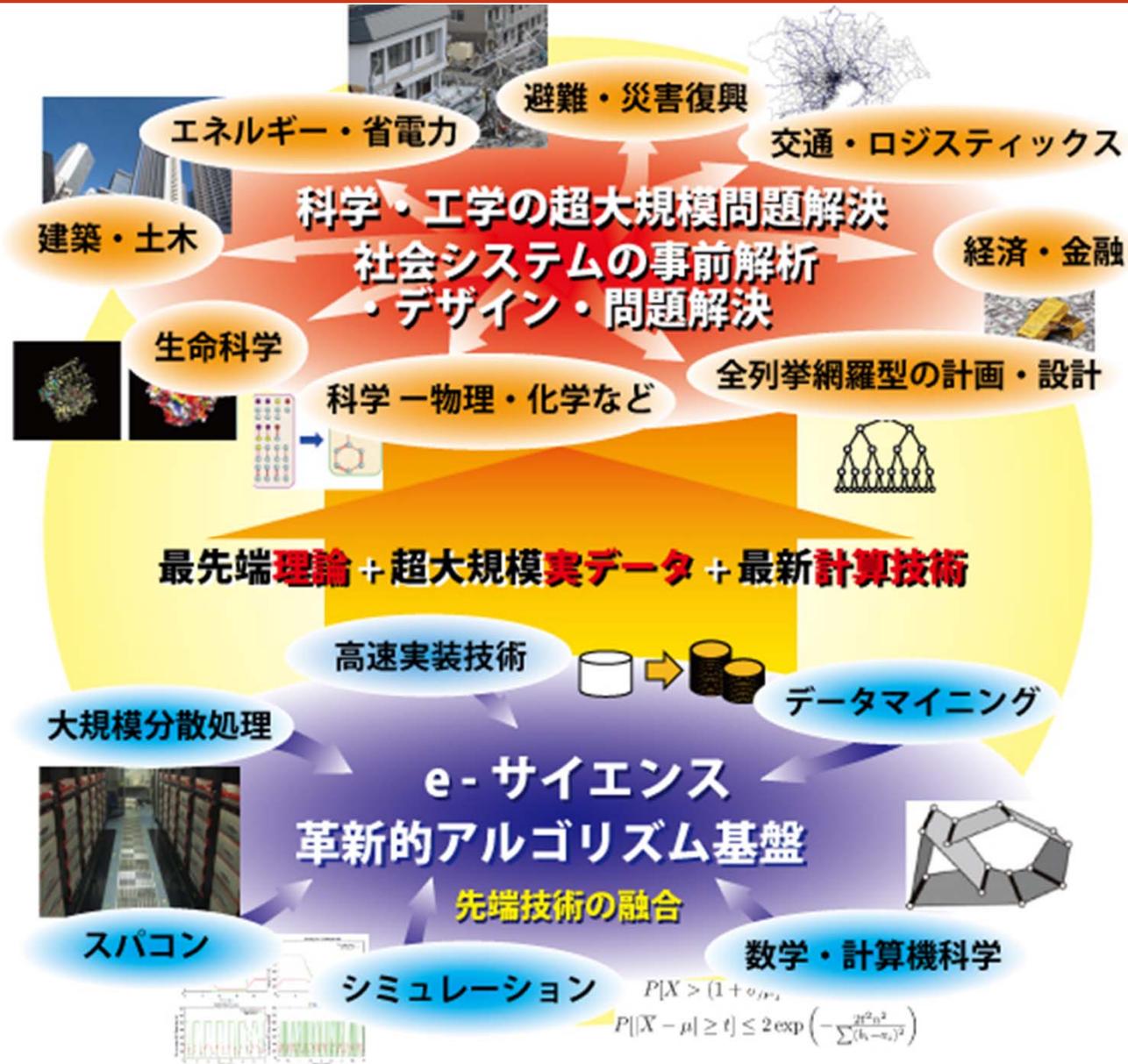
有機的な融合



革新的アルゴリズム(最適化、列挙、探索等)による社会システムのデザイン

- データマイニング、人工知能、機械学習、統計、HPC などの分野と結合し、**新しい学問体系の創生と人材育成**
- 物理、化学、生物などの科学、土木、建築、機械などの工学、環境・エネルギー、交通、経済・経営の**諸分野への適用**
- 地球的規模の諸問題(環境、エネルギー、バイオ)及び**突発的事態**(防災、避難及び復興計画の策定)への対応

# 革新的アルゴリズム基盤の整備と現実的な諸問題の解決

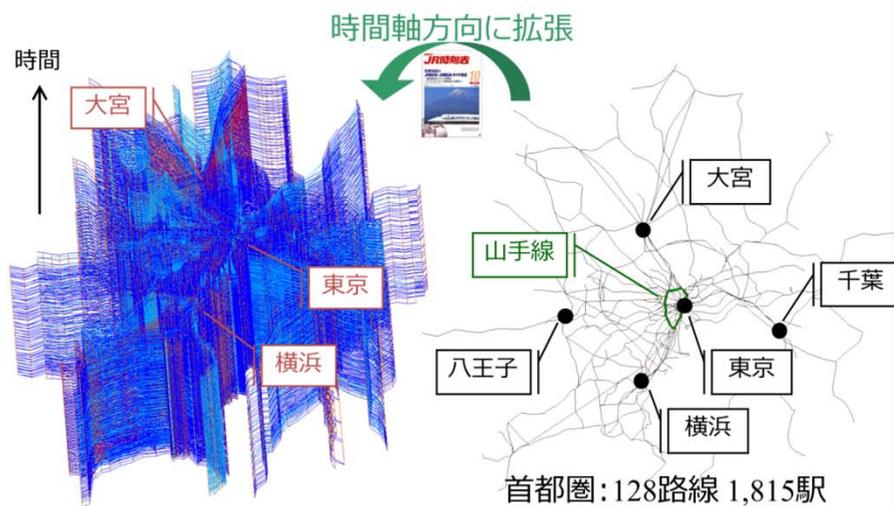


問題の解決が期待される大規模かつ複雑な諸問題の例

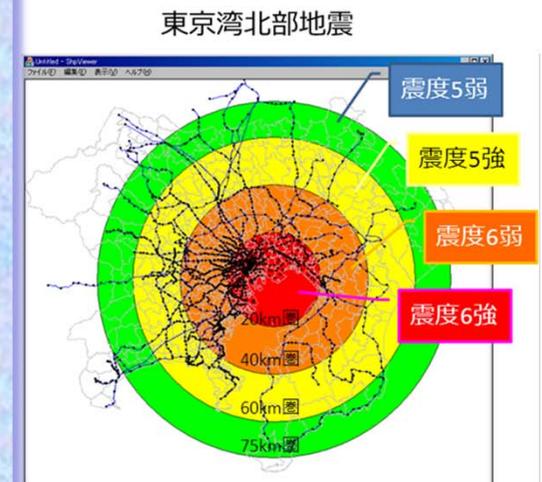
- ・ 数十億点の大規模ネットワークの探索アルゴリズムと数十億規模のデータに対するクラスタリングアルゴリズムの高速実装



- ・ 防災計画の策定、災害時の避難と誘導及び情報収集と解析、復興計画の策定、スマートグリッドによる高度かつ安定な電力供給
  - 交通データに対する経路探索。動的に変化する交通量等から高速な重要度判定を行うことによって、**交通管制等に活用**する
  - ソーシャルネットワークデータに対する動的な重要度、影響度の判定。**各点の周辺、及び広域内における影響**(情報の伝播力)を推定する
  - 全経路(天文学的な個数)の列挙による最悪時の動作保証



- ・ **災害時における大規模な物流網及び輸送計画の高速デザイン**
- ・ 首都圏鉄道ネットワーク
  - 時空間ネットワークによる詳細な表現
  - 渋滞・事故情報を考慮した動的データ更新
  - パーソントリップ調査等による帰宅困難者の推定

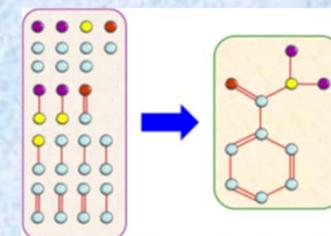
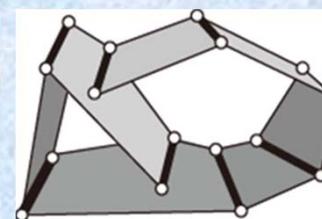
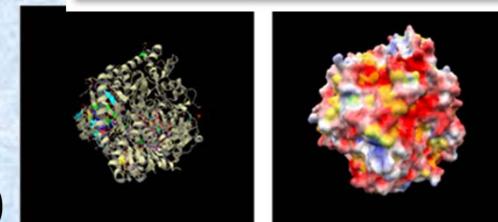
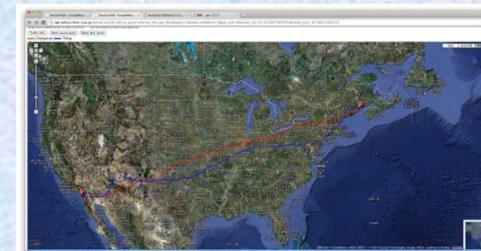


## 主な準備状況

- 社会システムに対する安心できるデザインや計画を考えるには、すべての場合を想定した**全列挙網羅型の計画・設計が重要**
  - これまでのアルゴリズム技術、データ収集及び計算機的能力では極めて困難
  - 全列挙網羅型計画を目指した革新的なアルゴリズム論が必要
  - 要素技術としては、**最適化技術**、**超コンパクトデータ表現技術**、**データ圧縮技術**など ⇒ これらは日本が世界に先駆けて研究を進めている
  - これまでの実績からいつでもプロジェクト研究を開始できる準備状況

## 準備研究の具体例

- 超大規模ネットワークにおける高速探索技術の開発
  - 超大規模ネットワークデータ(**点数10億以上**)が対象
  - 実行時間(最短路問題)
    - 全米グラフ **3秒**(2400万点、5800万枝)
    - 超大規模グラフ **870秒**(10億点、20億枝)
- 高速列挙アルゴリズム
- 3次元構造の探索と予測
  - (組合せ論に基づく蛋白質立体構造高速探索・予測)
- 大規模構造・メカニズム創生
  - (組合せ剛性理論の応用)
- 新規化合物生成のための構造探索・列挙



## 研究組織と予算計画等

- **京都大学**に諸分野における未解決課題を革新的アルゴリズムと最新の計算技術により解決する**中核的研究拠点**を構築し、国内外の研究者との共同研究、研究交流を行う
  - 東工大(超大規模データ最適化問題に対する高速計算システム)、東北大学(耐災害情報システム基盤)にもサテライト型研究拠点を配置
- **世界的な研究連携体制**
  - (Yahoo研究所(米国)、マックスプランク情報研究所(ドイツ)、ZIB国立研究所(ドイツ)、Reny研究所(ハンガリー)、Intractabilityセンター(米国)、中国清華大学、韓国KAIST等)
  - 研究拠点形成と、日本が学術的な優位性を持つ数理科学分野の力をコアに、最先端情報処理基盤を構築し、新たな計算とアルゴリズムのパラダイムを創生する。これにより、**次世代の基盤研究において世界をリードする地位を確立する。**
- **予算計画**
  - 初期投資 共同研究拠点施設整備:1億円、ソフトウェア開発や大規模な検証実験環境のための計算機クラスターとネットワーク及びインターフェース:5億円、海外拠点等との連携経費:1億円
  - 運営費 年間5億円 [内訳 施設賃借料および設備運営費:1億円、人件費:2億円(博士研究員、プログラマー、事務経費等)、旅費1.5億円(海外著名研究者招聘費用を含む)]、その他:5千万円(会議費、消耗品費、計算機設備更新費等)]
- **スケジュール**
  - 平成23年度には各応用分野において現状の技術では解決不可能かつ重要な問題の抽出を行う
  - 抽出された問題に対して、平成24年度から4,5年間で、最先端のアルゴリズムを開発し、個別分野での問題解決を行う。**ただし社会情勢により特定の諸問題を早期かつ優先的に扱う。**
  - 平成29年度までに、個別に開発されたアルゴリズムの共通部分をアルゴリズム基盤として整備する