

3. 分野連携による新しい科学の創出

従来は異なる研究分野と見なされていた諸分野を有機的に結合することにより、新しい科学を創出することが期待できる。ここでは、新たな学術フロンティアを切り開く「基礎科学の連携と統一理解」、近未来に重要性が増すと考えられている「ビッグデータの有効利用」、そして、大規模データ処理が発生するため計算科学との連携が必須となる「大規模実験施設との連携」について記述する。

分野	連携例
基礎科学の連携と統一理解	基礎物理における連携 宇宙科学と地球科学の連携による惑星科学 生命科学、物質科学、ものづくりの分野横断連携
ビッグデータの有効利用	計算科学基盤技術の創出と高度化 有効利用例①：衛星・観測データの有効利用 有効利用例②：ゲノム解析
大規模実験施設との連携	X 線自由電子レーザー施設 SACLA 等の大規模実験施設との連携

3.1 基礎科学の連携と統一理解

基礎科学におけるあくなき知の探求は、どの時代においても科学技術を発展させる強い動機づけとなっている。一見すると応用には結びつかないような基礎科学的研究が、次世代の応用につながり我々の生活を豊かにしてきた例も少なくない。これまで、細分化され分野ごとに発展・深化してきた基礎科学の各分野を、計算科学の手法により強く連携させ、新たな学術フロンティアを切り開いていくことが期待されている。ここでは、宇宙・物質・生命に関連する 3 つの連携事例を代表例として示す。

(1) 基礎物理における連携

素粒子・原子核・宇宙天文物理などの基礎物理分野では、究極的な物理法則を見つけること、そしてその物理法則を用いて多様な物理現象を理解することを目指してきた。

従来は、各分野に細分化され、計算機性能の限界などの要因もあり個別分野において基礎法則の確立が行われてきた。しかし、科学技術の進歩により、各分野の基礎法則のみでは理解できない新たな実験事実や観測事実が明らかになってきている。そこで、分野ごとに発展・深化してきた各分野を計算科学の手法により強く連携させ、新たな学術フロンティアを切り開いていくこと（未知への挑戦）、そして各分野の基礎法則のみでは理解できない新たな実験事実や観測事実を明らかにし、理解すること（森羅万象の統一理解）を目指している。そのためには、さまざまな未解決問題を克服する新しい理論を見つけ、スーパーコンピュータによる大規模な数値シミュレーションを用いて実際に計算していくことが必要である。近年の大規模計算によりこれまで定性的な理解にとどまっていた事象につい

分野連携における新しい科学の創出

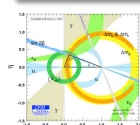
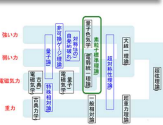
基礎科学の連携と
統一的な理解基礎物理(素粒子・原子核・宇
宙天体物理)における連携

目標・目的、克服すべき学術的課題

- 細分化し分野ごとに発展・深化してきた基礎物理の各分野を計算科学の手法により強く連携させ、新たな学術フロンティアを切り開いていくこと(未知への挑戦)
- 各分野の基礎法則のみでは理解できない、新たな実験事実や観測事実を明らかにし、究極的な物理法則の解明と多様な物理現象を理解すること(森羅万象の統一理解)

従来の研究

- 個別分野での物理法則の理解(計算機性能の限界による単一の時間または空間スケールにおける物理学)

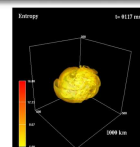
小林・益川行列
要素の決定※1究極の統一理論
の探求

分野連携の方策

- 実験不可能なマルチスケール物理現象に対する基礎物理の連携シミュレーションによる新しい法則の発見
- 精密観測や精密実験による高精度観測値と大規模シミュレーションによる精密理論計算の比較による新しい法則の発見

大規模計算で実現されること

- 究極的な物理法則の解明と多様な物理現象の理解
- 重大未解決問題解明に向けた知的フロンティアでの研究の促進
- 大規模計算により、定性的理解から定量的理解へ

超新星爆発と物質の起源
超新星爆発の3次元流体シミュレーション(瀧田 et al 2013年)

※1: CKMfitter Group (J. Charles et al.), Eur. Phys. J. C41, 1-131 (2005) [hep-ph/0406184],
updated results and plots available at: <http://ckmfitter.in2p3.fr>

でも定量的理解への転換が可能となりつつある。

基礎法則の連携と計算科学の手法を仲立ちとした分野連携により、重大未解決問題の解明に向けた知的フロンティアでの研究が促進され、究極的な物理法則の解明と多様な物理現象の理解につながっていく。

基礎物理における連携において今後、必要となる計算機性能を下表に示す。

課題	要求性能 (PFLOPS)	要求メモ リバンド 幅 (PB/s)	メモ リ量/ ケース (PB)	ストレ ージ 量/ケー ス (PB)	計算時 間/ケー ス (hour)	ケース 数	総演算量 (EFLOP)	概要と計算手法	問題規模	備考
カイラル対称性とQCD に基づく有効バリオン 間相互作用の決定と その応用	510	390	0.066	0.5	880	10	16000000	格子QCD (カイラル5次 元型)、ハイブリッドモンテ カルロ法、CG法	問題規模 格子点: 128 ⁴ ×32、格子間隔: 0.1 [fm] 以下	ノード数を16 ⁴ ノードを仮定し、 ノードあたり性能を、オンチップメモ リ容量 200MB、オンチップメモ リバンド幅 6TB/s、ネットワークレ イテンシ 1 μ sec程度、ネットワ ークバンド幅128GB/sを想定。
閉殻を仮定しない殻 模型計算	100	10	※ 1	※ 0.0001	28	100	1000000	モンテカルロ殻模型法に よる原子核の構造計算、 軽い核	空間を調和振動子基底で展 開し、7~8主殻までを考慮 10 ⁹ ステップ数。	メモリ量は10000ノード×100GB で計算
相対論的輻射流体計 算による超新星爆発メ カニズムの探究	18	70	1.6	1.3	1200	10	780000	ニュートリノ輻射輸送計 算(超新星爆発)	空間512×64×128 位相空間 24 ³ で1秒分の時間を計算	100Tflops/ノード × 10000ノード、 主記憶100GB/ノード、主記憶バ ンド幅 20TB/s/ノード、オンチッ プメモリ容量 2MB/コア、1000コ ア/ノード、通信速度60GB/s/ノ ード

※印の値は未だ精査中である。より精度の高い数値は Web で随時公開する。

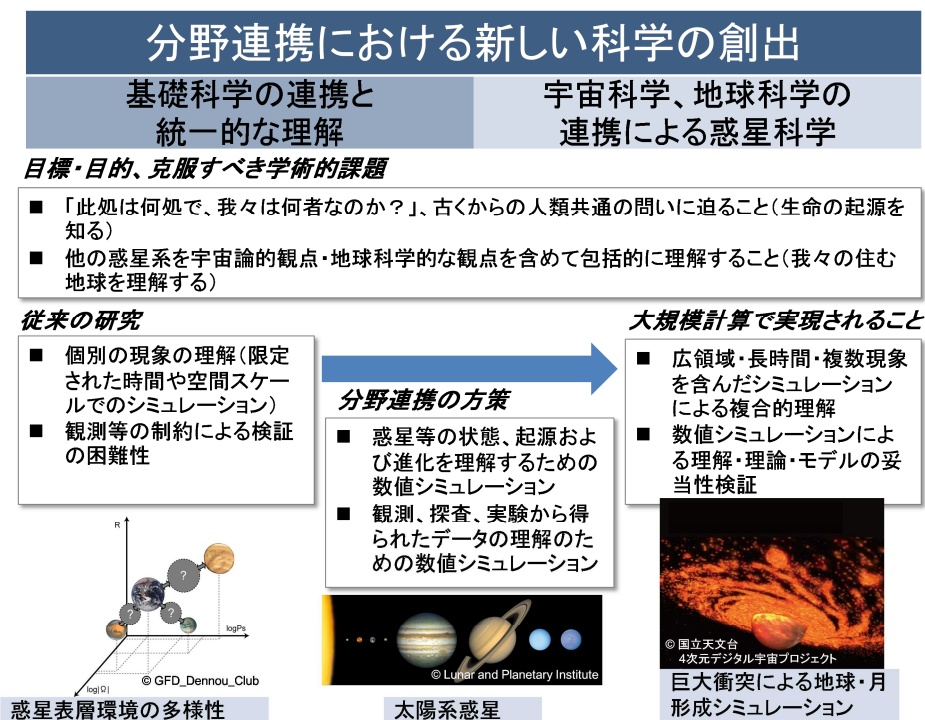
3 分野連携による新しい科学の創出

(2) 宇宙科学、地球科学の連携による惑星科学

「此処は何処で、我々は何者なのか?」、古くからの人類共通の問いに迫ることが宇宙科学、惑星科学の究極の目標である。近年の太陽系探査ならびに太陽系外惑星観測などにより、この問いを具体的かつ科学的課題として考えることが可能になりつつある。更に、他の惑星系を宇宙論的観点・地球科学的な観点を含めて包括的に理解することは、我々の住む地球を理解することにつながる。今後、国内外で、多くの惑星探査が計画・実施されていくが、計算科学であるシミュレーション技術や、現実世界の影響をモデリングに適用するデータ同化技術がこれらの企画立案と実施には必須であり、深く寄与・貢献することが期待されている。

宇宙科学および地球科学と計算科学との連携により広領域・長時間・複数現象を含んだシミュレーションが可能となり、複合的な理解が可能となる。また、数値シミュレーションによる理解・理論・モデルの妥当性検証も可能となり、我々の住む地球を理解し、我々生命の起源を知る大きな一歩へとつながっていく。

宇宙科学、地球科学の連携による惑星科学において今後、必要となる計算機性能を下表に示す。

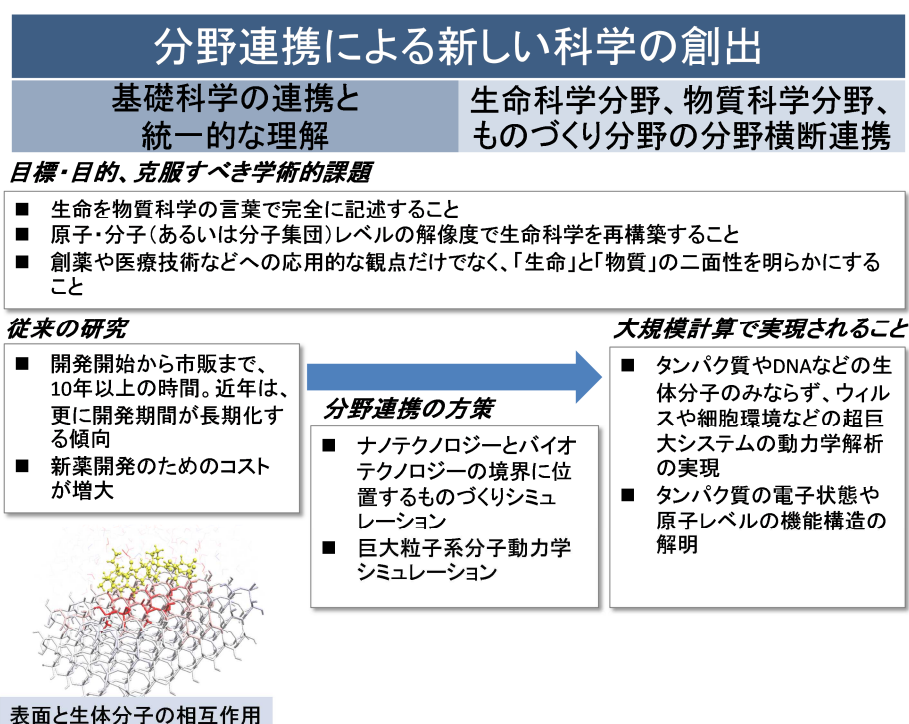


課題	要求性能 (PFLOPS)	要求メモリ バンド幅 (PB/s)	メモリ量/ ケース (PB)	ストレージ量/ ケース (PB)	計算時間/ ケース (hour)	ケース数	総演算量 (EFLOP)	概要と計算手法	問題規模	備考
惑星系形成のシミュレーション	4	0.02	0.00001	0.05	1000	100	1500000N	N-体 シミュレーション	粒子数: 1億体 積分時間: 1億年 (ステップ数: 10G)	論文で報告されているアルゴリズムとGRAPE における計測結果から算出。1ステップ1粒子あたり1万5千演算、グループ内粒子数128。メモリアクセスは6000 演算あたり32 バイト
地球・惑星の形成シミュレーション	520	29.0	0.001	1	24	100	4500000	SPH 計算	粒子数: 10億体 積分時間: 数ヶ月 (ステップ数: 100M), 演算量 $N \log N$	演算量、メモリ転送量、メモリ使用量は、TSUBAMEでのプロファイルを元に外挿
惑星表層環境の形成と進化シミュレーション	6	25	0.01	4	100	1000	2000000	流体計算 + 輻射計算 (スペクトル法+差分法)	格子数: 3840x1920x192, 100 ケース x 10 惑星, 積分時間: 10年 (ステップ数: 30M), 1ステップ1格子あたりの演算量: 50K	演算量、メモリ使用量は、TSUBAMEでのプロファイルを元に外挿

3 分野連携による新しい科学の創出

(3) 生命科学分野、物質科学分野、ものづくり分野の分野横断連携

タンパク質や DNA などに代表される生体分子は、生命科学的側面から見ると「生命」を構成する基本単位に位置づけられるが、物質科学的側面から見ると対称性の少ない非常に複雑な「物質」であり、「生命」と「物質」の二面性を持つ存在である。したがって、タンパク質や DNA などの生体分子の研究、特にその立体構造に基づく解析は、生命科学と物質科学という大きな 2 分野の境界に位置する課題であり、それぞれの分野で培ってきた方法を横断的に集約し駆使することで、大きなブレークスルーが期待できる。



「生命科学的側面」と「物質科学的側面」は相補的であり、この二つを融合することによって、創薬や生体分子を活用したものづくりなど社会的に重要な課題において飛躍的な発展が可能になる。例えば、薬剤のターゲットであるタンパク質の生命活動における働きは生命科学的な研究対象であるが、薬剤とタンパク質の結合は物理化学的相互作用であり、そこでは物質科学的方法論が有効である。今後のスーパーコンピュータの高度化と合わせ、物質科学分野で培われた精度の高い計算方法をタンパク質-薬剤結合解析に応用していくことで、創薬分野でのブレークスルーが期待できる。更には、タンパク質や DNA などの生体分子のみならず、ウイルスや細胞環境などの超巨大システムの動力学解析が実現し、タンパク質の電子状態や原子レベルの機能構造を解明することが想定される。

生命科学分野、物質科学分野、ものづくり分野の分野横断連携において今後、必要となる計算機性能を下表に示す。

課題	要求性能 (PFLOPS)	要求メモリ リバンド幅 (PB/s)	メモリ量 /ケース (PB)	ストレージ量/ ケース (PB)	計算時間/ ケース (hour)	ケース数	総演算量 (EFLOP)	概要と計算手法	問題規模	備考
創薬などMD・自由エネルギー計算	1000	400	0.0001	0	0.0012	1000000	43000000	全原子分子動力学シミュレーション	ケース数: 10万化合物×10標的蛋白質(10万原子程度)	B/F=0.4, 数百から数千ケース同時に実行することを想定しているため、実行時に必要な全メモリ量、各ケースの実際の実計算時間は、表の値の数百～数千倍となる。メモリ量/ケースは100ノード実行時を想定。
高精度創薬	0.83	0.1	1	0.001	1	100	300	薬品とタンパク質間相互作用の量子化学計算を統計的ゆらぎを含めた複数サンプルで行う	フラグメント分子軌道法で～500残基程度までの計算	計算要求は「物質科学」のフラグメント分子軌道法のところを参照
バイオデバイス設計	1.1	0.2	1	0.001	1	100	400	200～500残基程度のタンパク質の分光計算	電子軌道数10万超	計算要求は「物質科学」のフラグメント分子軌道法のところを参照
細胞環境・ウィルス	490	49	0.2	1.2	48	10	850000	全原子/粗視化分子動力学シミュレーション	～1億粒子	B/F=0.1

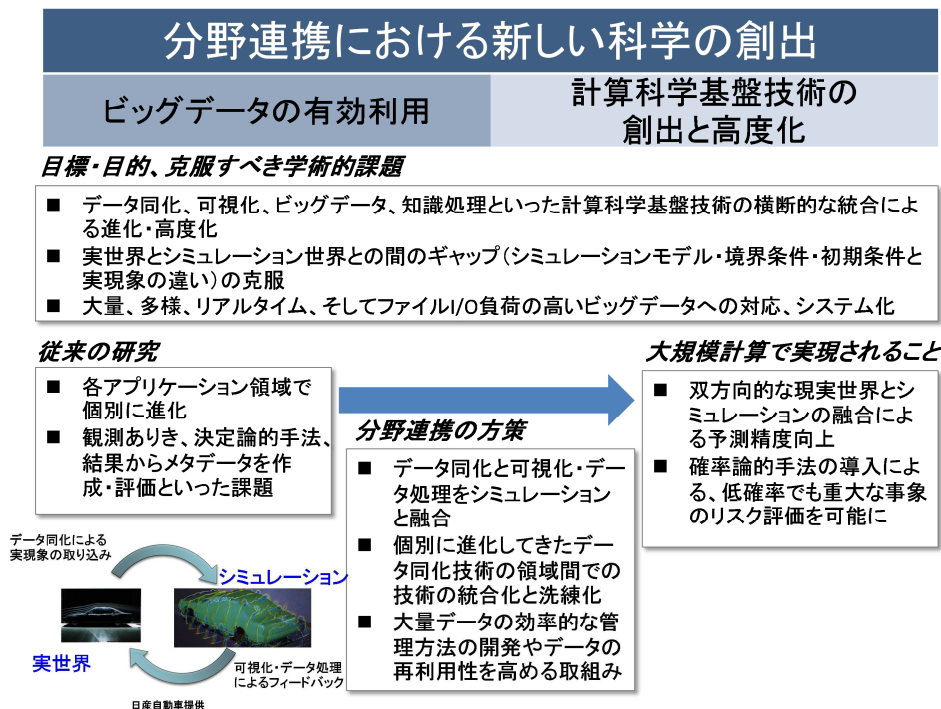
3 分野連携による新しい科学の創出

3.2 ビッグデータの有効利用

社会的課題の解決を促進する研究や、従来の計算技術の限界を超えた知の探求を行うためには次世代の計算科学技術基盤が必要である。特に近年、大規模数値シミュレーション、衛星・観測データ、個人のゲノム情報、大規模実験施設で得られる実験データなどの量が飛躍的に増加しており、いわゆるビッグデータの効率的な利用技術などの計算科学技術基盤が求められている。ここでは、冒頭に必要な計算科学技術基盤のあり方と高度化について述べ、その後、2つの代表的な連携事例を示す。

(1) 計算科学基盤技術の創出と高度化

実世界で起こる現象は現象自体が複雑であり、その因果関係を正確に把握することは難しい。多くの場合、現実世界のありのままを模擬することはできず、シミュレーションは対象とする実現象（問題）を理想化して計算することになる。したがって、現実世界で起こる現象とシミュレーションする現象との間には、少なからずギャップが存在する。現実世界の問題の特徴をうまく再現できるような方程式群の形式を選定し、関連するパラメータを決めるプロセスをモデリングと呼ぶ。モデリングは観測や実験結果に対する洞察から得られた知見を基にしているが、シミュレーションの正しさや精度に影響する非常に重要なプロセスである。



従来、モデリングおよびそれに関わる各種技術は各アプリケーションの分野ごとに個別の技術的進化を遂げてきたが、共通部分も多い。例えば、現実世界の影響をモデリングに

適用する技術であるデータ同化は、気象をはじめ、石油掘削、制御、ものづくり、創薬や分子シミュレーションなどの分野でも研究・活用されている。また、結果の解釈の一助となる可視化技術もさまざまな研究領域で利用され、個別の現象に適した表現方法が研究されている。

計算科学基盤技術の創出と高度化において今後、必要となる計算機性能を下表に示す。

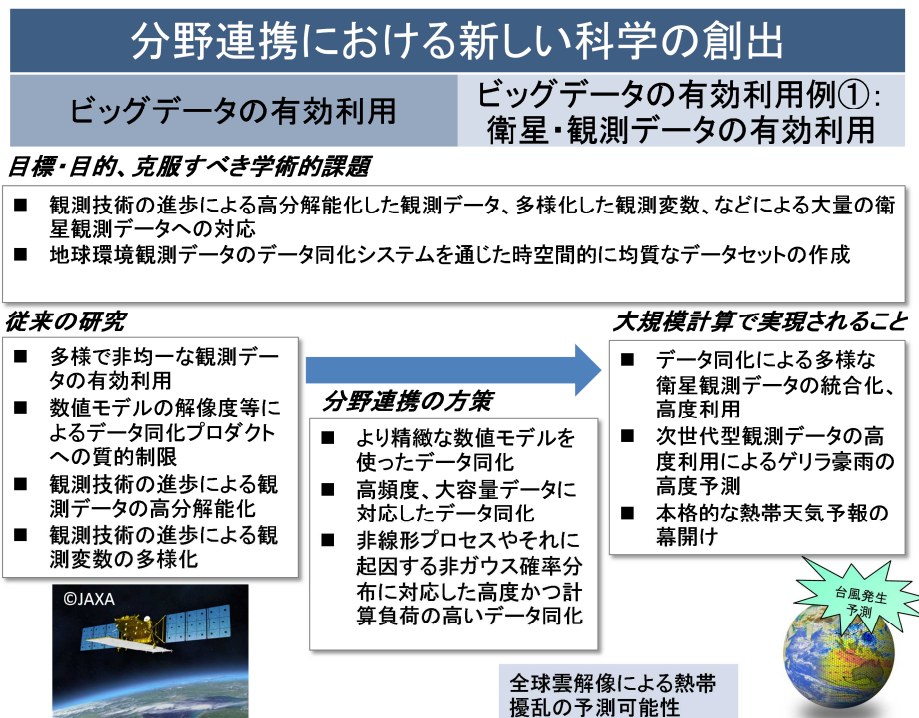
課題	要求性能 (PFLO PS)	要求メモリ バンド幅 (PB/s)	メモリ量 ケース (PB)	ストレージ 量/ ケース (PB)	計算 時間/ ケース (hour)	ケース 数	総演算量 (EFLOP)	概要と計算手法	問題規模	備考
並列レンダリング	200	61	0.8	10	0.5	1	360	ボリュームレンダリング(レイキャスト、ファイルベース)		対象によって問題規模等は異なるため、典型的な例で概算
並列レンダリング	200	61	2	1	0.5	1	360	ボリュームレンダリング(In situ)		対象によって問題規模等は異なるため、典型的な例で概算
データ圧縮	500	25	8	10	0.5	1	900	POD圧縮(ファイルベース)		対象によって問題規模等は異なるため、典型的な例で概算

3 分野連携による新しい科学の創出

(2) ビッグデータの有効利用例①：衛星・観測データの有効利用

大気、海洋、陸域の物理・化学および生物環境に関する観測データは、環境変動の監視・検出や影響予測のための基礎データとして重要な役割を果たしており、日々の天気予報から地球規模の気候変動まで幅広い環境問題に適切に対処することに貢献している。これらのデータは、現場における直接観測に加え、人工衛星からのリモートセンシングなどさまざまな方法により観測が行われ、その利用分野も広い範囲に及んでいる。ただし、それぞれの観測データは異なる場所、時間で得られており、また観測される物理量変数や精度等もさまざまであるため、単に既存の観測データをまとめただけではそのまま実際に利用することは難しい。そこで、データ同化と呼ばれる、数値モデルを用いて異なった観測データを統合する手法により、より使いやすいデータセットに加工されている。

大規模なデータ同化は、これまで気象・気候分野での技術的発展が顕著であった。しかし、最近では、設計制御、石油掘削、分子シミュレーションなど他の分野においても適用が図られてきており、各アプリケーション分野のコミュニティに閉じていたデータ同化技術は、コミュニティを越えて共通化する動きが起きている。また、観測技術の進歩による高分解能化した観測データ、多様化した観測変数などによる大量の衛星観測データおよびそのデータを用いたシミュレーションの双方の桁違いの大容量化に対応するために、ビッグデータを扱うデータ同化技術のイノベーションが起こりつつある。



データ同化技術の進化にともなって、シミュレーションが高分解能化・精緻化することにより、ゲリラ豪雨や熱帯天気予報と言った社会的インパクトの高い局所的な気象現象の

予測などに正面から取り組むことが可能となっていく。

衛星・観測データの有効利用において今後、必要となる計算機性能を下表に示す。

課題	要求性能 (PFLOPS)	要求メモリ バンド幅 (PB/s)	メモリ量/ ケース (PB)	ストレージ 量/ ケース (PB)	計算 時間/ ケース (hour)	ケース 数	総演算量 (EFLOP)	概要と計算手法	問題規模	備考
局所的・集中的大雨、熱帯気象の高度予測	220	270	0.7	5	580	2	900000	大気モデル:NICAM(有限体積法)、アンサンブルデータ同化:LETKF	水平解像度3.5km、鉛直100層、1000アンサンブルメンバー、3時間おきの同化サイクル 2ヶ月積分 格子点: 大気 640x320x150, 海洋 3600x1800x150 Δt: 大気1min, 海洋 30sec, 結合 10min 100イタレーション 積分時間: 3month	10万ノードを仮定(大気モデルのノードあたり隣接通信1GB/s) 演算量、メモリ転送量、メモリ使用量は、京でのプロファイルを元に外挿
統合地球環境再解析	3.1	13	0.018	0.022	18	240	480000	4次元変分法		B/F値: 大気4.66, 海洋4.24 演算量、メモリ使用量は、ES2のプロファイルを元に精査、メモリ転送量は、ソースから見積もったB/F値をもとに、演算量から算出(キャッシュは考慮していない)。

3 分野連携による新しい科学の創出

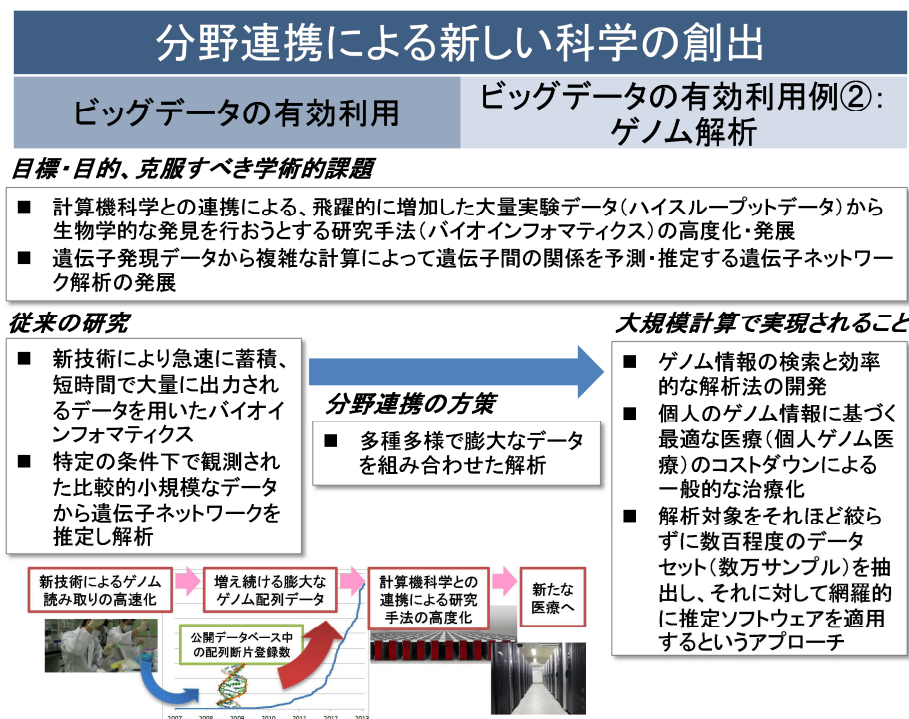
(3) ビッグデータの有効利用例②：ゲノム解析

ヒトゲノム計画は、分子生物学におけるマイルストーンの一つであり、たった一人の DNA 解読に世界の多くの分子生物学者が関わり長い年月を要した。しかし、近年になって次世代 DNA シークエンサーと呼ばれる超高速かつ低コストで DNA を解読する装置が開発され、個人のゲノム情報を含むさまざまな細胞のゲノム解析が容易に行うことが可能となった。

分子生物学では、細胞内の大量の遺伝子発現を同時計測可能な DNA マイクロアレイなどの登場により、実験で得られるデータの量が飛躍的に増加している。それらの大量データ、いわゆるハイスループットデータから計算科学を用いて生物学的な発見を行おうとする研究手法（＝バイオインフォマティクス）が行われるようになっており、計算科学との連携が必要な重要な 1 分野として認識されている。

生命科学におけるデータ解析では、ゲノム配列データだけではなく遺伝子発現データや DNA 修飾のデータ（エピゲノム）、タンパク結合など多種多様で膨大なデータを組み合わせ解析が行われる。今後開発される観測技術によってより多くのデータが蓄積されることが予想され、観測技術の進展にしたがって多様な解析ソフトウェアが組み合わされて利用されていくことになる。

本分野の研究の進展により、ゲノム情報の検索と効率的な解析法が開発され、現在は極めて高価な治療である個人のゲノム情報に基づく最適な医療（個人ゲノム医療）が、一般的な治療へとコストダウンを図ることができるようになる。



ゲノム解析において今後、必要となる計算機性能を下表に示す。

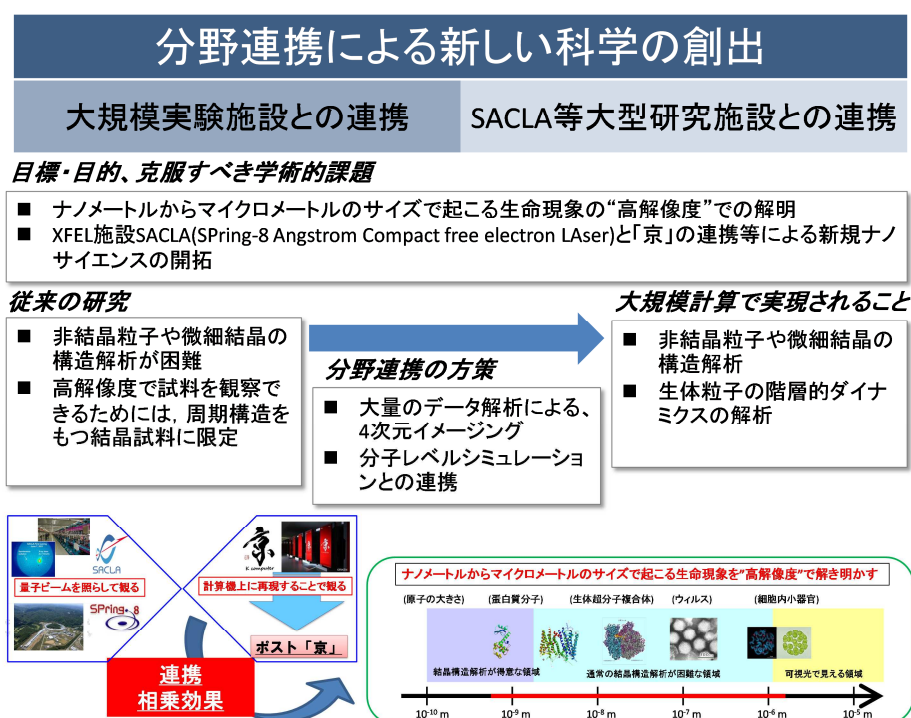
課題	要求性能 (PFLOP/s)	要求メモリ バンド幅 (PB/s)	メモリ量 /ケース (PB)	ストレージ 量/ ケース (PB)	計算 時間/ ケース (hour)	ケース 数	総演算量 (EFLOP)	概要と計算手法	問題規模	備考
個人ゲノム解析	0.0054	0.0016	1.6	0.1	0.7	200000	2700	シーケンスマッチング	がんゲノム解析200,000人分のマッピングおよび変異同定	1人分の解析を1ケースとした。入力データを分割することで、細かい単位での実行、拠点をまたいだ実行も可能。整数演算中心のため「総演算量」はInstruction数とした。総浮動小数点演算量は45.864EFLOPとなる。
疾患遺伝子発見のための統計的解析	10	0.0002	200	2	140	5	25000	ゲノムワイド連鎖解析(GWAS)	ヒトゲノム3Gbp x 200,000人分、1ケース4万人	メモリ量は800GB/node、ノード数25万を仮定

3 分野連携による新しい科学の創出

3.3 大規模実験施設との連携

(1) X線自由電子レーザー施設 SACLA 等の大型研究施設との連携

XFEL (X線自由電子レーザー) は、波の位相がきれいにそろったレーザーの性質を持つ超高輝度の X 線を発生させることのできる光源であり、これまで構造を解くのが難しかった非結晶粒子や微細結晶の構造解析に威力を発揮すると期待されている。2011 年 3 月に完成した XFEL 施設 SACLA (SPring-8 Angstrom Compact free electron LAser) では 1 日に最大で約 500 万枚の回折パターンが得られるため大量のデータ処理を行う必要があり、計算機科学技術との連携の重要性が謳われている。特に、生体粒子の動態をも含めた 4 次元イメージングに期待が高まっている。



また、SACLA による非結晶試料のイメージング実験では、10 ナノメートルからマイクロメートルのサイズの生体粒子の姿を、ナノメートル程度の解像度で捉えることが可能である。この空間スケールは、粗視化 MD 法⁵など分子シミュレーションのよいターゲットとなる。大規模分子シミュレーションを用い、SACLA の実験データを解析することで、生体粒子の階層的ダイナミクスの研究が進み、この空間スケールで起こる生命現象の理解につながる。

大型研究施設と計算科学の連携により、新規ナノサイエンスが開拓され、ナノメートルからマイクロメートルのサイズで起こる生命現象の高解像度での解明へとつながっていく。

⁵ 数から数十の原子を 1 つの単位として表現したモデルを用いて計算を行う手法。目的に応じて、どのくらいの原子数を粗視化の一単位とするかが決定される。

大規模実験施設との連携において今後、必要となる計算機性能を下表に示す。

課題	要求性能 (PFLO PS)	要求メモリ バンド幅 (PB/s)	メモリ量/ ケース (PB)	ストレージ 量/ケース (PB)	計算時間/ケース (hour)	ケース 数	総演算量 (EFLOP)	概要と計算手法	問題規模	備考
大量実験データ解析による4次元イメージング	2	0.01	0.000001	0.000001	2.8E-11	1E+12		構造分類、3次元構造構築、200時間軸推定のための統計処理	10 ⁷ イメージ×10 ⁵ イメージの相関計算で、合計10 ¹² 回の相関計算。1回の相関計算を1ケースとした。京ではローカリティを考慮したI/Oにより最適化している。	通信、ファイルI/Oは引き続き調査必要。特にファイルI/Oがボトルネックとなる可能性あり。
実験解析結果に基づく動的構造モデリング	490	49	0.2	1.2	48	10	850000	全原子/粗視化分子動力学シミュレーション	～1億粒子	B/F=0.1

※印の値は未だ精査中である。より精度の高い数値は Web で随時公開する。