

#### 4.4.6 可視化・データ処理

##### (1) 現在行われている課題

ペタスケールからエクサスケールの計算機環境が生み出すデータから、科学的な発見や物事の本質を引き出し、創造や改善につなげ、研究の成果を社会に還元することは非常に重要である。しかしながら、スパコン上でシミュレーションを実行するだけでは科学的な成果は得られない。生成されたデータを可視化・分析し、そこから意味のある情報を抽出し、人が理解できる形式に昇華させ、利用可能な情報となつて、初めてシミュレーションの意義がある。この点で、可視化とデータ分析は、HPCを支える重要な基盤技術として位置づけられる。

可視化研究は、米国では90年代から政府の研究費が継続的に投資され、研究をリードする拠点が多数形成されてきた[1]。特に、大規模データに対する研究開発は、並列レンダリング処理、並列可視化法の研究など、米国が中心となっている。また、研究成果とともにオープンソースソフトウェアも数多く提供されている。米国エネルギー省が開発した可視化アプリとして、VisIt[2]、ParaView[3]の二つが存在するが、両方ともそれ以前に同じく政府系ファンドで開発された可視化ツールキット VTK[4]を用いて構築されている。可視化研究については、IEEE Visualization[5]、SuperComputing[6]、Siggraph[7]、EuroGraphics[8]などの学会で議論されている。

大規模なデータ可視化に関する具体的な研究項目としては、大規模並列可視化システムの構築技術、3次元データに対してレイトレーシングやボリュームレンダリングなどの画像（図4.4.6-1 参照）を生成するレンダリング技術[9]、生成された画像を高精細ディスプレイや立体視などに投影する表示デバイス技術[10]、時間とともに変化する非定常データ可視化[11]、相互に関連する多くの変数に対する多変量データ可視化[12]、可視化と分析を統合する Visual Analytic 技術[13]、計算と同時に可視化やデータ処理をファイル出力なしで行う In situ 技術[14]、データベースと統合した QueryDriven 可視化技術[15]、可視化プロセス自体の再現やマイニングを対象とする Provenance 技術[16]など多岐にわたる研究が展開されている。

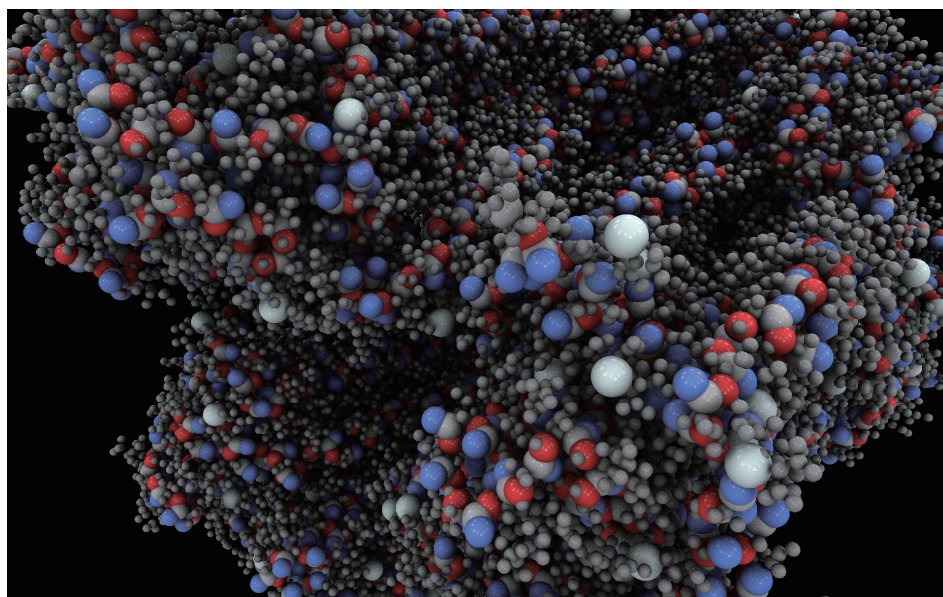


図 4.4.6-1 分子シミュレーションの並列レイトレーシング画像

データ：横浜市立大学大学院生命医科学研究科

池口満徳 博士/ 山根 努 研究員

可視化：理化学研究所計算科学研究機構 可視化技術研究チーム

一方、国内の可視化研究の状況として、可視化とデータ処理のコミュニティの規模が小さく連携は比較的疎であるが、各々主要な領域に特化した研究が行われている。例えば、大規模データの総描的な画像生成に適したポイントベースのボリュームレンダリング法の開発[17]、大規模データのデータレイアウトに適した階層的な表示方法の研究[18]、ボリュームデータの位相骨格に着目した表現[19]やテンソル場の解析[20]、大規模な時系列データの着目点のビジュアル化[21]、大規模な海洋シミュレーションの多変量データの効果的な可視化法[22]、ビジュアルデータマイニング[23]の研究が進められている。表示系の研究としては、タイルドディスプレイでの可視化[24]、没入型環境での立体視研究[25]、大規模粒子の可視化技術[26]の研究がある。また、可視化のシステム化研究および開発としては、RVSLIB[27]、日本原子力研究開発機構での非定常可視化システム[28]、海洋研究開発機構における CAVE VR[29]、理化学研究所で開発中の LSV システム[30]、国立天文台の Zindaiji3[26]などの可視化アプリケーションなどがある。

2011 年からエクサスケールに向けた可視化研究のロードマップが検討されており、エクサスケールの可視化とデータ処理に向けて、ワークショップの開催、共通コンポーネントの特定、開発計画の策定、開発システムの適用データ規模を段階的に拡大する方針でシステム構築を進める計画が記載されている[31]。

## (2) 長期的目標

大規模なシミュレーションに加え、時間とともに変化するデータ、相互に関連性を持つ多変数の分析、センサーから得られるリアルタイムデータなど、いわゆるビッグデータの中から有用な情報を抜き出し、科学的発見や設計に役立てることが可視化とデータ処理の目標である。

データの規模や複雑性が高まる中で、迅速な処理、優れたデータ処理アルゴリズムの開発などに加え、さまざまな分析・可視化プロセスの中から適切な手法を選択適用するなど、高度な使い方により生産性を高める取り組みにより、ヒトの理解や発見を促進することが望まれる。

ペタスケールのデータに対しては、現状の可視化・分析技術を改善することにより対応可能なものもあり、中期的な視点としては有効である。しかしながら、エクサスケールのデータに対してはデータ量の巨大さから、別の手法が必要となる。また、複雑な階層構造を持つ新しいアーキテクチャのハード機能を生かせる新しい可視化・分析アルゴリズムなど、エクサスケールで有効な新しい可視化手法の研究も必要である。この過程においては、アプリ分野とのコデザインが有効である。

研究開発の継続性の点からは、まず、エクサスケールでも適用可能な可視化フレームワークの骨格を構築し、既存手法を機能として組み込み、ペタスケールでの可視化サービスを提供する。それから、エクサスケールでの新しい可視化手法の研究を行い、その成果を可視化フレームワークへ統合していくというアプローチが採用されると思われる。

個別の可視化技術については、計算により生成された大規模なデータの移動やファイル入出力が不要な *In situ* 可視化や可視化と分析を統合した *visual analytic* などの技術が取り込まれていく。インタラクティブ性は重要であるが、エクサスケールにおいては、現状の技術の外挿では実現が難しいため、領域毎に特化したメタデータ化（データの特徴を表現する抽象的なデータ）によるデータ削減のアプローチなどと併用することになるだろう。また、データ量が膨大になり、人間が認識できる量を超えた場合には、プロセスマイニング（可視化方法やデータ処理方法自体の探索）による後処理の効率化や自動化も有効である。

可視化して抽出した情報をユーザにわかりやすく提示するという観点からは、ヒトが持つ複数のコミュニケーションインターフェースを有効に活用して情報を提示する技術も有効と考えられ、認知心理学などの観点も必要になる。このように、基本的な技術を積み上げてシステム化し、新しい技術を生み出していくアプローチが採用されるであろう。

可視化・データ分析分野が他の研究分野と大きく異なる点は、データ構造と手法を軸として、アプリケーション分野を横串で貫き展開できる点である。また、研究成果を基にソフトウェアを開発しサービスにつなげることが重要であり、研究拠点を設け、集中して推進していくべきである。

エクサスケールで役立つデータ可視化と分析には、多くの革新的な技術的なブレークスルーが必要で、かなりの開発とコストが予想される。したがって、継続的なファンディングが必要であり、開発資金面へ配慮しておくことが成功の鍵の一つとなる。また、研究開発を通して、コミュニティの活性化、人材育成、産業分野への展開などが期待される。

可視化・データ分析処理の要素技術を研究する分野には、応用数学、情報科学、アプリケーション、計算機科学などがある。可視化・データ分析の研究開発は、このような幅広い分野の教育や研究の機会につながる。また、研究開発された技術は、直接産業分野へ波及し、ソフトウェアの製品化にも貢献し得る。

### (3) ターゲットとする研究対象

可視化とデータ分析の研究は、方法論の研究成果を具体的なソフトウェアとして整備し、ユーザに対して可視化サービスを提供し、実用的な成果を創出するシナリオを考慮すべきである。この点から、可視化システムの開発とエクサ向けの可視化・分析技術の研究という両輪がある。

可視化システム開発については、ペタからエクサにわたり利用可能な可視化フレームワークの設計・開発が挙げられる（図 4.4.6-2 参照）。ここでは、可視化に関するさまざまなユーザ要求に応えられる設計が必要であり、可視化システムを使うアプリケーションユーザとのコデザインが望ましい。大規模なデータに対応するため、分散並列環境に対応したクライアントサーバ方式のリモート並列可視化システムの構築が必要である。要素技術として、大規模並列ファイルの入出力、レンダリング、画像重畳（各ノードで生成した部分画像を集約し1枚の画像にする処理）などの研究開発が必要である。

データ分析の基本システムは、大規模なデータに対してさまざまな分析メソッドを独立に適用し、得られた結果の集約処理を更に並列処理する MapReduce 型の枠組みを利用し、大規模データを効率的に処理するベースシステムを構築する。また、シミュレーションの後処理として、データベースの活用も考慮する必要がある。

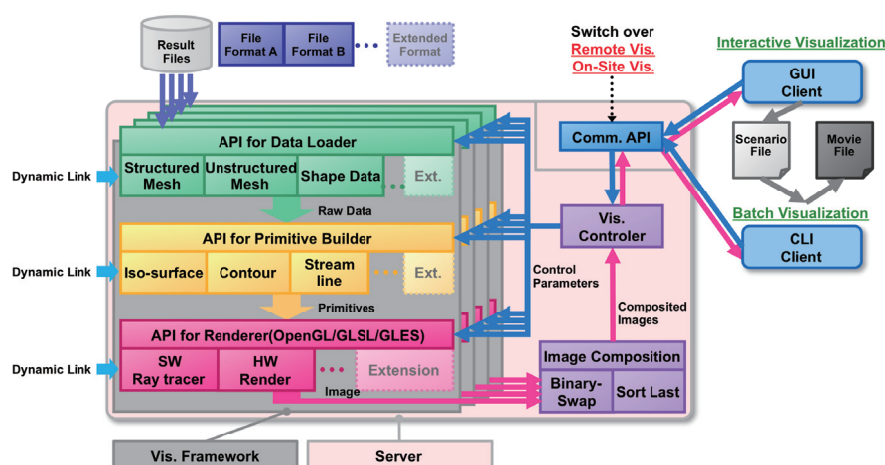


図 4.4.6-2 並列分散可視化システムの概念設計

可視化：理化学研究所計算科学研究機構 可視化技術研究チーム

<http://labs.aics.riken.jp/ono.html>

エクサスケール向けの可視化研究としては、可視化については In situ 可視化、特徴抽出などが研究対象となる。データ探索のためには、インタラクティブな可視化技術が重要で、データ



ストリーミング技術の応用が必須である。また、データ分析研究としては、開発したベースシステムを用いて、さまざまな分析を行う処理の研究がある。

データ表示技術としては、拡張現実や仮想現実の表示デバイス、超高解像度、没入型可視化環境、ホログラムなどが研究対象となる。

#### (4) 次世代に解決すべき課題

可視化・データ分析技術はさまざまな要素技術から構成される。このため、一概には言えないが、基本的な特徴として、下記の点が挙げられる。

##### (i) 大規模データの扱い

多数の大規模データファイルの管理、高速なファイル I/O 技術。時系列のデータを扱う場合には、圧縮やリオーダーリングなどの前処理も行う。

##### (ii) 並列データマネジメント・データ処理・スケジューリング技術

分散並列処理が必要なため、データ管理とスケジューリングアルゴリズムを用いる。

##### (iii) 並列レンダリング技術

大規模並列環境でのスケーラブルなレンダリングは、領域分割された担当空間のデータに対して画像を生成し、最終的に一つにまとめて表示する方法が基本的である。画像の生成過程では、レイトレーシングやボリュームレンダリングなどの個別の描画方法に最適な並列処理方法がある。

##### (iv) 並列データ重畳技術

分散環境でレンダリングした部分画像データを高速に重畳し、マスターノードに集約して、一枚の画像として生成する。binary-swap、binary-tree、direct send などの複数の重畳アルゴリズムを用いて、最適化する。

##### (v) エキスパートシステム構築フレームワーク

可視化とデータ分析のいわゆるポスト処理は、最終的にシミュレーションの結果から各領域の現象把握や設計に役立つ情報を抽出し、提示することが一つの目標である。これらの作業を効率化するため、特定領域の可視化・データ処理を行うための汎用エキスパートシステム構築フレームワークを整備する。

#### (5) 課題を解決するために必要なアプリケーション群（要求性能）

可視化は、大別して、計算と同時に行う In situ 可視化と計算で出力したファイルに対して行うファイルベース可視化がある。これらを単一の性能要求では記述できないため、それぞれの場合について必要性能を述べる。この場合、単一ノード性能は固定して考え、必要なバンド幅を見積もっている。

表 4.4.6-1 可視化要求性能

	ファイルベース可視化	In situ 可視化
想定する演算性能 (ノード当たり)	100TFLOPS	100TFLOPS
ネットワークバンド幅 (ノード当たり)	500GB/s	2TB/s
メモリ容量 (ノード当たり)	50GB	100GB
メモリバンド幅 (ノード当たり)	0.2TB/s	20TB/s
ストレージ容量	各シミュレータが生成する最大容量×2	各シミュレータが生成する最大容量×2
ストレージの帯域	0.1 PB/s	0.01 PB/s

インタラクティブな可視化にはグラフィクスハードウェアやアクセラレータがあると都合がよい。

In situ 可視化の場合は、組み込むアプリケーションと同程度の計算機性能が要求される。

#### (6) 他分野との連携

今後の大規模な計算機環境において重要な可視化技術となる In situ 可視化は、計算と同時にスケーラブルな可視化と分析を実施し、画像や統計情報などのメタデータを出力する技術である。In situ 可視化ツールの実装として、汎用的なライブラリとしてリンクする方法と解析領域に固有の分析コードを用いる方法がある。前者は、計算で利用するデータ構造との親和性が重要になる。一方、後者は、その処理機能のプログラム記述が必要になる。このように可視化技術は、分野横断的に切り出せる汎用的な技術と個々のアプリケーション分野に固有の手法とがある。分野毎の知識を使い効率的な手法を採る場合には、各分野と融合的に進める必要がある。

さらに、可視化は、アプリケーション分野だけに限らず、入力や表示デバイス技術などハードウェアとも密接に関連している。個別の計算手法には、行列演算や周波数分析などの数値ライブラリ、フレームワークなどの計算機科学分野との連携も必要となる。

#### (7) 公的な意見交換とそのサマリー

国内の可視化コミュニティは小規模で、適切な学会が存在しないため、個別のシミュレーション分野のアクティビティとなっている。このためグループミーティングや個別のディスカッションになっている。

公的な意見交換としては、2013 年の 6 月に開催される計算工学会において、可視化を含む大規模シミュレーションのシンポジウムが開催された。

準公的には、2012 年 10 月に理化学研究所計算科学研究機構（AICS）可視化チームが 3rd International Large-Scale Visualization Workshop を開催し、エクサスケールでの可視化環境についてのディスカッションを行った。参加者は、20 名程度の有識者（神戸大、京都大、宇宙航空研究開発機構、AICS など大規模データに取り組む機関が中心）で、ローレンスバークレイ国立研究所、Intelligent Light 社、米国エネルギー省関連の海外機関からも参加があった。ここでは、大規模データ可視化の課題と今後の取り組みについて意見交換、研究協力を確認した。

#### (8) ロードマップ

可視化・データ処理分野のロードマップを図 4.4.6-3 に示す。

年代 課題	2012 ～	2014 ～	2016 ～	2018 ～	2020 ～	2022 ～
フィージビリティスタディー	大規模データ管理 Interactive 技術	Visual Analytic Framework InSitu Vis Framework		エキスパートシステム機能		
要素技術開発	データストリーミング 大規模並列 レイアウト 階層データ管理ライブラリ TWD 表示	オンライン設計				
システム設計開発	Interactive 可視化		InSitu 可視化 Visual Analytic Framework		エキスパートシステム 統合可視化	

図 4.4.6-3 可視化・データ処理分野のロードマップ

#### 参考文献

- [1] McCormic, B., et al., Visualization in Scientific Computing, Computer Graphics, 21-6, 1987.
- [2] <https://wci.llnl.gov/codes/visit>
- [3] <http://www.paraview.org>
- [4] <http://www.vtk.org>
- [5] <http://visweek.org>
- [6] <http://www.supercomp.org>
- [7] <http://www.siggraph.org>
- [8] <http://www.eg.org>

- [9] Ma, K.-L., et al., Parallel volume rendering using binary-swap compositing, *Computer Graphics and Applications*, IEEE, vol.14, no.4, pp. 59-68, 1994.
- [10] Chao Li, et al., A survey of multi-projector tiled display wall construction, *Multi-Agent Security and Survivability*, 2004 IEEE First Symposium, pp. 452- 455, 2004.
- [11] Lum, E., et al., A hardware assisted scalable solution of interactive volume rendering of time-varying data, *IEEE Transactions on Visualization and Computer Graphics* 8, 3, pp. 286–301, 2006.
- [12] Sukharev, J., Correlation study of time-varying multivariate climate data sets, In *Proc. IEEE Pacific Visualization*, pp. 161–168, 2009.
- [13] Scholtz, J., Beyond Usability: Evaluation Aspects of Visual Analytic Environments, *Visual Analytics Science And Technology*, 2006 IEEE Symposium, pp.145-150, 2006.
- [14] Ma, K.-L., In Situ Visualization at Extreme Scale: Challenges and Opportunities, *Computer Graphics and Applications*, vol. 29, no. 6, pp. 14-19, 2009.
- [15] Stockinger, K., et al., Query-driven visualization of large data sets, In *Proc. IEEE Visualization*, pp. 167–174, 2005.
- [16] Silva, C. T., et al., Provenance for Visualizations: Reproducibility and Beyond, *Computing in Science & Engineering*, vol. 9, no. 5, pp. 82-89, 2007.
- [17] Sakamoto, N., et al., Visualization of Large-scale CFD Simulation Results Using Distributed Particle-Based Volume Rendering, *International Journal of Emerging Multidisciplinary Fluid Sciences*, Vol. 2, No. 2, pp. 73-86, 2010.
- [18] Itoh, T., et al., Hierarchical Data Visualization Using a Fast Rectangle-Packing Algorithm, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 10, No. 3, pp. 302-313, 2004.
- [19] 竹島, 高橋, 藤代, 位相ボリューム骨格化アルゴリズムの改良, *情報処理学会論文誌*, Vol. 47, No. 1, pp. 250-261, 2006.
- [20] Bi, C., Takahashi, S., and Fujishiro, I., Interpolation of 3D Diffusion Tensors by Locating Degenerate Lines, In *Poster Proc. of IEEE Pacific Visualization 2011*, pp. 9-10, March, 2011.
- [21] 宮村ほか, 時空間データの可視化, *可視化情報*, Vol. 30, Suppl.1, pp. 269-272, 2010.
- [22] Matsuoka, D., et al., Visualization for High-Resolution Ocean General Circulation Model via Multi-Dimensional Transfer Function and Multivariate Analysis, *International Conference for High Performance Computing, Networking, Storage and Analysis (SC12)*, November 10-16, 2012.
- [23] 白山, 知的可視化, 丸善, 2006.
- [24] 坂井ほか, 高精細タイルディスプレイを用いた並列ボリュームレンダリングシステムの実装, *情報処理学会論文誌:コンピューティングシステム*, Vol. 5, No. 3, pp. 124-136, 2012.
- [25] Kageyama, A., et al., Immersive Visualization in PI-CAVE, *Proc. International Conference on Modeling and Simulation Technology (JSST 2011)*, pp. 286-290, 2011.
- [26] 武田, 大規模粒子シミュレーションデータの PC による可視化, *可視化情報*, Vol. 32, No. 127, pp. 156-161, 2012.



- [27] 武井ほか, 大規模非定常数値シミュレーションのリアルタイム可視化 : 並列計算サーバによる可視化方式の実用化に向けて, 情報処理学会論文誌, ハイパフォーマンスコМП्यूティングシステム, Vol. 41(SIG\_8(HPS\_2)), pp. 107-118, 2000.
- [28] 村松ほか, 並列計算機上での流体解析のための実時間可視化システムの開発, 日本原子力研究所, JAERI-Data/Code 98-014, 1998.
- [29] Ohno, N. and Kageyama, A., Region-of-Interest Visualization by CAVE VR System with Automatic Control of Level-of-Detail, Computer Physics Communications, Vol. 181, Issue 4, pp. 720-725, 2010.
- [30] [http://www.csrp.riken.jp/application\\_j.html](http://www.csrp.riken.jp/application_j.html)
- [31] <http://hpci-aplfs.aics.riken.jp>