

S-H19-167

大規模スーパーコンピュータ技術と統合する連続波テラヘルツ光

応用技術の調査研究

(財)高度情報科学技術研究機構 飯塚 幹夫

1. 調査研究の概要

日本が世界を先導する高温超伝導素子による連続波テラヘルツ光技術と高性能スパコン利用技術を統合し、バイオ分野における画期的な分子情報を提供する日本独自の次世代テラヘルツ光応用技術を目指し、その技術課題、解決策、競争力のある研究方法を調査した。

2. 調査研究の背景・目的

近年、生化学、たんぱく質、微生物、遺伝子解析、先進医療、ナノバイオ工学などバイオ、ライフサイエンス分野は、画期的に能力を増すコンピュータ技術との協調作業により、著しく急速な進歩を遂げている。生命系の本質的理解の根底をなす生体高分子の特性の把握は最も重要な基軸で、こうした分野の基幹的な計測技術としてテラヘルツ分光分析技術が重要である。テラヘルツ波は電磁波と光の中間の周波数帯域(0.3~10THz)にあり、その周波数は生体高分子の骨格振動、振動・回転準位、分子間相互作用などの振動数帯にも相当し、かつ透過能力も高い特性がある。したがって、この特性を活かせば、化学物質、生体高分子、微生物、細胞、人体の病理組織などの動態の特性分析、さらには被曝リスクの低い医療用光源への応用も考えられる。しかし、こうした連続波テラヘルツ光を用いた分析観測から得られる観測情報量は莫大となる。その莫大な情報を扱いつつ、分子情報を逆解析し、分子の動的挙動を明らかにする作業に従来方法を用いるとしたら、それは不可能である。そこで、従来の構造生物学がコンピュータの急速な発達に助けられて今日のたんぱく質の静的構造を明らかにしたが、それらを遥かに上回る莫大な量の動的情報を解析するには、計測と大規模な高性能スーパーコンピュータの融合的なシステム化が必須である。そこでは、高速データマイニング、データハンドリング技術を活用し、莫大な情報を多面的かつ高速に分析する。さらに、その分析情報から分子の動的像を再構成するには、第一原理または量子論的、さらには経験的パラメータを用いる各分子動力学の階層的かつ高度な大規模シミュレーションを駆使する必要がある。

このため、連続波のテラヘルツ光源と大規模高性能スーパーコンピュータを利用し、世界に先駆けて、わが国独自の生体高分子の動的性質を統合的に把握可能な分析システムの開発を視野に、わが国独自の連続波テラヘルツ光応用分析システムの実現を目指すにあたり、解決すべき技術課題、準備すべき解析法、大規模シミュレーションソフトウェア、さらに必要な計算資源量等を調査し、また、その分析システムの実現を支援する技術研究、体制、計画等の研究方法も調査し今後の開発に繋げていくことを調査研究の目的とした。

3. 調査結果

日本が世界を先導する高温超伝導素子による連続波テラヘルツ光技術と高性能スパコン利用技術を統合し、バイオ分野における画期的な分子情報を提供する日本独自の次世代テラヘルツ光応用技術を目指し、その技術課題、解決策、競争力のある研究方法を調査した。その結果、以下の結論を得た。

3.1 連続波テラヘルツ光応用分析技術開発に関する動向調査

連続波テラヘルツ光応用分析技術は、光源の取り扱いが難しいことから主に正確な周波数精度が要求される実験室分光で用いられている程度が現状である。取り扱いが容易で高出力・周波数可変な連続波テラヘルツ波光源開発が連続波テラヘルツ光応用分析技術の課題である。現在、原理的に有望な連続波テラヘルツ光源は、高温超伝導体を使った発振素子であり、特に1~4THzの範囲では最も自然な光源である。

3.2 連続波テラヘルツ光応用分析システムでの計算処理内容の想定

コンピュータシステム要件は、そこで行われる計算処理の特性により決まるため、はじめに、計算処理の対象となる計算内容について想定した。図-1 に、連続波テラヘルツ光応用分析システムでの計算内容とその関係を示す。そこで調査研究に当たり、主な計算処理内容は以下の2つを想定した。

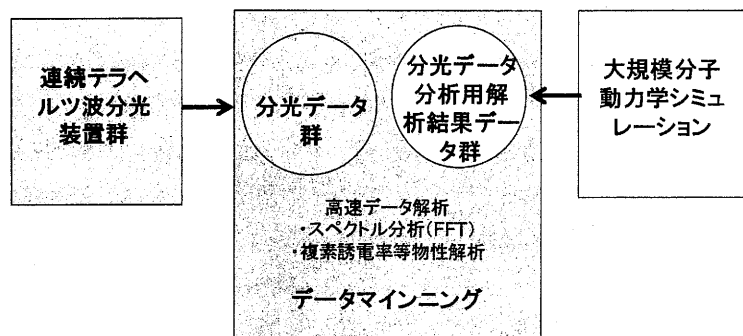


図-1 連続波テラヘルツ光応用分析システムでの計算内容

- (1) 分光データ情報の高速データ解析:分光データ処理(スペクトル解析、吸収バンドの同定のためのデータマイニング)による生体分子、構造変化等の同定、および莫大なデータの中に埋もれた有益な情報のデータマイニング解析を行なう。
- (2) 大規模分子動力学シミュレーション:テラヘルツ波との相互作用による立体構造の基準振動モード励起分布、立体構造変化による基準振動モード励起分布の変化を解析する。また、テラヘルツ波との相互作用下での電子状態、分子間相互作用の変化を解析する。解析データを、分光データ分析用データとして提供する。

3.3 連続波テラヘルツ光応用分析システムと融合した大規模高性能スーパーコンピュータシステムに関する調査

既存の分析装置との統合システムは見られなかった。分析装置との統合システムを構成する大規模高性能コンピュータシステムの概念的設計を検討した。本システムの以下の主要な2つ計算処理内容について課題と解決法を検討した。

1) 分光データ情報の高速データ解析のためのコンピュータシステム:解析ケース数は莫大であるものの、各ケースについては最先端の計算機能力を必要としないためグリッドコンピュータシステムの利用が有効である。しかし解析ノウハウが重要なため、分光データ情報の高速データ解析の幅広い分野の研究者、技術者が簡単に知識を集積し、必要な知識(ソフトウェア等)を利用し、さらに高度化して、循環、発展できるソフトウェアシステムの構築が必要である。このようにソフトウェアシステムを構築するため、分光技術分野のコミュニティをサポートする次世代分光データ情報の高速データ解析用コミュ

ニティフレームワークをグリッド上に構築することが有効である。

2)大規模分子動力学シミュレーションのためのコンピュータシステム:大規模分子動力学シミュレーションのための、数百ペタフロップス級の大規模高速並列計算機の開発とそのコードの高速最適化が課題である。そのためには、大規模シミュレーション基盤(並列モデリング(データ並列、タスク並列等)、設計パターン(データ並列のための分割法:領域分割法、原子分割法、エネルギー分割法、スペクトル分割法等)、並列プログラミング技術、並列コンパイラ、並列プログラミング構築支援システム等の並列処理技術等の共通技術)も発展させ整備することが有効である。

3.4 大規模データ解析、大規模分子シミュレーションに関する技術動向調査

・大規模データ解析については、グリッドコンピュータシステム上に、次世代分光技術コミュニティフレームワークを構築し、そのフレーム内に大規模データマイニングソフトウェア群を整備することが有効である。大規模データマイニングソフトウェアの開発においては、分光データに関する情報を距離空間へマッピングする一般的なモデルを開発する必要あり、データマイニング手法と最適化手法の大規模高速化が必要である。

・大規模分子シミュレーションについては、強結合近似モデルの大規模分子動力学シミュレーションのソフトウェアの開発が必要である。強結合モデルとリカーゾン法の組み合わせは、 $O(N)$ の計算負荷軽減性と大規模高速性を有し効果的である。またメモリ・演算器が階層化し、数万から数十万コアの超並列構造化した大規模計算資源への、そのコードの高速最適化が課題である。

3.5 世界に対し日本が先導性を保つための戦略的研究のあり方の検討

連続波分光分析技術、スパコン技術、材料技術等の有効な連携方法を検討した。連携は、日本が先導性を保つように、それぞれが世界レベルでの競争力を強めあうような連携法が必要である。例えば、図-2に示すような連携法を推奨する。この連携の相乗効果は大きく、他分野への波及効果も大きい。

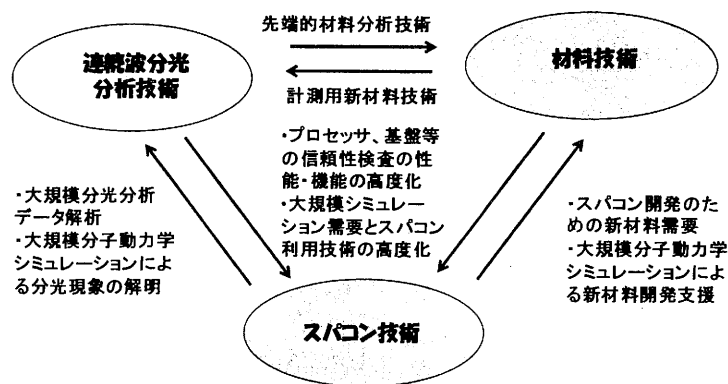


図-2 連続波分光分析技術、スパコン技術、材料技術等の有効な連携方法

4. まとめ

連続波テラヘルツ光応用技術において、世界に対し日本が先導性を保つためには、大規模計算機システムとの連携を強化した戦略的研究のあり方が、競争力強化において必須である。