

[2005]九州大学情報基盤センター一年報 : 2005年度

<https://doi.org/10.15017/1467612>

出版情報 : 九州大学情報基盤センター一年報. 2005, 2006. 九州大学情報基盤センター
バージョン :
権利関係 :

第6章 プロジェクト紹介

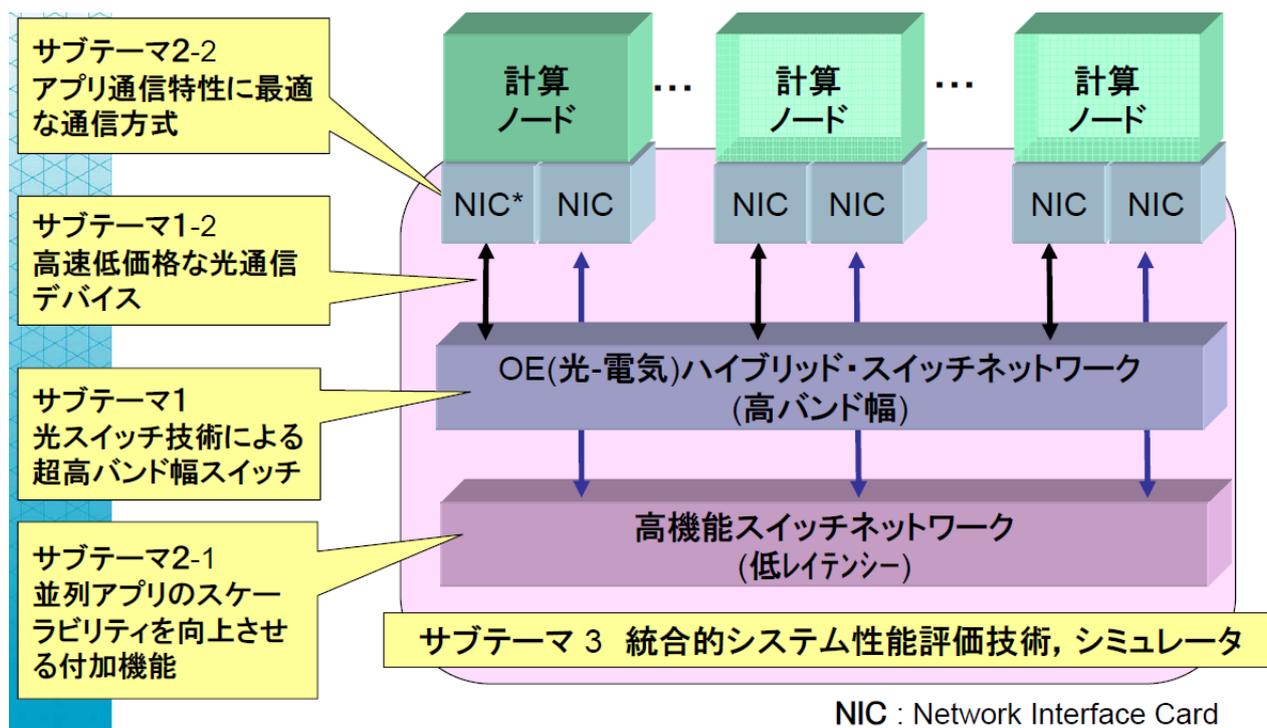
6.1 ペタスケール・システムインターコネクト技術の開発

研究代表者: 村上 和彰 (九州大学情報基盤センター長, Email: murakami@i.kyushu-u.ac.jp)

6.1.1 プロジェクトの概要及び 2005 年度の進捗状況

PSI プロジェクトとは、文部科学省「次世代 IT 基盤構築のための研究開発」、研究開発領域「将来のスーパー コンピューティングのための要素技術の研究開発」(平成17年度～19年度)に採択された研究開発課題「ペタスケール・システムインターコネクト技術の開発」である。

本プロジェクトでは、ペタフロップス超級スーパーコンピュータシステムの構成において数千～数十万規模の高速計算ノードを相互結合するシステムインターコネクト技術を対象に、現状のシステムよりもコスト対性能比で1桁上を目指して、高性能化、高機能化、低コスト化を同時に達成するための3つの要素技術、すなわち、光パケットスイッチと超小型光リンク技術、動的通信最適化による MPI 高速化、システムインターコネクトの総合性能評価技術を、九州大学の他に財団法人福岡県産業・科学技術振興財団、財団法人九州システム情報技術研究所、富士通株式会社と共同で開発する。



サブテーマ 1: 光技術を用いた超高バンド幅スイッチング技術の開発

● サブサブテーマ 1-1 「光パケットスイッチング技術の開発」:

波長多重 (WDM: Wavelength Division Multiplexing) 技術によるテラビット/秒級の超高バンド幅インターコネクと WDM 信号の一括切り換えが可能で 32~64 ポート程度の光パケットスイッチの要素技術を開発する。これらの技術により高バンド幅化した転送データ (ペイロード) を電気信号に変換することなくナノ秒オーダーでの高速一括切り換えが可能となる。その結果、光ファイバケーブルと光スイッチ機器が各々数千本、数十台にまで大幅に削減され、それに伴い占有面積、消費電力、コストの大幅な削減と信頼性の向上が可能となる。3年間の研究期間内で 10 ポート程度の光パケットスイッチの試作、動作検証、課題抽出を行い、本技術の確立を目指す。

本年度はインタコネク制御部及びリーフスイッチ部について回路設計を行った。さらに、光送受信部のモジュール設計を行った。

● サブサブテーマ 1-2 「光電気変換部集積化技術の開発」:

ボード (プリント基板) 間インターコネクの光化のためにはノードのボード端部に光送受信部を多数配置する必要があるが、従来技術ではその大きさのため実現が困難である。本研究ではボード間インターコネクを数十ギガビット/秒に高バンド幅化でき、かつ、現状最新の光送受信部 (XFP: 10 Gigabit Small Form Factor Pluggable) に比べ 1/10 以下の占有面積とコストを可能にする光電気変換部集積化の要素技術を開発する。本技術では数チャンネルを 1 モジュールに集積化し、それを多数配置することで数十~百数十ポートの光送受信部をボード端部に実装可能とする。3年間の研究期間内で 10 ギガビット/秒 × 4 チャンネル以上の超小型光送受信部モジュール (幅 10mm × 長さ 10mm × 高さ 5mm) を設計試作し、動作検証、課題抽出を行い、本技術の確立を目指す。

本年度は、光電気変換部のモジュールの小型集積化に関する基本設計を行い、13mm × 15mm × 5mm のサイズでの部材設計を完了した。

サブテーマ 2: 高機能・高性能システムインターコネク技術の開発

● サブサブテーマ 2-1 「コレクティブ通信をサポートする高機能スイッチの開発」:

ノード間の通信経路上で通信データに対して種々の演算を施すことが可能なスイッチのハードウェアを設計し、これを搭載した高機能スイッチ装置の開発を行う。本高機能スイッチ装置を用いて、総和をはじめとする種々のコレクティブ通信機能をハードウェアにオフロードし、アプリケーション全体の高速化を狙う。本研究では、当該高機能スイッチを用いた数十ノード規模の評価システムを構築し、コレクティブ通信を従来比で 5 倍以上高速化することを目標に試作機での性能検証を行う。

本年度は、ノードマッピングによらず同じ Collective 演算の結果を導出可能な 1 パスアルゴリズムを開発した。また、ペタスケールシステムに必要な信頼性、安定性を確保するためのプロトコルを設計した。

● サブサブテーマ 2-2 「動的最適化を用いた MPI 高速化技術の開発」:

適応的に通信パラメータを調整する動的通信最適化技術の開発を行う。本技術は、MPI レベルで得られるノード数やプロセス割り当ての情報を取得する機能、ネットワークインタフェースに送出されるデータサイズをプロファイル (収集、分析) する機能、および、それらによって得られる情報から通信パラメータやプロセスの配置を最適化する機能で構成される。これらの

機能を高機能スイッチやそれに接続する通信インタフェースに実装し、動作検証、課題抽出を行い、MPI 通信における動的最適化技術の有効性を実証することを目指す。

本年度は、Collective 通信の実装アルゴリズムによる性能の変化について解析するとともに、実行環境の状態に応じた性能予測モデルを開発した。

サブテーマ 3: ペタスケール・システムインターコネクトの性能評価環境の構築

- サブサブテーマ 3-1 「ペタスケール・アーキテクチャの開発」:

代表的な大規模科学技術計算アプリケーション・プログラムの計算や通信パターンを分析し、その主要計算処理部の超並列分散化を検討する。また、これらの主要計算部を効率よく実行できるペタスケール・システム用計算ノードアーキテクチャを検討する。さらに、計算ノードのシミュレータを開発し、主要計算部の実行時間をシミュレーションにより評価し、サブサブテーマ 3-2 のシステムインターコネクトの性能評価環境と結合して、システム全体のアプリケーション実行性能評価を行う。また、性能評価の過程では、性能ボトルネックを解消するようアーキテクチャの改善を検討し、大規模科学技術計算アプリケーションの実行に適したペタスケール・アーキテクチャを開発する。

本年度は、対象とするアーキテクチャの仕様を定義し、さらに計算ノード内の SIMD 機構を活用するコンパイラを実装した。

- サブサブテーマ 3-2 「ペタスケール・システムの性能予測技術の開発」:

小規模システムでのプログラム実行によって得た各種統計情報に基づき、ペタスケール・システムの性能を現実時間内で精度良く見積もる技術を開発する。また、特にシステムインターコネクトに焦点を当て、システム設計時に利用できる性能評価環境を構築する。さらに、実際のアプリケーション・プログラムを用い、開発した性能見積り手法の評価を行う。最終的には、実在する高性能コンピュータ・システム（例えばテラスケール・システム）の性能予測を想定した評価実験を実施する。本技術を確立することで、今後のペタフロップス超級、さらには、エクサ級システムの開発に資することを目指す。

本年度はシミュレーション環境の基本構成をとりまとめるとともに、シミュレーション用プログラムコードの作成方法に関する検討とその支援ツールを開発した。

- サブサブテーマ 3-3 「ナノアプリケーションのコード分析と超並列化」:

代表的なナノ分野アプリケーションを選定し、実行プロファイル情報および 3-2 による性能評価ツールを用い、アプリケーションコード中の高負荷部の抽出と当該部分の超並列化を行う。更に高精度の性能評価が必要なコードについては 3-1 による計算ノードシミュレータを利用した詳細なコード解析を行う。

本年度は二つの代表的なナノアプリケーション (PHASE, GAMESS) の分析を行い、実行プロファイル情報から高負荷部の抽出を行った。

6.1.2 PSI シンポジウム 2005 の開催

本プロジェクトの研究計画などを紹介するシンポジウムを、2005 年 12 月 19 日に日本科学未来館で開催した。

プロジェクトの概要や各サブテーマの目標等に関する講演の他に、本プロジェクトの背景でもある次世代スーパーコンピュータの開発に関する基調講演を日本電気株式会社の渡辺 貞氏より、さら



に本プロジェクトに求められる役割等についての総括を、筑波大学の朴 泰祐教授より、それぞれ頂いた。

参加人数は162名であり、講演後の質疑や当日配布したアンケートの結果より、本プロジェクトへの関心の高さをあらためて認識できた。

6.2 NAREGI(National Research Grid Initiative) プロジェクト

6.2.1 NAREGI プロジェクトの概要

「超高速コンピュータ網形成プロジェクト (National Research Grid Initiative : 通称 NAREGI)」は文部科学省が進めるプロジェクトで、2003 年からスタートしました。プロジェクトの目的は広域分散型の最先端研究教育用大規模計算環境 (サイエンスグリッド環境) を実現するために、世界標準に準拠した実運用に耐える品質のグリッド基盤ソフトウェアを研究し、開発することです。

NAREGI プロジェクトでは国立情報学研究所(NII)におけるグリッド基盤ソフトウェア研究開発、及び分子科学研究所におけるナノアプリケーションを用いた開発グリッド基盤ソフトウェアの実証研究、の両面から研究開発が推進されています。

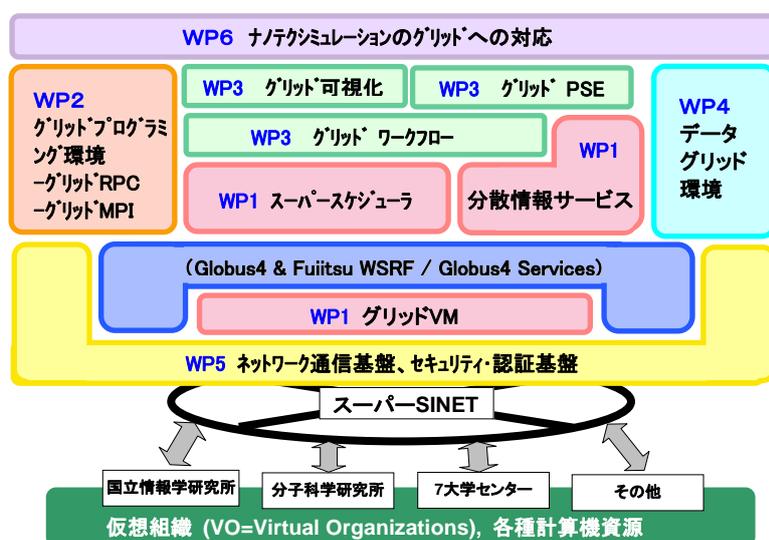


図1 NAREGI プロジェクトグリッド基盤ソフトウェア開発における各 WP の関係

NII におけるグリッド基盤ソフトウェアの研究開発体制は Work Project(WP)1-6 から成り(図 1)、九州大学の青柳が責任者を務める WP6 はグリッド基盤を最大限に活用するためのナノアプリケーションのグリッドへの対応を行っています。以下では WP6 の主な研究開発内容について説明します。

6.2.2 ナノアプリケーションの連成計算に関する研究開発

ナノサイズの複雑な物理現象の解析においては空間・時間スケールや物理法則が異なるシミュレーション間の連携(マルチスケール・マルチフィジックス連成解析)がしばしば重要です。WP6 ではグリッド上での大規模な連成解析を支援するツールとして連成に関わるアプリケーション・コンポーネントの独立性を保持しながら、幅広いナノアプリケーション群を簡単な操作で連成可能なグリッド連成ミドルウェア (Mediator) を開発しました。Mediator の使用によりコンポーネント毎に異なる方式で離散化された物理量のデータ交換を効率良く行うことが可能です。連成計算の例として無限個の溶媒溶質相互作用を取り込む RISM 法^{*1}とナノ分子の電子状態を解析する FMO 法^{*2}の連成による水溶液中におけるタンパク質の電子状態計算の流れを図2に示します。

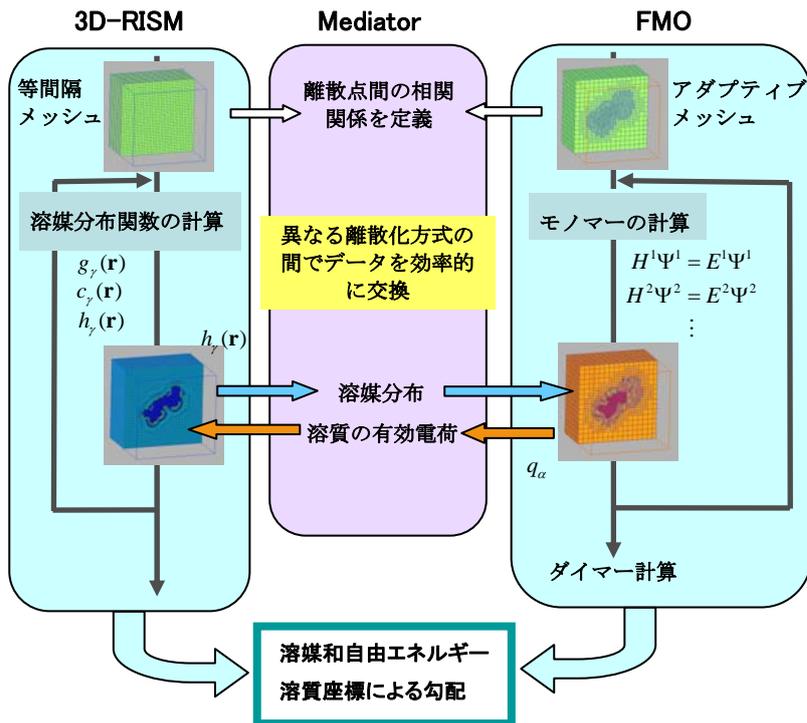


図2 RISM-FMO 連成計算の流れ

FMO法はナノ分子を小さな部分系(フラグメント)に分割し並列計算により効率的に電子状態を解析する方法です。一方RISM法では分子性液体の統計力学に基づく積分方程式(RISM方程式)を解くことにより溶質-溶媒間のサイト-サイトペア相関関数 $h_\gamma(\mathbf{r})$ を求めます。連成計算では初期値から出発して3D(3次元)-RISMによる溶媒分布とFMOによる溶質分子のモノマー計算が自己無撞着となるまで繰り返し、各繰返しの

最後に溶媒分布と溶質原子上の有効電荷を交換します。RISM法により3次元等間隔メッシュ上の値として得られた溶媒分布のデータはFMOの計算量節約のために溶質の近傍で密、遠方で疎なアダプティブメッシュ上の値へと変換(マッピング)されます。変換に必要なメッシュ間の相関関係を計算開始時にMediatorに登録することにより、連成アプリケーション開発者は離散化手法の異なる個々のアプリケーション内でそれぞれのデータ構造と交換データの格納場所等を頭に意識せずに、効率的なデータ交換が可能です。

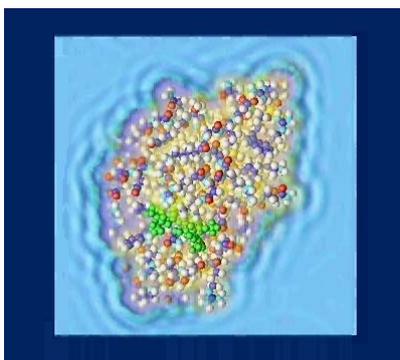


図3(a) リチウムの水和構造

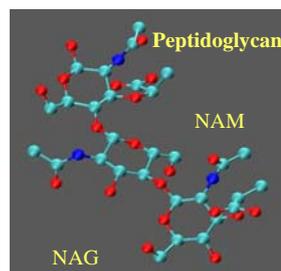
NAM: N-Acetylmuramin
NAG: N-Acetylglucosamin

図3(b) Peptidoglycanの構造

RISM-FMO 連成計算の応用例としてリゾチウムの溶菌作用を解析するために水溶液中のリゾチウムの電子構造計算を行った結果を図3に示します。この生体内酵素反応では活性中心から細胞壁を構成する糖タンパク(Peptidoglycan)(図3(b))へのプロトン移動が重要である

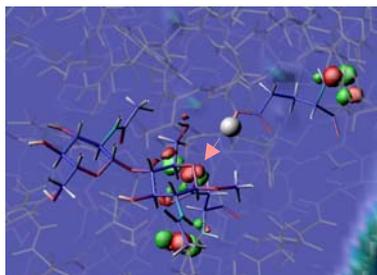


図 3(c) 活性中心付近の電子構造

といわれてきました。分子動力学シミュレーションによる水和構造(図 3(a))を用いて行った計算ではプロトンの位置が移動するにつれて Peptidoglycan の電子構造(図 3(c))に大きな変化が認められ、糖鎖の結合乖離のミクロの過程を RISM 法と FMO 法の大規模な連成計算によって解明することに成功しました。そのほか我々の研究グループでは RISM-FMO の他にも分子動力学(MD)と有限要素法(FEM)の連成計算やMDとPB(ポアソン-ボルツマン方程式)解法の連成計算等にも Mediator を応用し成果を上げています。

6.2.3 ナノアプリケーションの「コンポーネント化」に関する研究開発

グリッド環境におけるジョブの効率的な分散実行を行うためにナノアプリケーションを機能単位に分割し、分散効率および各々の計算粒度やデータ転送容量、遅延許容度等に応じてグリッド基盤上の最適な計算資源を選択できるようにアプリケーションやツール類の「コンポーネント化」を行っています。例として上述の FMO をコンポーネント化した場合の計算の流れを図 4 に示します。

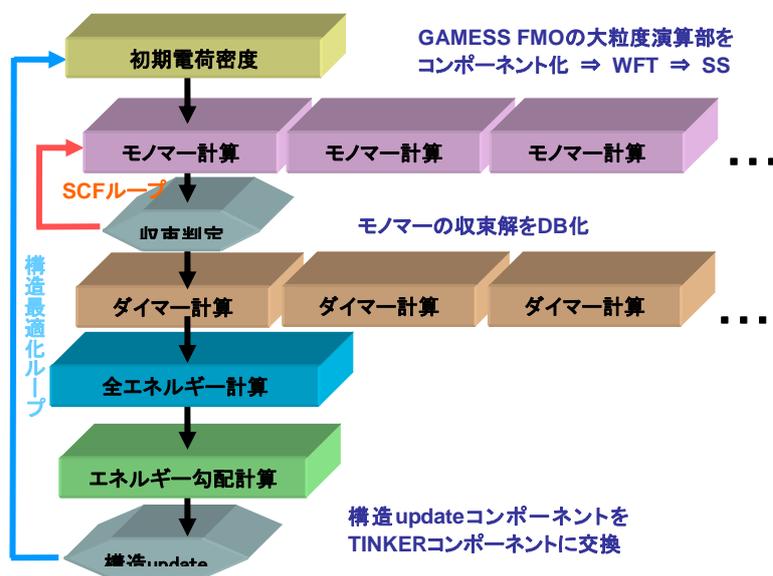


図 4 疎結合 FMO 計算の流れ図

流れを図 4 に示します。

FMO 法ではナノ分子をフラグメントに分割し、全フラグメント(モノマー)およびフラグメント対(ダイマー)に対して電子状態計算を行い、それらの結果からナノ分子全体のエネルギーとエネルギー勾配を計算します。各フラグメント(対)の計算は他のフラグメントの影響(クーロン相互作用)下で行われ、モノマー計算は初期値から始めて結

果が無撞着となるまで繰り返されます(SCF ループ)。これらの計算部分(初期値、モノマー、ダイマー、全エネルギー計算部分)をそれぞれ、独立の演算モジュールとして分離(疎結合化)しファイル I/O によりデータ交換を行うようコンポーネント化しました。これらのコンポーネントを P S E (WP3 開発の Problem Solving Env.) ツールのデプロイメント機能によりグリ

ッド基盤上に配置し、ワークフローツール(WFT)により各コンポーネントの接続(図5の線)を行い全体の FMO による計算シナリオを作成します。WF による実行では WP1 開発の IS(information Service)機能及びスーパースケジューラ(SS)と GridVM の機能により、各モジュールに対する最適な資源の割り当てや WP3 開発のグリッド可視化システムによる結果表示等が利用可能です。WF の他に GridRPC によるグリッド環境を利用したより大規模な並列計算を行うことも可能です。また疎結合 FMO のために電子状態計算結果のデータベース (DB) 化システム、構造最適化エンジンとして分子動力学法プログラム Tinker を利用した構造最適化計算と計算結果の DB 化システムを開発しています (図6)。

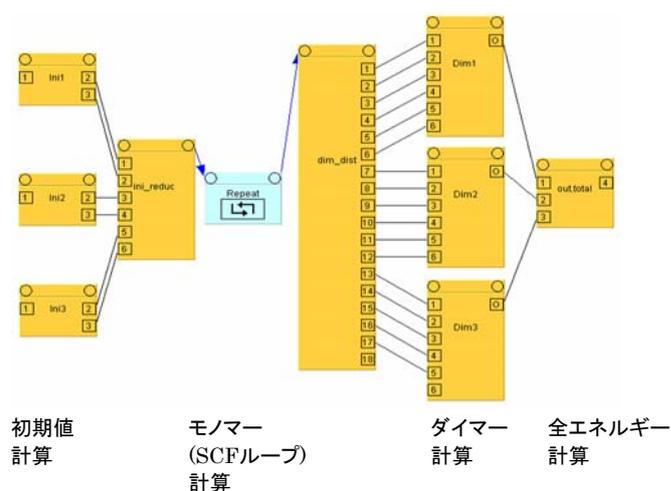


図5 疎結合 FMO のエネルギー計算部分のワークフロー



図6. グリッド上のコンポーネント連携によるタンパク質構造解析のフレームワーク

※1 RISM法：オリジナルコードは 分子科学研究所 平田教授グループにて開発された。

※2 FMO法：オリジナルコードは 産業技術総合研究所 北浦博士グループにて開発された。

6.3 拠点大学交流事業：次世代インターネット技術のための研究開発と実証実験

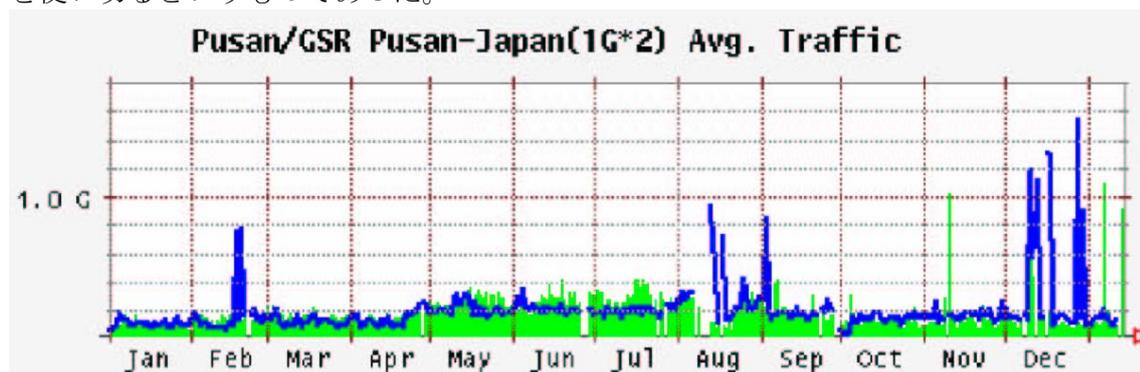
6.3.1 拠点大学交流事業とは

インターネットは、現在、社会の生活基盤になっている。例えば、水や電気と同じように、その利用ができなくなると、もはや生活がなりたたなくなっている。このインターネットの構成要素は非常に多岐にわたり、多くの技術が結集されて、今のインターネットが維持されている。しかし、水や電気と違う点は、その構成要素や技術が今もなお、性能のよいもの、機能の高いものに進歩している点である。

このような、インターネットに関連する多くの技術に対して、この拠点大学交流では、九州大学を中心としたネットワーク関係の研究者および韓国の忠南大学を中心としたネットワーク関係の研究者が共同／結集して研究を行いこれらの次世代のインターネット技術に総合的に取り組んでいる。

6.3.2 平成 17 年度の主な研究活動

現在、本事業には、7 つの共同研究テーマがあるが、この中で、次世代インターネット基盤技術の研究開発および、次世代インターネットを用いた遠隔医療の開発と臨床応用の活動は非常にアクティブである。次世代インターネット基盤技術の研究開発共同研究は、主に、日本側が北部九州を中心とした次世代インターネット研究コミュニティである九州ギガポッププロジェクト(QGPOP, 代表: 岡村耕二)と、韓国側が KESG (Korean Engineering Study Group, 代表: Choong Seon Hong)の共同会合が中心になって推進されている。QGPOP と KESG による共同研究は、年間を通して、ほぼ毎月、日本と韓国交互に行われ、各メンバーによる最新の研究状況の報告や、実証実験に関する打ち合わせを行った。平成 17 年度に行った次世代インターネット基盤技術の研究開発共同研究の中のテーマのうち、最も際立っていたのは、日韓での非圧縮 HD ビデオ(超高精細動画像)の伝送実験で、現在の 2.4G の回線のうちの 1.5G を使い切るといったものであった。



上記のグラフは、2005 年 1 月から 12 月までの日本と韓国の研究ネットワークのトラフィックを示すが、回線容量 2.4G に対して、通信容量が 1Gbps を超えることは我々の実証実験がはじまる 11 月以前にはなかったことを示している。我々の実証実験が本格的に始まった 12 月以降は、たびたび 1Gbps を超えるトラフィックが発生しており、本事業による実証実験で、初めて次世代的な通信の研究が開始され始めたことを示している。なお、高精細画像伝送実験という立場では、韓国の釜山にあるネットワーク処理装置の性能不具合がこの実証実験で発見され、完全な品質での伝送実験は平成 17 年度には完了していない。よって、平成 18 年度も引き続き、この実証実験を継続する予定である。次世代インターネット基盤技術の研究開発共同研究の他の興味深いテーマとしては、日本の九州大学と韓国の忠南大学それぞれで、

学外とのトラフィックを記録・解析し比較するというもので、著作権保護という観点から P2P 利用が厳しく制限されている日本の大学と、トラフィックの浪費という観点から P2P 利用が抑制されている韓国の大学のトラフィックパターンの違いが明らかに異なるという興味深い結果が得られている。この研究は今後、論文などにまとめられ公開される予定であるが、この研究は本事業による深い日韓交流共同研究のたまものであるといえる。

また、次世代インターネットを用いた遠隔医療の開発と臨床応用共同研究の一部のメンバーは QGPOP あるいは KESG のメンバーでもあるため、次世代インターネット基盤技術の研究開発共同研究にも参加している。また、次世代インターネットの基盤研究を行っている共同研究に対して、次世代インターネットの応用研究を行っている共同研究ががっちりスクラムを組んで実用的な研究を進めているよい例であるといえる。平成 17 年度は、5 月にソウルで開催された韓国医学会全国大会、11 月に横浜で開催された日本医療情報学秋季大会さらに 1 月のセミナー時と、次世代インターネットの先進的な技術を用いたおおがかりなデモンストレーションを共同で行ったが、いずれも成功し、特に日韓あるいは世界レベルの医療サイドから非常に高い評価を得ることができた。特に、11 月のデモでは、ソウル大学校附属病院の最先端の電子カルテシステムの紹介・デモをライブで行なえるまでに我々の事業が技術的、また、体制的に整えられていることが示された。なお、この韓国・ソウルからの電子カルテシステムは日本ではほとんど導入が進んでいないため、その紹介、デモは非常に好評で、本来ならば、会場に集まった数百名の日本人医師が旅費を負担してソウルに見学に行かなければならないところを、このようなシステムを利用できたことで画期的な医療教育が実践できたといえる。

右の図は、セミナー時行ったデモの一部である。これは、高度な内視鏡手術の映像を患者の完璧なプライバシー保護を行なった状態で、東京の会場にライブで伝送した映像を用いて次世代インターネットを用いた遠隔医療の開発と臨床応用の共同研究者メンバー及び APAN に参加した各国の医療メンバーによる活発な議論が行なわれ、我々の取り組みがほとんど実用化レベルに達していることが示された。



6.3.3 拠点大学交流事業セミナー

拠点大学交流事業では、年に 2 回基本的に全メンバーが参加するセミナーを開催している。平成 17 年度は、9 月に韓国で、1 月に日本で、合計 2 回のセミナーを開催した。韓国で開催したセミナーは、九州大学大学院システム情報科学研究所と忠南大学で行われているワークショップとジョイントする形で開催した。本事業では、テーマはインターネットのみに限られるが、大学のワークショップとジョイントすることで、参加者の幅を広げることができた。さらにこの韓国で開催されたセミナーには、九州大学および忠南大学の両総長が出席され、格式の高いものとなった。

1 月に日本で開催されたセミナーは、より広く国際的な見地から本事業を評価するために、日本・東京で開催される APAN (Asia Pacific Advanced Network) 国際会議と共催した。拠点大学のセミナーに、APAN 参加者の日韓以外の国の方に参加してもらい、よりグローバルなセミナーを開催し、本セミナーの国際的なプレゼンスを高めることを一つの目的とした。さらに、通常のセミナーよりも大規模になることにより、各研究発表者の発表の質のさらなる向上を目指した。APAN 会合には 20 カ国以上から約 400 名の参加があり、その中で行われた

拠点セミナーも従来のセミナーに加えて、日韓以外の国の研究者また、企業からの発表もおりませ、大盛況であった。例えば、次世代インターネットを用いた遠隔医療の開発と臨床応用の共同研究のセッションでは、日本、韓国に加え、中国、香港、インド、フィリピン、タイ、シンガポール、オーストラリアからの医療関係者の発表、参加があった。また、前述した、インターネット上での高度プライバシー保護技術を用いたライブ手術映像伝送デモ、また、韓国からの非圧縮手術映像の伝送デモなどを行い、本事業の取り組みが実用段階であることを世界中に示すことができた。結果として、このセミナーでは、本拠点方式メンバー以外の方に我々の日韓次世代インターネットに関するプロジェクトを十分認知させ我々の国際的なプレゼンスを十分高めることにも成功し、本セミナーの目的を十分果たしたといえる。