次世代HPCにおける大規模データの可視化環境

小野謙二
理化学研究所

VCADシステム研究プログラム
機能情報シミュレーションチーム

次世代計算科学研究開発プログラム
生命体基盤ソフトウェア開発・高度化チーム

SS研HPCフォーラム2008

TOC

- 可視化システム構築上で考慮すべき背景
  - 計算機環境
  - アプリケーション，データ構造
  - 問題点
  - 方向性
  - ケーススタディ
- 提案システムの概要
  - システム全体のイメージ
  - コンセプト
  - 開発項目
- 大規模データ可視化システム
  - 可視化シナリオ
  - 機能設計と機能モジュール群
  - インタラクティブ性の確保の方針
  - 開発の方針

可視化システム構築上で考慮すべき背景

- 計算機環境
- アプリケーション，データ構造
- 問題点
- 方向性
- ケーススタディ

計算機環境　TOP500 JUNE 2008

Rank	Site	Processors	RMax	Processor	System Model
1	DOE/NNSA/LANL	122400	1026000	PowerXCell 8i	BladeCenter QS22 Cluster
2	DOE/NNSA/LLNL	212992	478200	PowerPC 440	BlueGene/L
3	Argonne National Laboratory	163840	450300	PowerPC 450	BlueGene/P
4	Texas Advanced Computing Center/Univ. of Texas	62976	326000	AMD x86_64 Opteron Quad	Sun Blade x6420
5	DOE/Oak Ridge National Laboratory	30976	205000	AMD x86_64 Opteron Quad	Cray XT4 QuadCore
6	Forschungszentrum Juelich (FZJ)	65536	180000	PowerPC 450	BlueGene/P
7	New Mexico Computing Applications Center (NMCAC)	14336	133200	Intel EM64T Xeon 53xx (Clovertown)	SGI Altix ICE 8200
8	Computational Research Laboratories, TATA SONS	14384	132800	Intel EM64T Xeon 53xx (Clovertown)	Cluster Platform 3000 BL460c
9	IDRIS	40960	112500	PowerPC 450	BlueGene/P
10	Total Exploration Production	10240	106100	Intel EM64T Xeon E54xx (Harpertown)	SGI Altix ICE 8200
16	The University of Tokyo	12288	82984	AMD x86_64 Opteron Quad	Hitachi Cluster
20	University of Tsukuba	10000	76460	AMD x86_64 Opteron Quad	Appro XtremeServers
24	Tokyo Institute of Technology	12344	67700	AMD x86_64 Opteron Dual	Fire x4600 Cluster
34	Kyoto University	6656	50510	AMD x86_64 Opteron Quad	Fujitsu Cluster
49	The Earth Simulator Center	5120	35860	NEC	SX6

ヘテロ環境 > 開発環境の複雑化
Peta-scale/2008, Exa-scale/2018（SciDAC2008）

HPCとアプリの動向・課題

- 計算機環境
  - MPP
  - マルチコア
  - アクセラレータ
  - ヘテロ環境
    - 開発環境
- アプリケーション
  - プリプロセス
    - モデル作成
    - 境界条件設定
  - シミュレーション
    - ヘテロな計算機環境
  - ポストプロセス
    - 分散並列可視化環境
- 主要な可視化パターン
  - パラメータスイープ
  - 時系列可視化
  - 比較可視化，最適化空間可視化
  - 多変数可視化

システムの大規模複雑化

データ爆発

- Petaプロジェクトの分野
  - ライフサイエンス，宇宙・航空，天文，ナノマテリアル
  - 防災，地球物理，ものつくり
- データ構造
  - 直交系
  - 非構造系
  - 粒子系
  - AMR
- 規模
  - TB~PB
- シミュレータ
  - 大規模並列計算
  - 連成解析　マルチフィジックス，マルチスケール

大規模並列計算の後処理における問題点

- データコピー，移動が高コスト
  - 処理時間，MMU/HDD容量
- データは動かせない
  - 適切なツールなしにはデータアクセスさえ不可
  - 動かせたとしても，処理手続きが煩雑，面倒
  - ユーザの心理的ハードルは高い

- 本当にやりたい解析以外の処理が少なくない
  - ファイル操作，処理準備
  - 「考えること」に集中できる環境整備が必要

問題点の要因

- データの大規模性
  - 空間規模，時系列データ，多変数
- データの分散性
  - 分散並列，GRID
- システムの複雑性
  - ヘテロ環境，ファイルシステム，ネットワーク
- more…

に起因し，

- 作業環境，データマネジメント，知識発見の枠組みが未整備
  - storage, retrieve, archive, indexing, summarize, sharing

大規模計算の後処理システムの方向性

- データを動かさない
  - その先には．．．
    - データ共有・協調・協同作業
    - リモート，仮想組織

- 集約的なデータサービスが中心となる
  - アクセス，分析，処理
  - データ，結果，知識，リソースの共有
    - グループ単位のデータリポジトリ
    - データのブラウズ，検索，分析

↓

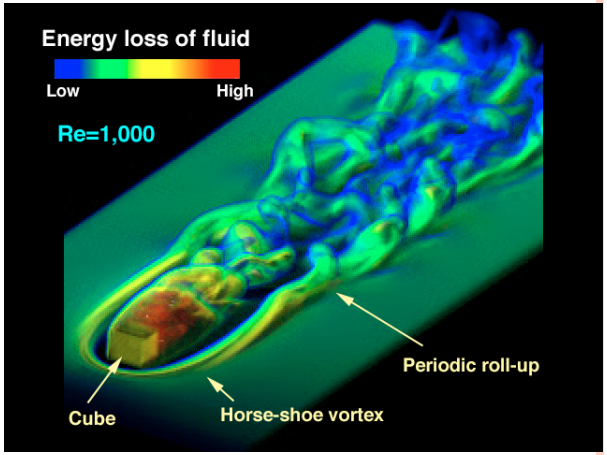
  - 現象理解，知識の共有

ケーススタディ　ある研究者の一週間

- パラメータ計算（パラメータスイープ）
  - シミュレーションの計画を練る
  - パラメータ設定を繰り返す
  - 計算モデルを用意し，ジョブスクリプトを投入
- 数日経過．．
  - ジョブの状況を確認し，結果をDL．．．数日
  - あるいは，ローカル処理のためデータを間引く
  - 比較検討のためのデータ可視化準備
  - 統計量をチェック
  - インタラクティブに結果を観察
  - ムービーを作成
  - 作業の繰り返し．．．
- 結果を眺めて，スナップショットを作成，並べて比較
  - 考える，現象の理解
  - 論文作成のための図を作成

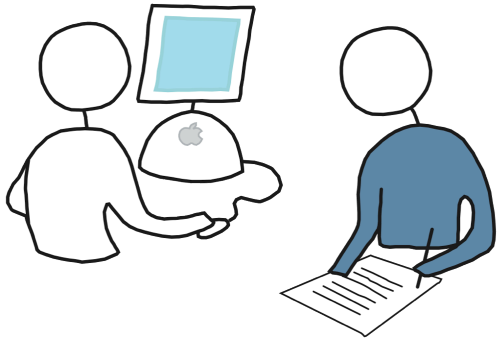
ケーススタディ　可視化分析のストーリー

- 場の全体構造の把握
  - 流線やボリュームレンダリングなど
  - 非定常挙動，ROIの特定
- 写実的イメージによる観察
  - 人間の認識力を利用したディテールの把握
- 簡素化した表現
  - 空間構造を記号的に把握
  - 時系列の変化など複雑なものを抽象的に把握
- 詳細な分析
  - 例えば，
    - ROIからの流線の逆積分
    - 温度勾配と渦との関係
- 知識への昇華
  - 様々な分析結果を統合的に見て高次の知識へと昇華
  - アノテーションによる知識表現



考える過程を支援
観察と分析を同時に

リモート可視化における
ポストフェーズのユーザシナリオ１



- ログイン，ファイル操作
  - rlogin, ssh, (s)ftpなどで操作可能
  - シングルサインオン
- 定型作業のスクリプト自動処理
  - クライアント上で作業の実行シナリオを記述し，プレビューで確認する
  - プロセス実行シナリオをサーバーで実行するために対応するスクリプトを書き出し，スクリプトキューに投入する
  - 実行状態をモニタする．オフラインのときは状態を保持し，再びオンラインになったときにステータスを取得し反映させる
  - 自動処理は，ファイル検索，ファイルIO，処理，可視化アプリ駆動などポストフェーズに関わるものを対象とする

リモート可視化における
ポストフェーズのユーザシナリオ２

- リソース管理
  - リソース情報を表示する
  - キューの情報を表示する
  - インタラクティブ（dedicated）モードの時間予約を行う
  - キューにジョブを投入する
  - 認証管理を行う
  - きめ細かなグループ管理制御
- アプリケーションを起動
- PCのアプリケーションと協調的に作業をすすめる
  - カット＆ペースト
- パラメータスイープ
  - パラメータ空間の設計を行う
  - バッチスクリプトを書き出す
  - バッチ投入
  - 処理プロセスのモニタ
  - 各パラメータの計算結果の可視化処理を自動化する

リモート可視化における ポストフェーズのユーザシナリオ３

- データベースによる支援
  - クライアントとサーバーは協調してデータを処理
    - サーバーとクライアントでDBが動作
  - オンライン/オフラインのタイミングに合わせて同期
  - ファイル群のなかから目的のファイルを検索する
  - 計算結果ができると，エージェントが動き，結果のDBへの登録と関連づけを自動で行う
    - データ記述のルールを定める必要がある
    - メタ情報は自動でXML表現
    - ディレクトリ毎にインデクスファイルを自動生成
- データをアーカイブする
  - システムアーカイバーとの整合性を考慮
- 大規模データを可視化する
  - サーバークライアント型のシステムとの整合性
  - 可視化の自動処理　ー　可視化システムのスクリプト
  - 重要な可視化
    - 時系列，多変数，比較

ポストフェーズの特徴

- 計算と異なり主観的作業
  - ユーザの数だけ処理のシナリオがある
- インタラクティブ性が効率に大きく影響
  - 計算処理システムと同じ運用では，ユーザの利便性・生産性が高まらない
- ポストフェーズのリソース運用が効率化の鍵

- 専用リソース（CPU/GPU cluster）の有用性
  - 柔軟な処理への対応
  - 多様なソフトウェアシステムの利用が想定される
  - 標準的なLinux clusterシステムが既存資産を利用可能で利便性が高い

LLNLのシステム

	BG/L	Gauss
CPU	131,072	512 (Dualx256)
MMU(TB)	32.8	3
Disk (TB)	900	900
Peak (TFLOPS)	280.6	-
	IBM	AMD Opteron (2.4GHz) nVidia QuadroFX4500 (512MB >128GB, total) (17GB/s, duplex)

巨大リソースを利用したシミュレーション

- チームによるプロジェクト推進
  - 複数研究者のコラボレーション

- 結果の共有化の仕組みが必要
  - コミュニティ群（アプリケーション毎ほか）
  - 共有する情報
    - データ，メソッド，プログラム，結果，知識，．．．

  - きめ細かなprivilege制御がシステムに必要

提案システムの概要

- システム全体のイメージ
- コンセプト
- 開発項目

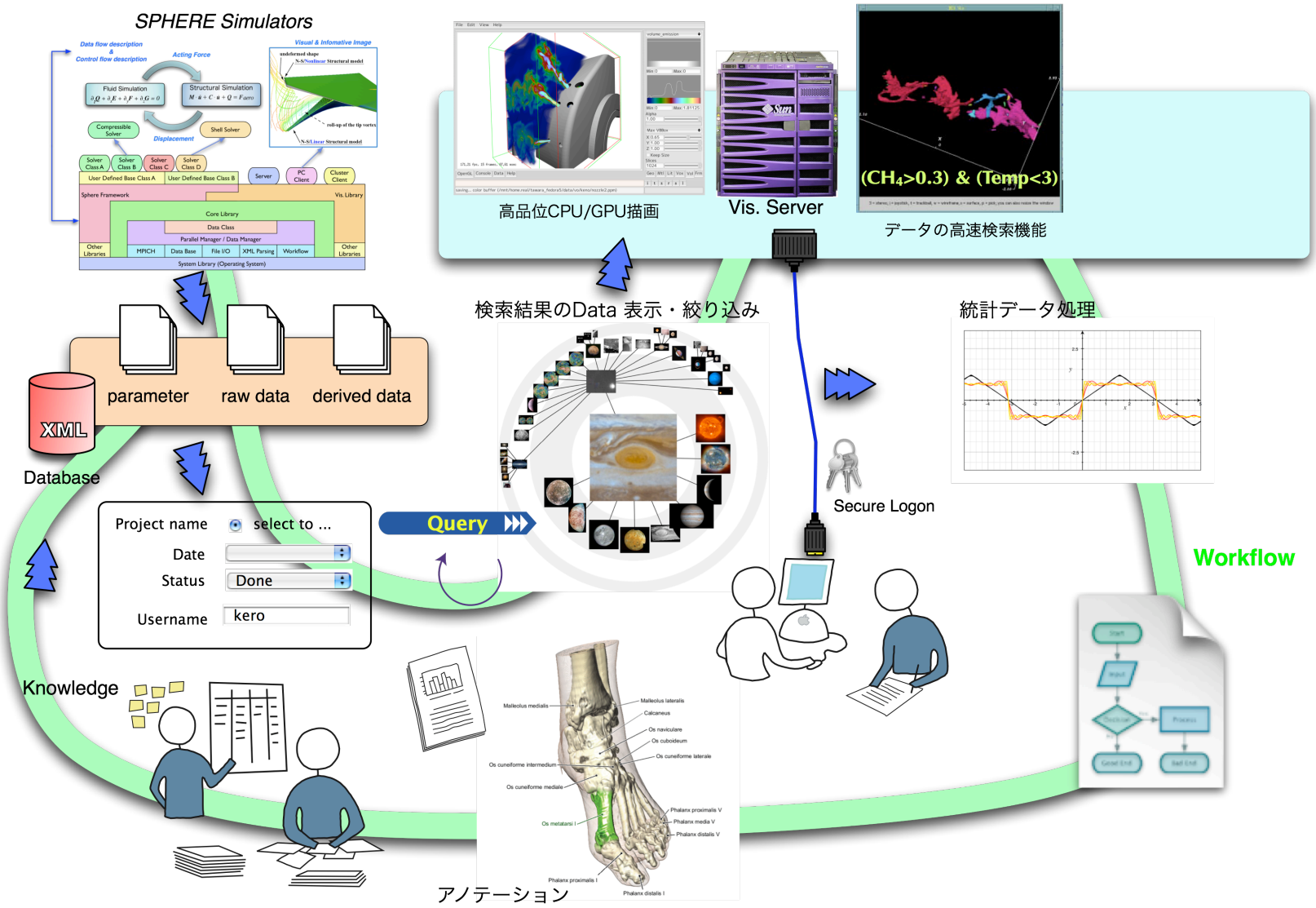
提案するポスト処理

- 後処理の各プロセスを緩やかに有機的に結合したシステム
- スクリプトベースのツール・アプリ群
- 共同・協同作業のサポートの織り込み

- 研究者のリクエストのボトムアップ

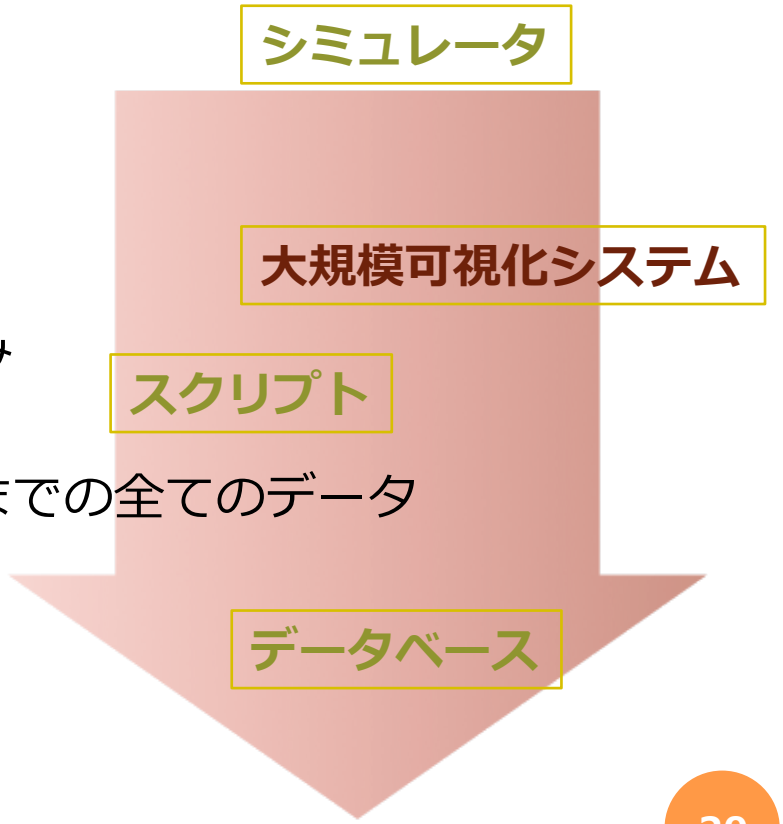
- 研究者の「考えること」を支援
  - システムの大規模化によって生じる複雑性が問題

ポスト処理のイメージ


SPHERE Simulators
高品位CPU/GPU描画
Vis. Server
(CH4>0.3) & (Temp<3)
データの高速検索機能
検索結果のData 表示・絞り込み
統計データ処理
parameter
raw data
derived data
XML
Database
Project name
select to ...
Date
Status
Done
Username
kero
Query
Secure Logon
Workflow
Knowledge
アノテーション

ポスト処理のコンセプト

- シミュレーション結果から，有用な知識・情報を効率的に引き出す
  - 様々な可視化手法による場の把握
  - 生成された2次データの利用
  - イメージ化
  - データ分析
  - 多様・大量なデータの管理のしくみ
- データ管理
  - シミュレーション開始から分析後までの全てのデータ
  - 個々のデータの関連づけ
- 処理の効率化
  - 定型作業の自動処理
- ユーザ作業環境
  - 単独で動作するコマンド，アプリの連携・緩やかな結合



シミュレーションの価値を増幅させる，知識を得るためのツール群

提供機能

- ユーザ作業環境の提供
  - rlogin, ftp
    - ファイル操作，転送
- 自動処理機能
  - プロセス実行シナリオ作成，スクリプト生成
  - スクリプト処理（シェルベース）

既存ソフトウェアの流用を行い，コンセプトの実現に足りない部分を独自開発

- データサービス機能
  - データ検索，表示，関連づけ自動化
- ツール群
  - フィルタ処理，変換，分析，ユーザプログラム
- リソース管理機能
  - 認証，予約，ステータス，ジョブ管理
- 大規模データ可視化

主な開発項目

管理対象のデータ候補

- ユーザ情報
- プロジェクト情報
- ジョブ実行情報
  - 使用計算機
  - 並列実行数
  - ジョブクラス
- ファイル情報
  - パラメータファイル
  - 初期入力データ
  - シミュレーション結果ファイル
  - ポスト処理パラメータファイル
  - 可視化結果ファイル（画像）
  - 解析結果ファイル
  - 研究ノートファイル
- パラメータ情報
  - シミュレーションプログラム名
  - 入力データファイル名
  - シミュレーション設定パラメータ

研究データのライフサイクル全体を管理

大規模データ可視化システム

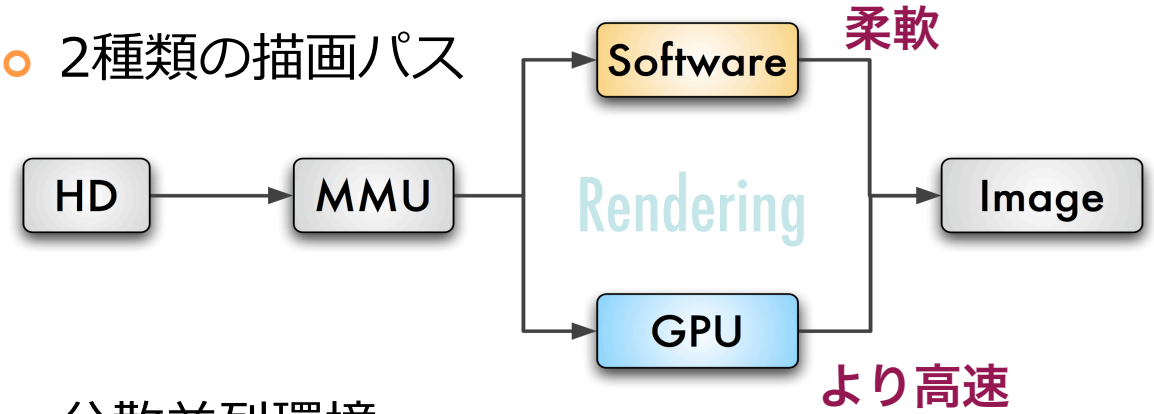
- 可視化シナリオ，キーワード
- 開発中の可視化システムの基本要件
- 機能設計と機能モジュール群
  - 基本可視化
  - データ分析
  - 知的可視化
  - 先進レンダリング
  - インテリジェントデータリダクション
  - Time-Varyingデータハンドリング
- インタラクティブ性の確保の方針
- 開発の方針

可視化

- データを人の目により認識し，理解力を促進する
- 多角度・側面からの検討により，本質の理解を助ける

- インタラクティブ性（スピード）とUI（作業効率）が重要な鍵
- リモート環境でも，5fps程度の性能が必要

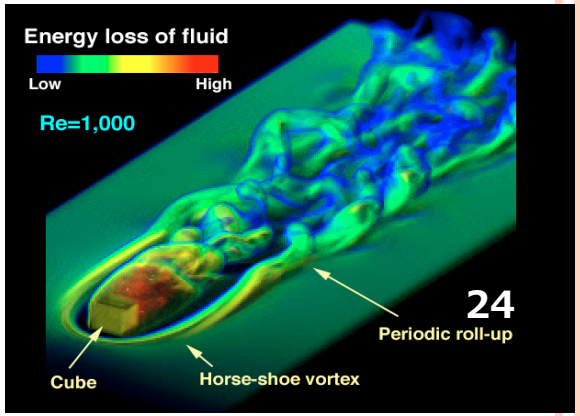
- 2種類の描画パス



- 分散並列環境
  - レンダリングパイプラインの効率的な並列処理
  - イメージ/データ通信のための高速ネットワーク

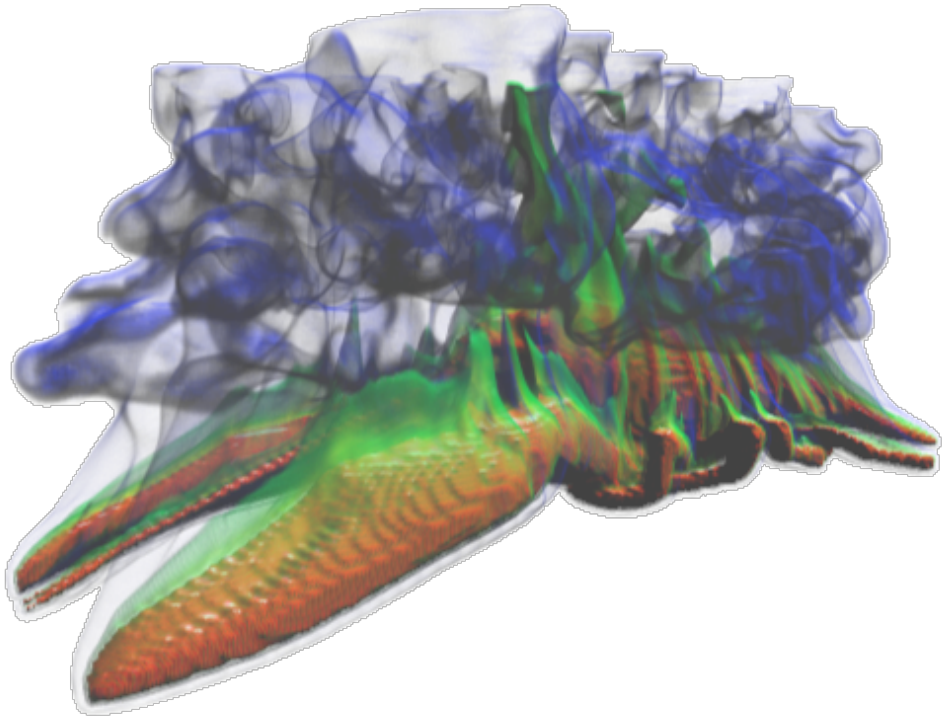
高品質イメージは認識能力を補助

インタラクティブ性は空間認識力を高める



既存の可視化システム

- 商用
  - AVS
  - FieldView
  - EnSight
  - VisLink
  - RVSLIB



- 非商用
  - Visit https://wci.llnl.gov/codes/visit/
  - ParaView http://www.paraview.org/
  - V-Isio http://vcad-hpsv.riken.jp/

可視化シナリオの分類　ー　３つの軸

- サーバープログラムの実行場所

	クライアント	可視化プロセス
ローカル	ユーザPC	ユーザPC
リモート	ユーザPC	サーバー

- ユーザインタラクション
  - インタラクティブ ：　ユーザの操作に応じて可視化
  - バッチ ：　定義されたシナリオに基づき可視化

- 実行タイミング
  - リアルタイム ：　計算途中結果を可視化
  - ポスト ：　終了後に結果を可視化

可視化システム - KEYWORD

- 可視化基本機能
- インテリジェントデータリダクション機能
- データマネジメント機能
- ファイルフォーマット設計
- マルチプラットホーム対応クライアント
- リモート機能，バッチ機能

必須機能：
基本システム機能
＋
ベーシック機能モジュール

- 並列レンダリング
- 知的可視化機能
- データ分析，マイニング機能
- インタラクティブ機能
- パラメータスイープ

大規模データ可視化の特徴的機能

- 先進レンダリング機能
- Time-Varyingデータハンドリング機能
- デバッグ機能

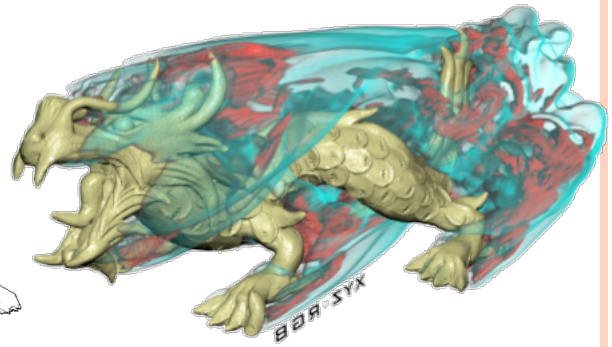
先進的・魅力的な機能

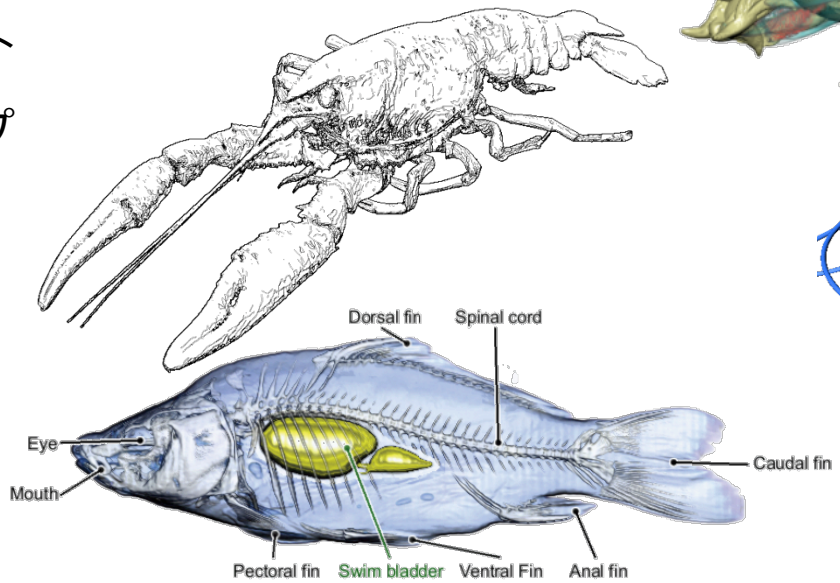
開発中の可視化システムの基本要件

- リモート可視化とローカル可視化
  - 共通クライアントによる統一環境の提供
- リアルタイム可視化とポスト可視化
  - ファイル経由の可視化を原則
- インタラクティブ可視化とバッチ可視化
- SWレンダリングとHWレンダリング
- 大規模データハンドリング
- 並列可視化
- 可視化基本ライブラリと機能モジュール群の構造
  - 将来の機能拡張と実装の容易性を担保
- 移植性
  - Linux, Windows, Max OSX, PC cluster, Supercomputer
- 他の可視化アプリとの連携
  - リモート可視化プロトコルの共通化

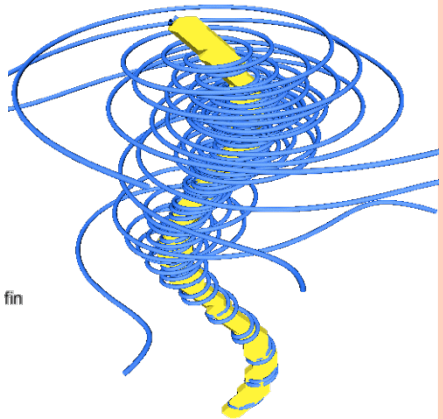
可視化システムの機能モジュール群

- 基本可視化メソッド
- インテリジェント・データリダクション
- データマネジメント
- パラメータスイープ
- 知的可視化機能
- 統計処理
- デバッグ
- バッチ
- アノテーション
- シナリオオーサリング
- 立体視

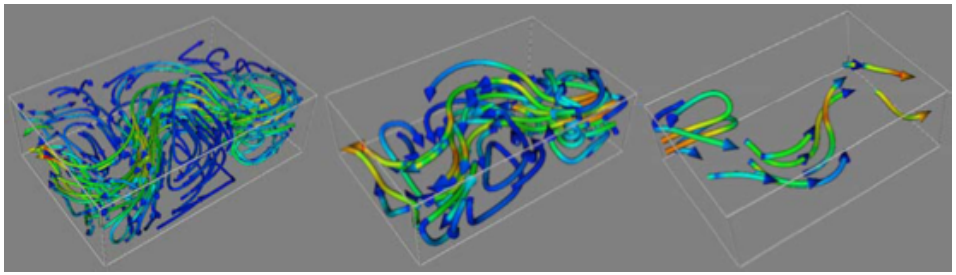




Vis2005 Course note



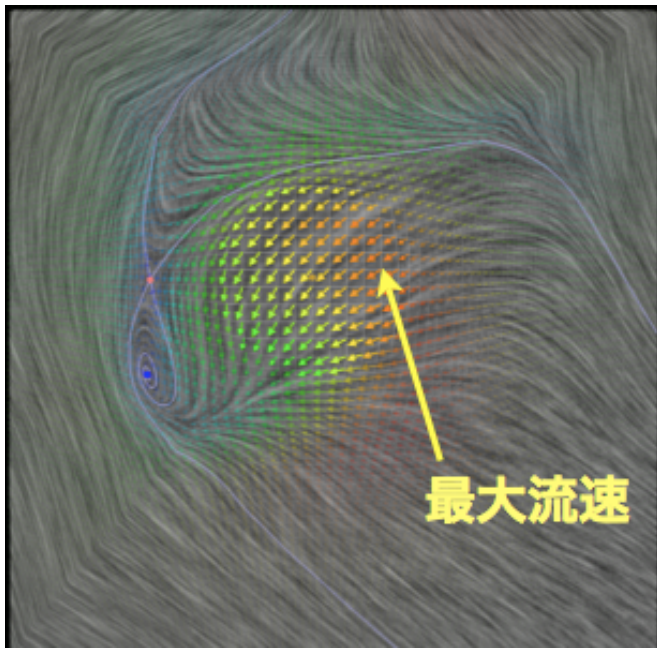
Jiang, Vis2004



Griebel, Vis2004

基本可視化機能

- スカラ
  - コンタ，ペイント，VR，等値面
- ベクトル
  - ベクトル図，流線，タイムライン，流跡線，LIC
- データプローブ，グラフ
  - 値の表示
- オブジェクト生成
  - 等値面ポリゴン，グリフ
- ビュー・ウインドウ操作系
  - UI，ウイザード

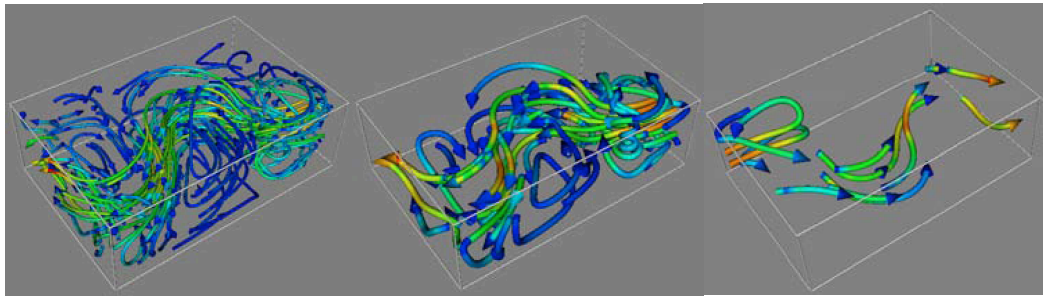


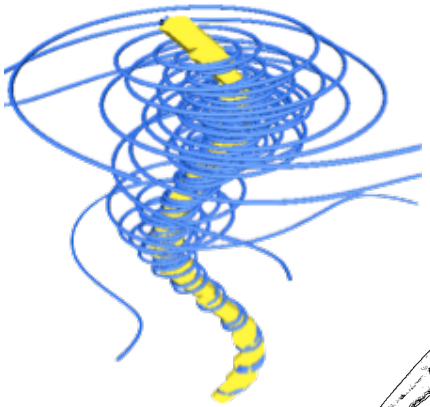
データ分析機能

- 統計分析
- 時間変動解析
  - FFT, Wavelet
- 乱流解析用モジュール
- 固有値解析

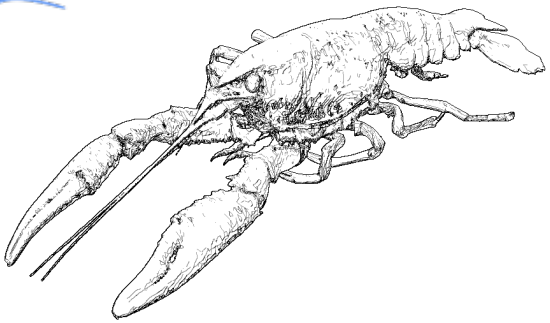
知的可視化機能

- 特徴抽出
  - 渦管，特異点
- 比較
- Illustrative rendering
- グリフ
- メタデータ化
- 手順の自動化
- データに対する演算
- 推奨パラメータ抽出
- より具体的な手法は問題依存





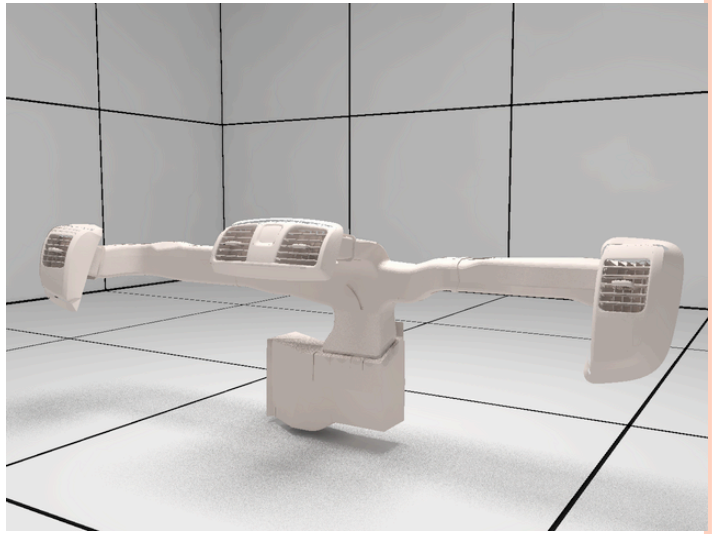
Jiang, Vis2004

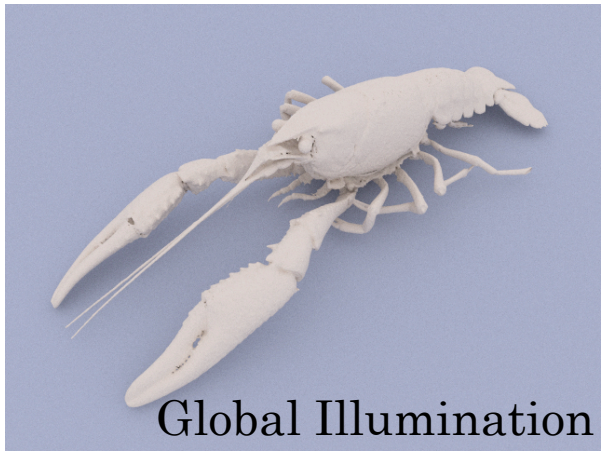


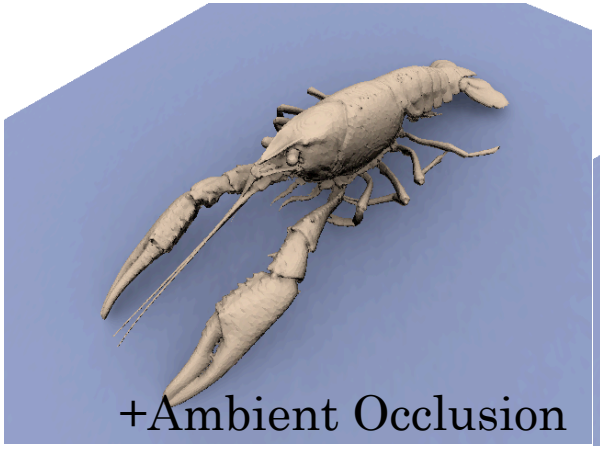
Griebel, Vis2004

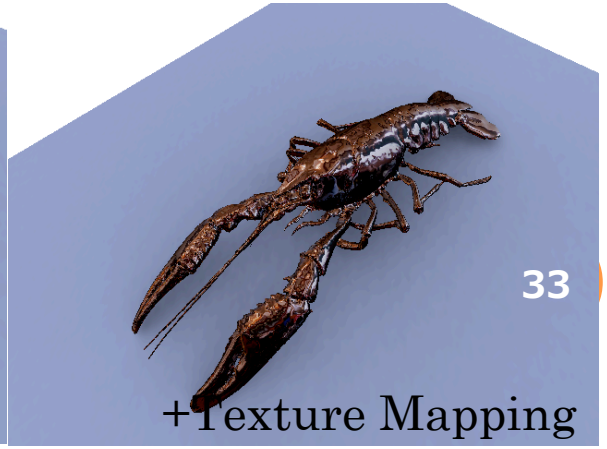
先進レンダリング機能

- 高品位レンダリング
  - Photo realistic rendering
  - Global illumination
- マルチコアレンダリング
  - CPUベースの並列アルゴリズム
- 並列GPU
  - GPUを使った並列処理，イメージ重畳









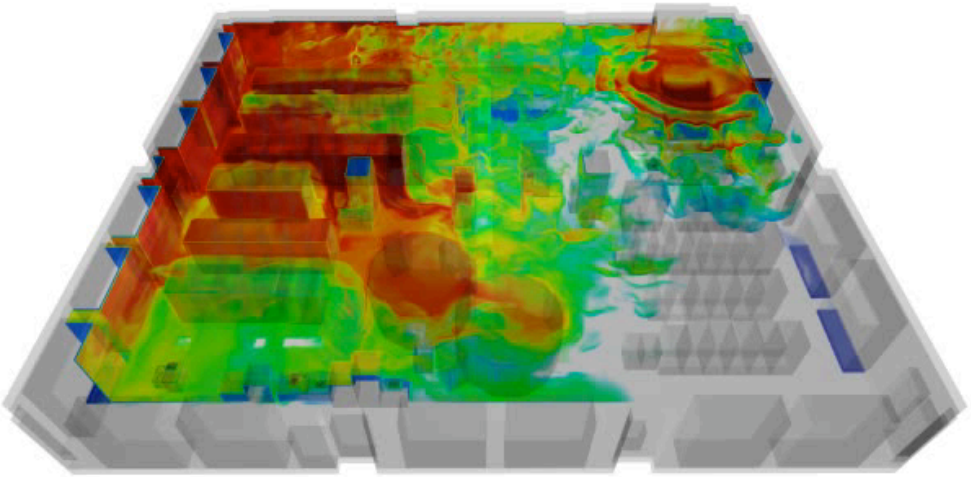
インテリジェントデータリダクション機能

- データ圧縮技術
  - 符号処理による圧縮・伸長
  - 量子化
  - 間引き
  - LOD技術
- 観察対象となるデータの抽出
  - 領域選択（ROI, Region Of Interest），ROIの移動（Roaming）
  - 特徴抽出（Feature Extraction）により，小規模のメタデータに変換
- データベースの利用
  - パラメータ並列の分析作業時のデータ管理
  - 可視化結果の管理
  - 生データとメタデータとの関連づけ

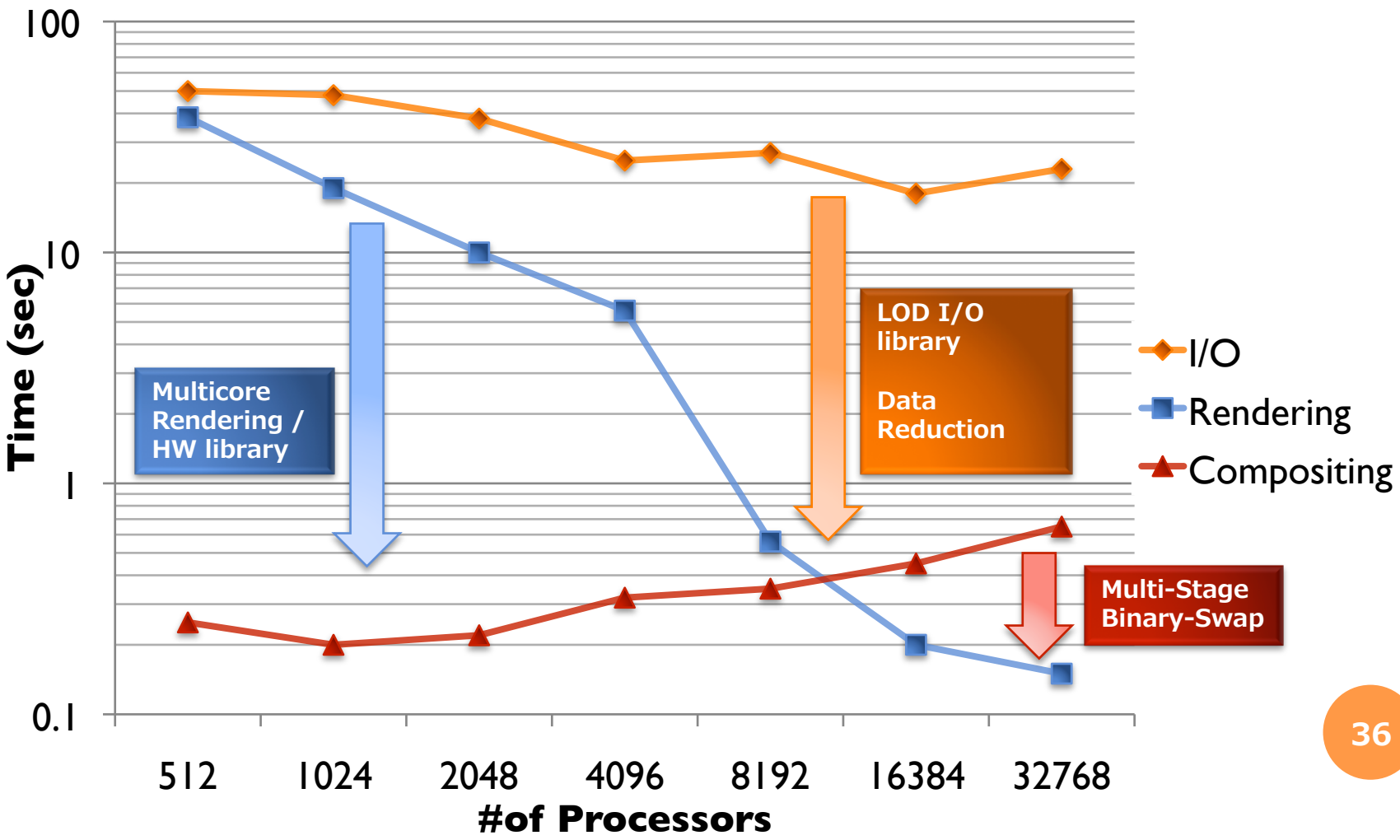
- インタラクティブ可視化の前処理の機能も提供

TIME-VARYINGデータハンドリング機能

- Out of Coreデータの効率的な処理
  - プリフェッチ
  - LODを利用したインデクシング
    - Global Static Indexing for Real-time Exploration of Very Large Regular Grids, SC2001, V. Pascucci(UCD, LLNL)
  - 並列IO
  - データベース
- データ削減技術



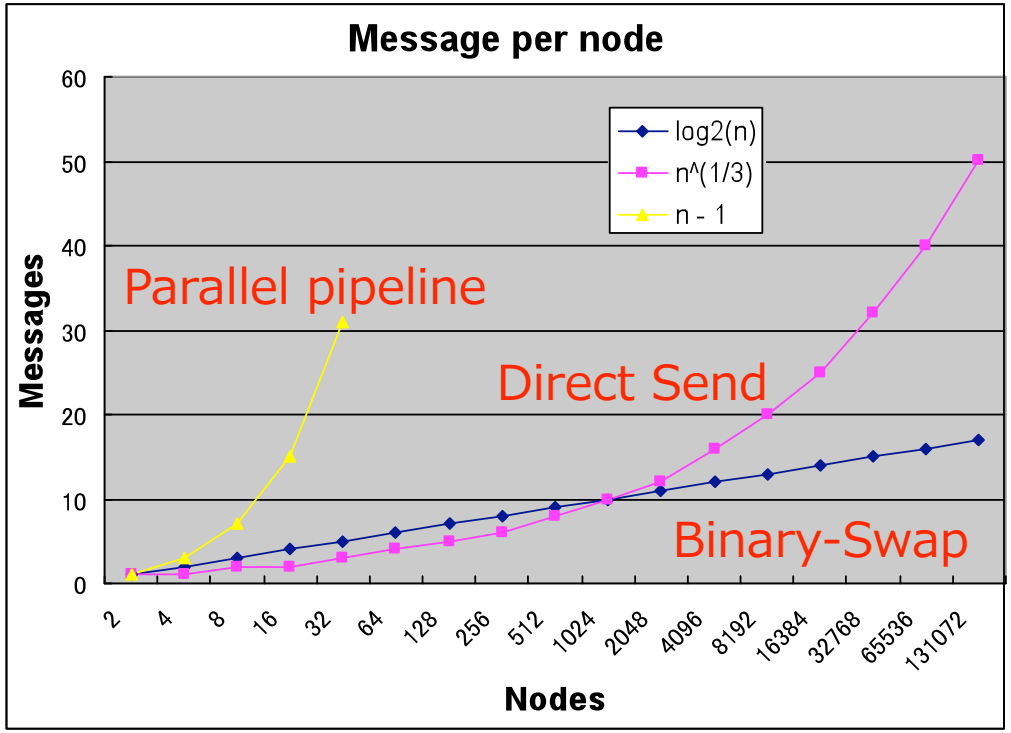
インタラクティブ性確保の戦略



インタラクティブ機能

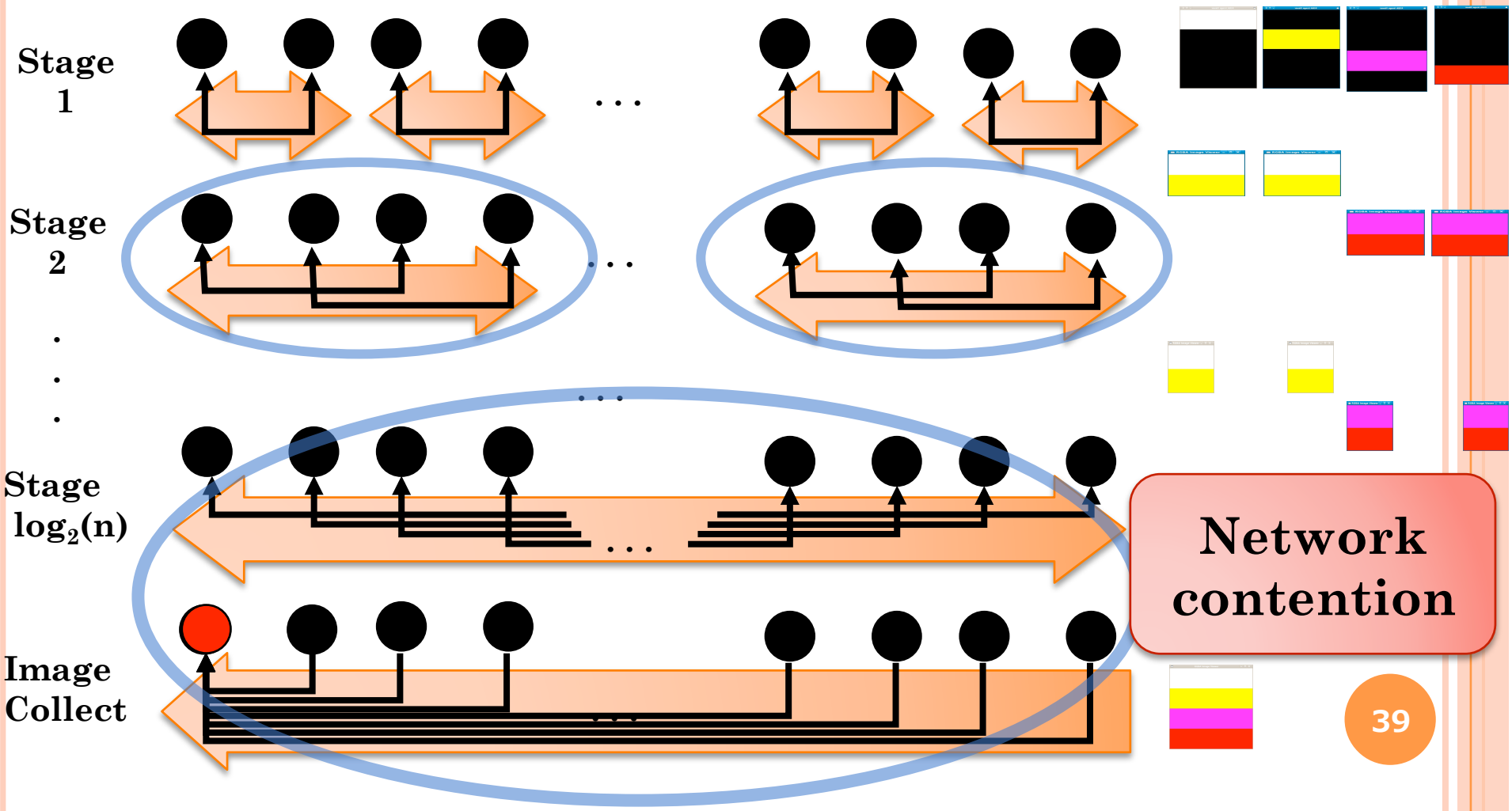
- 高速IO
  - LOD技術
  - DCT圧縮
- 高速画像生成
  - マルチコアレンダリング
  - GPUレンダリング
- イメージ重畳
  - Direct Send, Binary-Swap, SLIC
- データ配信
  - イメージパケット
  - ストリーミング
- 性能予測
  - ネットワーク大域，レイテンシ
  - 各プロセスの所要時間

イメージ重畳方式の性能比較

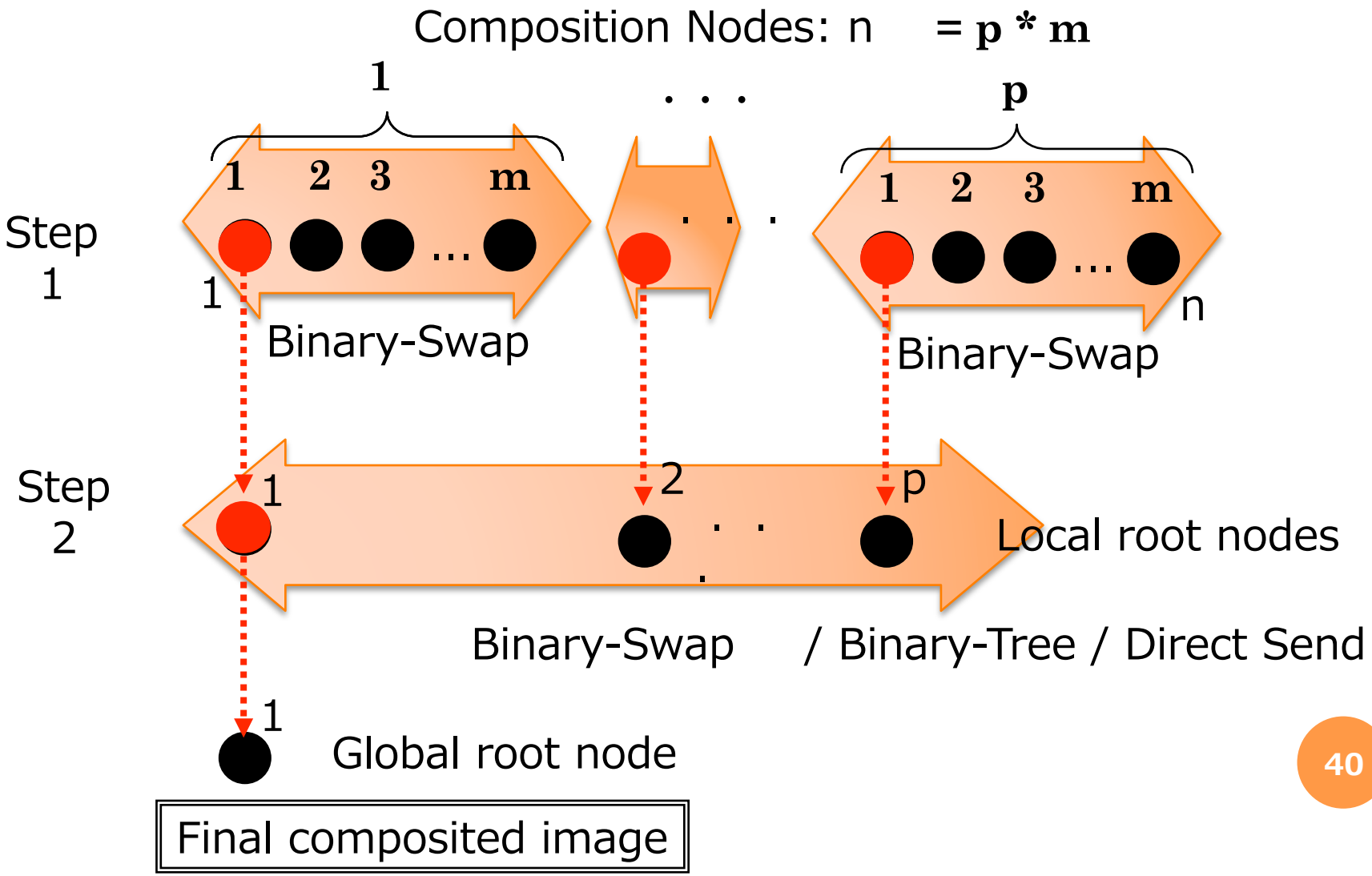


- Binary-Swap
  - High scalability
  - Static scheduling
  - Several optimizations
    - Easy combination
- MPP
  - RIKEN One-Rack BG/L
    - 2048 processors

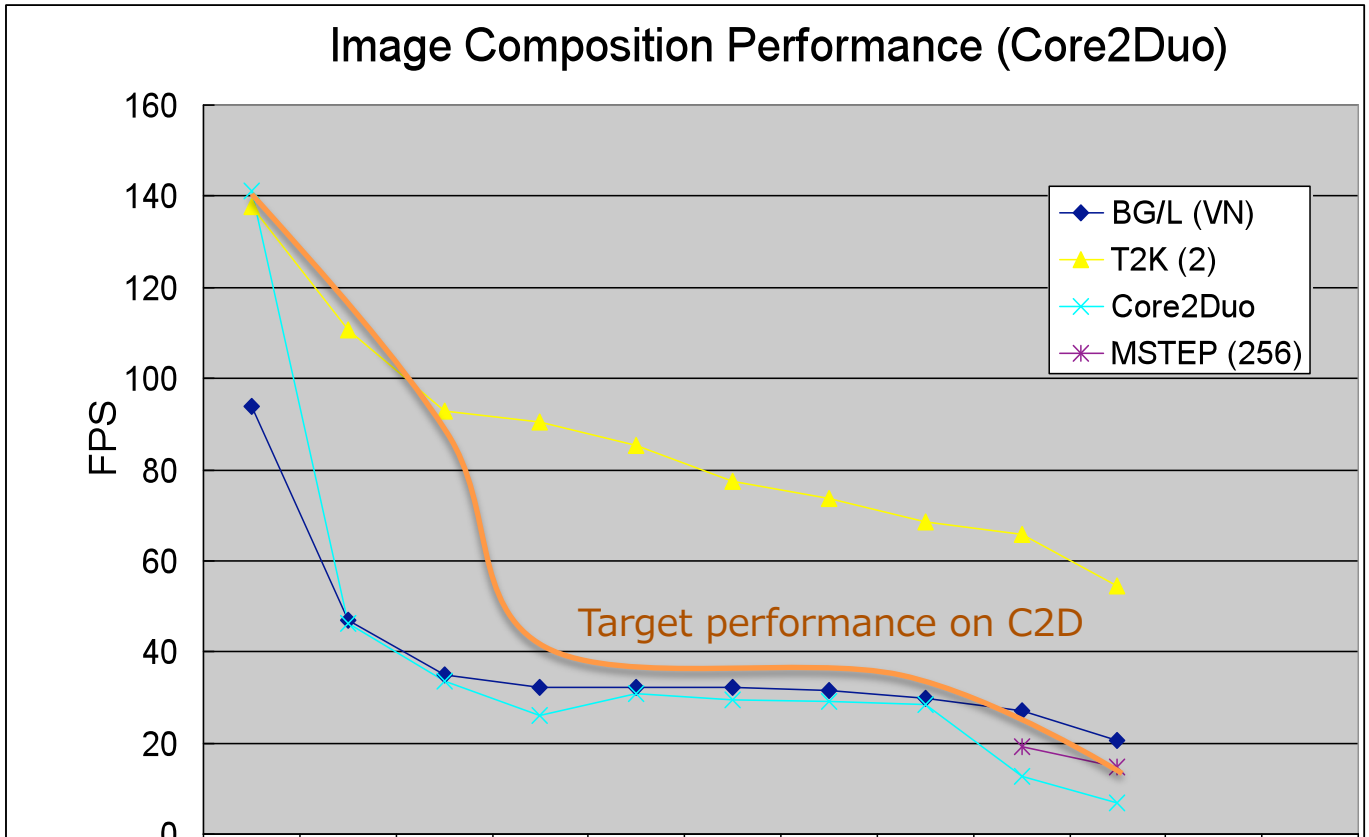
Binary-Swap Image Composition



Multi-Step Image Composition



イメージ重畳の測定結果



	2	4	8	16	32	64	128	256	512	1024		
BG/L (VN)	93.9	46.9	35.1	32.3	32.2	32.1	31.5	29.7	27	20.7		
T2K (2)	137.6	110.7	93	90.5	85.4	77.6	73.7	68.6	65.8	54.5		
Core2Duo	141.3	46.1	33.5	26	30.9	29.3	29.2	28.5	12.7	6.7		
MSTEP (256)									19.3	14.9		

検討している可視化ライブラリ

- 本体計算機上で稼働する可視化ライブラリ
  - ラフな検討とROIの絞り込み，データ量削減
  - 基本的機能を備えたサーバークライアントシステム
  - バッチ処理のみ

- 専用クラスタによる可視化ライブラリ
  - 主記憶容量は小さい（事前にデータ削減の必要）
  - インタラクティブ性の提供
  - コモディティ利用，低コスト，開発リスクの低減
  - 本体の占有時間の短縮
  - 最新のGPUが利用可能

ポスト処理技術開発：２つの軸

- 研究支援
  - ポスト処理プランを国内外の既存ツールベースで構築し，早期のサービス提供を目指す
    - 基本プラン
    - データハンドリング
- 研究開発
  - 研究的要素をもつ高機能・高性能なポスト処理技術については研究をすすめ，その成果を研究支援へ
    - LSVシステム

まとめ

- ポスト処理環境の提案
  - ユーザの可視化シナリオベース
- 大規模データ可視化システム
  - インタラクティブ性能の分析
  - 要素技術の検討
  - システム設計