

LHCアトラス実験のための 国際広帯域ネットワークの活用

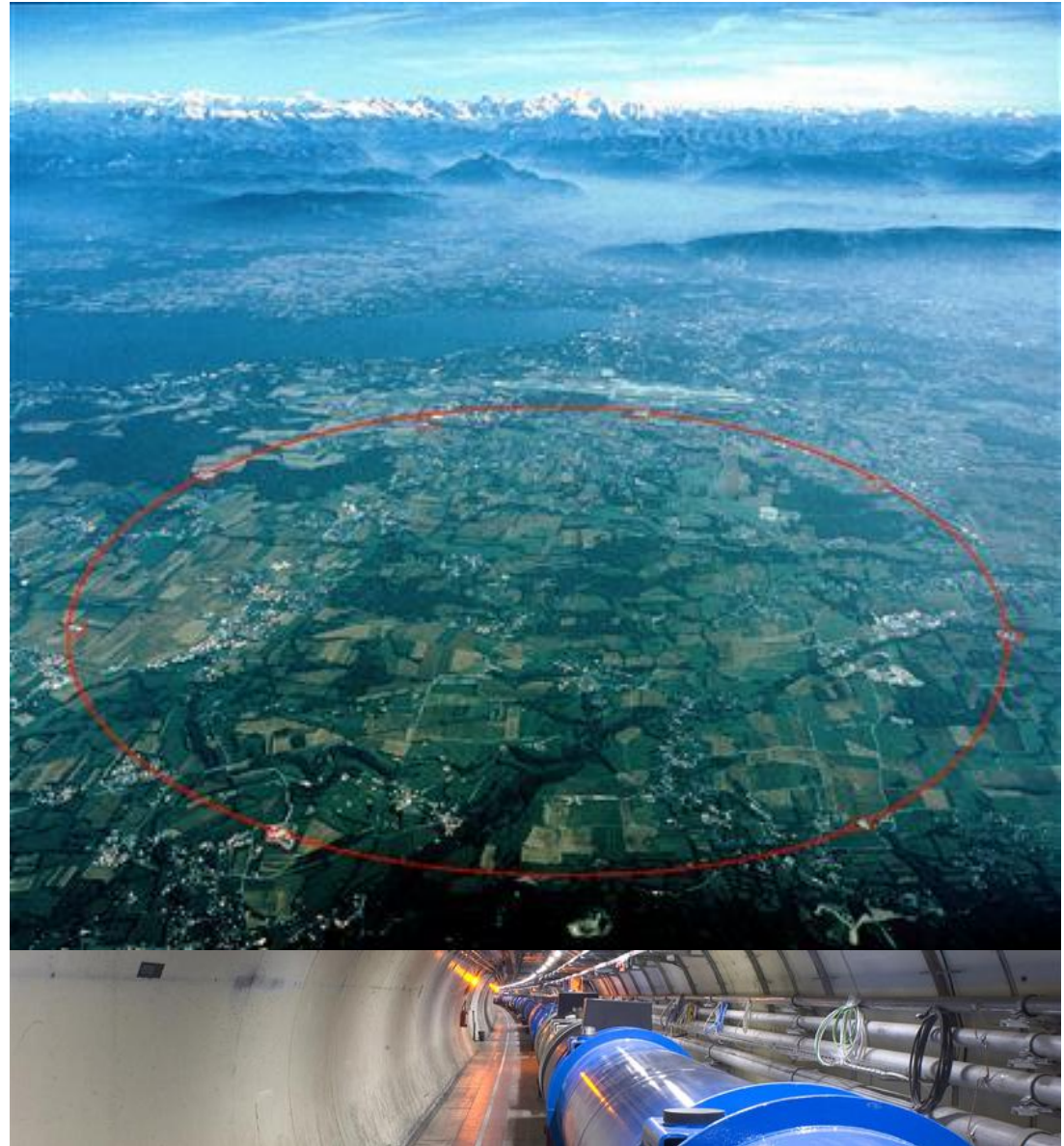
東京大学 素粒子物理国際研究センター(ICEPP)
真下哲郎, 松永浩之, 磯部忠昭, 坂本宏,
上田郁夫, 田中純一, 松井長隆

広帯域ネットワーク利用に関するワークショップ
(ADVNET2009)
2009年6月30日 @ 東京大学 小柴ホール

LHC加速器

Large Hadron Collider

- CERN(欧州原子核研究機構, ジュネーブ郊外)
- 世界最大(周囲27km)
地下約100mのトンネル中
- 陽子・陽子衝突型加速器
- 最先端の素粒子物理実験
(Higgs粒子、超対称性粒子の発見、...、Black Hole?)
- 2008年9月稼動、直後に事故
- 2009年秋に再稼動
- 4つの実験:
ATLAS, CMS, ALICE,
LHCb(いずれも巨大な国際共同実験グループ)

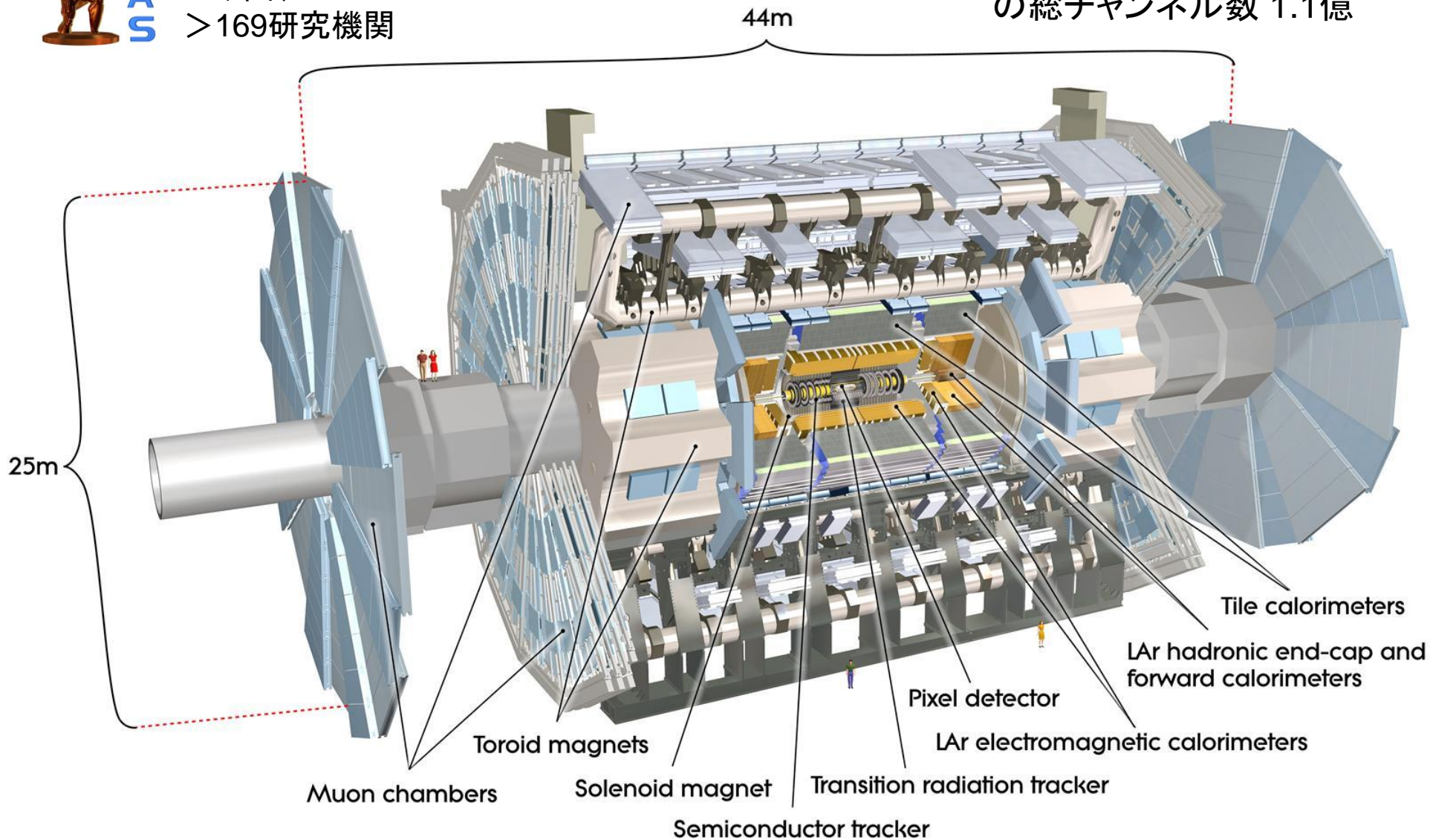




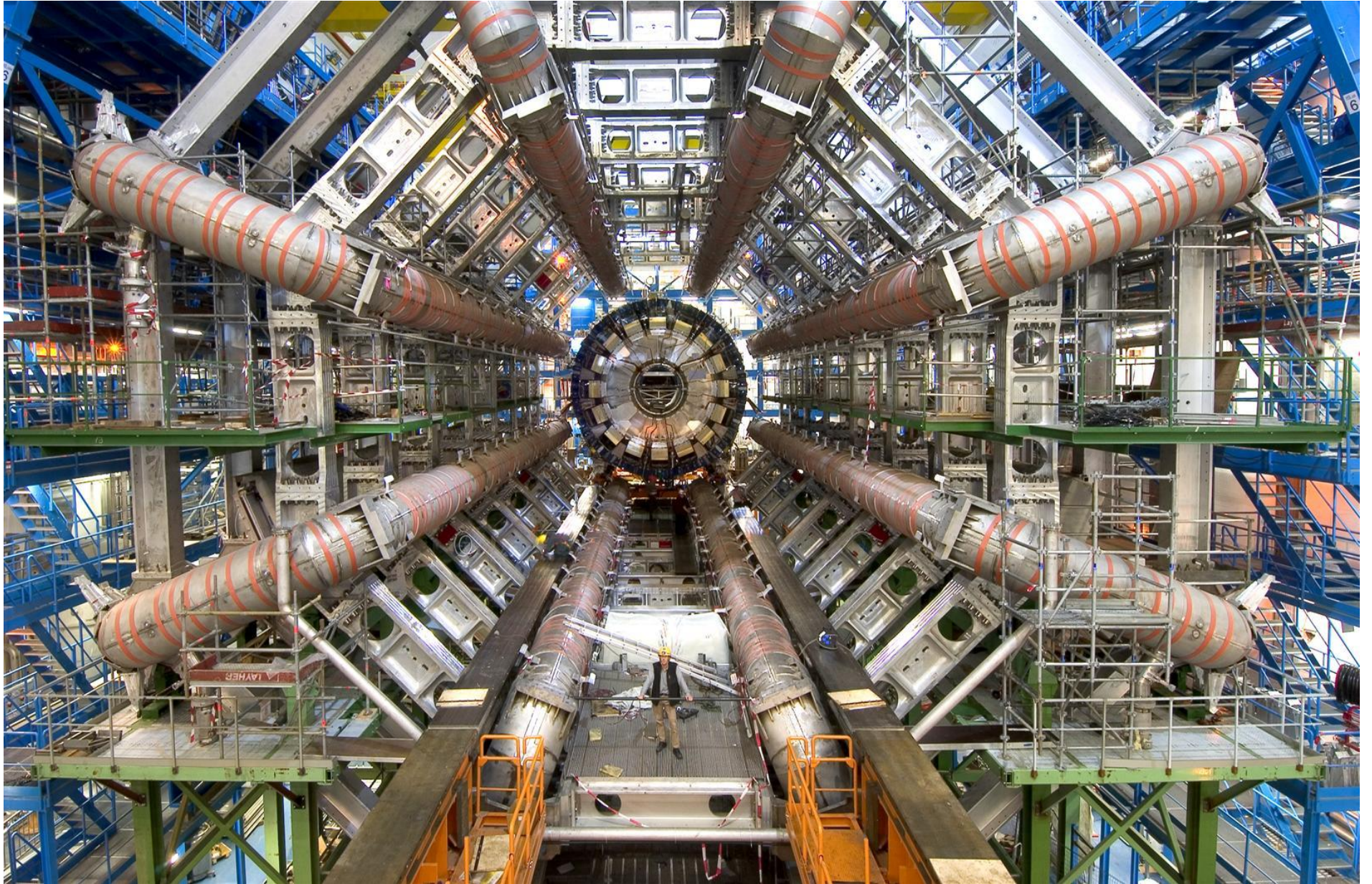
~2500の物理屋
37ヶ国、
>169研究機関

ATLAS測定器

総重量 7000トン、センサー
の総チャンネル数 1.1億



建設中のATLAS測定器(2005年)



ATLAS日本グループ

- 15研究機関：KEK、筑波大、東大、早稲田大、東工大、首都大学東京、信州大、名古屋大、京都大、京都教育大、大阪大、神戸大、岡山大、広島工大、長崎総合科学大
- スタッフ約80名、大学院生約60名

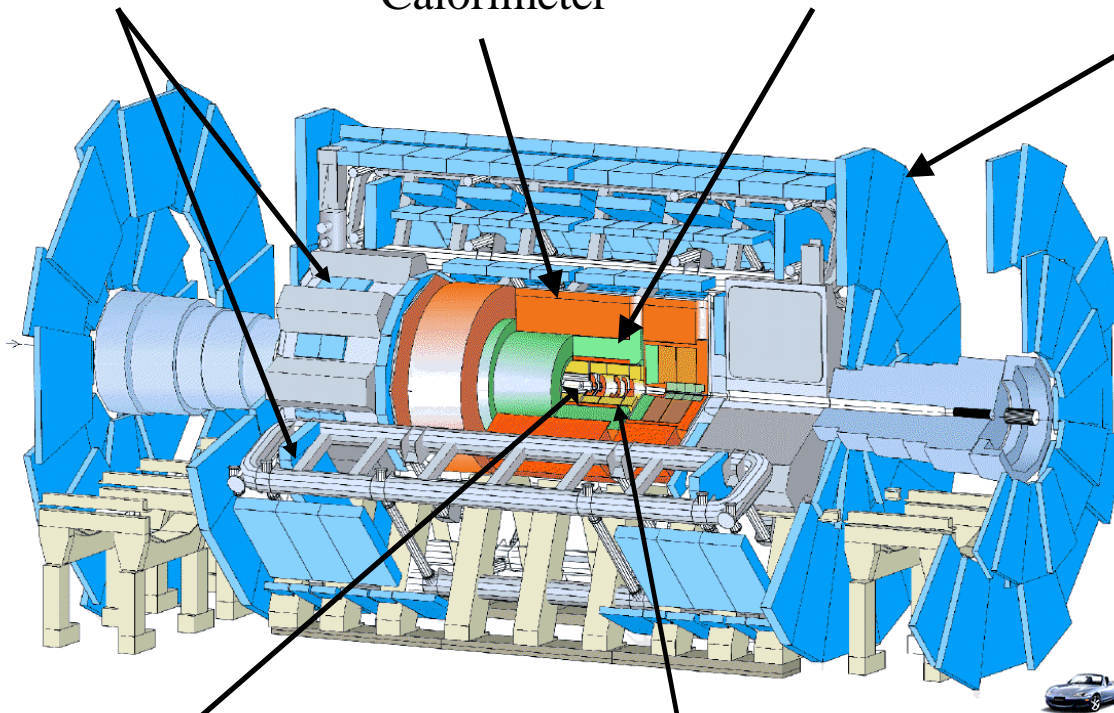
様々な部分に貢献：

Toroid Magnets
(Air-Core)

Hadron
Calorimeter

EM Calorimeter

Muon Spectrometer



Inner Detector

Solenoid Magnet

+ DAQ, Trigger
+ Software
+ Regional Center

WLCG (Worldwide LHC Computing Grid)

- LHC実験から発生するデータは未曾有の量
ATLAS実験の場合：
検出器から発生する「生データ」：年間数PB
物理解析の便のための二次的な集約データ(“ESD”、
“AOD”など)やシミュレーションデータ、それらのreplica
を含めると、さらに数倍
他の3実験についても同程度の量のデータが発生
- 現地CERNの計算機資源だけでは処理は不可能
- 世界的な分散計算機環境を用いて対応：WLCG

WLCG (cont'd)

計算機センターの階層構造

- Tier-0 (CERN)
 - 生データのすべてを保管、データの一次処理 (“ESD” の作成)
- Tier-1 (世界に10ヶ所程度) (役割は実験によって違うが)
 - データの基本的処理と保管を行なうインフラの中核
 - 生データの2ndコピー、一次処理データを分担して保管
 - “AOD” (二次処理データ) の作成・保管、Tier-2への転送
 - データの再処理
 - シミュレーションデータの保管
- Tier-2 (国・地方レベル) (役割は実験によって違うが)
 - 物理解析の拠点
 - シミュレーションデータの生成 (Tier-1へ転送)
- Tier-3 (通常の大学の研究室レベル)
 - 物理解析の最終段階

日本の「地域解析センター」

- ATLAS実験から発生する大量のデータを物理解析するための日本国内の拠点、WLCGのサイトとしては、Tier-2(+Tier-3)の役目
- 東京大学素粒子物理国際研究センターに設置(東大本郷キャンパス内)
- 2001年からR&D開始、2006年末に本番の計算機システムを導入(2009年末更新予定)
- 実験現場であるCERNに滞在する日本人のための「CERN分室」も構築(比較的小規模)



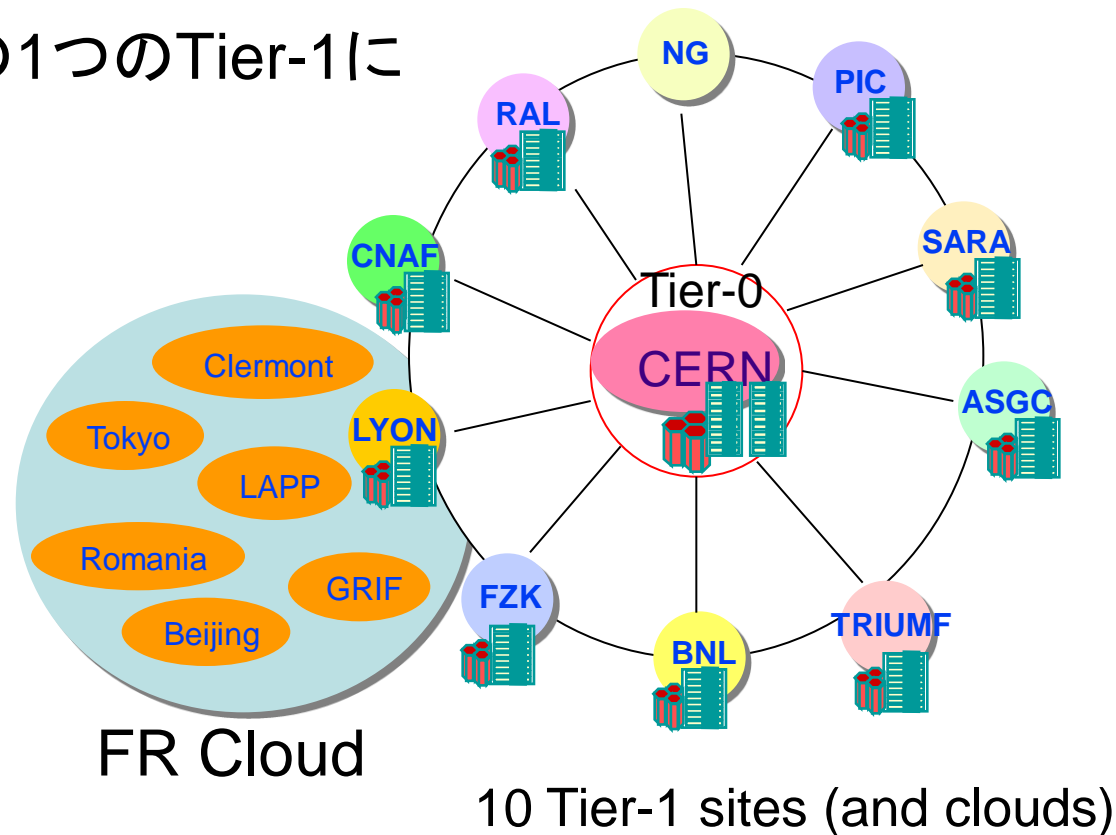
ATLAS distributed computing

- “Cloud”モデル

- 1つのcloudは、1つのTier-1と複数のTier-2から成る
- Tier-2は特定の1つのTier-1に

- ICEPP

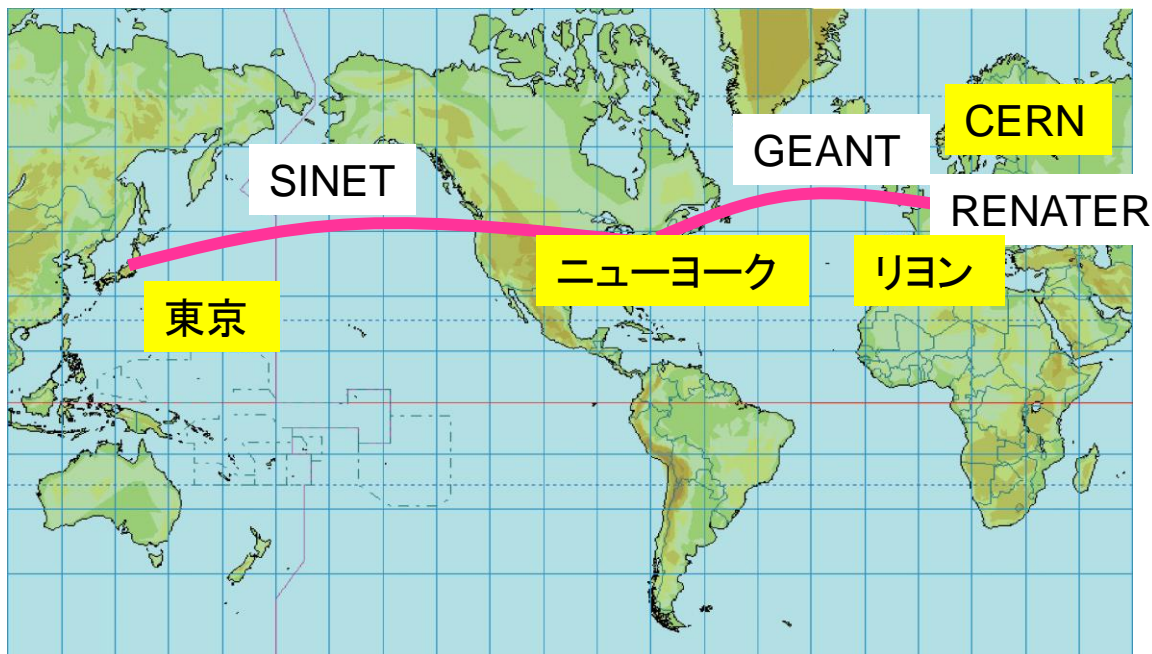
- フランスのcloud
- LyonのIN2P3計算センターがTier-1
- Lyonから最も遠隔地



東京とCERN/CC-IN2P3間の経路

経路すべてが10 Gbps

- 2007年6月、CC-IN2P3(Lyonの計算センター)とRENATERの接続が10 Gbpsに
- 2007年7月(SINET3運用開始後)、当地域解析センターから東大情報基盤センター経由でSINETへ10 Gbpsで接続できるように
- 2008年2月末、ニューヨークMANLANのSINET-GEANTの接続が10 Gbpsに



RTT ~ 290 msec

- 1 Lyon-OPN (193.48.99.100) 0.237 ms 0.630 ms 1.042 ms
- 2 Lyon-INTER (134.158.224.4) 0.242 ms 0.187 ms 0.193 ms
- 3 vl3114-paris1-rtr-021.noc.renater.fr (193.51.186.178) 5.433 ms 5.428 ms 5.433 ms
- 4 te0-1-0-3-paris1-rtr-001.noc.renater.fr (193.51.189.41) 38.231 ms 41.017 ms 22.606 ms
- 5 renater.rt1.par.fr.geant2.net (62.40.124.69) 7.626 ms 5.529 ms 5.534 ms
- 6 so-3-0-0.rt1.lon.uk.geant2.net (62.40.112.106) 12.921 ms 12.865 ms 12.871 ms
- 7 so-2-0-0.rt1.ams.nl.geant2.net (62.40.112.137) 21.007 ms 21.003 ms 21.007 ms
- 8 nyc-gate1-RM-GE-7-2-0-207.sinet.ad.jp (150.99.188.201) 104.622 ms 104.565 ms 104.573 ms
- 9 tokyo1-dc-RM-P-2-3-0-11.sinet.ad.jp (150.99.203.57) 314.833 ms 296.275 ms 296.274 ms
- 10 UTnet-1.gw.sinet.ad.jp (150.99.190.102) 296.808 ms 296.834 ms 297.311 ms
- 11 bwtest1.icepp.jp (157.82.112.61) 297.760 ms 296.757 ms 296.613 ms

10 Gbps の帯域を有効に利用できるか自明ではない

国際ネットワークの利用

- ほとんどがファイル転送（大部分GridFTP）
- 東京の地域解析センターは、WLCG用に最低 2 Gbpsの帯域を約束
- 他の使用も（特にCERN分室と東京）
- 合わせて平均4 Gbps程度の国際線の帯域の利用を想定しネットワークの整備をお願いしてきた
- LHCの実験：15年くらいは続く（データ量は実験開始以降次第に増加、利用するネットワーク帯域もおそらく増加）

プロトコルなど

- 当面 IPv4 のみ
- Unicast のみ
- TCP のみ
- 大部分GridFTP
 - LCG middlewareのwrapper utilityやFTSとよばれる転送の自動化ソフトウェアを通して利用
 - 標準のソフトウェア/ミドルウェアの利用：転送性能を上げるための特殊なテクニックは使えない
 - 両サイドのマシンは、特定のサイトとの転送にだけ最適化されているわけではない(TCP window sizeなど)

主要なファイル転送

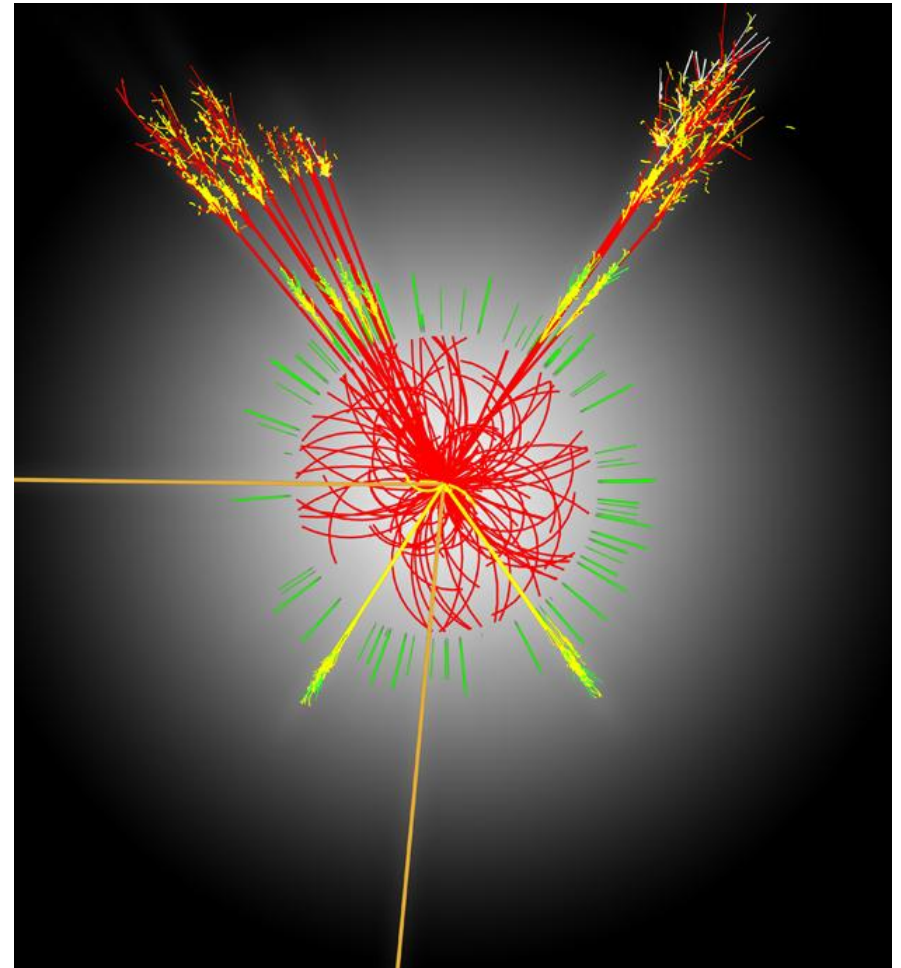
- (Tier-0(CERN) ⇒) Tier-1(Lyon) ⇒
Tier-2(東京の地域解析センター):
実験装置からのデータを集約したデータのファイル
- Tier-2(東京) ⇒ Tier-1(Lyon):
東京で生成したシミュレーションデータ
- 東京 ⇔ 「CERN分室」:
実験装置からのデータを集約した独自の集約データや
シミュレーションデータなど

シミュレーションデータの役目

- 現在まで、LHCのビーム同士が衝突して発生した“事象” (“event”) はまだ1つも無い (宇宙線によるeventは取得されているが)
- しかし「“実データ” (“real data”) が無いのでしかたなくモンテカルロ・シミュレーションを行なって時間をつぶしている」のでは**ない**
- シミュレーションデータは物理解析に必須：
 - 実データが出てくる前に、解析の準備をするため
 - 実データが出てきた後に、物理の結果を出すため

シミュレーションデータの役目（続き）

- 探している新粒子が発生しているかどうかは見ただけではわからない
- 既知の粒子・現象も含めたシミュレーションデータと実データを比較し、統計的に処理して初めて物理の結果が出せる
- 検出器の設計段階から、物理解析手法の最適化、物理結果の導出にシミュレーションは必須



ヒッグス粒子発生事象のシミュレーション

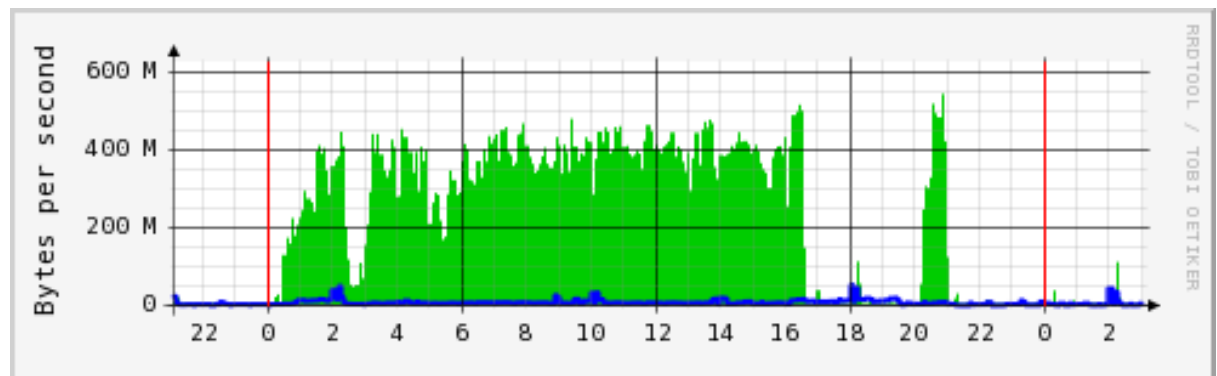
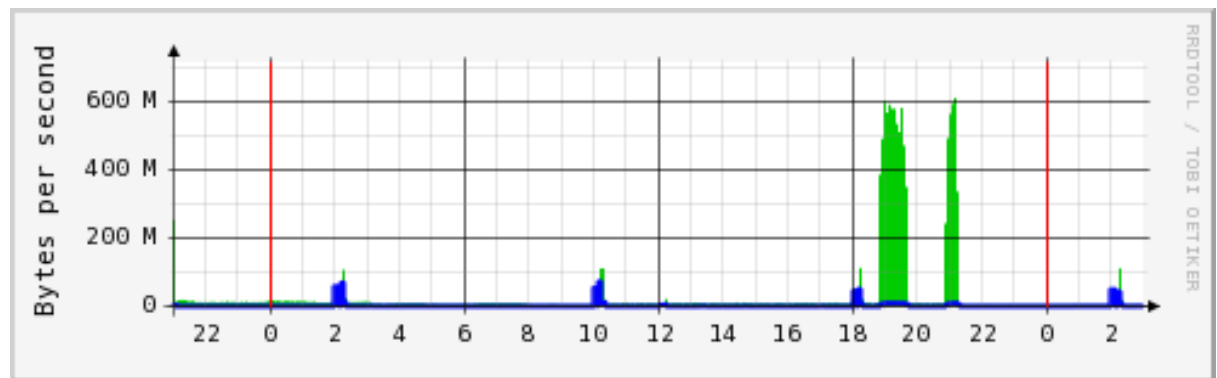
ファイル転送性能の向上

- 2004,5年頃から、東大とCERN間、東大とLyon間のデータ転送のテスト
 - 実質1 GbpsのWAN接続
 - 両サイドに複数のテストマシン(GbE NIC)
- 一方で、WLCGのproduction用マシンを用いての「Service Challenge」と称するデータ転送テストが2004年末から
- 国際線接続の安定性も次第に向上
- Production用マシンの性能向上
 - 標準でBIC TCP
 - Grid middlewareの向上
 - 64ビット化
- 同時転送ファイル数(20)と各ストリーム数(10)に増加(東京-Lyon間)

Lyon→Tokyoのデータ転送

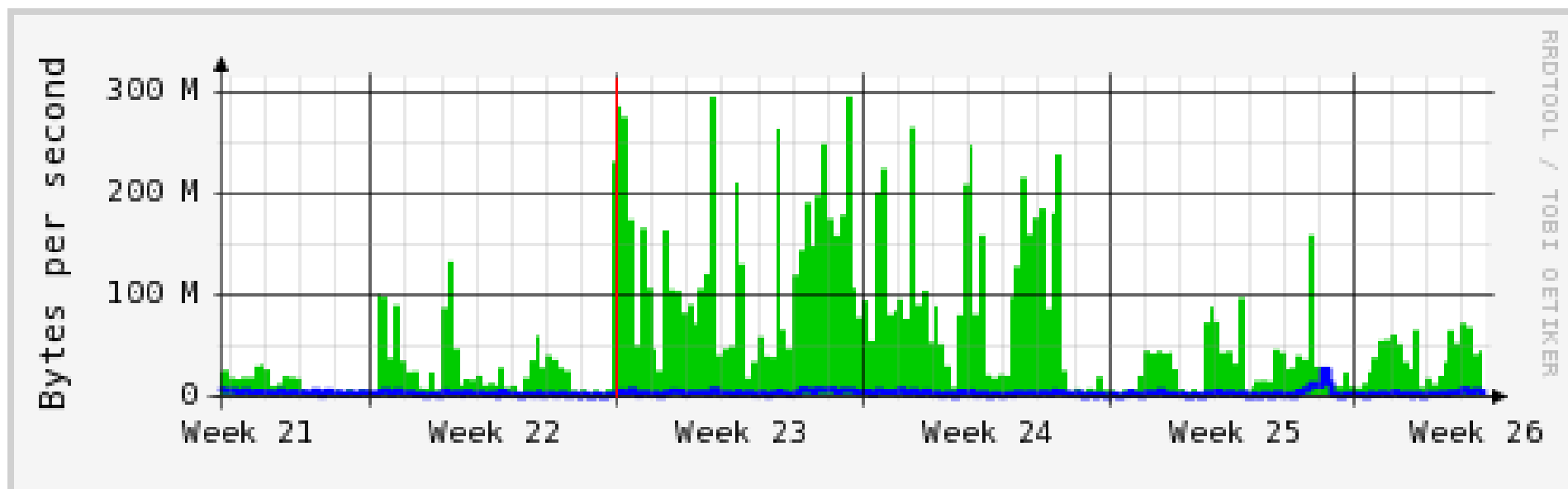
disk-to-disk の throughput

- 2008年5月
 - “CCRC08”という特別演習で観測
 - >500 MB/s で数十分
 - 東京のファイルサーバ数：6
-
- 2009年5月
 - 普段の状態
 - ~400 MB/s で数時間
 - 東京のファイルサーバ数：13



STEP'09

- “Scale Testing for the Experimental Programme”
- LHC4実験の合同演習
- 実データが出た時をrealisticにシミュレーション
- 2009年6月前半(2週間弱)



今後はこれが“常態”に

まとめと今後

- ネットワークの帯域の活用は、これまでのところほぼ満足すべきレベルに達している
- 今年の秋からLHCの運転再開、本格稼働へ
 - 2010年にかけては通常の冬季のshutdownはせず、2010年末まで連続運転の予定(5 TeV vs. 5 TeV? 設計値: 7 TeV)
- アトラス日本グループも、激しい競争の中 LHCの最初の物理の成果をめざす
 - 特に最初の1,2年が非常に重要
- 引き続きネットワークの安定的な運用とサポートをお願いしたい