

CMS Computing 2001

Paolo Capiluppi, Bologna

Paolo Capiluppi
September 2000



Strategia per il Calcolo



- ◆ **Contribuire in modo sostanziale (e proporzionato all'impegno INFN in CMS, ~20%)**
 - alla definizione degli “algoritmi” di Trigger di Alto livello (2/3),
 - alla definizione del TDR del DAQ (2001),
 - alla definizione del TDR del Computing (2002)
 - e alla definizione del TDR della Fisica (2003).
- ◆ **Contribuire in modo sostanziale allo studio (R&D), disegno ed implementazione del “Computing” di CMS**
 - Partecipazione ed uso dei tools di GRID
 - Implementazione e misura delle risorse distribuite, con coordinamento italiano ed internazionale (Data Challenges inclusi)
 - Preparazione delle risorse e strumenti che permetteranno l'analisi dei dati e la “fisica” ad LHC
- ◆ **CMS ha un chiaro “commitment” sulle Tecnologie di GRID, il modo migliore di risolvere la complessita' del Calcolo ad LHC.**



Come perseguire la strategia per il Calcolo?



Preparare CMS Italia alle milestone precedenti, ovvero

- ◆ **Produrre ed analizzare gli eventi per gli studi di HLT e gli studi di Fisica**
- ◆ **Prototipizzare la distribuzione delle risorse ed il relativo coordinamento (tecnico e di persone), ovvero gerarchia alla Monarc (Tier-n) e GRID tools (distributed data access and distributed computing)**
- ◆ **Utilizzare il personale (Tecnici e Ricercatori) di ogni sede di CMS nelle attività che porteranno all'analisi dei dati (coinvolgimento personale di interessi e di responsabilità)**



Stato e prospettive



◆ **Nella Primavera 2000 in Italia sono stati generati piu' di 300k eventi (segnale e fondo)**

- **Bari, Bologna, Padova, Pisa, Roma1**
- **Gia' dall'Autunno 2000 nelle stessi Sedi verranno "Orchizzati" gli eventi (nuovi) generati, con lo schema completo, inclusa l'analisi (user DataBase)**

◆ **Le Sedi coinvolte cercano di utilizzare le "competenze" che si stanno sviluppano/consolidando nel personale coinvolto (DataBase management, Farm management, Production tools, Physics channels, ...)**

Riferimenti nazionali per le attività tecniche



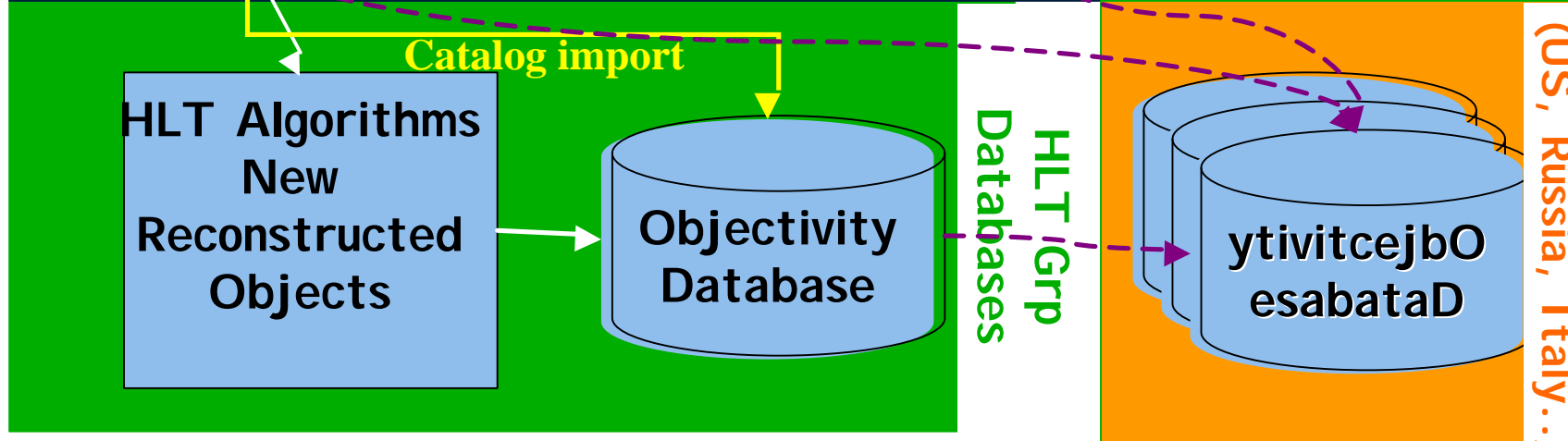
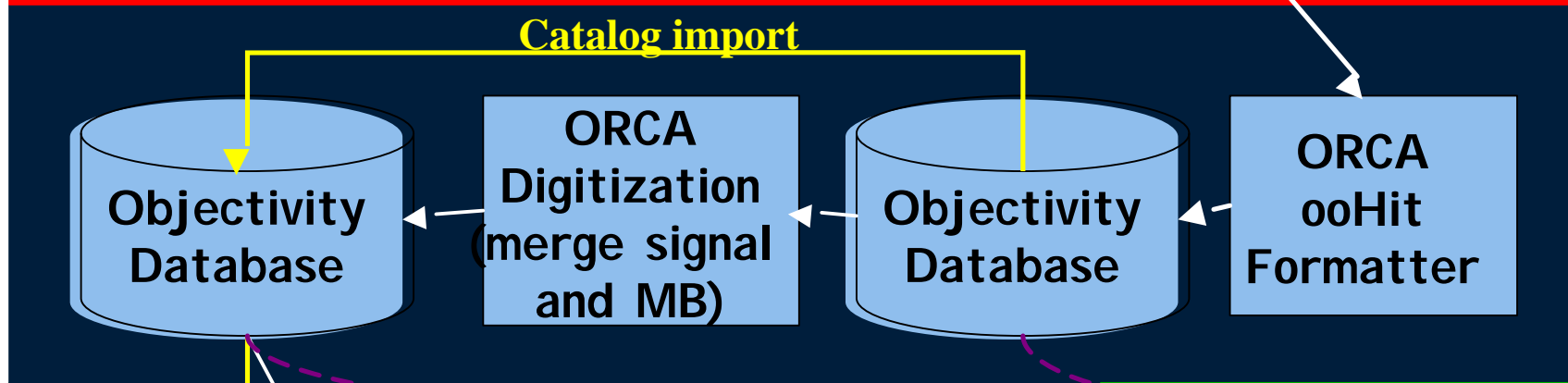
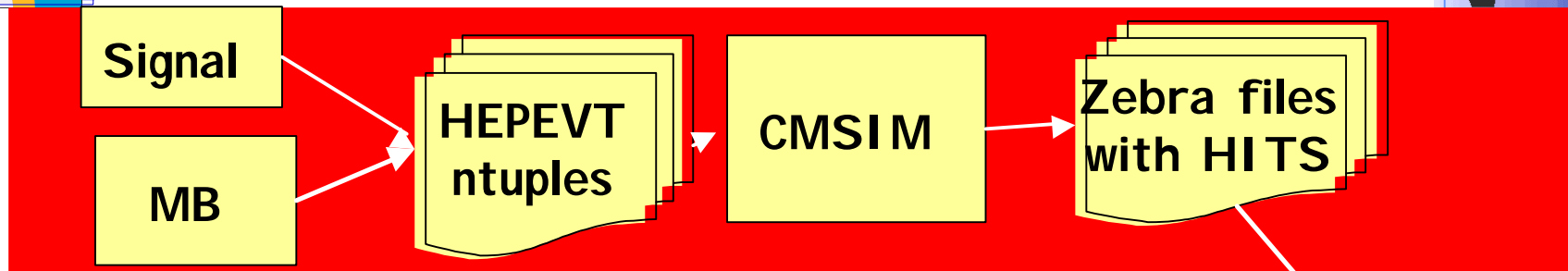
- **SCRAM e ORCA:** Claudio Grandi (Claudio.Grandi@bo.infn.it) tel +39 0512095240 (CERN 71623) - Ivano Lippi (Ivano.Lippi@pd.infn.it) tel +39 0498277098 (CERN 78527) - ([Installation tips](#))
- **Pythia e CMSIM:** Stefano Lacaprara (Stefano.Lacaprara@pd.infn.it) tel +39 0498277100 (CERN 71556) - Alessandra Fanfani (Alessandra.Fanfani@bo.infn.it) tel +39 051209532 (CERN 71623) - (Ugo Gasparini Ugo.Gasparini@pd.infn.it tel +39 0498277051 (CERN 71556,72078))
- **Objectivity (installazione e gestione):** Massimo Sgaravatto (Massimo.Sgaravatto@pd.infn.it) tel +39 0498277029 (CERN 72078) - Lucia Silvestris (Lucia.Silvestris@cern.ch) tel +39 0805442433 (CERN 71501) - (Annalina Vitelli (Annalina.Vitelli@cern.ch) tel ? (CERN 71555,72059)) - ([Installation tips](#))
- **Altri items LHC++ /CERNlib:** Massimo Sgaravatto (Massimo.Sgaravatto@pd.infn.it) tel +39 0498277029 (CERN 72078) - Riccardo Veraldi (Riccardo.Veraldi@bo.infn.it) tel +39 0512095232 (CERN 71623) - ([Installation tips](#))
- **Ambiente su piattaforme Linux** Giovanni Organtini (Giovanni.Organtini@roma1.infn.it) tel +39 0649914328 (CERN 71643)
- **Ambiente su piattaforme Sun** Riccardo Veraldi (Riccardo.Veraldi@bo.infn.it) tel +39 0512095232 (CERN 71623)

Responsabili di riferimento nelle sedi

- **Bari:** Lucia Silvestris (Lucia.Silvestris@cern.ch) tel +39 0805442433 (CERN 71501)
- **Bologna:** Claudio Grandi (Claudio.Grandi@bo.infn.it) tel +39 0512095240 (CERN 71623)
- **Legnaro:** Gaetano Maron (Gaetano.Maron@lnl.infn.it) (*da confermare*)
- **Padova:** Ivano Lippi (Ivano.Lippi@pd.infn.it) tel +39 0498277098 (CERN 78527)
- **Pisa:** Andrea Sciaba' (Andrea.Sciaba@pi.infn.it) tel +39 050880296
- **Roma1:** Luciano Barone (Luciano.Barone@cern.ch) tel +39 0649914332 (CERN 72087)



Production 2000





PRS work: reminder

- **Current work of the four groups: High-Level Trigger**
 - ◆ Lvl-1 trigger max. accept rate $R_{\max}=100$ kHz
 - Current Lvl-1 is 30 kHz; the extra factor 3 is "safety"
 - ◆ HLT accept rate 100 Hz; need 1:1000 selection
- **Develop reconstruction algorithms & software implementing them, as well as selection in steps:**
 - ◆ "Level-2" trigger: anything one can do without using the tracker
 - ◆ "Level-3" trigger: include the tracker
 - ◆ "Level-4" trigger: full event analysis/cuts on physics
- **Planning:**
 - ◆ Lvl-2 step: measure rejection factor (factor $\sim 10?$) using only muon/calor data ($\sim 1/4$ of total?)
 - ◆ Lvl-3 step: ditto (factor 10?) including tracker
 - Overall: complete Lvl-3 step by end 2000 (term of PRS groups)

Muons

Goals

Deliver results on full L1 + L2 chain, in the full η range of the muon detector



- develop/validate L1 simulation in ORCA

- prove **rate reduction** of **factor 10** between L1 and L2:

- L2.1 : muon reco., p_T redefinition, L1 seeded

- L2.2 : apply vertex constraint => improve resolution

- L2.3 : non-prompt muon rejection (χ^2, \dots)

- L2.4 : use calo info, isolation cut

=> apply to **single muon** & **di-muon** topologies

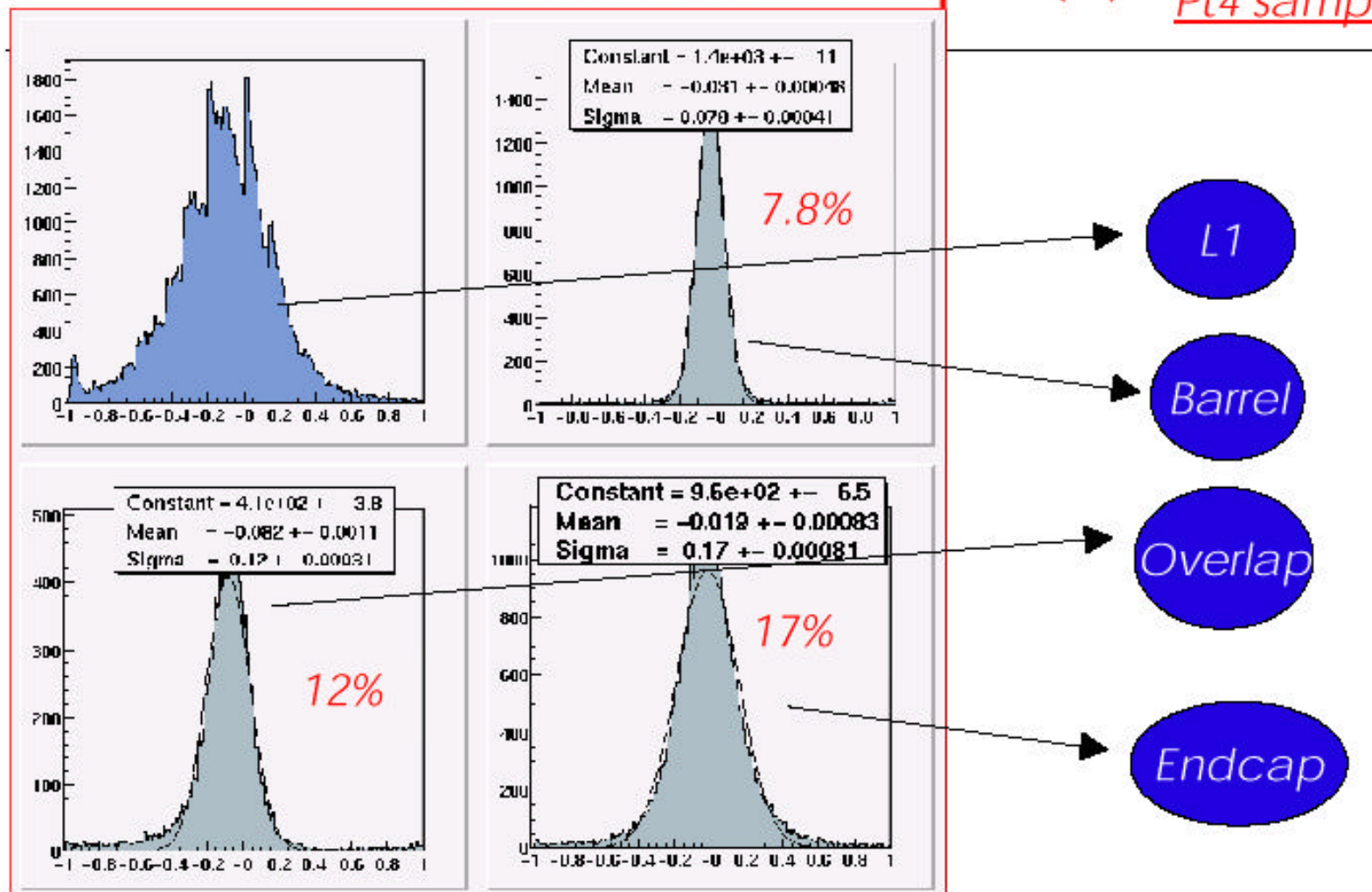
- prove good **efficiency** in (some) “signal” topology (e.g. $Z \rightarrow \mu\mu$, $H \rightarrow WW$)

- provide **L2 filter code** at ORCA level for **single muons** and **di-muons**,
(\rightarrow trigger Tracker digitization)

- have a first look to **muon+calo trigger objects**

L2.2 Resolution 1/pt (II)

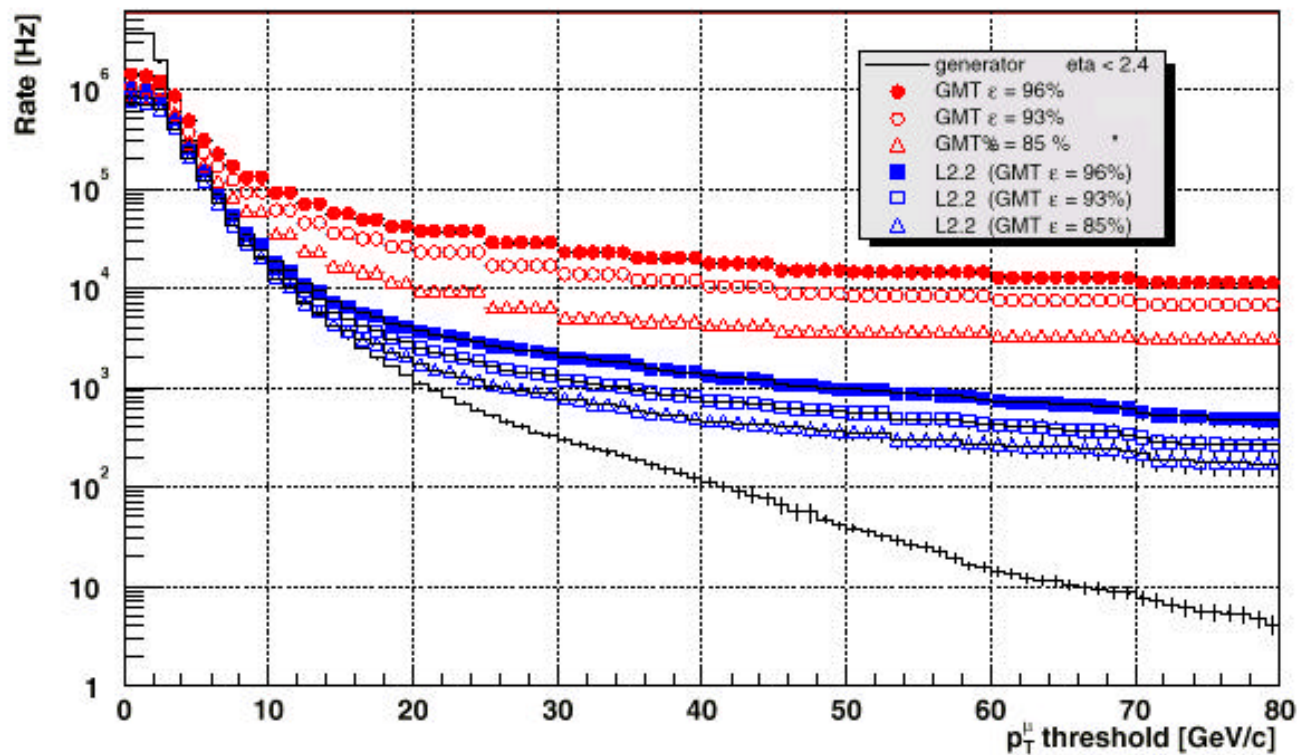
Pt4 sample



Annalina Vitelli - INFN Torino

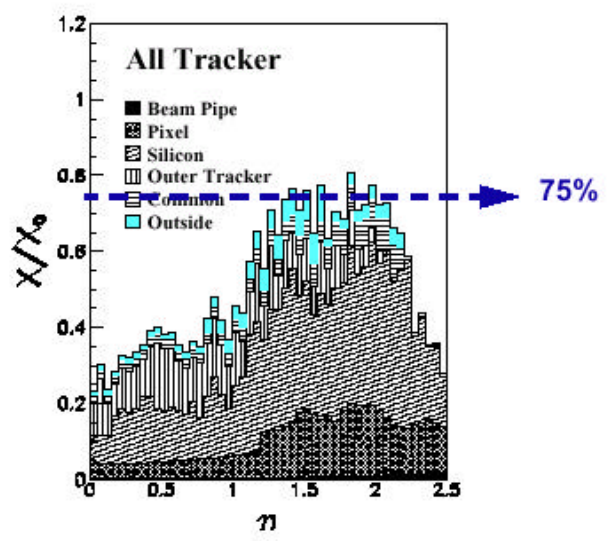


Single muon:L1&L2 (prel.CSC&GMT code)

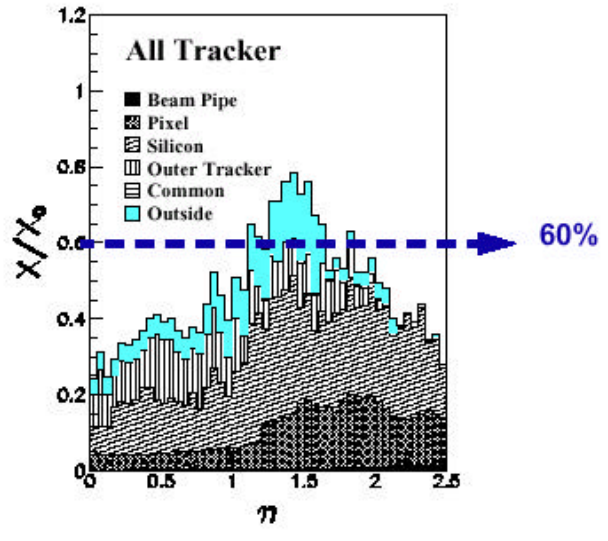


Tracker b/tau

Material



February 2000 - ATDR



April 2000 - New service routing

July 2000 - Cable material increased by ~ factor 2... More realistic description

Alessandra Caner - CERN

Strawman Trigger Selection

- **Level 1**
 - Single electron L1 threshold @29.5 (95% at 35 GeV)
 - Double electron L1 threshold @12.5 (95% efficient at 17.5 GeV)
- **Level 2**
 - Single electron cuts for 85% efficiency at threshold Pt (35 GeV)
 - Double electron cuts for 85% overall (event) efficiency at 17.5 GeV
- **Rate out of Level1: 10.6+/-0.2 kHz**
- **Rate out of Level2: 1.63+/-0.05 kHz**
 - Contributed by single: 1.48 kHz
 - Contributed by double: 0.15 kHz
- **Overall rejection: 6.65; Rejection on double electron: 28**

Emilio Meschi - Egamma Level 2, HLT Workshop, July 21st 2000. [34]



Conclusion

Software Tools

- “cmsim - orca4 - ntuplemaker” finally working! (almost)

L2 Tau

- Isolation cut provided a factor of 10 reduction over L1 with efficiency ~65%.
- Next - use tracker at L3.

Jets / MET

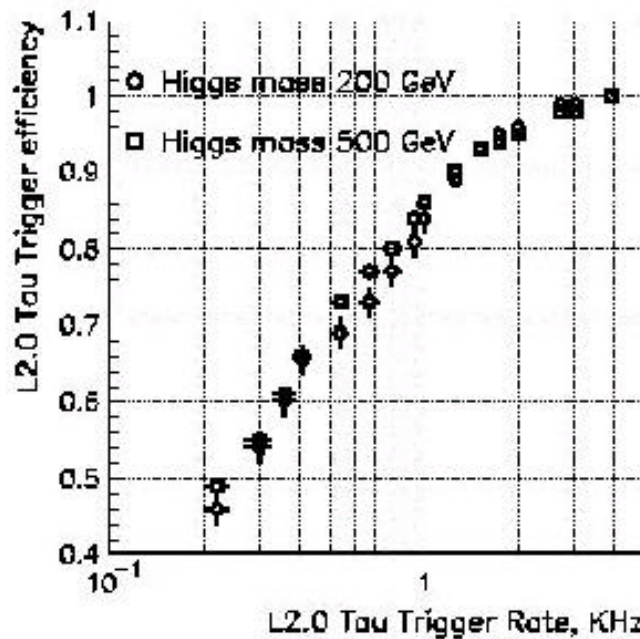
- Analysis of ORCA4 data just started.
- Follow the plan shown on June 6th.
 - Many parameters and algorithms to test.
 - Threshold, LSB, pedestal&pile-up energy subtraction, cone & window sizes, energy calibration, use of tracker/pixel etc.

Low Et Trigger and MET trigger are real challenge at $10E34$,
Hope to contribute to triggers on low mass scale physics.



L2 Tau Trigger

Efficiency v.s. L1 rate reduction



L2:
reduction = 10 at 65% eff.

4kHz -> 400Hz

L3:
Use tracker information

for $gg \rightarrow bbA/H$, $A/H \rightarrow 2\tau \rightarrow h^+ h^- X$ events passed
L1 and "off-line" cuts: $E_T^{j \rightarrow \text{jet}} > 60 \text{ GeV}$, $|\eta^{j \rightarrow \text{jet}}| < 2.4$



Ruolo dei Prototipi di Tier 2001



◆ Come detto il ruolo e' duplice:

- “Bottom-up”: risorse necessarie alla **produzione ed analisi delle simulazioni**
- “Top-down”: prototipizzazione e quindi studio delle architetture e tools di Tier-n integrati nella Griglia italiana ed internazionale

◆ Il coordinamento tra i Centri e' assicurato dal comune interesse per lo sviluppo dei tools, per la fisica associata all'esperimento e dal coordinamento di attivita'; questo coordinamento ha gia' prodotto:

- Istruzioni per le installazioni “uguali” nelle sedi CMS
- Sviluppo di tools (prototipi di Grid) per garantire la complementarita' degli sforzi
- Definizione dei responsabili e dei contatti (anche problem solving, vedi anche sopra)

◆ Inoltre CMS Italia ha gia' prodotto (primavera 2000) e produrra' (autunno 2000) simulazioni coordinate in molte sedi (es. Produzione di eventi anche nelle sedi non direttamente coinvolte nell'analisi di quegli eventi)

PLEASE READ, OUTLINES FOR CMS DATA PRODUCTION IN AUTUMN 2000

Production coordinator: Hafeez R. Hoorani, (NCP).

See CMSIM production meeting on 26.07.2000: [minutes](#)

In this production following centres will take part:

- **INFN**, Contact Person: Ivano Lippi, Ivano.Lippi@pd.infn.it.
- **Moscow University**, Contact Person: Olga Kodolova, Olga.Kodolova@cern.ch.
- **Caltech**, Contact Person: Vladimir Litvine, Vladimir.Litvin@cern.ch.
- **FNAL**, Contact Person: James Amundson, amundson@fnal.gov
- **HIP**, Contact Person: Veikko Karimaki, Veikko.Karimaki@cern.ch.
- **IN2P3**, Contact Person: Philippe Mine, pmine@poly.in2p3.fr.
- **Bristol**, Contact Person: Dave Newbold, Dave.Newbold@cern.ch.

Total # of events requested : 4320k

**Status of the CMSIM
Production requests:**

[egamma](#)

[muon](#)

[jetmet](#)

[Level 1 trigger](#)

- [Naming scheme](#), datasetname **_RUNNUM**, where **RUNNUM** is a unique run number assigned to the job. Following are the ranges of run numbers assigned to each production centre;

CERN	100001-200000
INFN	200001-300000



Produzioni e GRID



◆ **Gia' la produzione in Autunno 2000 usera' alcuni tool di base di Grid in CMS: GDMP (Grid Data Management Pilot).**

Deliverables and Milestones

- First Prototype
 - August 2000
 - Basic infrastructure
 - Security Done
 - Control Messaging Done
 - File transfers Done
 - Request Manager Done
 - Replica Manager
 - Replica synchronisation In Progress (OBJY, Catalog)
 - Database Manager In Progress
 - Multithreaded server Done

◆ **CMS (Italia) ha gia' una serie di pagine e tools per l'installazione coordinata in ogni sito del software. L'installazione comprende anche Globus e presto GDMP.**

◆ **Non c'e' problema per imporre i tools di Grid in CMS, li stiamo sviluppando insieme agli altri e li useremo appena pronti (come stiamo gia' facendo).**



Core Grid CMS contributions



◆ Dal Progetto INFN-Grid e dal Proposal EU DataGrid è chiaro che CMS in generale e CMS Italia sono coinvolte in modo determinante:

- “Core Grid” e’ inteso come WPs da 1 a 5 di DATAGRID
- Test Bed WP e HEP Application WP comprendono le attività di simulazione/produzione e valutazione del Modello a Tier-n alla GRID.

FTE	Ba	Bo	Ct	Fi	Pd	Pg	Pi	Rm1	To	
Core	1.2	0.5	(1.6)		0.5		1.9	0.8		4.9
Test Bed & App	1.9	3.6	1.2	1.7	4.6	1.6	2.4	1.6	0.9	19.5



Risorse richieste e strategia



◆ Un esempio di produzione (2000) Ottimistica!:

- CMSIM: 120 sec x 30 SI95 / event = 3600 SI95xsec per event
- ORCA: 180 sec x 30 SI95 / event = 5400 SI95xsec per event
- Track reconstruction: 100 sec x 30 SI95 / event = 3000 SI95xsec per event
- Autumn 2000 production is about 1800 CPU (30 SI95 each) weeks
- About 200 CPU at CERN and at least 100 CPU outside will take about 6 weeks production time (better estimate is of the order of 500 CPU total)
- **And Analysis!?**

◆ Un esempio di produzione (2001), sempre ottimistica!:

- Sara' richiesto un altro "magical factor of 10" per gli HLT
- Track finding, vertexing, signals : all critical and different Luminosities study
- Un fattore due nella necessita' di risorse e' veramente minimale = circa 1000 CPU (CERN and outside)



Breakdown costs for 2001



2001(Klit)	CPU (SI95)	CPU cost	CPU #	Disk (TB)	Disk cost	Tape Lib	Tape Lib cost	LAN Unit	LAN cost	Tapes (TB)	Tapes cost	TOTAL
Legnaro	3,100	223,000	89	5.0	390,000	1	100,000	1	102,000	20	54,000	869,000
Bari	700	50,000	20	1.0	78,000	0	0	1	21,000	5	14,000	163,000
Bologna	700	50,000	20	1.0	78,000	0	0	1	21,000	5	14,000	163,000
Padova	700	50,000	20	1.0	78,000	0	0	1	21,000	5	14,000	163,000
Pisa	700	50,000	20	1.0	78,000	0	0	1	21,000	5	14,000	163,000
Roma1	700	50,000	20	1.0	78,000	0	0	1	21,000	5	14,000	163,000
Catania	350	25,000	10	0.5	29,000	0	0	0	0	2	5,000	59,000
Firenze	350	25,000	10	0.5	29,000	0	0	0	0	2	5,000	59,000
Perugia	350	25,000	10	0.5	29,000	0	0	0	0	2	5,000	59,000
Torino	350	25,000	10	0.5	29,000	0	0	0	0	2	5,000	59,000
Totale	8,000	573,000	229	12.0	896,000	1	100,000	6	207,000	53	144,000	1,920,000

+ Pavia: 1PC and backup unit= 10 Mlit

**Common Fund Computing 2001:
70kCHF = 87 Mlit CORE (Bologna)**



Breakdown costs for 2001 reduction



2001(Klit)	CPU (SI95)	CPU cost	CPU #	Disk (TB)	Disk cost	Tape Lib	Tape Lib cost	LAN Unit	LAN cost	Tapes (TB)	Tapes cost	TOTAL
Legnaro	2,700	194,000	77	5.0	390,000	1	100,000	1	85,000	10	45,000	814,000
Bari	500	36,000	14	1.0	78,000	0	0	1	17,000	0	0	131,000
Bologna	500	36,000	14	1.0	78,000	0	0	1	17,000	0	0	131,000
Padova	500	36,000	14	1.0	78,000	0	0	1	17,000	0	0	131,000
Pisa	500	36,000	14	1.0	78,000	0	0	1	17,000	0	0	131,000
Roma1	500	36,000	14	1.0	78,000	0	0	1	17,000	0	0	131,000
Catania	200	14,000	6	0.0	0	0	0	0	0	0	0	14,000
Firenze	200	14,000	6	0.0	0	0	0	0	0	0	0	14,000
Perugia	200	14,000	6	0.0	0	0	0	0	0	0	0	14,000
Torino	200	14,000	6	0.0	0	0	0	0	0	0	0	14,000
Totale	6,000	430,000	171	10.0	780,000	1	100,000	6	170,000	10	45,000	1,525,000

+ Pavia: 1PC and backup unit= 10 Mlit

**Common Fund Computing 2001:
70kCHF = 87 Mlit CORE (Bologna)**



Summary: CMS Computing 2001



Site (Mlire)	Inventario	Consumo	Totale
Bari	163		163
Bologna	149	14 + 87(core)	250
Catania	54	8	62
Firenze	54	5	59
Legnaro	815	54	869
Padova	149	14	163
Pavia	10		10
Perugia	54	5	59
Pisa	149	14	163
Roma1	152	14	166
Torino	54	5	59
Totale	1803	220	2023



Richieste aggiuntive 2000



◆Pd: 30 MLit.

- Spazio disco (al CERN) per circa 500 Gbyte; ricezione coordinata delle produzioni HLT in Italia

◆Ba: 15 MLit.

- 4 PC + 100 GB disco produzione HLT

◆Fi: 6 MLit.

- 1 PC server, front-end con la Farm di Sezione

◆Pg: 12 MLit.

- 4 PC produzione HLT

◆Rm1: 6 MLit.

- 100 GB disco produzione HLT

◆Ct: 5 MLit.

- 1 PC per sviluppo software