Computing in LHC experiments

LCG Launching Workshop

11-15 March 2002

The experiments have to get the best, most reliable and accurate physics results from the data provided by their detectors

Their computing projects are fundamental to the achievement of this goal

The LCG project was set up to help them all in this task

Outline

- Refining the Computing Model
- Schedule and high level milestones
- The experiments and the LCG
- Working with the GRID

Refining the Computing Model

- Update and refine parameters
- Incorporate constraints
- Re-optimise
- *LHC Computing Review (Hoffmann):*
  - Valuable basis for initial planning
  - Sizeable uncertainties in all numbers:
    - Event rates/sizes
    - Reconstruction time
    - Analysis access to data
    - Computing models
    - Many cycles of Moore's law (this is not true for Phase I)

Refining the Computing Model (2)

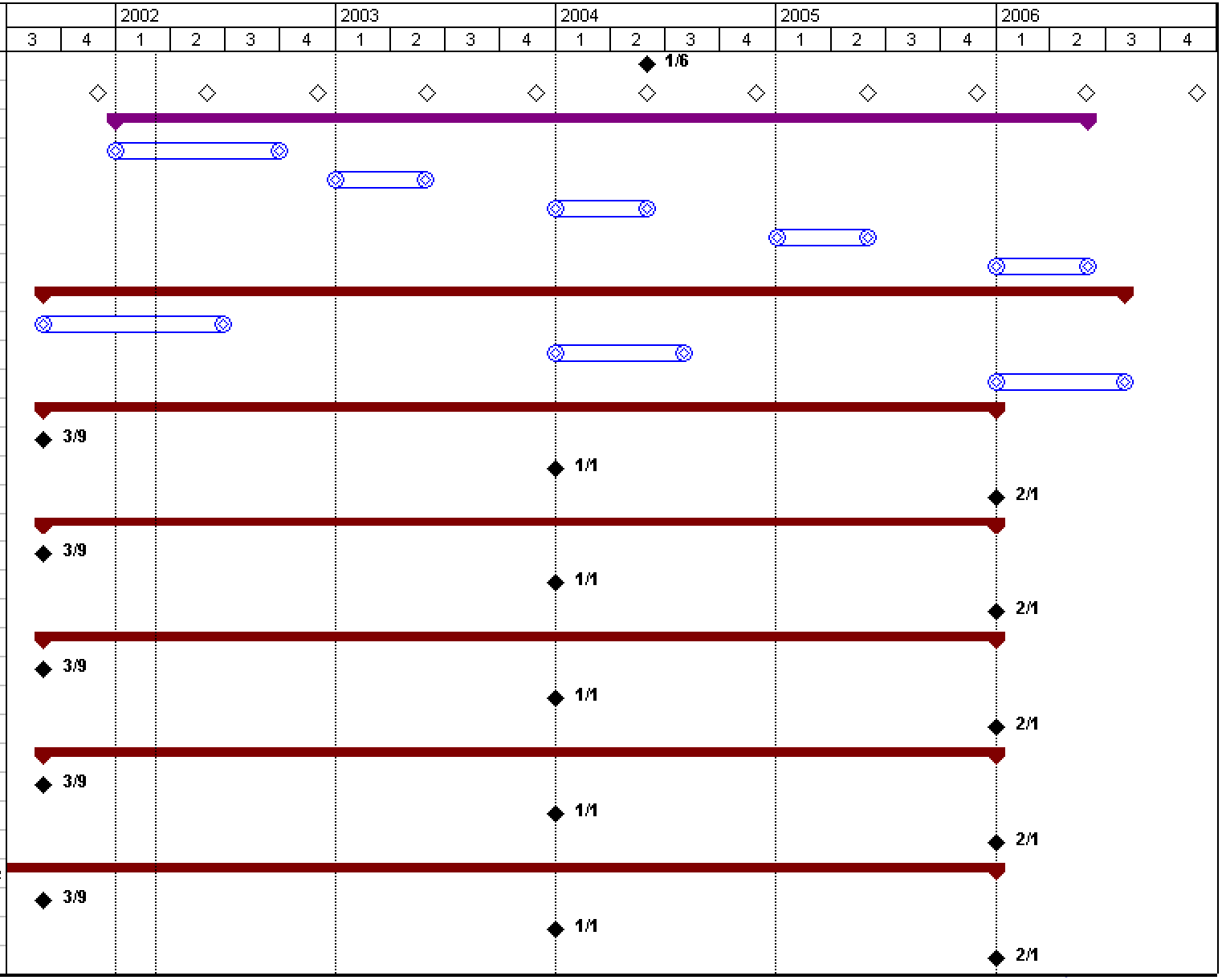
- A full iteration of the Computing Model is needed
  - 2002
    - Experiments and LHCC establish physics issues: events rates/size, etc.
    - PASTA Committee reviews cost trends (last revision 2 yrs. Old)
  - 2003
    - Experiments and LCG take into account probable cost constraints and iterate Computing Model
    - Computing TDRs due end of 2003
  - 2004
    - End of Phase I of the LCG Project
    - Deliverable: TDR with precise Computing Model, readjusted schedule and cost estimates

Schedule and high level milestones

- Experiments need a realistic track-able schedule (requested by the LHCC) geared at being fully ready at the time of the LHCC start
- The LCG Project has to provide accordingly a schedule which meets those of the experiments
  - All experiments are ready to work this out together with the Project
- The schedule is expected to be very detailed for the 1st phase of the project (end of 2004) and should be re-adjusted at the end of this period

ALICE high level planning

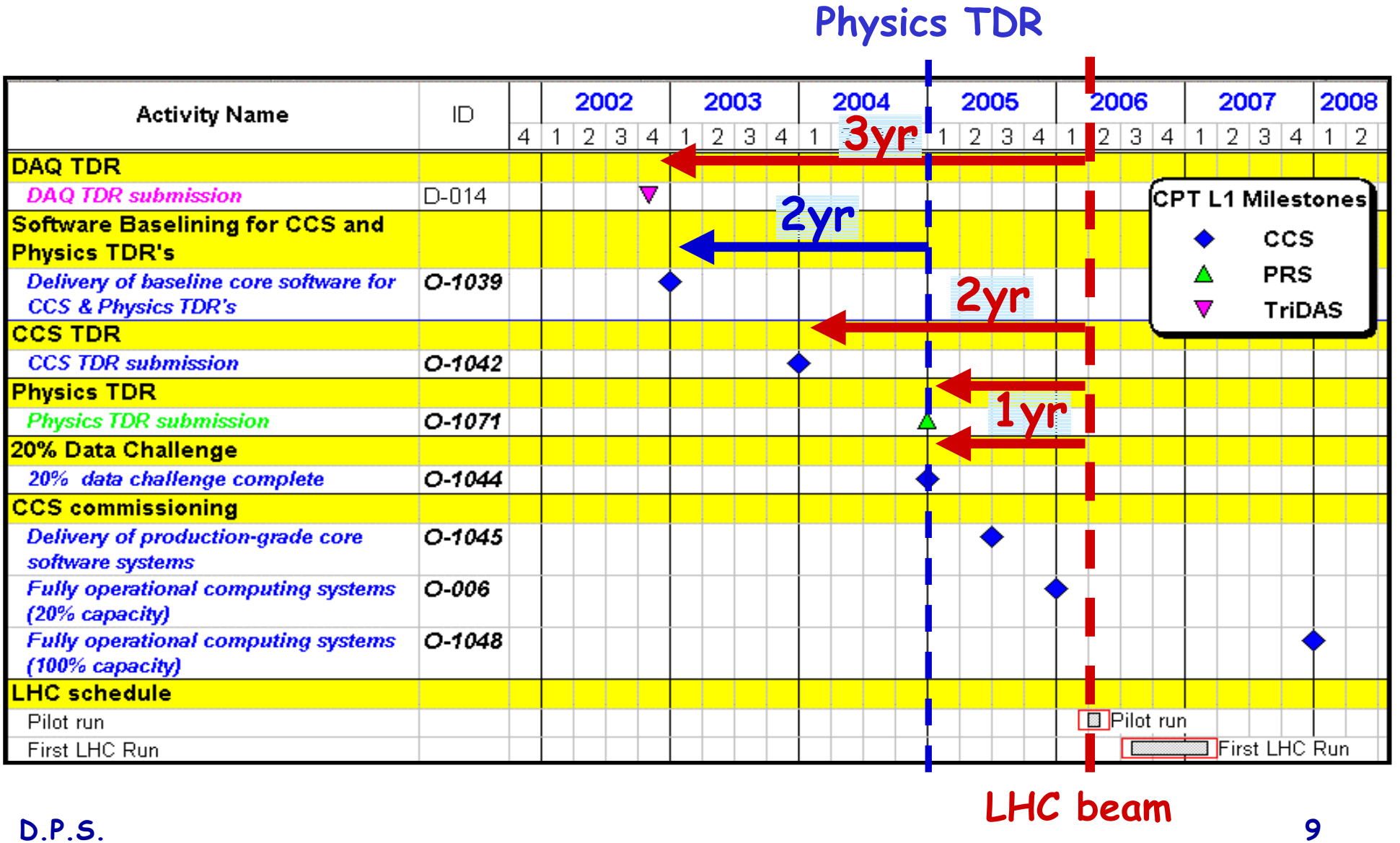
ID	Task Name
2	Computing TDR
3	**Releases of AliRoot**
16	**Data Challenges**
17	ADC4 200MB/s
18	ADC5 300MB/s
19	ADC6 450MB/s
20	ADC7 750MB/s
21	ADC8 1250MB/s
22	**Physic Challenges**
23	AP1 -- PPR 5%
24	AP2 10%
25	AP3 20%
26	**Simulation**
27	Version for AP1
28	Version for AP2
29	Version for AP3
30	**Production environment**
31	Version for AP1
32	Version for AP2
33	Version for AP3
34	**Reconstruction**
35	Version for AP1
36	Version for AP2
37	Version for AP3
38	**Physics Analysis Tools/Templates**
39	Version for AP1
40	Version for AP2
41	Version for AP3
42	**Basic Framework and Software Infra**
43	Version for AP1
44	Version for AP2
45	Version for AP3



ATLAS Major milestones

- **DC0**: "continuity" test on ~100k events, started in December 2001 and finishing around now;
- **DC1**:
  - Phase1 (April-July 2002): production and tests for High Level Trigger (HLT) TDR (end 2002). Scale: 10^7 events.
  - Phase2 (Sep-Dec 2002): focus on software tests, including Geant4, and Grid tools as available.
- DC1 **requires** use of resources outside CERN. (Very roughly 2:1 outside:CERN)
- **DC2**: First half '03: Scale up from DC2 to O(10%) of 1 year's data, more Grid focus. Foreseen as making extensive use of LCG TestBed.
- We anticipate **one DC per year** up to LHC start-up.
- **Computing TDR. Nov. 2003**
- **Computing MoU**. Originally for end 2001(!).
- **Physics readiness report**: 1 year before LHC start-up

CPT Level 1 Milestones
(V31. Not Including any LHC delays)

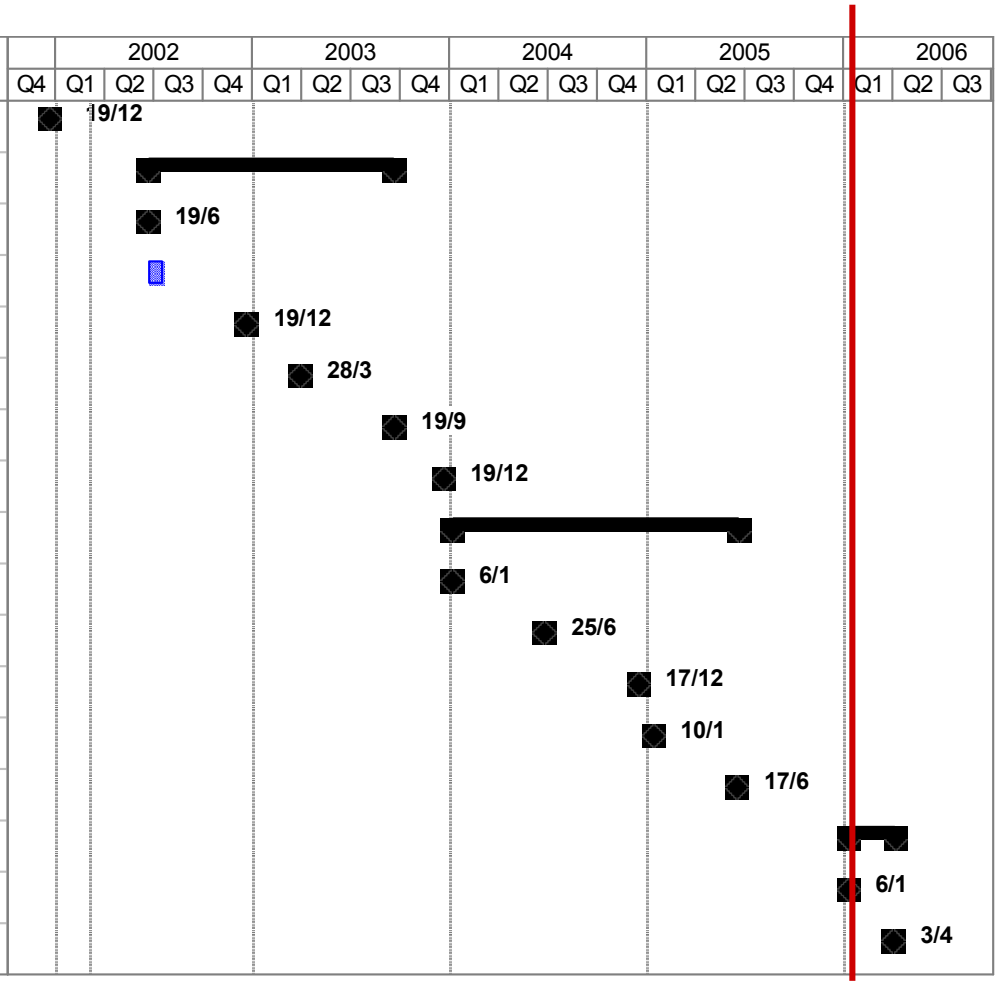


Activity Name	ID	2002	2003	2004	2005	2006	2007	2008
		4 1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2
DAQ TDR								
DAQ TDR submission	D-014							
Software Baselining for CCS and Physics TDR's								
Delivery of baseline core software for CCS & Physics TDR's	*O-1039*							
CCS TDR								
CCS TDR submission	*O-1042*							
Physics TDR								
Physics TDR submission	*O-1071*							
20% Data Challenge								
20% data challenge complete	*O-1044*							
CCS commissioning								
Delivery of production-grade core software systems	*O-1045*							
Fully operational computing systems (20% capacity)	*O-006*							
Fully operational computing systems (100% capacity)	*O-1048*							
LHC schedule								
Pilot run								
First LHC Run								

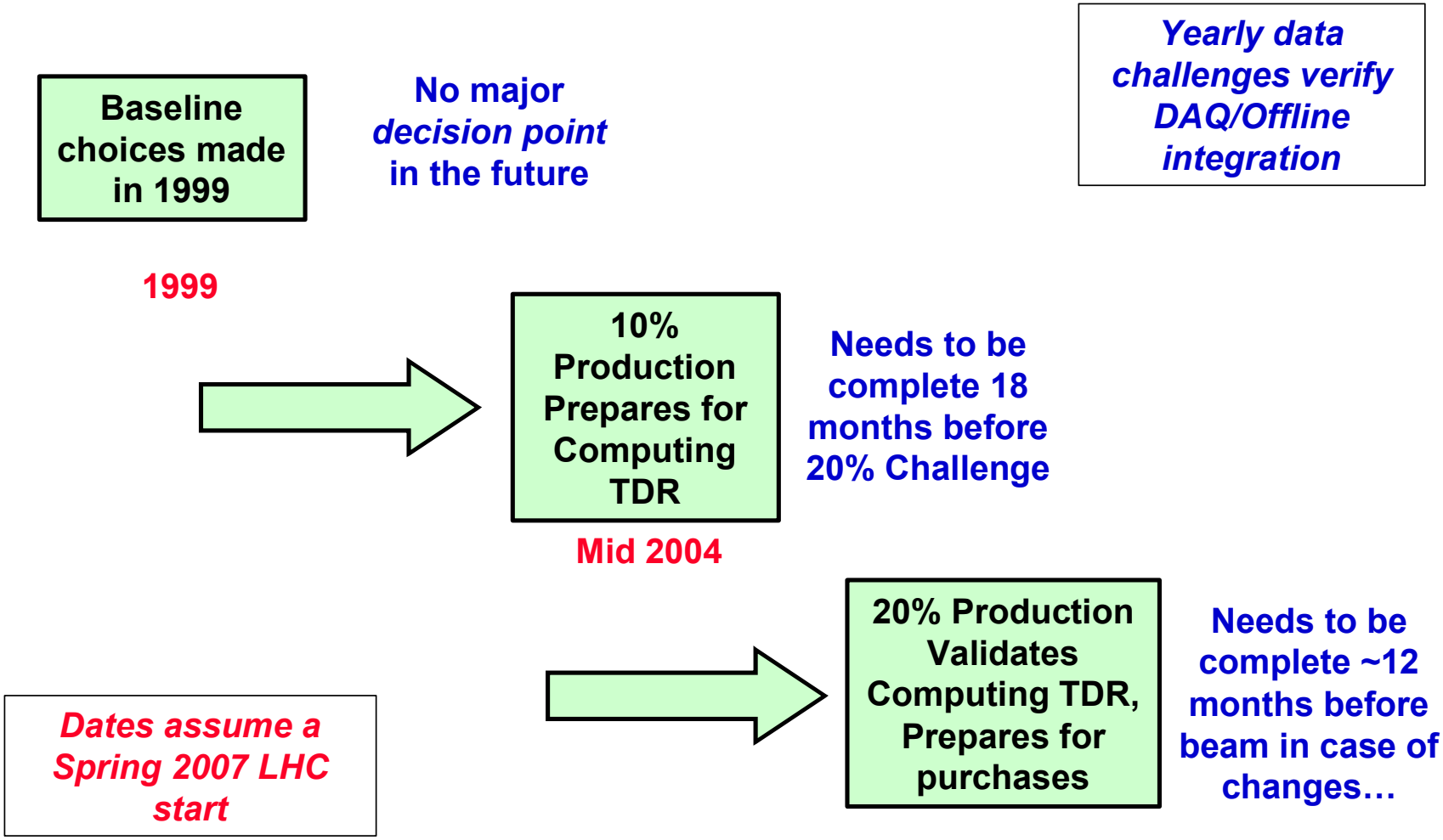
LHCb High Level milestones

(assumes LHC start April '06)

ID	Task Name
1	DAQ TDR (submitted)
2	Development - Prototype Cycle
3	DC1 - first tests using grid middleware
4	MC production 10**7 events
5	LHCb light and Trigger TDRs
6	Delivery of core software & applications
7	DC2 - functionality/performance test
8	Computing TDR
9	Development - Production Cycle
10	Computing - 10% capacity
11	DC3 - grid based production tests
12	Delivery of production quality software
13	Computing - 20% capacity
14	DC4 - production tests of computing model
15	Operation
16	Computing - 100% capacity
17	First colliding beams



Rationale of ALICE schedule



F.C.

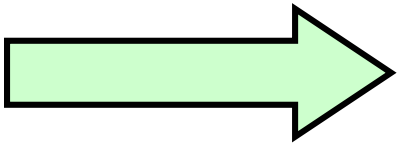
Rationale, and *possible* new schedule

Baseline Choices for Physics TDR (2 Yr Process)

Mid 2003

Needs to be complete One Year before Computing TDR

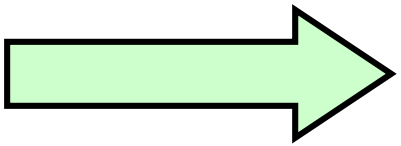
Try to gain a few months on V31 schedule which was too tight



5% Challenge Prepares for Computing TDR

Mid 2004

Needs to be complete One Year before 20% Challenge



20% Challenge Validates Computing TDR, Prepares for purchases

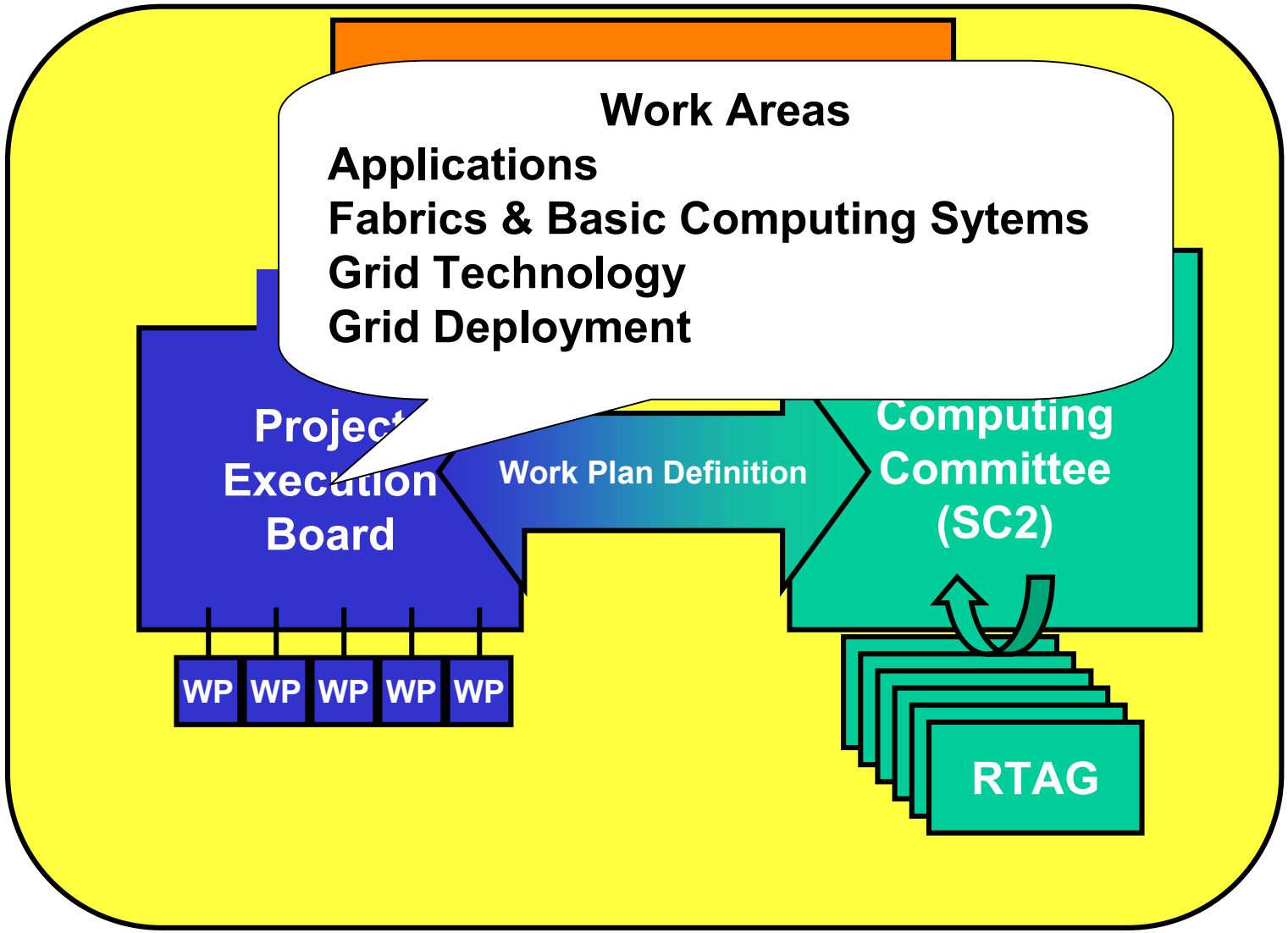
Mid 2005

Needs to be complete ~18 months before beam in case of changes…

Dates assume a Spring 2007 LHC start

D.P.S.

The LHC Computing Grid Project Structure



The Experiments and the LCG

- The success of the LCG Project is fundamental to the success of LHC computing
- All experiments are fully committed to collaborative development
- The experiments have the ultimate responsibility of their computing projects
  - The *RTAG SC2 PEB WP* mechanism has to work in practice
  - Experiments have to participate in the decision-making during execution of the project (*WP*)
- Timescales need to be clear
  - Impact on experiments project schedule
  - Tracking progress
- Distributed computing is fundamental
  - GRID technology seems to be the basis for successful deployment of the distributed computing model

The Experiments and the LCG
Common Projects

- Everything that is not an experiment-specific component is a potential candidate for a common project
- Experiments are in a period of important changes
  - reconsidering their persistency strategy
  - Integrating *Geant 4* and *Fluka* (Alice) into their frameworks
  - etc…
- Now is a good time to launch and implement common solutions
  - delays are probably inevitable, however one should gain in the long term from better support and maintenance
- Existing and new resources are best focused through LCG common projects

The Experiments and the LCG
Common Projects (2)

- A sound model should be implemented to achieve successful common solutions
  - Lessons should be learned from past experience
  - Tight coupling between active user requirements and developers goals should always be ensured
- The mechanism foreseen by the LCG Project
  - The SC2
    - sets the requirements
    - approves strategy and workplan
    - monitors progress and adherence to the requirements
  - The PEB
    - gets agreement on milestones, schedule, resource allocation
    - manages the progress and direction of the project
    - ensures conformance with SC2 recommendations

The Experiments and the LCG
Common Projects (3)

- Some of the common projects are expected to benefit by being hosted in the context of one experiment
  - These projects follow the same SC2 ◁-▷ PEB mechanism
  - Hosting experiment accepts responsibility for general requirements
  - LCG personnel work on the project in the experiment environment
  - All experiments agree to test prototypes and iterate requirements
- If this all works experiments commit resources to these common projects but gain directly from LCG and from the work of other experiments.

The Experiments and the LCG
Priority issues for 2002

- Resolve persistency strategy (RTAG intermediate report)
  - Should be integral part of the data challenges of 2003
  - Experience is needed well before the computing TDRs
- Define analysis strategy
  - Possibly affected by persistency choice
- Full deployment of Geant 4
  - For CMS this means retirement of CMSIM during 2002
  - For ATLAS and LHCb, validation of physics processes implies that Geant 3 based simulations have still to be maintained for some time (2003)
  - Alice will reconsider use of G4 if physics improves. They plan to integrate Fluka for full detector simulation. CERN support is considered essential for this. (Fluka is also needed by the other experiments)

The Experiments and the LCG
Priority issues for 2002 (2)

- Refine the WorldWide computing model
  - Clarify relations with GRID middleware
  - There is scope for common projects arising from integration of software framework to GRID (see T.W. talk to-morrow)
    - Production and bookkeeping tools
    - Job submission, control and monitoring
    - Standard installation kits
    - Etc…
- All experiments are fully committed to perform successful Data Challenges in 2002

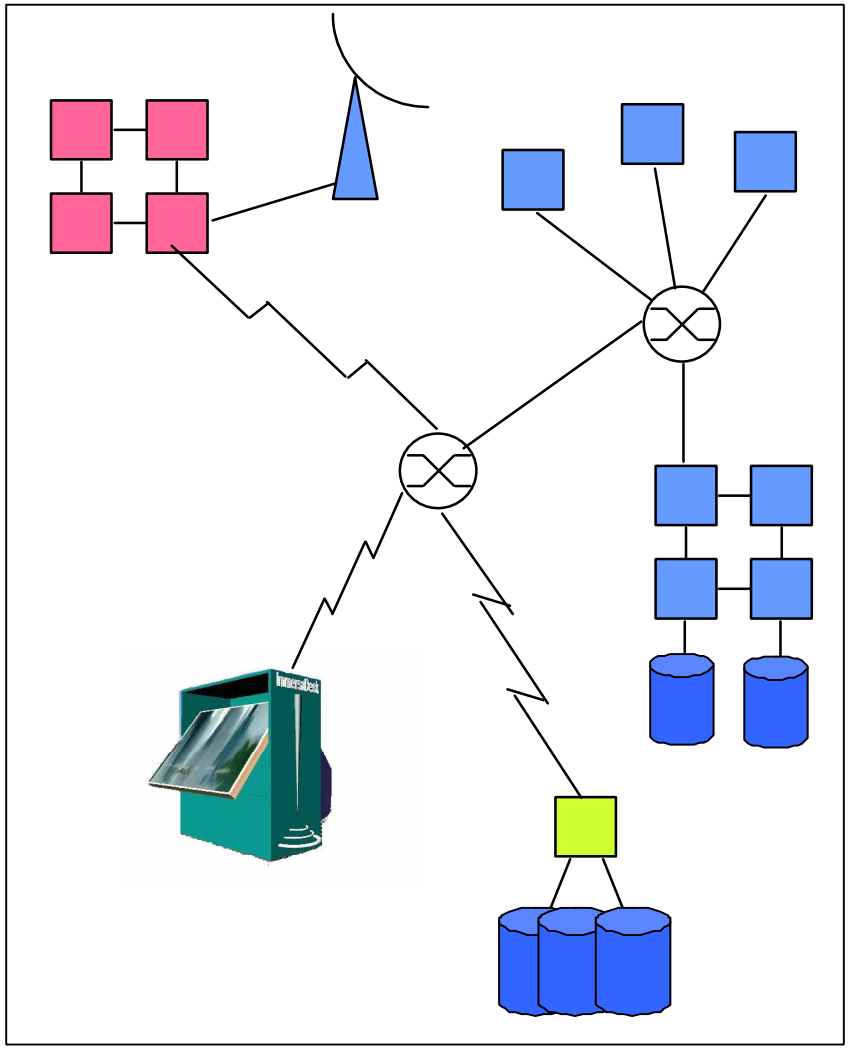
Working with the GRID

- The LCG has decided to adopt a WorldWide distributed computing model (MONARC)
- **Basic principle**: every physicist should have equal access to the data and to the resources … **very ambitious** …
- The system will be very complex
  - Number of components on each site, number of sites
  - Different tasks performed in parallel: simulation, reconstruction, scheduled and unscheduled analysis
- The GRID technology seems to be the basis for successful deployment of the distributed computing model
  - Dependable, consistent, pervasive access to high-end resources

The GRID

"Dependable, consistent, pervasive access to [high-end] resources"

- **Dependable:**
  - Can provide performance and functionality guarantees
- **Consistent:**
  - Uniform interfaces to a wide variety of resources
- **Pervasive:**
  - Ability to "plug in" from anywhere



Working with the GRID (2)

- Some words of caution:
  - GRID technology still "fluid"
    - Emerging standards
    - Several different middleware projects
      - DataGrid, DataTAG, CrossGRID, iVDGL, GriPhyN, PPDG etc., etc.
  - Experiment approach to the GRID not yet well defined
    - a "GRID use cases" RTAG is in preparation
- However major requirements are identified:
  - Tests have to run seamlessly across transatlantic testbeds
  - Need for horizontal components rather than just vertical solutions
  - LHC community will have to build experiment specific end-to-end application layer

Working with the GRID (3)

- GRIDification of the CERN computer centre
  - Necessary in order to gain real experience
  - A timescale and a strategy have to be elaborated
- The GRID deployment has to follow the schedule of the yearly Data Challenges to which all experiments are committed
- Convergence of the different projects
  - … cannot be delayed (waste and duplication)
  - … cannot be forced (all the opportunities of the GRID are not yet understood) there is a risk to converge on the wrong solution
- User support to the GRID testbed is very labour intensive (cf. DataGRID/WP8) yet necessary

CONCLUSIONS

The LCG Project must be a success

All four experiments are fully committed

- Make Common Projects successful by defining the correct mechanisms
- Detailed schedules and milestones are fundamental

The WorldWide distributed computing model is a real challenge