Version 0

G.N. Patrick

12th July 1999

SOME NUMBERS FOR A LHC(UK) SIMULATION FACILITY

1. Introduction

A previous document[1] summarised some of the requirements and rough costs for a central CMS(UK) computer centre over the period 2000 - 2007. The conclusion from a subsequent LHC-wide meeting[2] was that any bid to the Joint Infrastructure Fund (JIF) should be based on short-term simulation requirements over the period 2000 - 2003. This note looks at some of the basic numbers for such a bid.

2. Basic Parameters

The following assumptions are made about the basic design parameters for LHC and CMS:

$\sqrt{s}$	= 14	TeV
L	= 10^{34}	$cm^{-2}\,s^{-1}$
Bunch spacing	= 25	ns
Data taking time	= 10^7	s/year (duty cycle = 30%)
Level 3 trigger output rate	= 100	Hz

The total p-p cross-section at 14 TeV is ~ 100mb which gives an interaction rate of ~10^9 Hz and an integrated luminosity of 100 fb^{-1} over an "LHC year". The first-level trigger will reduce the event rate to 100 kHz and the Event Filter Farm will perform Level-2 and Level-3 trigger selections which will limit the output rate to ~100 Hz. The trigger must therefore achieve an overall rejection factor of 10^7 - this sets the initial scale.

3. Overlapping Events and Minimum Bias Background

With a bunch spacing of 25ns, it is anticipated that, on average, 25 minimum bias events will be produced per crossing. These overlapping events will be superimposed on the interesting events selected by the trigger and greatly complicate the simulation, reconstruction and analysis tasks.

In the *full* CMS simulation program (F77), CMSIM, the beam crossings are reproduced by mixing events at the digitisation level. Physics events and minimum-bias events are run separately through CMSIM on an event-by-event basis to produce hits which are stored in intermediate files. Data cards then control the subsequent event mixing and pileup, with the number of events in each bunch (usually) selected from a Poisson distribution.

Minimum bias events have been run through CMSIM115 on a HP C200 machine at the RAL CSF. From these, it has been estimated that to generate the hits for each overlap event requires ~580 SI95-sec and ~ 220kB of storage (although runs at CERN seem to indicate 500kB/event).

So, to simulate the hits from a sample of 10^8 overlap events in one month using current (F77) software requires:

CPU = 10^8 x 580 SI95-s/2419200s = 23975 SI95 - ie. 1,332 PII(450MHz) machines

and 22 - 50 TB of dedicated disk space to store the hits.

Once generated, this dataset of overlap events can be re-used many times and added to the various physics processes under study.

It is worth noting that the shower cuts were a factor of 10 smaller than the standard GEANT cuts and these can have a large influence on CPU time.

Using only "hard" overlap events can also make some savings. A *fast* Monte-Carlo, CMSJET, was used to generate both standard PYTHIA (MSEL=2) overlap events and hard overlap events (MSEL=1, $p_T > 5$ GeV). The hard events were effectively sampled from a ~45mb cross-section instead of the full QCD cross-section. This meant that ~10 events (elastic and diffractive) were excluded from the usual ~25 in each bunch crossing. It seems possible to make savings of a factor of 2-3 by employing such cuts at an early stage in the simulation chain.

4. Physics Processes

Approximate cross-sections have also been calculated at $\sqrt{s}$ = 14 TeV for a number of channels by running PYTHIA (via CMSJET). The events/year are calculated assuming full luminosity over a year (100 fb^{-1}).

Process	Cross-section	σ/σ_T	Events/Year (100 fb^{-1})	Comment
tt	594 pb	5.9×10^{-9}	6×10^{7}	Pythia (m_t = 175 GeV)
WH with W $\rightarrow \ell\nu$	0.46 pb	4.6×10^{-9}	4.6×10^{7}	Pythia (m_H = 100 GeV)
WW	71 pb	7.1×10^{-10}	7×10^{6}	Pythia
WZ	27 pb	2.7×10^{-10}	3×10^{6}	Pythia

The energy dependencies of the cross-sections for various physics processes with extrapolations to LHC energies are also shown in Figure 1. The H $\rightarrow \gamma\gamma$ decay is an example of a rarer process which has a branching ratio of only $\approx 2 \times 10^{-3}$ and a σ x BR of ~ 80 fb for a Higgs mass of m_H = 100 GeV. This means that only ~8000 events will be seen per year at full luminosity, but these will unfortunately sit on a large $\gamma\gamma$ continuum background, which is 10-20 times larger than the signal.

Based on the few physics channels above, the size of typical Monte-Carlo event samples will vary between 10^6 and 10^8 events. These will require:

- Simulation through the Monte-Carlo program.
- The addition of ~15-25 overlap events and pileup at the digitisation stage.

By processing events on RAL-CSF, it has been confirmed that for CMS this requires ~ 5000 SI95-s/event and storage (kine, hits & digi banks) of ~1.3 MB/event. This is to be compared with the 2MB/event estimated for a fully simulated event.

So, to simulate a sample of 10^7 physics events plus background in one month requires:

CPU = 10^7 x 5000 SI95-s/2419200s = 20,668 SI95 - ie. 1,149 PII(450MHz) machines

and 13 - 20 TB of storage.

In principle, these stages (given stable simulation software, background assumptions, etc.) will be run relatively infrequently for each physics process. In addition, the following tasks are likely to be run numerous times as code is written, algorithms are developed and physics studies undertaken:

- Reconstruction.
- Subsequent analysis.

These tasks can make use of the cpu power already installed for the simulation stages (assuming that the work can be scheduled in this way). The extra output (assuming ESD format of 100kB/event) would be ~1 TB for 10^7 events.

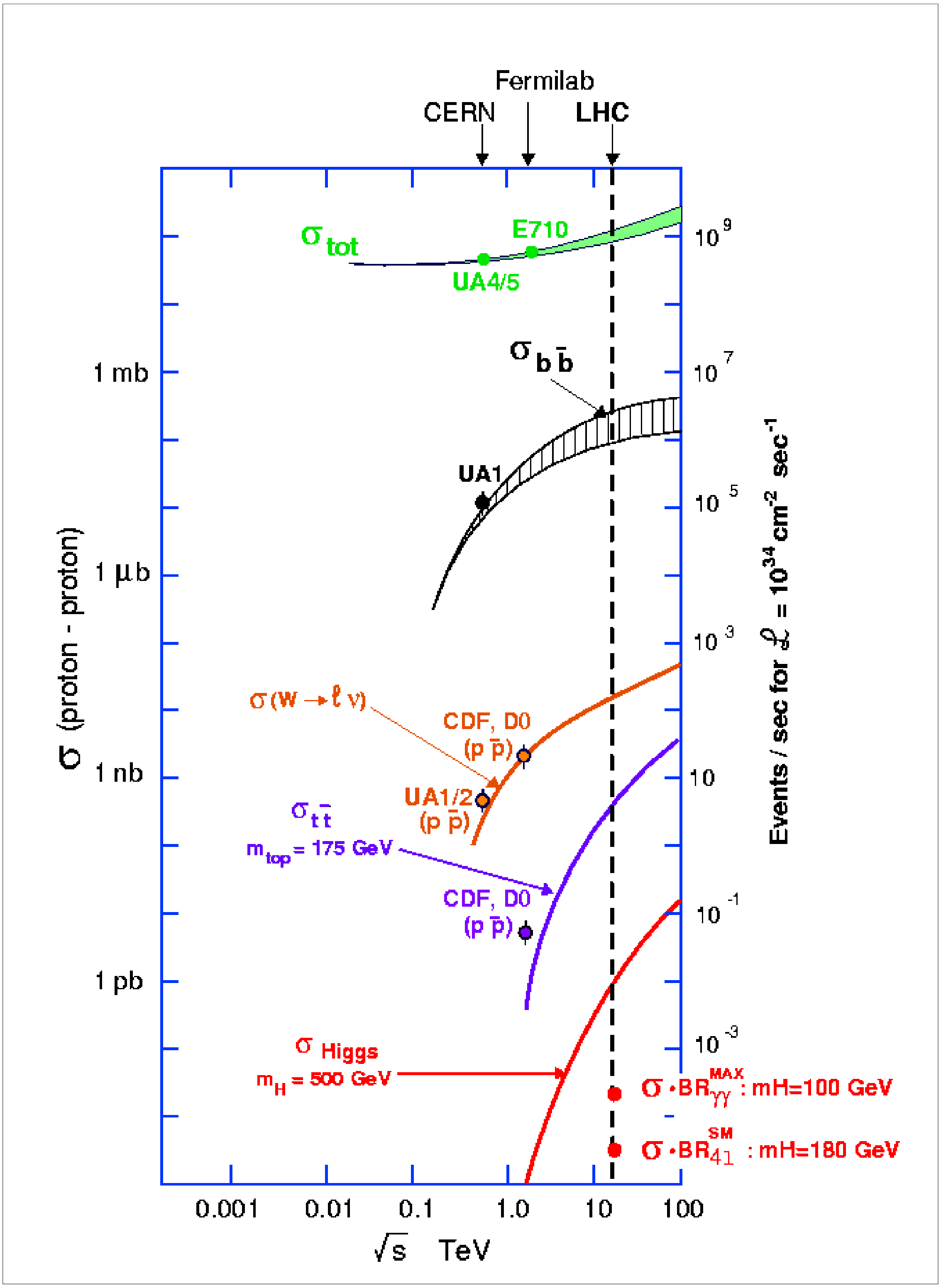

Figure 1 Energy dependence of some cross-sections at hadron colliders

[1] G. Heath et al: Proposal For A CMS(UK) Computer Centre, 18th June 1999
[2] P. Clarke: Minutes of first meeting of UK-LHC groups to discuss strategy for LHC wide JIF bid, 25th June 1999