International Collaborations in High Energy Physics and associated Technologies

G.P. Yeh

Fermi National Accelerator Laboratory, Batavia, Illinois

Academia Sinica, Taipei, Taiwan

Abstract

Large high energy physics experiments in the world now require very large scale international scientific collaborations. Each experiment may require hundreds of scientists from many countries, to build large, technically challenging detectors, and to efficiently collect and analyze petabytes of data. We describe some of the physics projects and their associated technologies at Fermi Lab in the U.S., CERN in Europe, and KEK in Japan. We also discuss some of their tools for working together, including the use of video conferencing, remote control rooms, high performance network, open source software, and cost-effective computing resources.

1 Introduction

Since the beginning of mankind, the quest for the most basic constituent "particles" (building blocks of matter) and the forces that govern them has been a great human endeavor. This intellectual challenge has led to many important discoveries in human history. By the beginning of the 20th Century, models of the atom were emerging. The electron, which was first named by the ancient Greek philosophers, was discovered in the year 1897. The nucleus and the proton were discovered in 1911 and 1919, respectively. It took nearly 50 more years until quarks, the constituents of the proton, were discovered in 1968. Knowledge of properties of the quarks and the leptons provide important steps towards understanding Nature and the origin of matter.

The experiments to study the quarks and leptons are carried out in large international laboratories such as Fermi Lab (Batavia, Illinois), CERN(Geneva, Switzerland), KEK(Tsukuba, Japan), SLAC(Stanford, California) and DESY(Hamburg, Germany). The accelerators and experiment detectors at these laboratories are technically extremely challenging. Detectors now (and in the future) often have millions of electronics readout channels, collecting data at rates of millions of Megabytes per second. The computing challenges in large data acquisition, storage, access, and analysis have resulted in inventions in the parallel computing "Farms" and many other aspects of computing.

The large high energy physics experiments now require very large scale international scientific collaborations. Each experiment may require hundreds of scientists from many countries. The need for information sharing for high energy physics led CERN to the invention of the World Wide Web. Another example of large scale collaboration is the development of Linux Operating System software and the Open Source software. Other issues for large collaborations include data sharing and analyses, Remote Control Room and data monitoring, video conferencing, e-Notebook, and high performance network. Large scale international scientific collaborations can be very productive and successful, as they have been in in high energy physics.

2 High Energy Physics

The most fundamental particles are six Quarks $\begin{matrix} u & c & t \\ d & s & b \end{matrix}$ and six Leptons $\begin{matrix} \nu_e & \nu_\mu & \nu_\tau \\ e & \mu & \tau \end{matrix}$. Since the discovery of the b quark in 1977, many labs searched extensively for the Top quark t, which was finally discovered in 1995. It is a tremendous human achievement that we have now discovered all of these most basic particles in Nature. The quarks and leptons interact with one another, governed by the gravitational, electromagnetic, nuclear weak, and nuclear strong forces. These four forces are carried by the graviton, the photon, the W and Z bosons, and the gluon, respectively. The Z boson of the nuclear weak force can decay into quarks and leptons. The more types of quarks and leptons there are in Nature, the faster the Z boson would decay. By measuring this decay rate precisely, we conclude that there are (at most) six types of leptons and six types of quarks.

The Top quark, because of its large mass, can only be produced and detected at Fermi Lab, and only at a rate of one Top-antiTop $t\bar{t}$ pair per 10^{12} collisions. To discover the Top quark required a collision rate of million collisions per second, resulting in few dozens of Top quarks. To achieve a much better understanding of the properties of Top quark, Fermi Lab and the CDF and D0 experiments are planning to improve the event rate and the data by 20 - 300 times in the next 2 - 5 years.

It is remarkable that the Top quark is so massive. The bottom quark, which is the heaviest among the first five quarks, has a mass of only 5 $\mathrm{GeV/c^2}$. The Top quark mass of 175 $\mathrm{GeV/c^2}$ (billion electron volts) means that the Top quark is heavier than nuclei of 77 of the 107 elements in the Periodic Table, as heavy as atoms of Iridium and Gold. Atoms have radii of 10^{-8} cm. The proton, with a mass of 0.9 $\mathrm{GeV/c^2}$, has an effective radius of 10^{-13} cm, while the quarks and leptons have much smaller radii of less than 10^{-17} cm. The Top Quark lives for only 10^{-24} seconds after being produced, while the Bottom quark b has a lifetime of 10^{-12} sec. In contract, the proton has a lifetime of greater than 10^{+40} seconds.

CP symmetry violation is an expected effect of the "Standard Model" with three pairs of quarks and three pairs of leptons, but so far only one process (involving the K^0 meson) has been measured. CP-violation relates to matter-antimatter imbalance. Many CP-violation effects are expected in B meson decays. In cosmology, one great puzzle is the disappearance of antimatter from the Universe. Studying the CP-violation in B decays will help us understand this very important but poorly understood effect.

The masses of the neutrinos are important numbers in physics. Since the discovery in 1956 of the ν_e neutrino, many exhaustive efforts to measure the mass of the neutrinos have obtained results consistent with 0. In the Standard Model, the masses of the neutrinos have been assumed to be 0. In astrophysics, neutrinos are produced soon after the Big Bang, during the explosion of stars to supernova, and in nuclear reactions including those in the stars and our sun. Neutrinos are also produced by cosmic rays which come into the earth's atmosphere. The number of neutrinos in the universe is extremely large. Non-zero neutrino masses may have great implications in physics. In 1998, the Superkamiokande experiment results seem to indicate that ν_μ produced in the atmosphere "oscillates" away to ν_τ or some other particle, thus suggesting that neutrinos may have non-zero mass. There will be many neutrino experiments in the future to measure these important parameters, and improve our understanding of the neutrinos.

To understand the empirical Standard Model of quarks and leptons with their non-zero masses, a well known explanation is the "Higgs Mechanism". If the universe is filled with a spin zero "Higgs field", then the quarks, the leptons, and the gauge bosons could interact with this field and acquire the particle masses. Physicists and experiments world-wide have searched extensively for (something like the) Higgs particle, which should have a mass less than 1000 $\mathrm{GeV/c^2}$. The discovery of the Higgs particle is a crucial milestone in physics.

Supersymmetry is a symmetry relating half-integer spin fermions and integer spin bosons. There are strong theoretical arguments which suggest that the Standard Model breaks down at energies above 1000 GeV. Supersymmetry provides a frame work that will accomodate all of the physics so far in high precision agreement with the Standard Model, but significantly improve over the Standard Model at higher eneregies. Supersymmetry suggests that there be a large number of new SUSY particles with masses between 100 GeV and 1000 GeV, which should be observable with the present and/or next accelerators/colliders [1].

3 Laboratories, Experiments and Detectors

The Fermi National Accelerator Laboratory, Fermi Lab, provides research facilities and programs for physicists from 98 U.S. institutions from 36 states, and 90 foreign institutions in 20 countries. The CDF Experiment, for example, now has collaborators from 50 institutions in 10 countries U.S., Japan, Italy, Taiwan, Canada, Switzerland, Germany, Russia, England, Korea. The Fermi Lab proton-antiproton collider, the Tevatron, with center-of-mass energy of 2000 GeV, is the highest energy accelerator in the world.

To produce particles such as the the Top quark, we must concentrate extremely high energy in extremely small space. Only the Tevatron accelerator at Fermi Lab has enough energy to produce the top quark. The Tevatron can accelerate 10^{12} protons and 10^{11} anti-protons in opposite directions to 1000 billion electron volts (GeV) per particle. The Tevatron achieves such high energy by using 1000 superconducting magnets. Bunches of protons and anti-protons travel at very close to the speed of light, going around the 6.3 km tunnel more than 45000 times per second, and collide inside an area with a radius of 0.005 cm.

The European Laboratory for Particle Physics, CERN, has 19 Member (European) countries with 5000 physicists from 309 institutions, and 35 non-member countries with 1800 physicists from 218 institutions. The Large Hadron Collider, with circumference of 27 kilometers and center-of-mass energy of 14,000 GeV, will be the highest energy accelerator in the world, and is scheduled to start in the year 2005.

The Particle and Nuclear Physics Organization KEK in Japan is the main laboratory for high energy physics in Asia. KEK is completing the B Factory for the Belle experiment, and has been extensively developing the Joint Linear Collider (http://www-jlc.kek.jp/). The JLC Physics and Detector Studies Group now has members and subgroups in China, India, Taiwan, Korea, Australia, Philippines, Vietnam, Singapore, and Japan.

There are strong, active Linear Collider study groups in the 3 regions of Asia, Europe, and North America. The Linear Collider, with length of up-to 25 kilometers and beam spot size of few nanometers, is technically very challenging. The hope is to complete by 2008 a Linear Collider with (Phase I) center-of-mass energy of 500 GeV, which can later be upgraded up-to 1,500 GeV. There are also study groups for the longer term Muon Collider with energy of 4,000 GeV (http://www.fnal.gov/projects/muon_collider/), and the Very Large Hadron Collider with energy of 100,000 GeV (http://www-ap.fnal.gov/VLHC/).

Modern particle detectors have been able to detect and study the Standard Model ingredients: the leptons (electron, muon, tau and the three types of neutrinos), photons, quark jets, gluon jets, and W and Z bosons. The large detectors now have millions of electronics channels, with detector data rates of millions of Megabytes per second, to measure accurately the position, momentum, energy, and type of the particles. The detector and the online data acquisition system must be able to process the data immediately, select from the millions of collisions per second, and record the most important types of events. With these detectors, we could find the Top quark, and candidates for other new physics, as soon as they are produced. The successful construction and operation of each large detector, and understanding the data and physics obtained with the detectors, all require a tremendous collaborative efforts for many years, by hundreds of physicists and technical experts world-wide.

A new useful device is the silicon detector, a precision charged particle tracking device near the collision point, which improves the detector's ability to identify the b-quarks. In additon to B physics studies such as the CP violation, b-quarks are expected in the decay products of the Top quark and other new physics, so that it is important to identify the b-quarks. The lifetime of bottom quarks is about 10^{-12} seconds. At nearly the speed of light, a bottom quark travels about (depdending on its energy) a millimeter in it's lifetime, before decaying into other particles. This distance is long enough to be detected by silicon detectors with spatial precision of microns.

In the next 10 to 20 years, the current and planned high energy physics laboratories and experiments world-wide will actively pursue the important new physics - Top quark studies, CP violation, neutrino mass, Higgs particle, Supersymmetry, and possibly other less anticipated new physics. The prospects are extremely exciting for great progress in our understand of Nature and in the advancement in technologies associated with high energy physics.

4 Computing

Many major experiments in high energy physics and nuclear physics will have large, challenging computing needs. The three large experiments at three B Factories, HERA-B(DESY), Belle(KEK), BaBar(SLAC) are on schedule to start in early '99 to collect large amounts of data for CP violation studies. The BRAHMS, PHENIX, PHOBOS, and STAR experiments at RHIC/Brookhaven plan to start heavy ion collisions in June, '99, with one of the main physics goals being to study quark-gluon plasma.

At Fermi Lab, CDF & D0 experiments will have data taking period Run II in 2000–2002, to study the newly discovered Top Quark, and search for B CP violation, Higgs, SUSY. The amount of data collected in Run II is expected to be 20 times that of Run I. The CPU requirement will be $N \times 10^5$ MIPS or 10^4 SPECint95 (Fermilab defines CPU equivalent to a VAX-11/780 to be 1 MIPS, and SPECint95 = 40 SPECint92 = 13 MIPS = 10 CERN_Units). The amount of data collected in Run III in 2003 –2005, is expected to be 15 times that of Run II. For the Large Hadron Collider starting in 2005, the Atlas & CMS experiments have estimated CPU needs of $N \times 10^7$ MIPS or 10^6 SPECint95.

4.1 Computing Systems at Fermi Lab

Most data at Fermi Lab are from the 10-12 large active physics experiments, with 1-150 TB of data and 30-500 physicist per experiment. These numbers are important in characterizing the computing problems and possible solutions. Big issues for hardware considerations are data processing, event reconstruction, data storage & access, data simulation, and desk top computing (analysis, administration, support). UNIX "Farms" which provide large CPU cycles have been cost-effective, parallel-processing loosely-coupled computing systems for Fermi Lab. The interactive and batch computing facilities are a mix of IBM, SUN, SGI, DEC systems, similar to other large computing centers. The extremely important high performance networks, with gigabit switches, provide 10/100 Mb/sec to every desktop. The Mass Storage Systems consist of 2 STK Silos (each can store up to 300 TB, an IBM 3494 robot (45 TB), and two new EMASS/GRAU robots with 1 PB capacity. Additional issues include huge Online data acquisition systems, desktops workstations/PCs, Business Systems, World Wide Web, Mail servers, extensive usage of Video Conferencing (including large collaboration meetings) and Remote Control Rooms [7] to monitor the data and the experiment from collaborating institutions/countries.

To fulfill the CPU needs at an affordable price, the ACP systems were developed at Fermilab in '87-'89 for CDF Level 3 and Offline, and also for other experiments. These systems were based on commercial (Motorola 68020) CPUs, and hundreds of custom boards. The ACPMAPS system was developed for more tightly coupled lattice QCD calculations (http://hppc.fnal.gov/acpmaps/acpmaps.html). The system had 5 Gflops with 256 Weitek XL-8032 processors in '88, and 50 Gflops since '92 with 612 i860 chips based on dual CPU boards.

UNIX workstation farms, have provided the main source of CPU at Fermilab and other labs since '91. The use of commercial UNIX workstations instead of the custom board based systems such as the ACP provided the ability to upgrade the systems easily. To satisfy the large increase in CPU requirements in the future, however, a change in hardware platform is necessary.

4.2 PC Farms

Since the introduction of the Intel Pentium Pro processors in 1995, the high performance and low cost of PCs (relative to RISC processors and UNIX workstations) have attracted many groups to test clusters of PCs for cost-effective parallel and distributed computing [2]. By April '97, a dozen groups from NASA, DESY, Fermilab, Sandia, NIH, other labs and universities reported [3] their performance benchmarks for various applications, some with plans to build clusters with thousands of processors.

Fermilab pioneered 10 years ago such parallel computing "Farms" as an alternative to the costly mainframe computers. The Fermilab UNIX Farms at their peak had a total of about 500 UNIX workstations. For such parallel computing on loosely-coupled processors, Fermilab developed CPS (Cooperative Processes Software). Other inter-processor communication software such as PVM and MPI have been developed, supported, and widely used by other labs.

With the severe CPU needs in the future, migration from UNIX workstations to

PCs, with their high performance and low price, is the most cost-effective solution. The advances in CPU chips are accompanied by similar great advances in commodity network and memory. The PC Farms and general PC computing also importantly provide an upgrade path.

4.3 Non-HEP "Beowulf" Projects

Among high energy physics and DOE labs, a loosely coupled computer system is known as a "Farm" Following the UNIX farms, the PC based systems are "PC Farms". The NASA Beowulf Project started in Summer '94 at CESDIS to produce software for off-the-shelf clustered commodity PC-class hardware, using a high-bandwidth low-latency network and open source Linux operating systems, and message-passing. An important contribution to the PC based parallel computing is the Linux device drivers for Ethernet by Donald Becker at CESDIS. There are now dozens of Beowulf projects/systems, and Pentium-Pro Cluster "Cookbooks". Software tools and manuals for clustered computing, both collected and developed by the NASA Beowulf Project, have been packaged as Extreme Linux (www.extremelinux.org, and www.redhat.com/extreme), available on a $29.95 CD-ROM. A person can have PCs for a "Personal Super Computer".

4.4 Linux Operating System

Linux [4] is the Operating System of choice for both the non-HEP Beowulf projects and the high energy physics PC Farms. Linux is "generic UNIX", and enables easy, quick migration [5],[6] from UNIX workstations to PC computing. A useful reference is Http://hepwww.ph.qmw.ac. uk/HEPpc/.

Linux is an open source, well supported, widely used Operating System with multitasking capability. Originally written by Linus Torvalds, Linux (and other Open Source software) has been developed by thousands of volunteer experts world-wide, and is available on the Internet. Most of the Linux system infrastructure (libraries, compilers, utilities) outside of the kernel comes from the Free Software Foundation (GNU).

Many of the software packages that physicists use as part of everyday work are available for Linux. These include TEX/LATEX, editors (Nedit, EMACS), TCP/IP, X11R6, Netscape, HTTP, FTP, Telnet, SLIP, CERNLIB, etc. Linux runs on many platforms, including Intel 386/486/Pentium/P-Pro/P-II, DEC Alpha, SPARC/Fujitsu(AP+) and MIPS R4x00.

In January '98, Fermi Lab's Computing Divison announced official support for Linux (www.fnal.gov/cd/unix/linux). Linux is officially supported by the leading companies providing database software, Oracle, Informix, Computer Associates, Sybase, Objectivity, and other companies such as Netscape, Corel, Interbase, Adaptec, Cygnus, DEC, Sun, SGI, and IBM. In September '98, Intel, Netscape, acquired equity positions in the Red

Hat Linux distribution company. The strong support from such leading companies greatly strengthens future development of Linux.

4.5 Price/Performance

The main reason for PC computing is its high performance at low price. By Summer '98, the price was less than $7 per MIPS (CPU equivalent to a VAX-11/780). The Price/Performance is decreasing at the rate of about 1% per week. By the year 2000, CDF & D0 will require CPU equivalent to 285 - 500 dual P-II 400 MHz PCs for Offline Production, and similar CPU for Online Level 3 event selection.

A dual processor PC running Linux, which has Symmetric Multi-Processing capability, is twice as fast as single processor PC. By February '97, the dual Pentium-Pro PC had the same performance, and more than a factor of 3 better in price, than a SGI R10000 workstation.

In CHEP '97 and the Pentium-Pro Cluster Workshop, the PCs had P-Pro 200 MHz and DRAM (66 MHz) at $4 / MB (vs. $40 / MB for UNIX Workstations). In CHEP '98, we see clusters with 400 MHz P-II and SDRAM (100 MHz) at $1.5 / MB. The price/performance came down by a factor of 2 between April '97 and Sept. '98. The doubling in speed every 12-18 months is expected to continue.

For Online "Level 3" event selection, CDF has demonstrated and adopted the solution of using ATM switch distributing data to (16) Linux PC Farms, with a total output rate of 240 MB/s (60 MB/s is required by the experiment). The open source nature of Linux has been particularly useful in testing, understanding and improving the Level 3 PC Farm.

Several local "white box" companies have participated in the Fermilab purchasing/bidding process for the PC Farms, with Fermilab specified hardware components such as the processor and motherboard. The PCs were delivered in 3 weeks, preinstalled with Fermi Linux, and other CERN, Fermilab, HEP software. The PCs were put together, including network, as a PC Farm within a few days of delivery.

In addition to the PC Farms, desktop computing is also migrating from UNIX workstations to PCs. Many related issues are being solved, e.g. clustering, batch system, I/O, even via wireless communication. Fermilab users can also order PCs using Fermi Linux web site. Vendor support is available (e.g. via contract with Red Hat). The software distributions are by free CD or download via network. Physicists will soon be able to easily analyze data using his/her Notebook PC. The development of PC Farms and PC computing has been important and holds promise for a great future.

5 Collaboration

To facilitate large collaborative efforts, the high energy physics community helps to develop and take advantage of the recent developments in high speed network, the World Wide Web and Internet, and computing. To help communication among collaborations Fermi Lab, for example, has 7 video conferencing rooms, each equipped with both packet (MBONE) and circuit switched systems, over the Internet and common carrier networks. To enable monitoring of the data being collected and the operation of the detector by collaborators and experts world-wide, CDF and other experiments will increasingly make use of Remote Control Rooms world-wide, e.g. at remote collaborating institutions. One tool that is being developed is *e*-Notebook to enable collaborators world-wide to record and share notes/information. To share and analyze the large volume of data (Petabytes) we also further develop the use of distributed computing. These issues for collaboration among laboratories are being studied and embodied into the emerging "collaboratories" [8]. To achieve these goals for collaborative work, the high energy physics community world-wide have had leading efforts in network [9] among collaborating laboratories and universities.

6 summary

Large scientific projects in the world now require very large scale international collaborations. In high energy physics, large scale international collaborations have been very successful. Each experiment may require hundreds of scientists from many countries, to build large, technically challenging detectors, and to efficiently collect and analyze petabytes of data. High energy physics and the experiments world-wide in the next 20 years will continue to be extremely important, exciting and challenging. Some of the tools have been developed for working together, including the use of video conferencing, remote control rooms, high performance network, open source software, and cost-effective computing resources. Great advances are expected for the future in the fundamental physics, our understanding of Nature, the technologies associated with high energy physics, and the tools for successfully working together in large collaborations.

References

[1] Fermi Lab (www.fnal.gov), CERN (www.cern.ch), KEK (www.kek.jp), DESY (www.desy.de).

[2] G.P. Yeh, Fermilab-Conf-98/361; Computing in High Energy Physics, CHEP '98

[3] Computing in High Energy Physics '95, (www.hep.net/conferences/chep95/). Super Computing '96, Pittsburgh (www.supercomp.org/sc97/)

Computing in High Energy Physics CHEP '97, Berlin (www.ifh.de/CHEP97/).
Pentium Pro Cluster Workshop, April '97 (www.scl.ameslab.gov/workshops/)
Super Computing '97, San Jose (www.supercomp.org/sc97/)

[4] www.linux.org and www.redhat.com

[5] M. Welsh, "Porting Applications to Linux", 1st International Dutch Linux Symp., 1994 (http://www.cl.cam.ac.uk/~mdw24/porting-linux/paper.tex).

[6] S. Goldt *et al.*, "The Linux Programmer's Guide", Linux Documentation Project (ftp://sunsite.unc.edu/pub/Linux/docs/linux-doc-project/programmers-guide)

[7] F. Abe, *et al.*, http://www.hep.net/conferences/chep95/html/abstract/abs_135.htm

[8] S. Loken, http://www.hep.net/chep98/

[9] Y. Karita, http://www.hep.net/chep98/