Journal of the Korean Physical Society, Vol. 55, No. 5, November 2009, pp. 2045∼2050

Heavy Flavor Physics Through e-Science

Kihyeon Cho and Hyunwoo Kim*
High Energy Physics Team, Korea Institute of Science and Technology Information, Daejeon 305-806

(Received 29 September 2008, in final form 20 February 2009)

Heavy flavor physics is an important element in understanding the nature of particle physics. Accurate knowledge of the properties of heavy flavor physics plays an essential role for the determination of the Cabibbo-Kobayashi-Maskawa (CKM) matrix. An asymmetric-energy e^+e^- B factory KEKB and its experiment Belle is under way. It will be upgraded to Super Belle in a few years, and the size of the B meson sample will be dramatically increased. Also, the data size of Tevatron experiments (CDF, D0) are on the order of Peta Byte (PB). Therefore, we need to use a new concept of e-Science in this area. This concept implies studying a huge size of data from heavy flavor physics anytime and anywhere even if we are not on-site in accelerator laboratories. The components of this concept include data production, data processing, and data analysis. We apply this concept to the current CDF experiment at Tevatron. We will expand this concept to Super Belle and large hadron collider (LHC) experiments, which will achieve an accuracy of measurements in the next decades.

PACS numbers: 29.85+c, 07.05.Bx, 14.40.Nd
Keywords: B physics, Heavy flavor physics, e-Science, Data process, Grid
DOI: 10.3938/jkps.55.2045

I. INTRODUCTION

Heavy flavor physics plays an important role for understanding the nature of particle physics [1,2]. There are three known generations of quark doublets, (u,d), (c,s), and (t,b). However, the origin of generations is unknown even in the Standard Model. Only charged current electroweak interactions can change flavors in the Standard Model. Electroweak eigenstates are not mass eigenstates, which introduced the Cabibbo-Kobayashi-Maskawa (CKM) matrix [3]. There are only the Standard Model connections between generations. CKM matrix elements are fundamental parameters of the Standard Model. They cannot be predicted theoretically but can be measured by experiments. The CKM matrix contains, in its simplest form, three rotation parameters and one phase parameter. Even if the non-zero phase parameter implies charge-conjugation-parity (CP) violation in flavor transitions, its origin is still one of the mysteries in elementary particle physics. The CP violation can manifest itself when multiple B decay amplitudes interfere with each other. B meson systems are known to be a good place for CP violation studies due to mixing, long lifetime, large production cross section, rich decay channels and heavy quark masses, the precision measurements of which provide confirmation of the CKM theory.

However, the Standard Model leaves many questions about the flavor sector unanswered, such as the origin of generation and masses, flavor mixing and disappearance of antimatter. The core properties of weak interactions provide parameters not predicted within the Standard Model. The Standard Model measurement provides the foundation of the theory while a deviation may give a hint on new physics beyond the Standard Model. The unitary triangle represents a graphical expression of unitary conditions. CKM unitary violation would imply new physics. We may test the Standard Model and the CKM theory by over-constraining angles and sides. Quantum effects beyond the Standard Model have very small probability as can be seen in the $b \rightarrow s$ transition, which means that any new physics effects will be revealed only by a sufficiently huge amount of B meson data.

Therefore, in order to handle more data and more collaborators efficiently, it is time to study heavy flavor physics through the use of e-Science. We apply this concept to the CDF (Collider Detector at Fermilab) experiment and show possible applications for future large hadron collider (LHC) and Super Belle experiments.

II. HEAVY FLAVOR PHYSICS EXPERIMENTS

1. Overview

Current heavy flavor physics experiments in lepton colliders are BESII, CLEOc, Belle, and BaBar. These ex-

---

*E-mail: hyunwoo@kisti.re.kr; Fax: +82-42-869-0789 periments can achieve excellent photon resolutions due to good calorimeters [4] and can exploit energy and momentum conservation to deduce presice information on missing energy. Recently, Belle reports a difference between direct CP violation in charged and neutral B meson decays [5].

In the meantime, heavy flavor physics experiments in hadron colliders produce large data sets and heavier B hadrons. The cross section for proton and antiproton is given by $\sigma(p\bar{p} \to b\bar{b}) \sim 150\ \mu b$ at 2 TeV while $\sigma(e^+e^- \to b\bar{b}) = 1nb$ at Υ(4S). The Tevatron also produces heavy hadrons such as Λ_b and B_s. Therefore, we call the Tevatron a full service B factory. The results from the lepton collider and the hadron collider are complementary. At the Tevatron, B hadrons are mostly produced in pairs. The main $b\bar{b}$ production mechanism is flavor creation through gluon fusion. Due to the excellent performance of the Tevatron, the CDF experiment had recorded 4.0 fb^{-1} data by September 2008. However, the Tevatron $b\bar{b}$ cross section is orders of magnitude smaller than the total inelastic cross section of around 50 mb. For this reason, the CDF experiment employs triggers that select events with signatures specific to various B decays. We use the dimuon trigger geared towards $B \to J/\psi X$ decays and the displaced track trigger used for measurement of B_s lifetime [6]. The heavy flavor physics results also show precision measurements of mass and lifetime for B_s and B_c mesons and of CP violation in the B_s system. The highlighted heavy flavor physics results from the CDF experiment are 1) observation of B_s with $\Delta m_s = 17.77 \pm 0.10$ (stat.) $\pm$ 0.07(sys.) ps^{-1} [7], 2) observation of new baryon states Σ_b [8] and Ξ_b [9], 3) single top quark observation (4.4 sigma) using 2.7 fb^{-1} data that the cross section of single top quark is 2.2 pb [10], 4) measurement of $\sin(2\beta_s)$ [11], 5) precision measurement of the top mass, $m_t = 172.4 \pm 1.0 \pm 1.3$ GeV/c^2 [12], 6) observation of new charmless $b \to hh$ states [13], and 7) evidence of $D^0 - \bar{D}^0$ mixing [14]. CDF has also presented recent results on B_c and B_s meson properties. The best B_c mass measurement was reported by CDF in fully reconstructed $B_c \to J/\psi\pi$ decays [15]. The measurements of the B_c lifetime in semileptonic decays agree with the measurements from the D0 experiment with a similar precision [16]. These measurements provide useful information on the unitarity angles that can be used to improve theoretical models for studying heavy mesons. Figure 1 shows the current measurement of the unitarity triangle by both lepton collider experiments (BESII, CLEOc, BaBar and Belle) and hadron collider experiments (CDF and D0) [17].

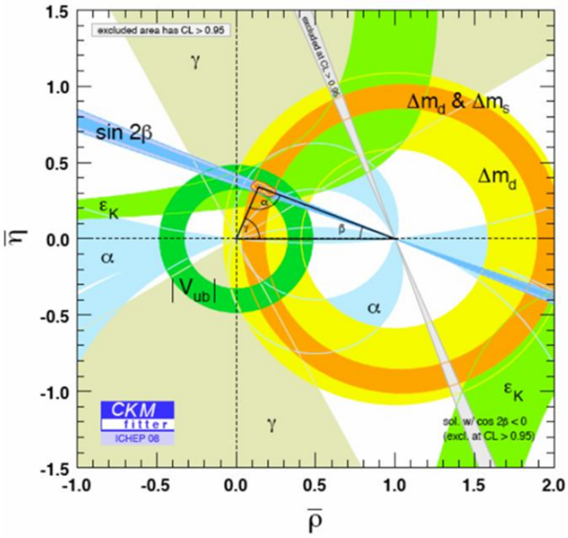

Fig. 1. Current measurement of the unitarity triangle by both lepton collider experiments (BESII, CLEOc, BaBar, and Belle) and hadron collider experiments (CDF and D0) [8].

2. Heavy Flavor Physics Experiments in the Next Decades

High energy physics will focus on three topics in the next decades. The first topic is energy frontier physics by experiments at the LHC and the International Linear Collider (ILC) to search for the Higgs particle, supersymmetry, dark matter, and a new understanding of space-time. The second topic is lepton physics by neutrino experiments to study neutrino lepton flavor violation, tau lepton flavor violation to find neutrino mixing and masses, and lepton number non-conservation. The third topic is heavy flavor physics by super B factory experiments to study CP asymmetry, baryogenesis, left-right symmetry, and new phenomena.

Heavy flavor experiments at LHC - CMS (Compact Muon Solenoid), ATLAS (A Toroid LHC ApparatuS) and LHCb - are in the energy frontier. For heavy flavor physics at CMS and ATLAS, some benchmark analyses are 1) cross sections for bottom, charm, and quarkonia, 2) correlation studies, 3) quarkonia analysis: polarization, production mechanism, 4) lifetime and properties of B hadrons, B^+, B_d, B_s, B_c, and Λ_b, 5) B_s oscillations and CP violation, 6) flavor-changing neutral current (FCNC) rare decays, $B_d \to K^{*0}\mu^+\mu^-$, $B_s \to \phi\mu^+\mu^-$ and $B^+ \to K^+\mu^+\mu^-$, 7) extremely rare FCNC decays, $B_s \to \mu^+\mu^-$ or $B_d \to \mu^+\mu^-$, and 8) lepton flavor violation (LFV), such as $\tau \to 3\mu$ [18]. For these analyses, triggers in CMS and ATLAS utilize the excellent muon system, trackers, and large acceptance. In CMS and ATLAS experiments, most analyses are focused on muons [19].

For heavy flavor physics at the LHCb experiment [20, 21], the luminosity is $2 \sim 5 \times 10^{32}$ cm^{-2}s^{-1}, and the cross section for $b\bar{b}$ is 500 μb. This produces many events up to 2 fb^{-1} for the first year. The LCHb experiment is a heavy flavor precision experiment searching for new physics in CP violation and rare decays. The LHCb detector is a unique forward detector as opposed to other experiments using barrel detectors. The LHCb experiment is designed to study CP violation in the b-quark sector at the LHC and to expand the current studies underway at the B factories (BaBar, Belle) and at the Tevatron (CDF, D0). The LHCb opens an opportunity to study B hadrons that cannot be produced at current B factories, and the energy of 14 TeV, much higher than that of the Tevatron, allows an abundant production of B particles (10^5 particles/s at the nominal luminosity) [22]. The $b\bar{b}$ production cross section is 2 orders of magnitude smaller than the total cross section visible in the detector, and the decay modes of the B hadrons that are interesting for CP violation studies have very low visible branching fractions, typically smaller than 10^{-4}. Hence, a very selective and sophisticated trigger is needed. LHCb is planning to operate a 3-level trigger system to select events of interest [22].

The heavy flavor physics program of Super Belle is expected to start in 2012. The integrated luminosity will be 70 ab^{-1} between 2012 and 2020. The two key measurements of the Super Belle experiment are non-Standard Model CP violation from B meson decays and LFV in tau decays. In detail, the experiment will provide 1) the non-Standard Model CP phase from high precision $b \to s$ penguin studies, 2) charged Higgs Bosons from searches in $B^+ \to \tau^+ v$ and $B \to D^{(*)}\tau^+ v$, 3) the non-SM right-handed current from $B \to K^*\gamma$ CP violation, 4) Inclusive measurements from $b \to s\gamma$, $b \to d\gamma$, $b \to sl^+l^-$, and V$_{ub}$, 5) understanding of loop vs. tree diagram from high precision unitarity triangle measurements, 6) LFV from searches in high statistics tau decays and 7) new physics searches in the up-quark sector from CP violation in $D^0 - \bar{D}^0$ mixing [23]. From the Super Belle experiment, we may find new particles, such as four-quark states, an endless list of rare B decays, B_S at Y(5S), and more D decays. For the physics states, we need to focus more on the early stage of the upgraded B factory at KEK to maximize physics output with 3 or 5 ab^{-1} data. The 10 ab^{-1} data will give us the direction of flavor physics which is the goal of current roadmap. The 50 ab^{-1} data will allow us to study flavor physics beyond the Standard Model towards the limits of systematics and theory expectation [23].

Now that LHC has turned on, the LHC experiments of high transverse momentum (p$_T$) will provide a unique effort towards the high energy frontier to determine the energy scale of the new physics. Meanwhile, collective efforts toward the high intensity frontier will provide flavor physics by rare K decays, universality tests in B and K, CP violation in the B_S system, improved CKM fits, rare B decays, LFV in muon and tau decays and g-2 to determine the flavor structure of new physics [24].

3. The Trend of Heavy Flavor Physics Experiments

Since the invention of the cyclotron by Ernest Orlando Lawrence at the University of California, Berkeley, in 1930 [25], accelerators and particle physics experiments have changed substantially. We will explain the recent trend in heavy flavor physics experiments from the perspective of data production, data processing, and data analysis.

First, from the data production point of view, the current center of mass energy is higher than that of the past, ranging from 10.56 GeV of $Y(4S)$ of B factory experiments to 14 TeV of LHC experiments. Second, from data processing point of view, the cross section is also higher than that of the past for both lepton colliders (e^+e^-) and hadron colliders ($p-p$). Table 1 shows the size of production data for heavy flavor experiments. The cross section of LHC is $\sigma(pp \to b\bar{b}) \sim 500\ \mu b$ at $\sqrt{s} =$ 14 TeV while that of B factories is $\sigma(e^+e^- \to b\bar{b}) = 1nb$ at $Y(4S)$. Therefore, the LHC produces 500,000 times more data than the current B factory experiments. The size of the data is on the order of PB per year.

In the meantime, the Belle experiment has measured fundamental parameters using a large sample that corresponds to an integrated luminosity of 0.9 ab^{-1}, or 900 million B anti-B meson pairs, and has consequently verified the framework of the Standard Model since its operation in 1999. Now the Belle collaboration is proposing an upgrade called Super Belle which starts from 2012 with an aim to probe new phenomena from physics beyond the Standard Model by using a much larger amount of data of up to 70 ab^{-1} or equivalently 70 billion B anti-B meson pairs.

Third, from data analysis point of view, we have more collaborators. The CLEO experiment, for instance, needed around 200 physicists whereas now about 2,000 physicists work for the CMS experiment or ALICE experiment. The number of collaborators has dramatically increased. Therefore, in order to cope with more data and more collaborators, as well as higher energy and higher cross section, it is time to adopt the e-Science paradigm for heavy flavor physics.

III. E-SCIENCE

1. The Definition of e-Science

Thousands of years ago, science was experimental to describe natural phenomena. For the last few hundreds of years, science has been theoretical, such as Newton's laws and Maxwell's equations. For the last few decades, science has been computational, focusing on simulation of phenomenological complexities. Today, science can be named e-Science, which is data-centric in nature to unify theory, experiment, and simulation. Contemporary scientists have to face many challenging problems that require large resources, particularly knowledge from many disciplines [26]. e-Science is a new research paradigm for science, which is computationally intensive and is carried

Table 1. The size of production data for heavy flavor physics experiments.

Collider	Current experiment		Next experiment	
	Experiment	Raw data size	Experiment	Raw data size
Lepton Collider (e^+e^-)	Belle(1999-present) BaBar (1999-2008)	~1 PByte (0.9 ab^{-1}) ~0.5PByte(0.5 ab^{-1})	Super Belle (2012-)	5 ~ 10 PByte/year
Hadron Collider $(p-\bar{p}/p-p)$	CDF (2001-present) D0 (2001-present)	~2 PByte (4.0 fb^{-1}) ~2 PByte (4.0 fb^{-1})	CMS (2008-) ATLAS (2008-) LHCb (2008-)	5 ~ 10 PByte/year 5 ~ 10PByte/year 0.2 ~ 1 PByte/year

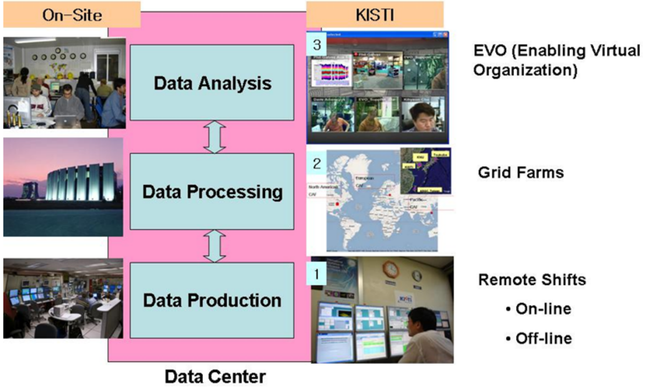

Fig. 2. Typical components of e-Science for high energy physics.

out in highly distributed network environments [27]. e-Science also relies on the grid technology to deal with huge data size. [27].

2. The Goal of e-Science

The goal of e-Science is to conduct scientific research "anytime and anywhere." High energy physics experiments are usually conducted at major accelerator sites (on-site) where experimental scientists carry out detector design and construction, signal processing, data acquisition, and data analysis on a large scale [28]. Therefore, the goal of e-Science for high energy physics is to study high energy physics "anytime and anywhere" even if we are not at accelerator laboratories (on-site). High-energy physics requires a particularly well-developed e-Science infrastructure due to its need for adequate computing facilities for the analysis of results and the storage of data [27]. To perform computing at the required high energy physics scale, we need data grid technology.

3. The Components of e-Science

As shown in Figure 2, the components of e-Science for high energy physics include 1) data production, 2) data processing, and 3) data analysis that can be accessed any time and anywhere even if we are not on-site. First, data production is to take both on-line shifts and off-line shifts anywhere. On-line shifts are taken through the use of a remote control room and off-line shifts are taken via a data handling system. Second, data processing processes data by using a high energy physics data grid. The objective of the high energy physics data grid is to construct a system to manage and process high energy physics data and to support the high energy physics community [29]. Third, data analysis is for collaborations around the world to analyze and publish the results in collaborative environments. We apply this concept to the CDF experiment at Fermilab.

IV. E-SCIENCE IN THE CDF EXPERIMENT

We take the CDF experiment as an application of e-Science to a realistic experiment. Table 2 shows the components of e-Science applied to Fermilab (on-site).

For data production, we have taken remote shifts both on-line and off-line. For the on-line shift, we have constructed a remote control room for the consumer operation (CO) shift at the Korea Institute of Science and Technology Information (KISTI). The CDF control room at Fermilab consists of many sections. One of them is a monitoring section to check the quality of data. CDF calls it CO, which does not affect the control and the data acquisition directly. Everything on the CO desktop at the CDF control room is available at a remote control room. The status and logs of 'Event Display,' 'Consumer Slide,' 'Consumer Monitors,' and 'Detector Check List' can be monitored via web browsers. The 'Consumer Controller' has to be sent to the remote site from the CDF control room [28]. Communications with the CDF on-site shift crews rely on the polycom system. For the off-line shifts, we take SAM (sequential access through meta-data) data handling shifts remotely [30]. The off-line data handling shifts monitor and respond to data transfer requests from ordinary SAM users and the online reconstruction farm in and out of the enstore tape libraries via the CDF d-Cache system.

For data processing, we use the grid technology. From the technical point of view, scalability should be guar-

Table 2. The components of e-Science compared to typical science at Fermilab.

Component	Typical Science	e-Science
Site (Location)	On-Site (Fermilab, USA)	Off-Site (KISTI, Korea)
Data Analysis	Conference Room	EVO (Enabling Virtual Organization)
Data Processing	CAF (Central Analysis Farm)	Pacific CAF (CDF Analysis Farm)
Data Production		
- On-line shift	CDF Control Room at Fermilab	Remote Control Room at KISTI
- Off-line shift	Off-line shift at Fermilab	SAM Data Handling shift at KISTI

Table 3. The progress of the CDF Grid [19].

Name	Starting date	Grid middleware	Job scheduling	Content	Site
CAF(Central Analysis Farm)	2001	-	Condor	Cluster farm inside Fermilab	USA (Fermilab)
DCAF (Decentralized CDF Analysis Farm)	2003	-	Condor	Cluster farm outside Fermilab	Korea, Japan, Taiwan, Spain (Barcelona, Cantabria), USA (Rutgers, San Diego), Canada, France,
Grid CAF (CDF Analysis Farm)	2006	LCG / OSG	Resource Broker + Condor	Grid farm	North America CAF European CAF Pacific CAF
CGCC (CDF Grid Computing Center)	2008	LCG	Resource Broker + Condor	Big Grid farm + Storage	Korea (KISTI) France (IN2P3) Italy (CNAF)

anteed [31]. Grid farms should be scalable to grid resources. The grid farms consist of O(100) sites, O(100k) CPU, O(10PB) disk, O(100k) jobs/day, O(10M) files, O(10Gb/s) transfer, O(10k) users and O(100) VO (virtual organization). From the collaborative point of view, accessibility should be guaranteed. In other words, the farms should be equivalently accessible to all collaborations and must have a global infrastructure for communication. From a political and financial point of view, they should have visibility. They should be recognized by institutes and countries. The grid technology can now be a common solution for future high energy physics programs. The CDF experiment makes good example. The CDF is using several computing processing systems, such as a central analysis farm (CAF) [32], a decentralized CDF analysis farm (DCAF) [29] and grid systems. Table 3 shows the progress of the CDF grid [28]. Based on the grid concept [33], CDF has constructed a CDF VO. Moreover, a significant fraction of these resources is expected to be available to the CDF even in the LHC era. The transition to a grid at the CDF experiment is in synchronization with the worldwide trend for high energy physics experiments. We have made a federation of the LHC computing grid (LCG) and open science grid (OSG) farms at the Academia Sinica in Taiwan, the LCG farm at the KISTI in Korea, and the LCG farm at the University of Tsukuba in Japan. We name this federation of grid farms as the 'Pacific CAF'.

For data analysis, we have configured and run the enabling virtual organization (EVO) servers at the KISTI. When Korean users use the EVO servers at the KISTI, the routing time is reduced by 60 ms without the network congestion inside USA, which gives a very stable research environment [28].

V. CONCLUSION

Heavy flavor physics plays an important role for understanding CP violation and decay mechanisms. Due to more data and more collaboration, as well as higher energy and higher cross section, it is time to study heavy flavor physics through e-Science. We apply this concept to the CDF experiment and show the possible applications for future LHC and Super Belle experiments. For the outlook of e-Science, we have a final dataset of around 500 million B decays from the BaBar experiment. The Belle experiment continues operations to order of

1,000 million decays. We expect to double the data set at the Tevatron. Therefore, high precision measurements are around the corner. We will also have data sets from the LHC experiments and the Super B factories in Japan near future.

ACKNOWLEDGMENTS

We would like to thank Minho Jeung (KISTI) for data processing at the LCG farm.

REFERENCES

[1] Abulencia *et al.*, Phys. Rev. Lett. **96**, 231801 (2006).
[2] K. Cho, Nucl. Phys. B **142** (Proc. Suppl.), 138 (2005).
[3] M. Kobayashi and T. Maskawa, Progress in Theoretical Physics **49**, 652 (1973).
[4] H. Ikeda and B. G. Cheon, J. Korean Phys. Soc. **50**, 1224 (2007).
[5] The Belle Collaboration, Nature **452**, 332 (2008).
[6] G. Giurgiu, *proceedings of the 8th International Conference on Hyperons, Charm and Beauty Hadrons* (Columbia, USA, 2008).
[7] A. Abulencia *et al.*, Phys. Rev. Lett. **97**, 242003 (2006).
[8] T. Aaltonen *et al.*, Phys. Rev. Lett. **99**, 202001 (2007).
[9] T. Aaltonen *et al.*, Phys. Rev. Lett. **99**, 052002 (2007).
[10] T. Junk, *In proceedings of the International Conference on High Energy Physics 08* (Philadelphia, USA, 2008).
[11] T. Aaltonen *et al.*, Phys. Rev. Lett. **100**, 161802 (2008).
[12] P. Mehtala, *proceedings of the International Conference on High Energy Physics 08* (Philadelphia, USA, 2008).
[13] A. Abulencia *et al.*, Phys. Rev. Lett. **97**, 211802 (2006).
[14] T. Aaltonen *et al.*, Phys. Rev. Lett. **100**, 121802 (2008).
[15] T. Aaltonen *et al.*, Phys. Rev. Lett. **100**, 182002 (2008).
[16] C. Liu, *proceedings of Flavor Physics and CP Violation FPCP2008* (Taipei, Taiwan, 2008).
[17] T.Hurth, *proceedings of the International Conference on High Energy Physics 08* (Philadelphia, USA, 2008).
[18] A. Kraan, *proceedings of the 34th International Conference on High Energy Physics 08* (Philadelphia, USA, 2008).
[19] H. C. Kim, R. J. Hu, S. H. Ahn, B. Hong *et al.*, J. Korean Phys. Soc. **52**, 913 (2008).
[20] The LHCb Collaboration, S. Amato *et al.*, Technical Proposal, CERN-LHCC/98-4 (1998).
[21] The LHCb Collaboration, R. Antunes Nobrega *et al.*, LHCb Reoptimized Detector Design and Performance TDR, CERN-LHCC/2003-030 (2003).
[22] R. Antunes Nobrega *et al.*, LHCb Computing Technical Design Report, CERN/LHCC 2005-019 (2005).
[23] M. Nakao, *proceedings of the 2nd open meeting for the proto-collaboration of Super B factory* (Tsukuba, Japan, 2008).
[24] G. Isidori, *proceedings of the 28th International Symposium on Lepton and Photon Interactions at High Energy* (Daegu, Korea, 2007).
[25] W. T. Chu, J. Korean Phys. Soc. **50**, 1385 (2007).
[26] K. Cho, Comp. Phys. Comm. 177, 247 (2007).
[27] The definition of e-Science is from wiki home page: http://en.wikipedia.org/wiki/E-Science.
[28] K. Cho, J. Korean Phys. Soc. **53**, 1187 (2008).
[29] K. Cho, Internat. J. Of Comp. Sci. and Network Sec. 7, 49 (2007).
[30] I. Sffligoi, Comp. Phys. Comm. **177**, 235 (2007).
[31] H. Sakamoto, *proceedings of the 2nd open meeting for the proto-collaboration of Super B factor*y (Tsukuba, Japan, 2008).
[32] M. Neubauer, Nucl. Instrum. Meth. A **502**, 386 (2003).
[33] I. Foster, C. Kesselman, and S. Tuecke, Internat. J. Of High-performance Computing Appl. **15**, 200 (2001).