

고 에너지 물리실험에서 대용량 데이터 처리 및 관리



안 보 영
고려대학교



제 소개...

- 학사 : 고려대학교 (1992년 – 1999년)
- 석사 : 고려대학교 입자물리실험 (1999년 – 2001년)
 - 졸업논문 : “Feasibility Study of the Light Higgs Search at Future Linear Colliders” , at KEK Japan.
- 박사 : University of Rochester (2002년 – 2008년)
 - 졸업논문 : “Measurement of W Boson Production Charge Asymmetry in ppbar collider”
 - Serving as a CDF Level–3 trigger expert.
- Post–Doc : 연재 고려대학교 입자물리실험
 - 3D track trigger development for super–Belle.

오늘 세미나의 내용은

- 연세 활발한 국제 연구활동이 진행되는 **고 에너지 충돌실험** :
 - 미국 **Fermi National Accelerator Laboratory (FNAL)** 에 **Collider Detector at Fermilab (CDF)** 실험.
 - 일본 고 에너지 공동연구소 (**KEK**) 에 **(super)-Belle** 실험.
- 각 실험에서 나온 대용량 데이터 생산과 filtering 시스템
 - **Data Trigger System** 소개.
- 그리고 각 실험에서 filtering 된 데이터의 저장과 관리
 - **Data Acquisition System** 소개

고 에너지 충돌실험은

입자

에너지

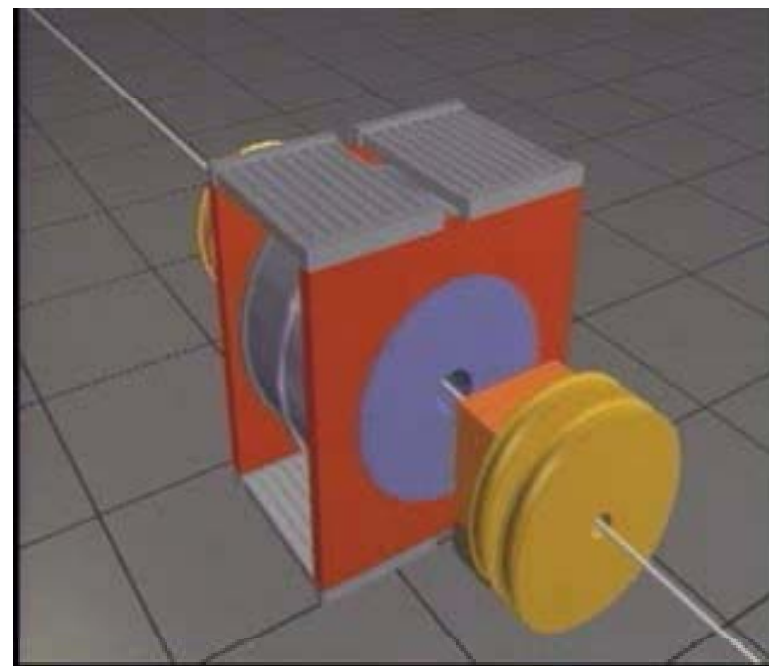
고 에너지 가속기 안



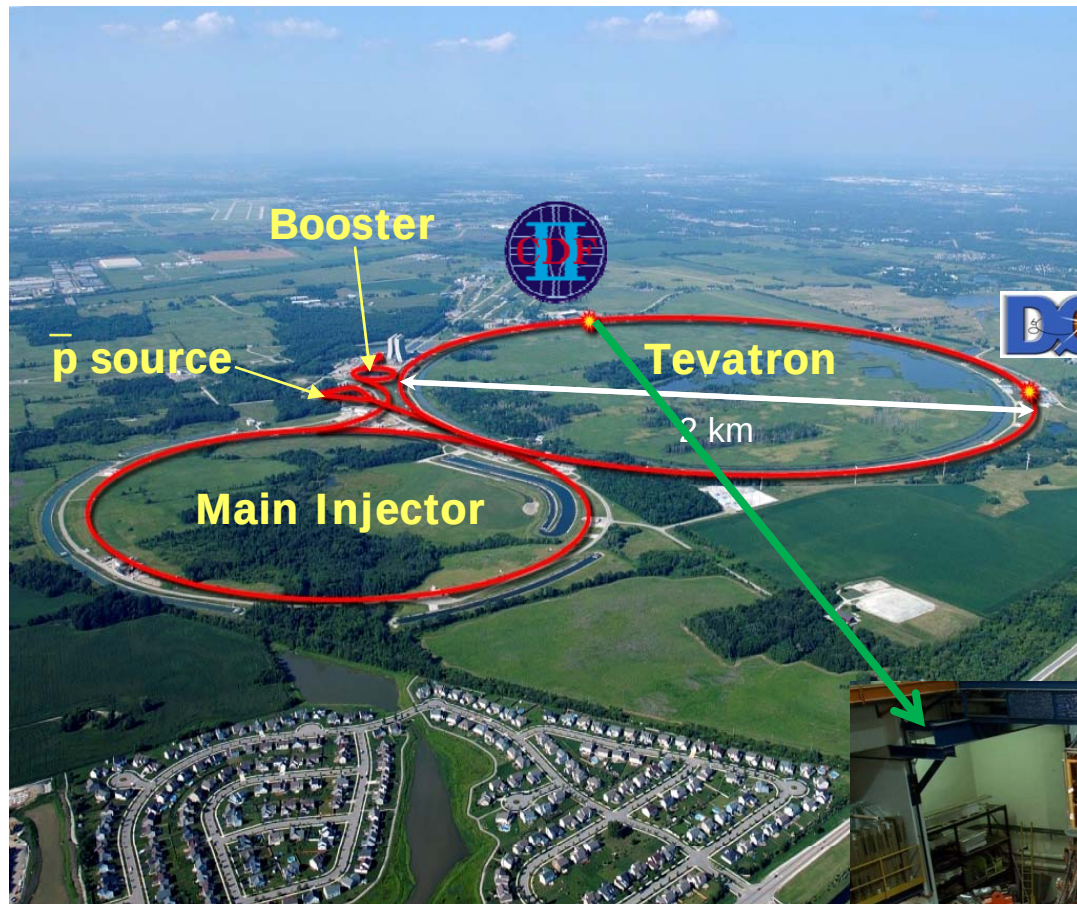
반입자

에너지

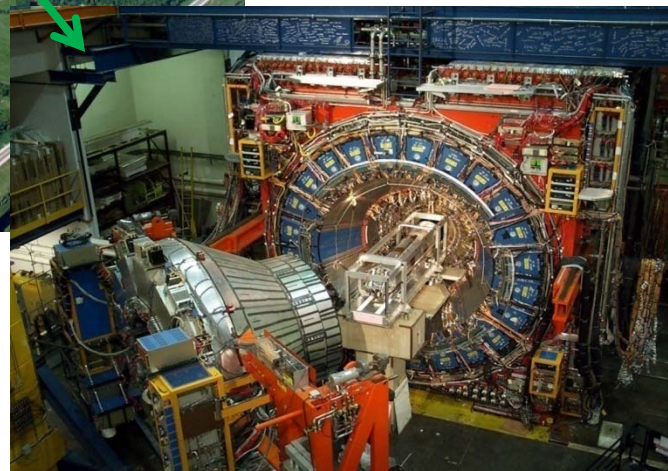
검출기 모형도



FNAL에서 CDF 실험은



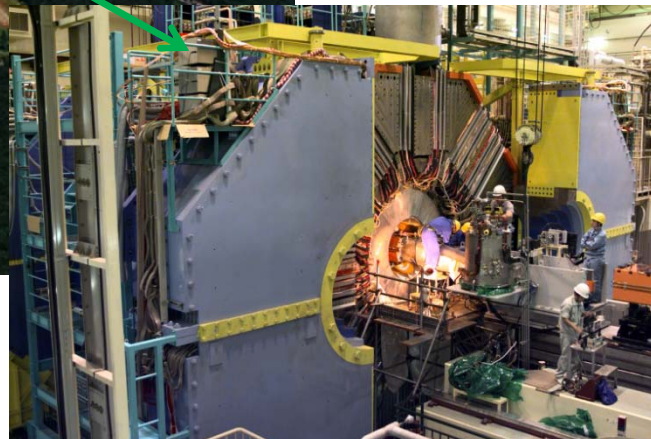
- 양성자와 반 양성자 충돌 :
 $\sqrt{s} = 1.96 \text{ TeV}$.
- 36x36 bunches,
collisions every **396 ns**.
- 연재 데이터 량 :
 4.2 fb^{-1} [$\times 10^{36} \text{ cm}^{-2} \text{ s}^{-1}$]
- 2009(?) shutdown 까지
총 데이터 목표량 :
 $6.5 - 7.0 \text{ fb}^{-1}$



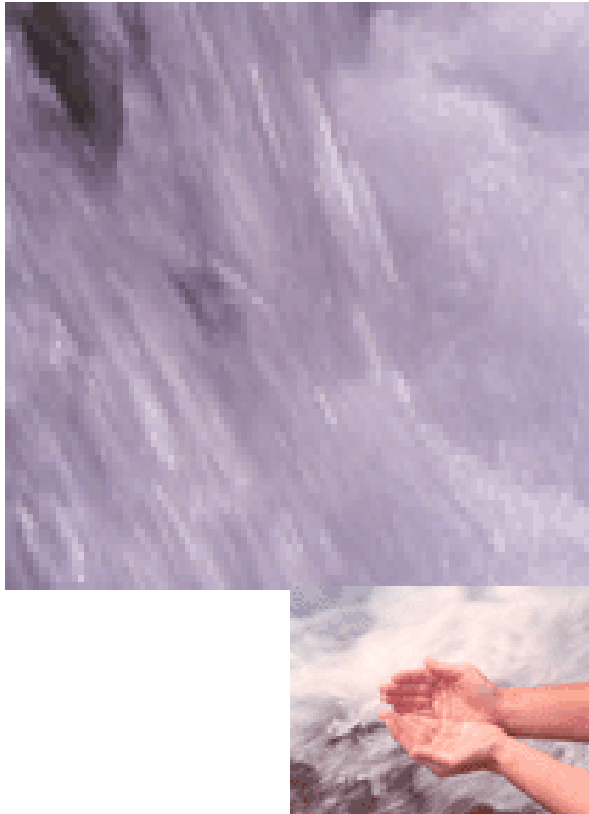
KEK에서 Belle 실험은



- 전자와 양전자 충돌 :
 $\sqrt{s} = 10.58 \text{ GeV}$.
- A bunch crossing rate :
 2.1 ns .
- 연재 데이터 량 :
 862 fb^{-1}
- 2009(?) shutdown 까지
총 데이터 목표량 :
 1.0 ab^{-1}
- 2012(?)부터 superBelle :
 $\sim 50 \text{ ab}^{-1}$



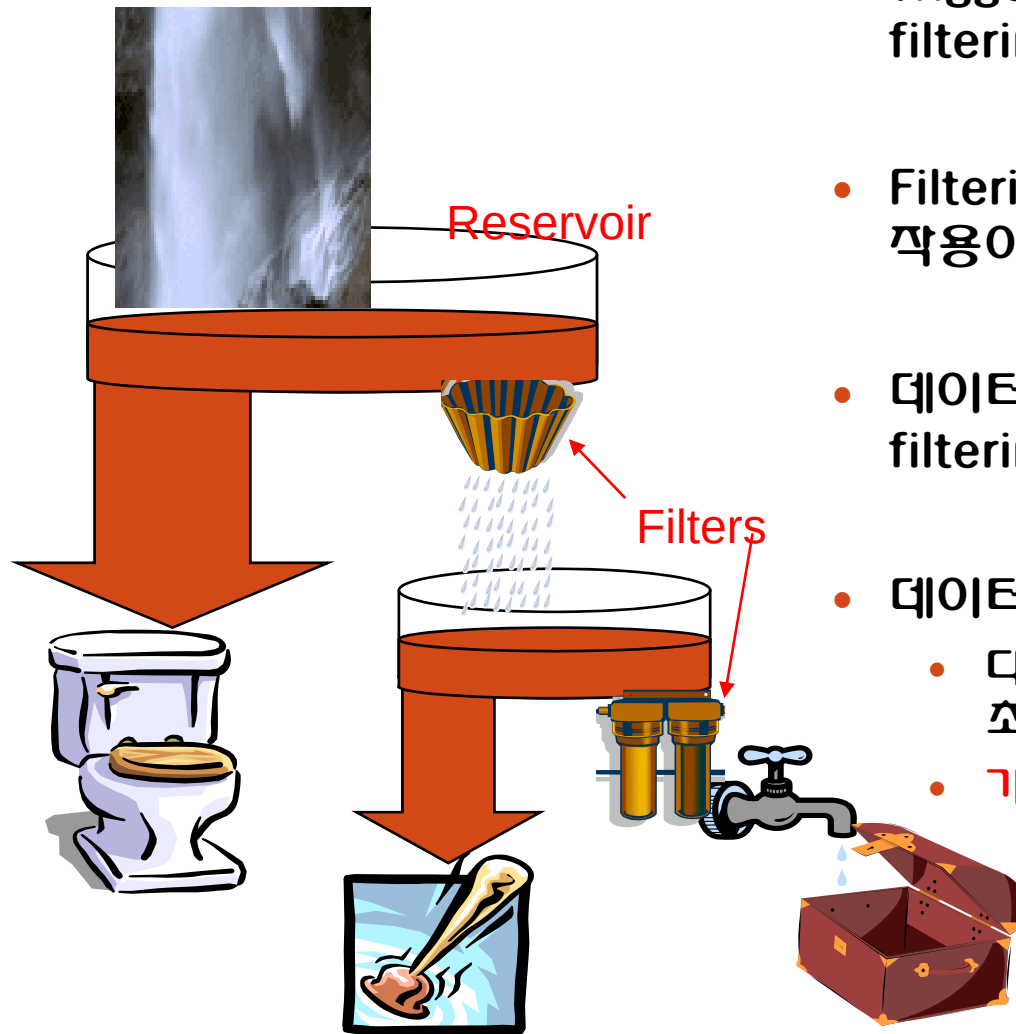
고 에너지 실험 데이터 흐름



- 고 에너지 충돌실험은 연속적으로 많은 양의 데이터를 빠른 시간 안에 생산.
 - 대용량의 데이터는 연구목적을 위해 필요한 데이터만을 골라내는 작업이 필요.
 - Filtering 작업은 빠르고 정확성을 요구함

➔ **Data Trigger System**
 - CDF 와 Belle, 각 실험은 각자의 물리연구 목적에 맞는 Trigger system을 구성함.
 - Filtering 후에 나온 데이터는 각 연구 주체별로 나누어서 Online 상에서 저장 관리함.
- ➔ **Data Acquisition System**

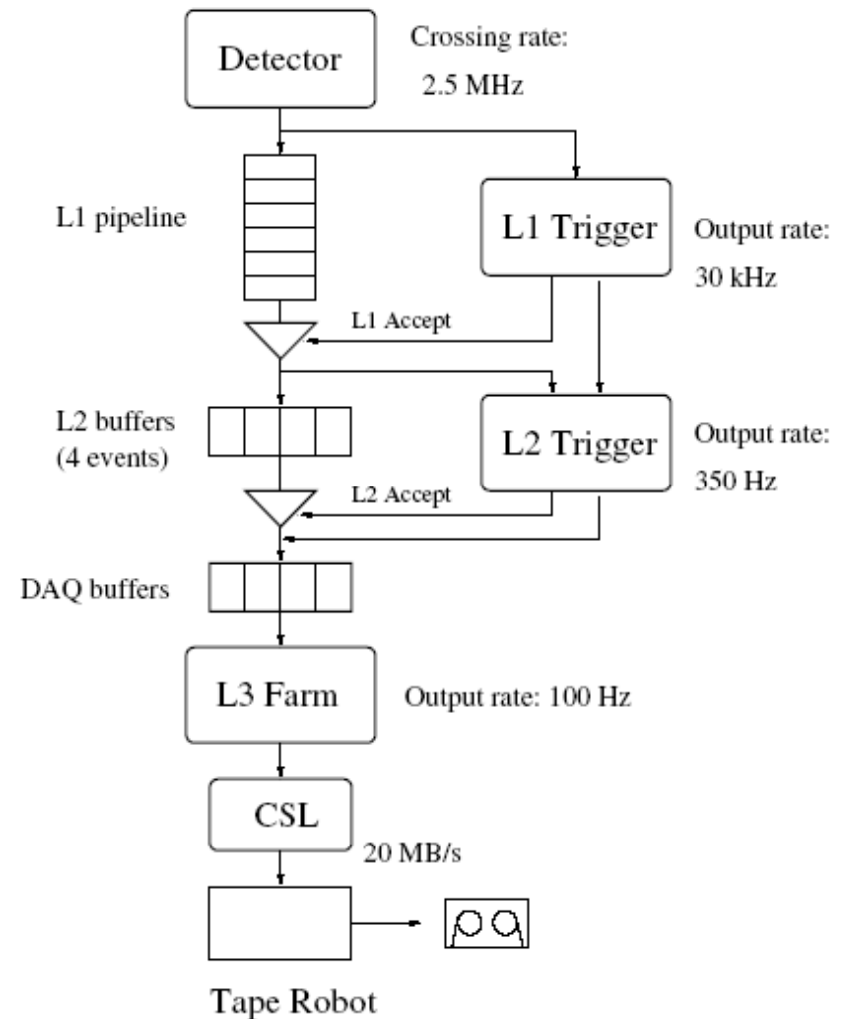
• Trigger 기본 개념 (Buffering and Filtering)



- Trigger System은 단계별(level) filtering으로 구성 : 3단계 (level 1, 2, 3)
- Filtering 동안 충분한 완충(buffer) 작용이 필요. (overflow → deadtime)
- 데이터 볼륨은 다음 단계의 더 정밀한 filtering을 위해 줄여준다.
- 데이터 filtering의 요건 :
 - 다음단계의 가용용량(시간과 볼륨)의 초과하지 않도록 함.
 - 가능한 많은 데이터를 저장하도록 함.

• CDF Trigger System

- CDF 실험은 3개의 trigger level을 가짐.
- Level 1: synchronous hardware trigger
 - 데이터 Input rate = 2.5 MHz .
 - L1 Accept rate = $\sim 30\text{kHz}$ (limited by L2)
- Level 2: a combination of hardware and software trigger (asynchronous)
 - 평균 processing time : $\sim 30\text{ms}$
 - L2 Accept rate = $\sim 350\text{Hz}$ (limited by event-builder)
- Level 3 : software trigger
 - Data reconstruction을 위해 massive PC farm을 이용함
 - L3 Accept rate = $\sim 100\text{Hz}$ (limited by tape writing)
- Data reduction rate $\rightarrow 1 : 25000$



• CDF DAQ System :

Event Builder, L3 PC farm, Consumer–Server/Logger

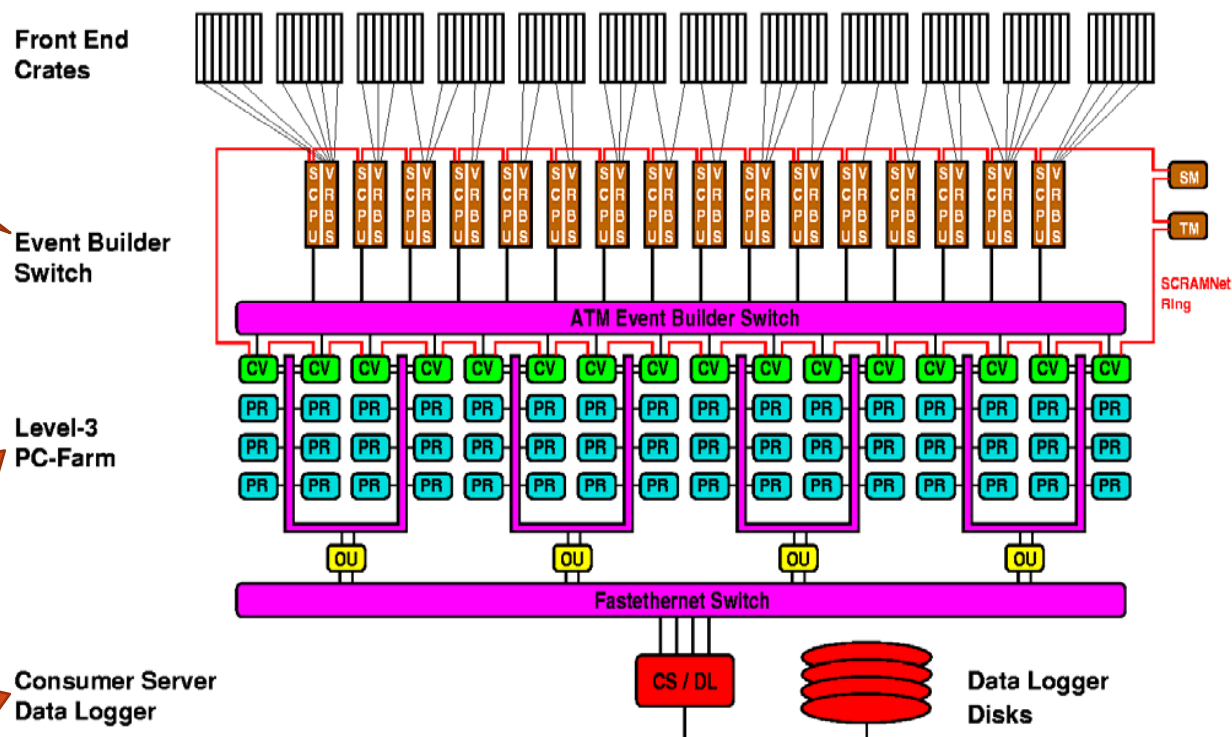
Level 2 read out crate으로 부터 “데이터조각” 를 읽고 level 3 PC farm으로 보냄.

- 18개의 subfarm에 약 300개의 dual processor Linux PC' s

- Event reconstruction and selection based on C++

L3로부터 data를 받아 analysis를 위해 sorting해서 저장

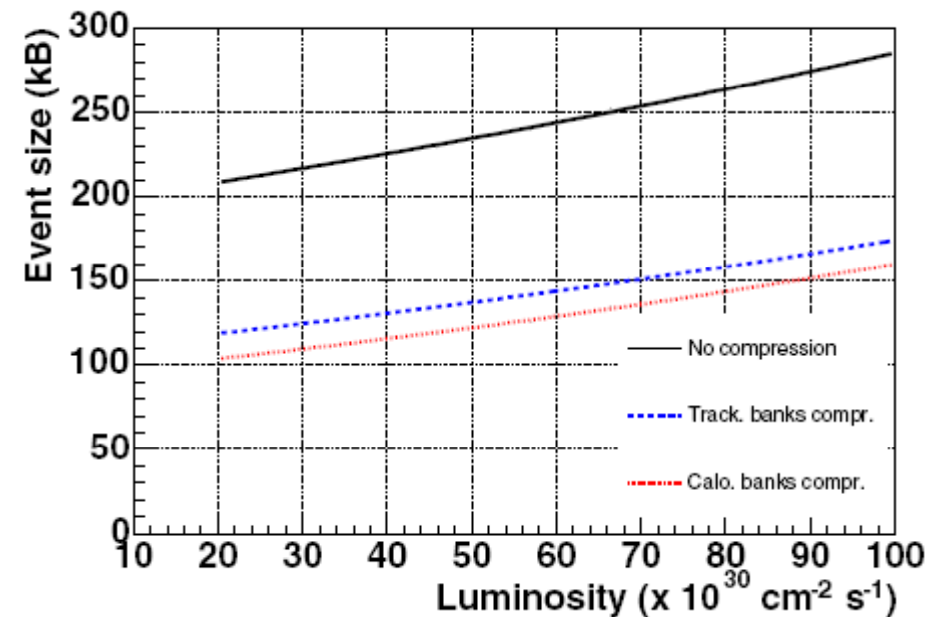
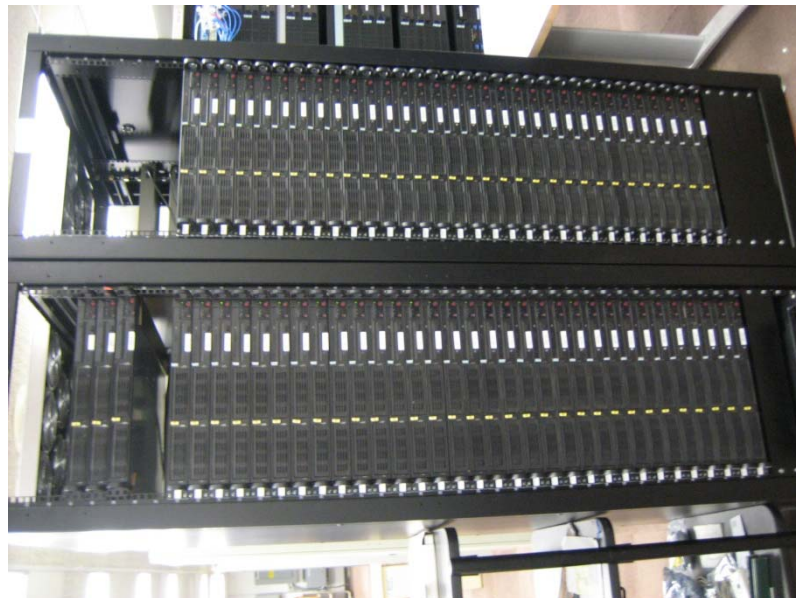
CDF Data Acquisition System: Engineering Run



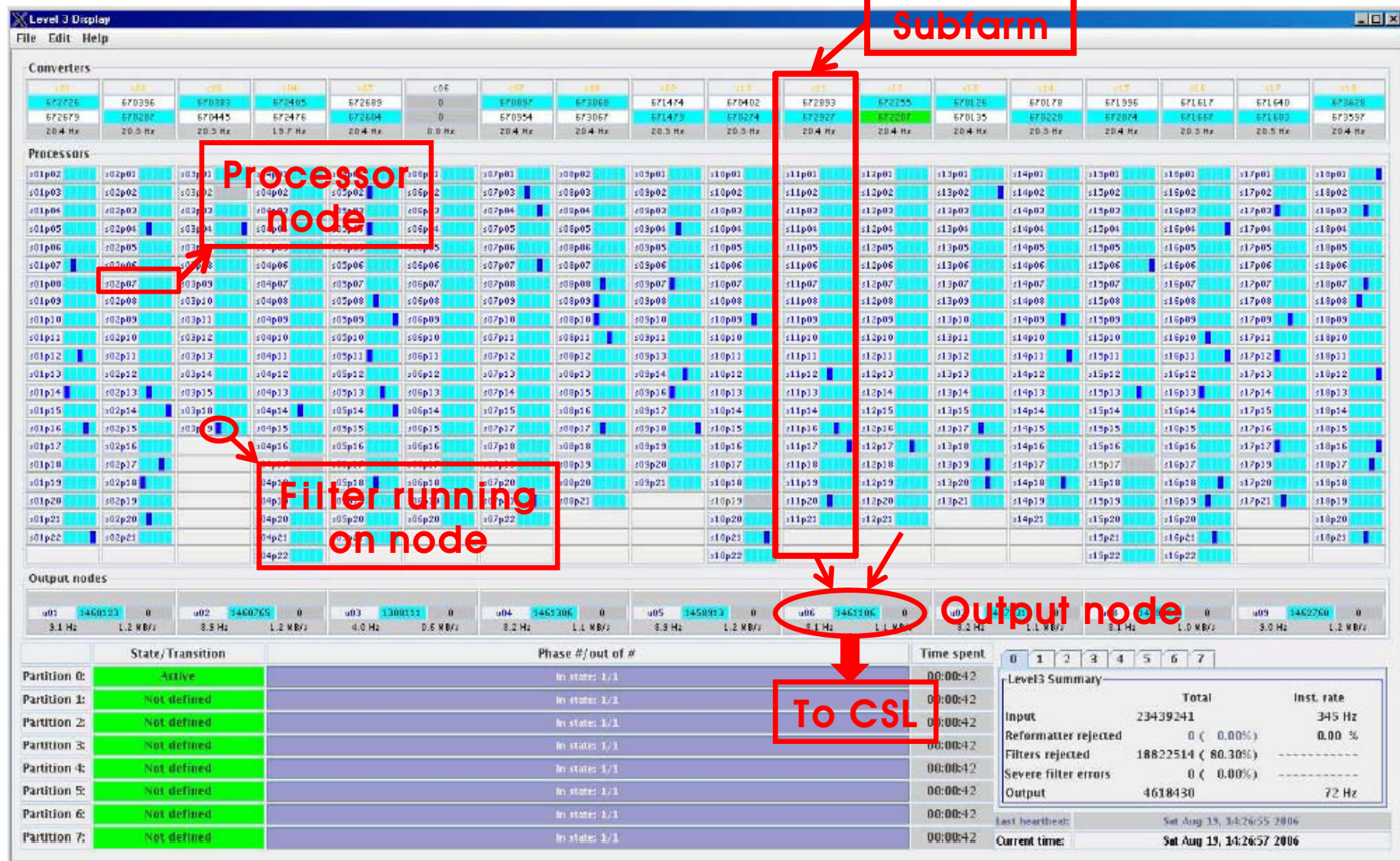
• CDF Level 3 Trigger

The Level-3 Trigger at the CDF Experiment at Tevatron Run II

Y.S. Chung¹, G. De Lentdecker¹, S. Demers¹, B.Y. Han¹, B. Kilminster¹, J. Lee¹, K. McFarland¹,
A. Vaiciulis¹, F. Azfar², T. Huffman², T. Akimoto³, H. Kobayashi³, H. Matsunaga³, M. Shimojima³, K. Tollefson⁴,
D. Torretta⁵, D. Waters⁶



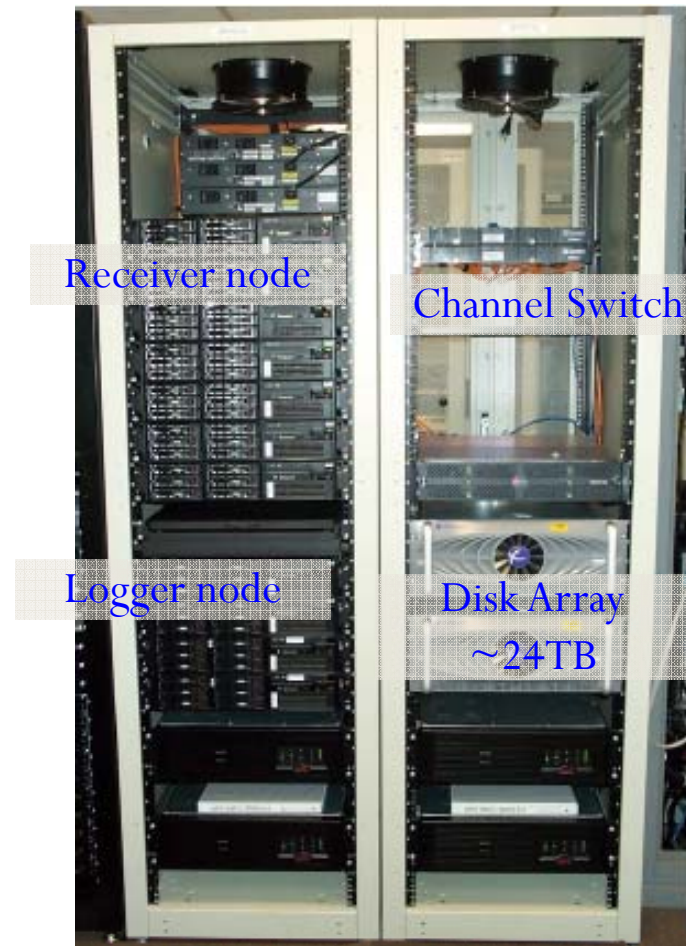
• CDF Level 3 Trigger Display



• CDF **C**onsumer–**S**erver/**L**ogger (CSL)

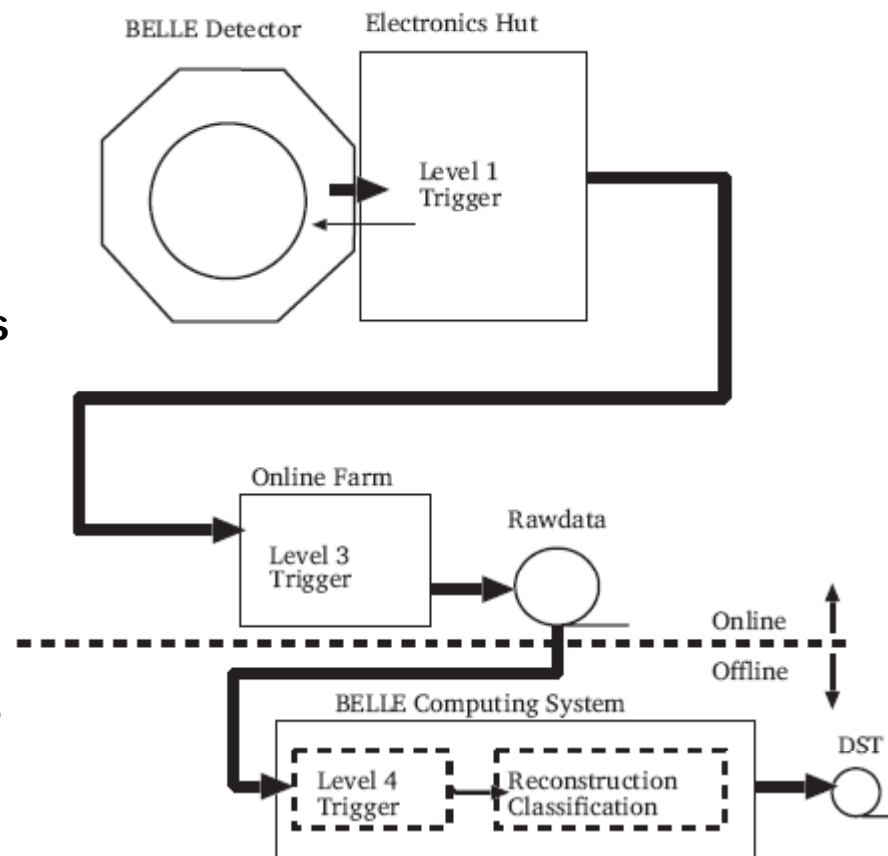
- Consumer–Server/Logger(CSL)은
 - Central online hub
 - Level 3 PC farm 으로 부터 events 받음.
 - Events are buffered onto disk
 - 그 events를 80 MB/s의 rate으로 disk에 기록함.
 - Handles as many consumer requests as possible
 - 그 events를 각 연구 카테고리에 따라 구분함.
 - Data files are moved to the Grid Computing Center and archived onto tape
 - 평균 event size : ~150kB

CSL Hardware



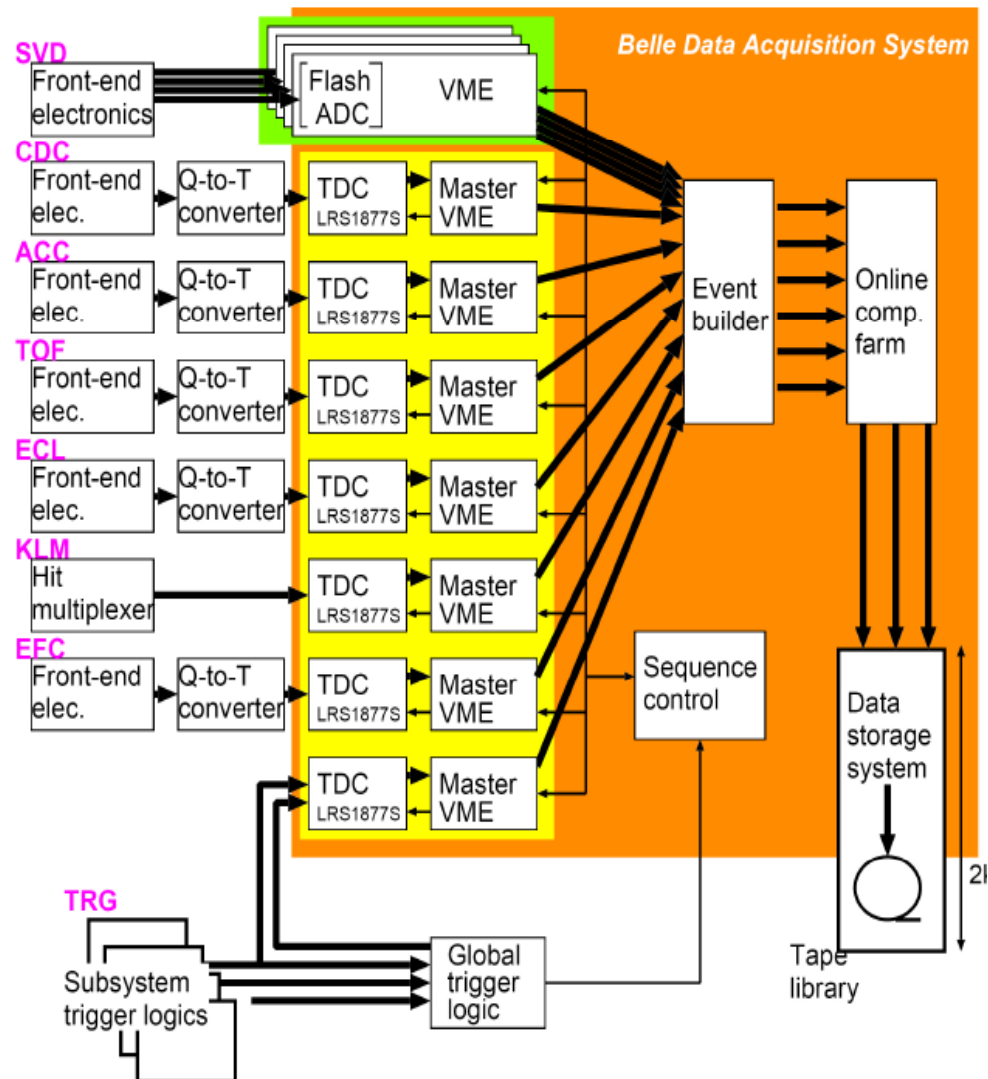
• Belle Trigger system

- Belle 실험은 3개의 trigger level을 가짐.
- Level 1: a hardware trigger
 - A central trigger system : 각 detector 에서 trigger signal을 수집하여 결정.
 - Event의 trigger 결정은 빔 충돌후 $2.2 \mu s$ 안에 이루어짐.
- Level 3 : a software trigger
 - Level 1 조건을 재확인.
 - Data size을 $\frac{1}{2}$ 배 줄이는 역할
- Level 4 : a software trigger in offline
 - a full reconstruction : 모든 detector 정보를 이용해서 실행



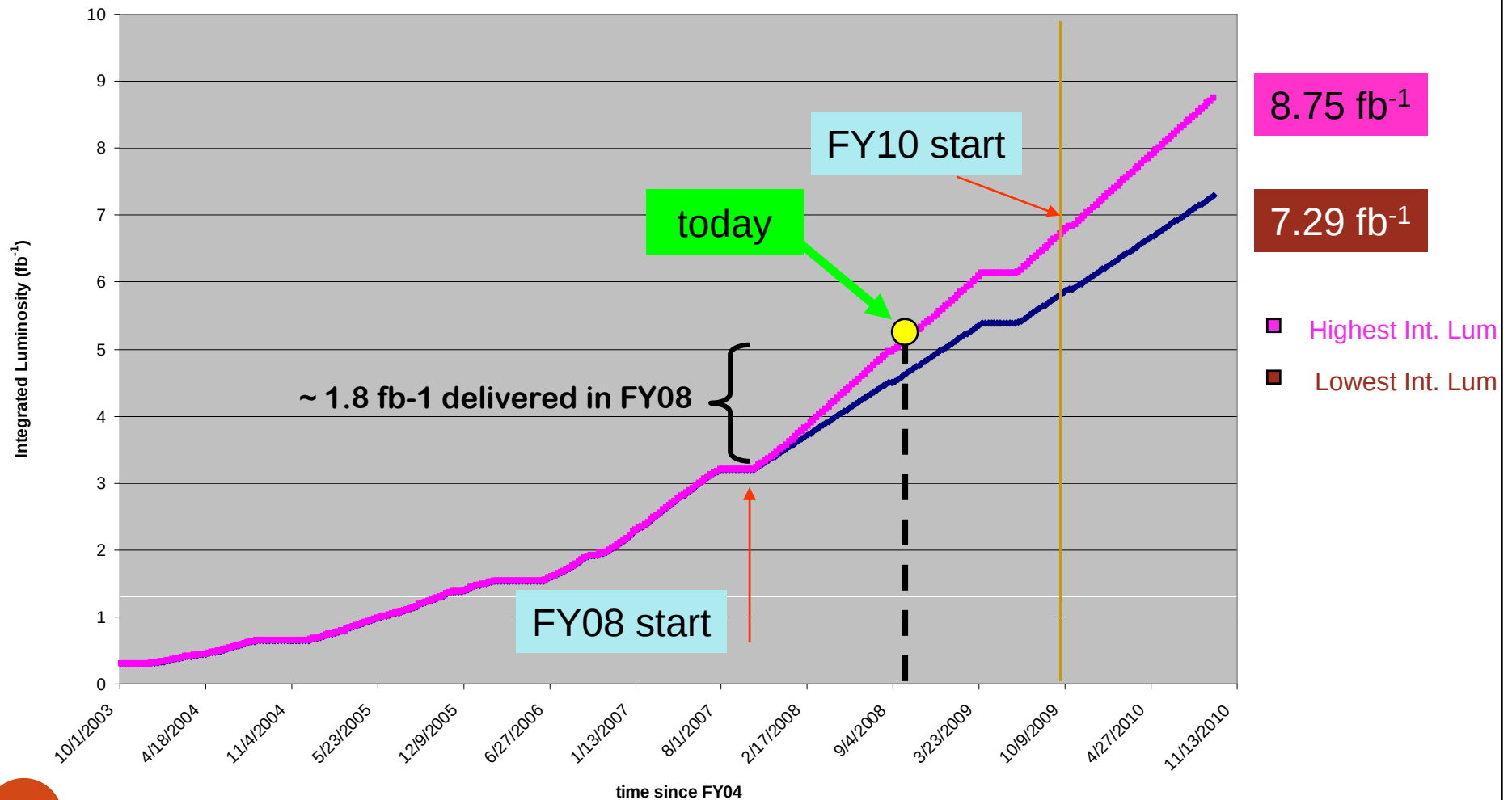
• Belle DAQ system

- Electric Signal 생성
 - Q-to-T : Charge to Time
 - TDC : Time to Digital converter
- Event Builder (CDF와 유사)
 - Electric signal Collection
 - Combined as single event
 - “detector-by-detector” → “event-by-event”
- Level 3 filtering : < 500Hz
- DAQ 시스템을 통과한 data는 약 30kB 크기로 15MB/s rate 으로 기록됨.



앞으로 CDF 실험은...

Projection curves



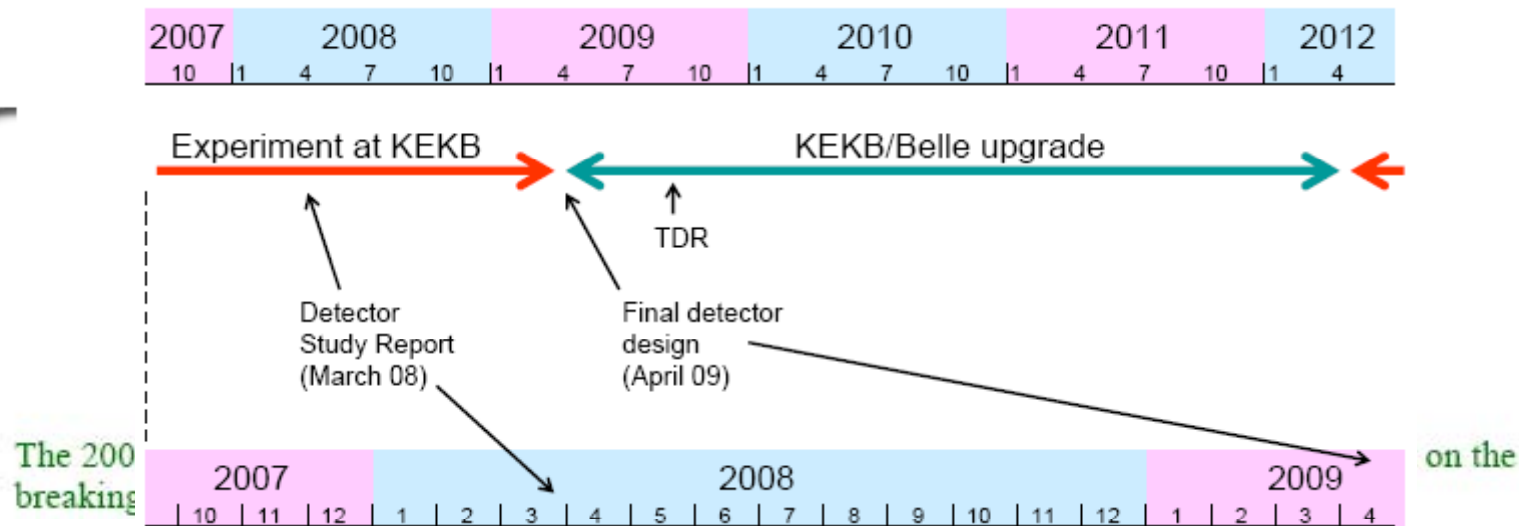
앞으로 Belle 실험은...



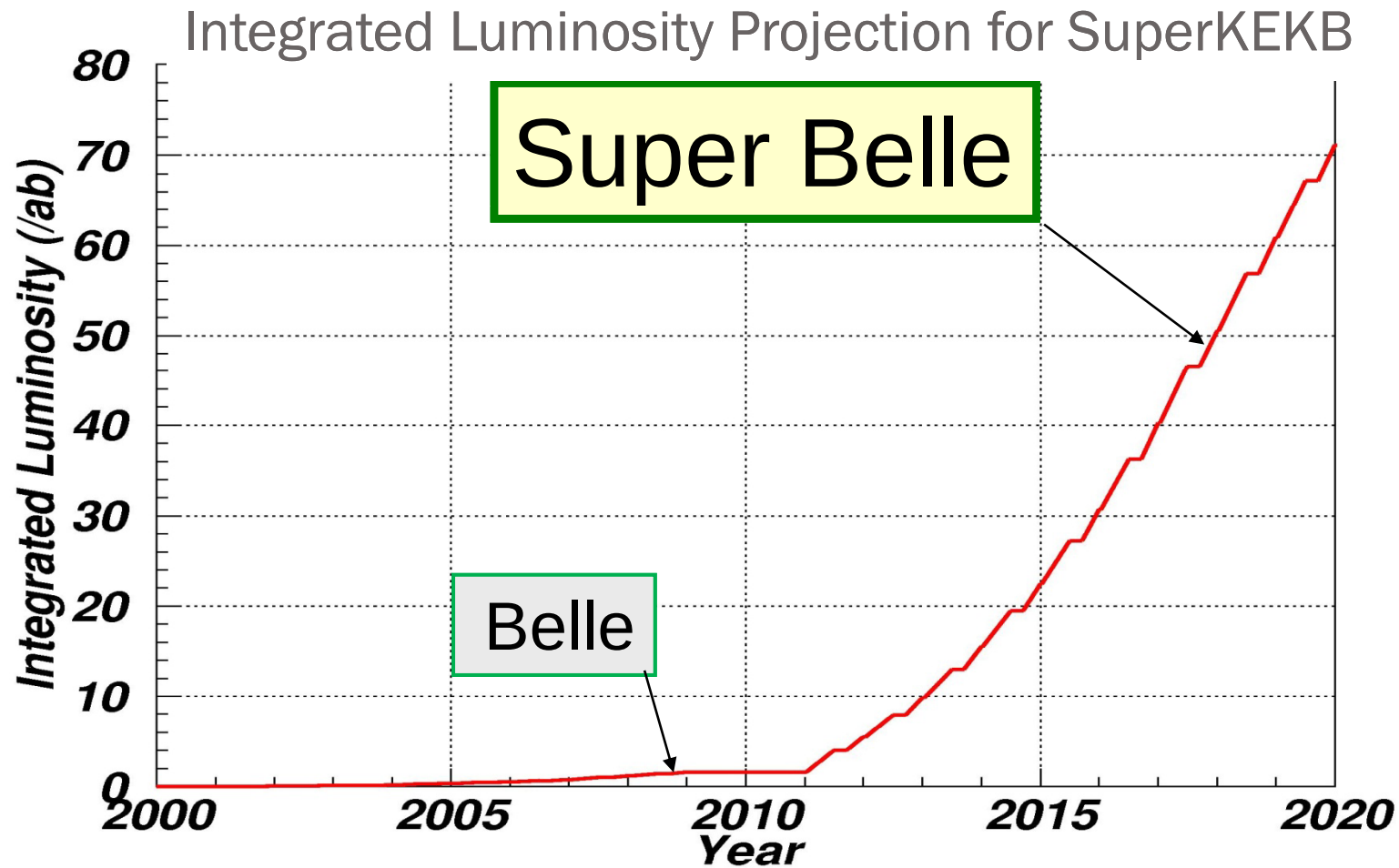
Belle results lead to Nobel Prize for

Belle group은 하나의 upgrade된 program 추진 → **Super Belle**

Kobayashi & Maskawa



Super Belle 실험은...



결론 및 요약

- 현재 활발한 국제 연구활동이 진행되는 **고 에너지 충돌실험**인 CDF 와 Belle 실험에 대한 소개.
- 두 실험 대용량 데이터 처리 과정과 시스템 소개.
 - Data Trigger System
 - Data Acquisition System
- 앞으로 고 에너지 실험의 데이터 사용 전망
 - CDF : 현재 4.1 fb^{-1} (150kB/event) $\rightarrow \sim 7 \text{ fb}^{-1}$ (2009)
 - Belle : 현재 1.0 ab^{-1} (30kB/event) $\rightarrow \sim 50 \text{ ab}^{-1}$ (super-Belle)
 - 따라서 앞으로 더 많은 computing resource (processor, disk 등등)와 효과적인 기술(processing technique)을 요구할것임.