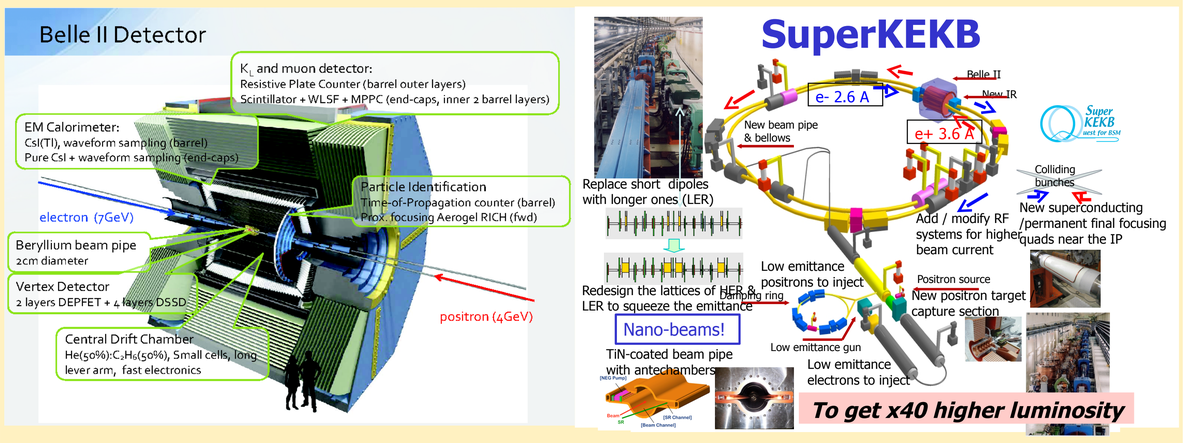
Implementation of parallel processing in ***basf2*** framework for Belle II

Ryosuke Itoh (KEK) and Soohyung Lee (Korea U.)
with
N.Katayama (Kavli IPMU), S.Mineo (U. of Tokyo), A.Moll (MPI), T.Kuhr, and M.Heck (KIT)

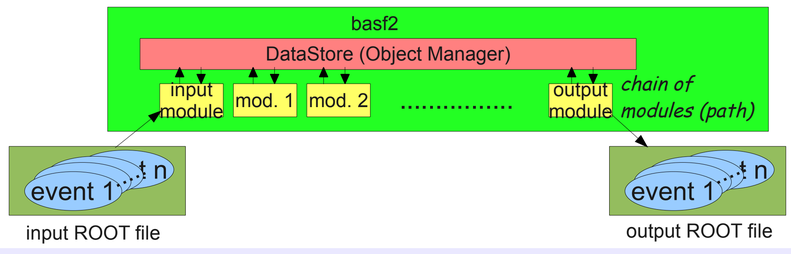
SuperKEKB and Belle II

- Belle II experiment @ SuperKEKB is a new generation B-factory experiment aiming at $\times 40$ higher luminosity.
- The main purpose of the experiment is the search for New Physics beyond the energy scale of LHC in the quantum effect in B meson decays.
- The data flow before the online storage is estimated to be more than 1GB/sec.
- The processing of such huge data flow is a challenge for Computing.
- For the massive production and HLT applications, the parallel processing of events (trivial event-by-event) is a must.



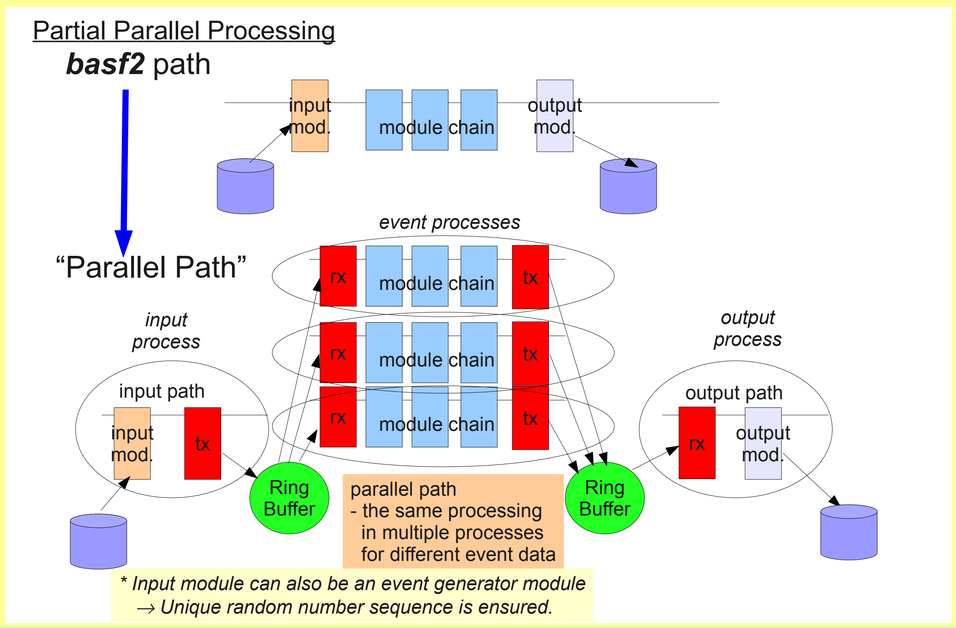
basf2 framework

- ***basf2*** is a software framework developed for the use in Belle II.
- ***basf2*** has a software bus architecture so that a large scale application is implemented by plugging a set of small-functioned modules into the framework (path).
- The objects are passed among modules by the DataStore object manager.
- The framework is designed to match with a wide range of the usage in MC/DST production, user analysis, and in the DAQ software.



Partial Parallel Processing

- Recent PC servers are equipped with multi-core CPUs and it is desired to utilize the full processing power of them.
- The parallel processing using the multi-core is included in the original design of ***basf2***.
- It is implemented so that the execution of the partial portion of the module chain can be parallelized in multiple Linux processes.
- It enables to use the same I/O modules in parallel processing without any modifications to them.



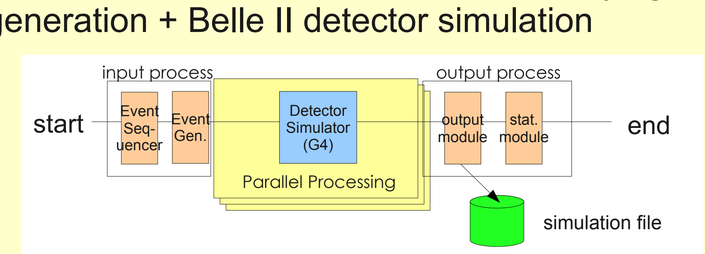
Object passing between processes

- To pass objects between processes, objects in DataStore are once streamed in a byte stream record by TMessage event by event.
- A transmitter(tx) module placed at the end of a path in a process streams the objects and place it in the ring buffer.
- The ring buffer is implemented using Linux IPC shared memory which is accessible by multiple processes.
- A receiver module(rx) picks up one event record from the ring buffer and then restores objects in DataStore.
- Multiple event processes are connected to a single ring buffer so that events are distributed/collected for the parallel processing. The load balancing of event processes is ensured by the ring buffer automatically.

Performance test

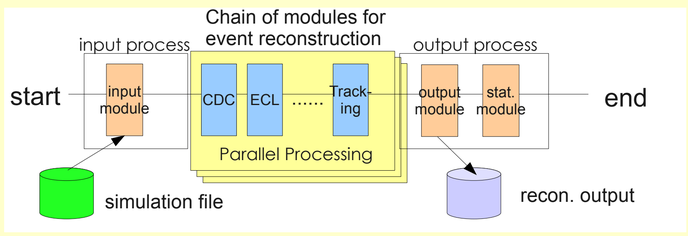
- The parallel processing performance is tested using a PC server with
  * 32 core CPUs (4 x Intel Xeon X7550 @ 2GHz),
  * 65GB memory,
  * Scientific Linux 5.5 (Kernel 2.6.18, 64bit).
- Two realistic benchmarks.
  * Elapsed time for 10,000 events is measured varying no. of event proc's.

1) Event generation + Belle II detector simulation

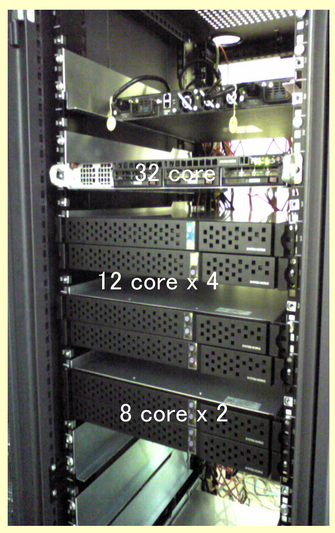


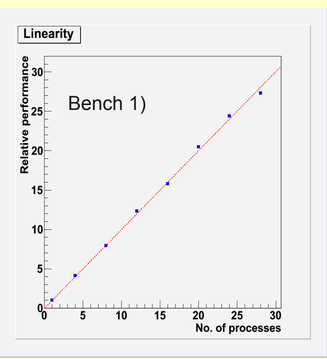
⇒ 0.66 sec/evt, ~500kB/event
→ max output rate = 23.2 MB/sec @ 28 processes

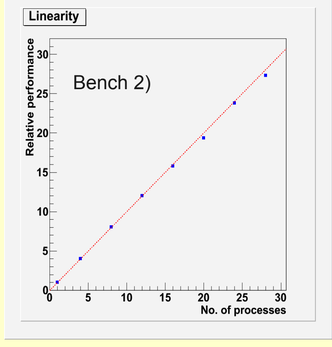
2) Belle II event reconstruction



⇒ 0.29 sec/evt, input ~500kB/ev, output ~750kB/ev
→ max output rate = 48.5MB/sec @ 28 processes







- It is confirmed that a good linearity is kept in both cases.
- Bottleneck in absolute performance :
  1) I/O bandwidth : two I/O subsystems for ***basf2***.
     a) ROOT TFile I/O
        Measured maximum read rate : 8.9MB/sec ← very slow.....
     b) SeqROOT I/O (specially developed for Belle II DAQ)
        * Event-by-event sequential collection of TStreamed objects
        Measured rate is 253.3 MB/sec. → used for performance test above.
  2) Object streaming/destreaming for the ring buffer
     a) Max. flow rate w/o parallel processing : 253.3MB/sec
     b) w parallel processing : 112.4MB/sec
     → Maximum flow rate becomes a half because of object streaming overhead but still good enough in "many core" parallel processing.

Use in Belle II HLT

- By replacing the ring buffer with the network socket connection (B2Socket), the objects can be transferred between different PC nodes.
- It enables the parallel processing utilizing both multicore CPUs and network-connected PC servers in a PC farm for HLT.
- The detail is reported in the poster by S.Lee in Track 1 (ID 390).