Single-chip Cloud Computer for Belle II DAQ

Kolja Prothmann

Max-Planck-Institut für Physik
(Werner-Heisenberg-Institut)

Excellence Cluster Universe

- Motivation and Requirements
- The SCC
  - Technical details
  - Programming example
- Application in the DAQ
- Conclusion

- Data from PXD is 10 times the data from rest of the detector
- Bandwidth 20 x 5Gb
- Data reduction is needed!
  - Bandwidth
  - Calculation time

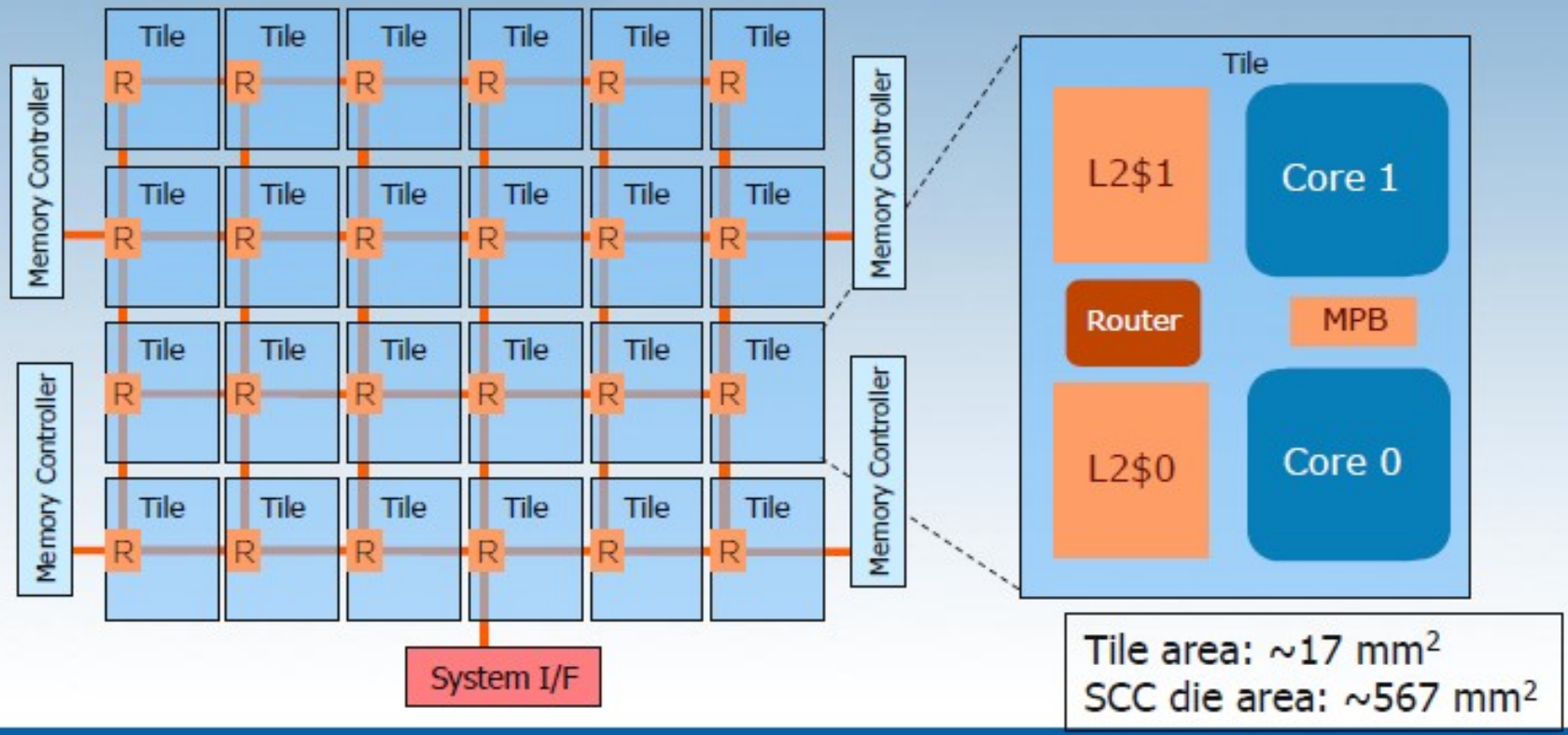
Single-Chip Cloud Computer (SCC) solves 'calculation time' issue!

→ Research proposal on the SCC handed in to INTEL

- INTEL terascale group
  - Aim is to provide scalable solutions for >100 cores
- Single-chip Cloud Computer
  - 48 cores
  - Advanced message passing (256 Gb)
- Modes of Operation
  - SMP (Symmetric MultiProcessing) with OS (e.g. Linux)
  - Message passing (library RCCE)
  - Baremetal (no OS)
  - Emulation on PC multicore CPUs (library RCCE)

Top Level Hardware Architecture

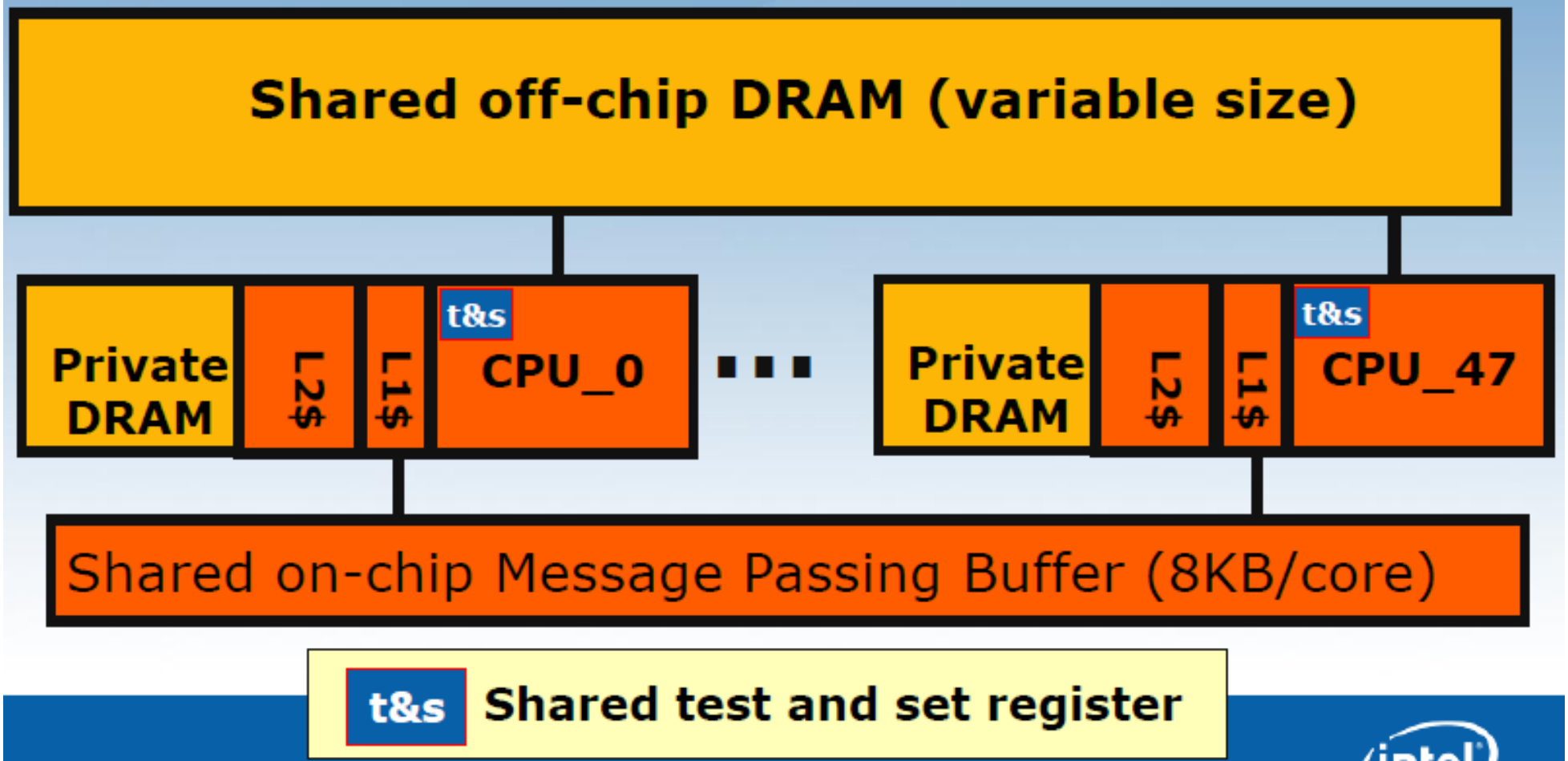
- 6x4 mesh 2 Pentium™ P54c cores per tile
- 256KB L2 Cache, 16KB shared MPB per tile
- 4 iMCs, 16-64 GB total memory



R = router, iMC = integrated Memory Controller, MPB = message passing buffer

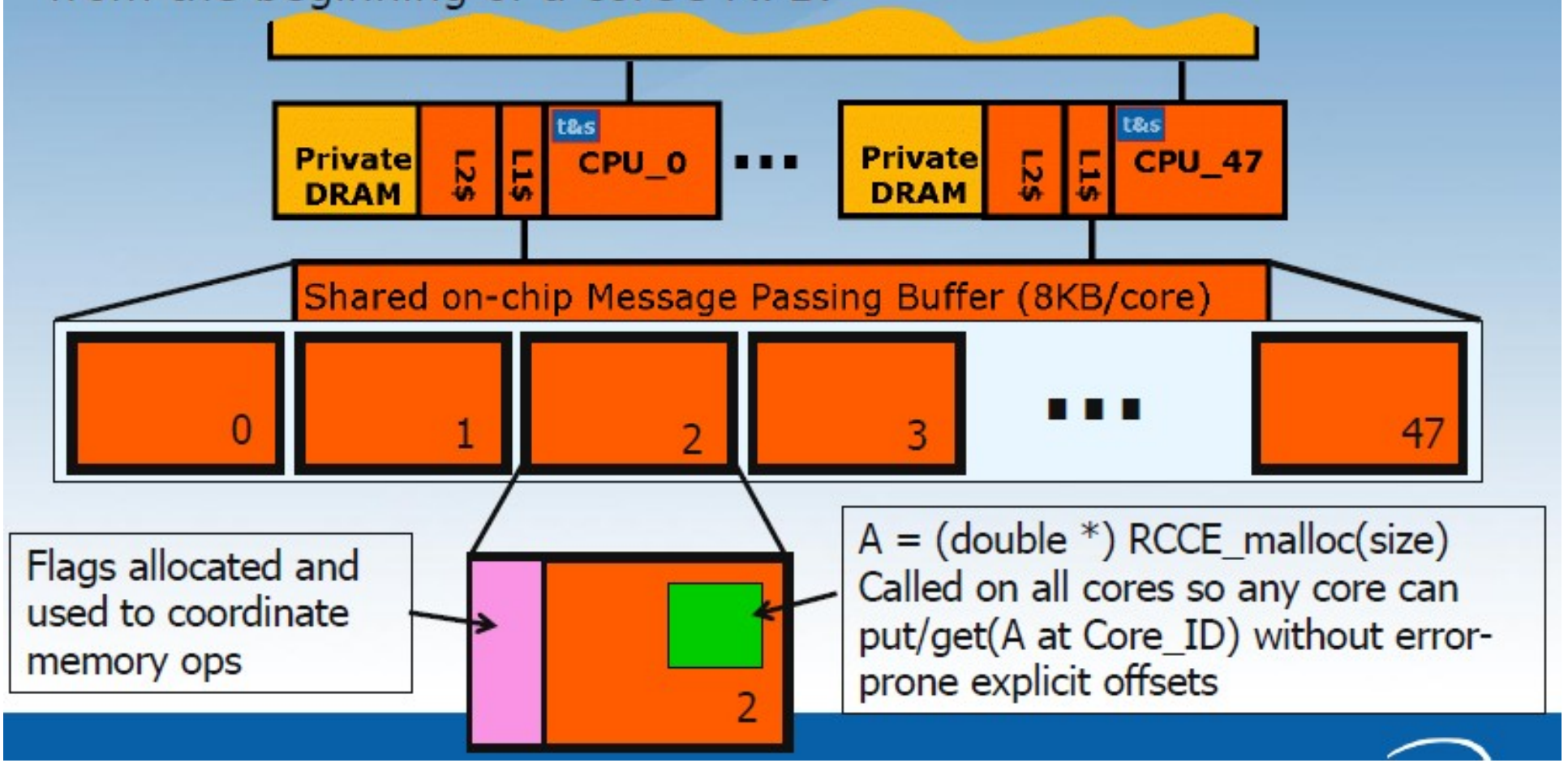
Programmer's view of SCC

- 48 x86 cores with the familiar x86 memory model for Private DRAM
- 3 memory spaces, with fast message passing between cores (■ / □ means on/off-chip)



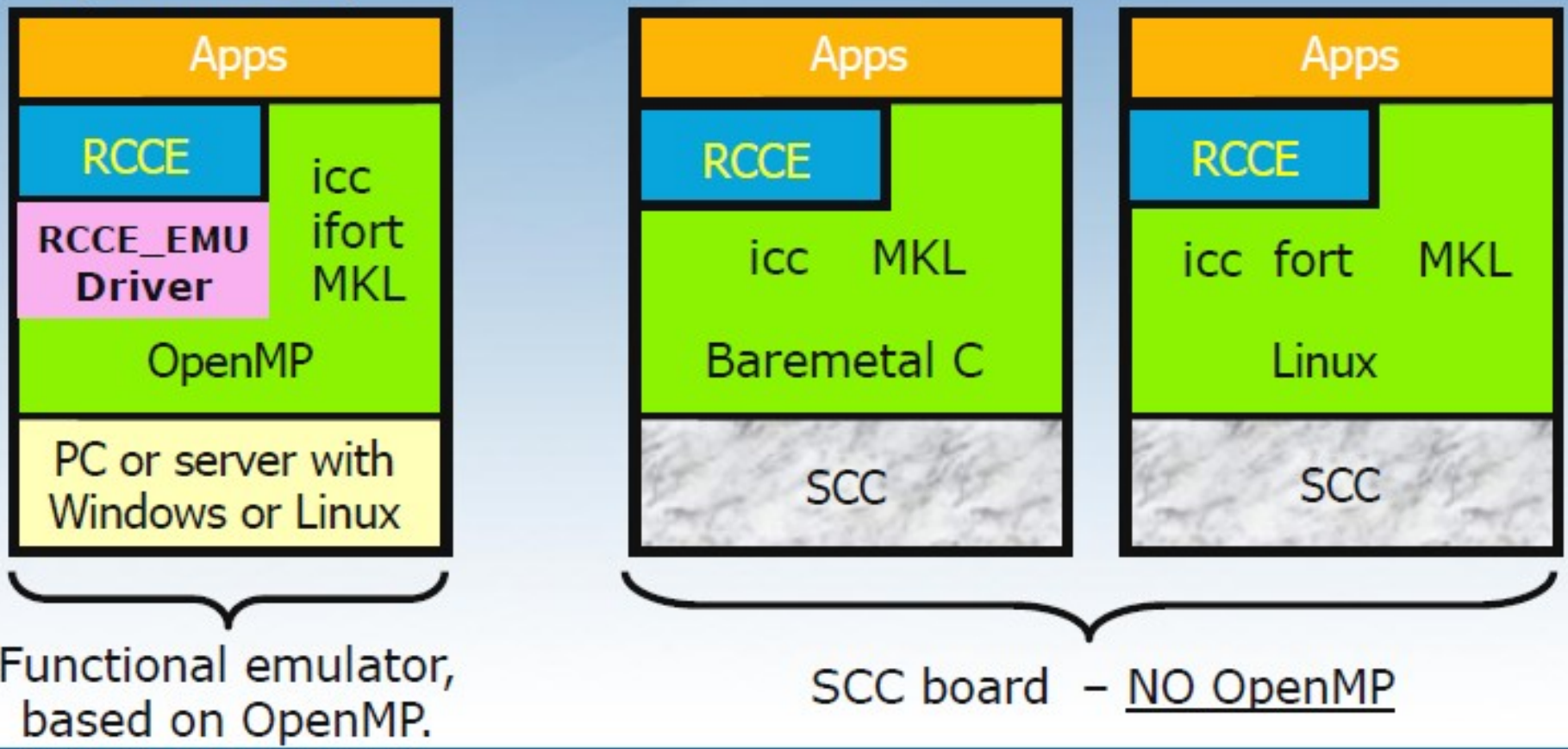
How does RCCE work? Part 2

- Treat Msg Pass Buf (MPB) as 48 smaller buffers … one per core.
- Symmetric name space … Allocate memory as a collective op. Each core gets a variable with the given name at a fixed offset from the beginning of a core's MPB.



SCC Platforms

- Three platforms for SCC and RCCE
  - Functional emulator (on top of OpenMP)
  - SCC board with two "OS Flavors" … Linux or Baremetal (i.e. no OS)



RCCE (speak: 'rocky') library

- Emulation (using OpenMP)
- SCC test board

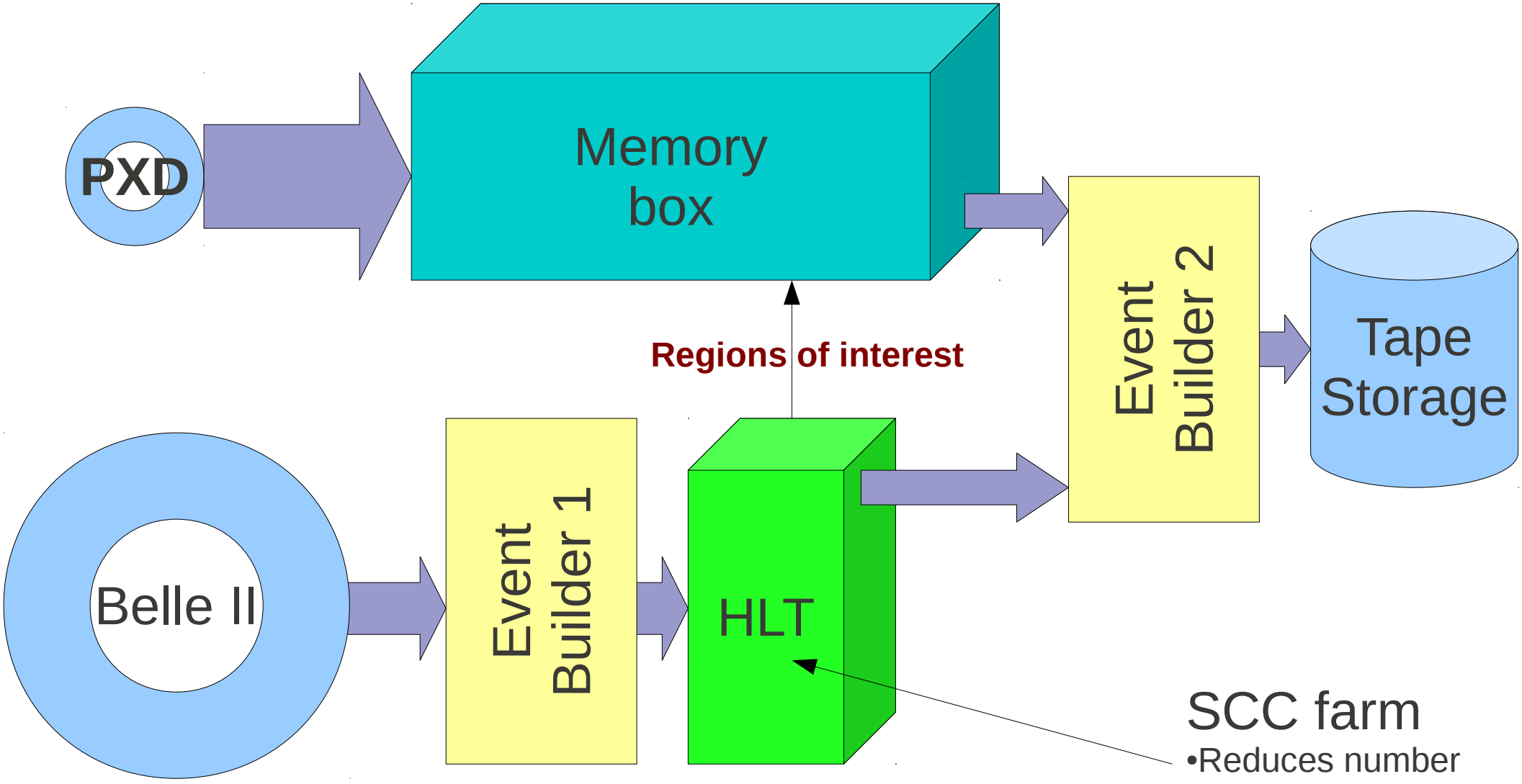
```
#include “RCCE.h”
…
RCCE_init(&argc, &argv);
buffer = (double *)malloc();
…
cbuffer = (double *)RCCE_malloc()
```

RCCE_send(buffer, size, ID);

Send private memory (buffer) to core ID

RCCE_recv(buffer, size, ID));

Receive into private memory (buffer) from core ID


PXD
Memory box
Regions of interest
Event Builder 2
Tape Storage
Belle II
Event Builder 1
HLT
SCC farm
•Reduces number of nodes
•Reduces memory in the Memory box

- Simulation
  - basf2
- Parallel fitting in analysis
  - Minuit2
- HLT without data reduction algorithm
  - 40 times more data means 40 times more computers

- Summary
  - SCC: 48 cores & advanced messaging hardware
    - SMP (symmetric multi processing)
    - RCCE
  - Looking forward to the outcome of the proposal!
  - Wide range of applications
- Conclusion
  - Alternative approach to a full FPGA implementation
  - Only small changes needed to the C++ code (if any)
  - No custom hardware needed

Backup

- InfiniBand
  - 2.5 Gb / 5 Gb
  - very low latency connection
  - Memory access without CPU usage (DMA)
  - linux drivers available
  - (copper cable)
- FiberChannel HBA (Host bus adapter)
  - 2 Gb / 4 Gb
  - Payload: 2112 Bytes
  - Header: 36 Bytes
    - Start of frame, 4Byte
    - FC frame header, 24Byte
    - CRC, 4Byte
    - End of frame, 4Byte