Center for Computational Sciences, University of Tsukuba
www.ccs.tsukuba.ac.jp

Power-aware, Dependable, and High-Performance Communication Link Using PCI Express: PEARL

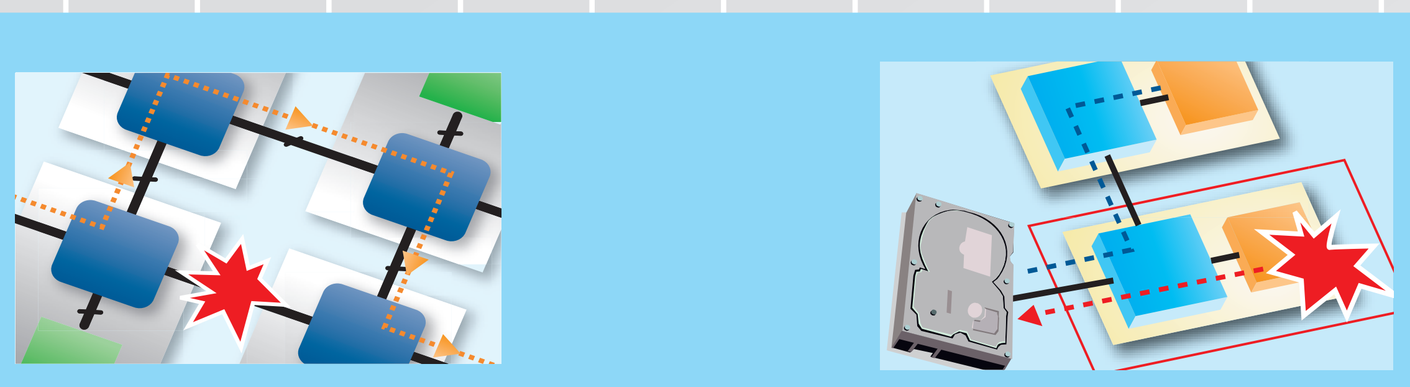
Concepts

Although parallel and distributed systems can provide some redundancy, they can only be truly dependable if the communications in such systems is also reliable.

We have created a high performance, power-aware network with redundancy for parallel and distributed systems ranging from high-end embedded systems to small scale HPC clusters.

PEARL:
PCI Express Adaptive and Reliable Link
PEACH chip:
PCI Express Adaptive Communication Hub

- Features high bandwidth and low power consumption using PCI Express® technology while still adhering to the PCIe standard.
- Intelligent control through use of an embedded processor for dependability and power-awareness.



(a) Fault link recovery (b) Device fail-over

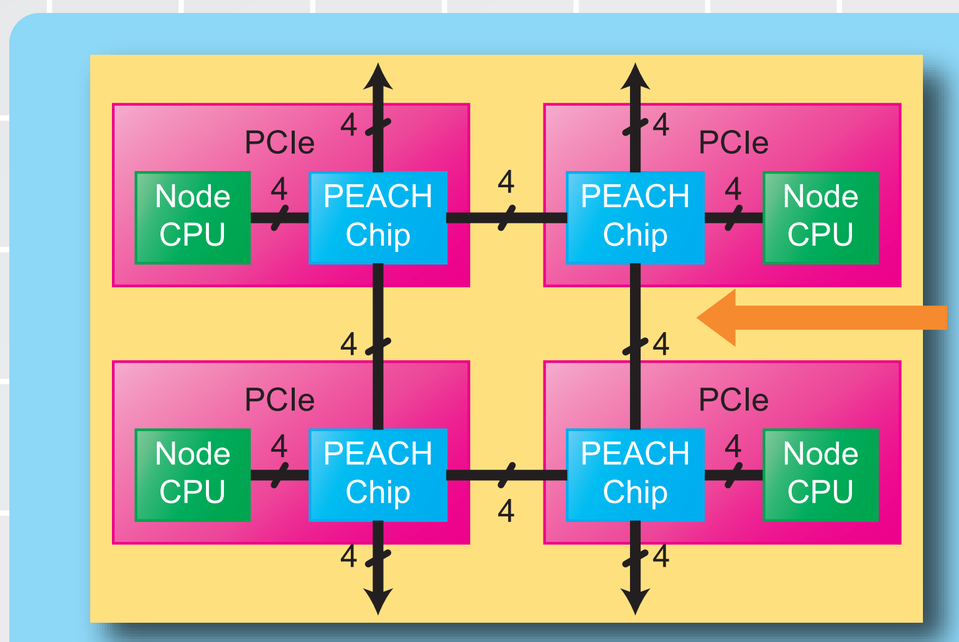
Dependability facilities of PEARL

PEARL: Communication Link Based on PCI Express Technology

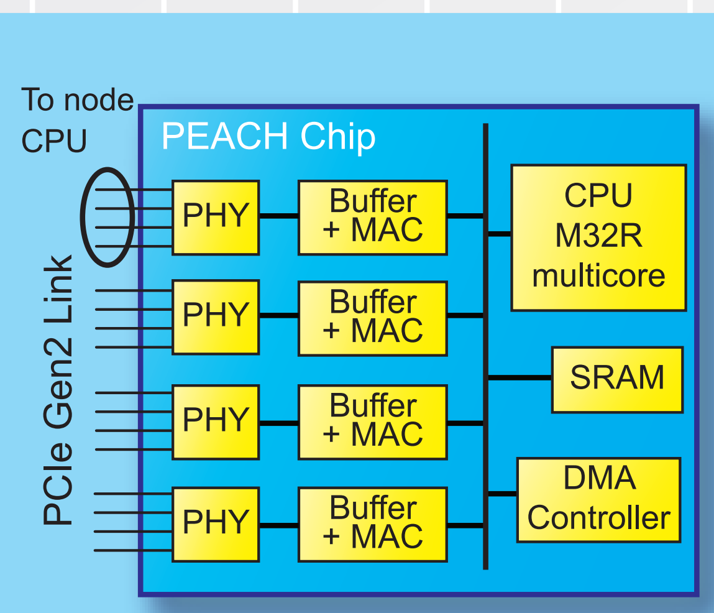
- Direct connection using PCI Express between nodes
  - Uses PCIe external cables with a range of several meters to connect between nodes.
  - PCIe Link connects between Root Complex (RC) and End-point (EP).
    - All PCIe ports are RC/EP switchable during initialization.
  - Any PCIe ready device can be directly attached to PEARL.

Features

- High-Performance
  - 20 Gbps, 2 Gbyte/s (Theoretical Peak) = InfiniBand® DDR 4x
- Energy efficient and Power-aware:
  - Consumes less power than conventional networks
  - Saves power by reducing the number of lanes and transfer rate
- Dependability
  - Error detection, flow control, retransmission by PCIe protocol
  - Fault tolerance through use of detour routing
  - Embedded processor monitors system and detects faults.



Overview of PEARL system



Block diagram of a PEACH chip

Selection of the number of lanes and lane speed
(Power consumption ratio obtained by 65 nm PCIe Gen2 PHY test chip)

# of lanes Lane speed	x1	x2	x4
2.5 GHz	2.5 Gbps (21)	5 Gbps (38)	10 Gbps (75)
5 GHz	5 Gbps (28)	10 Gbps (50)	20 Gbps (100)

PEACH Chip & Board

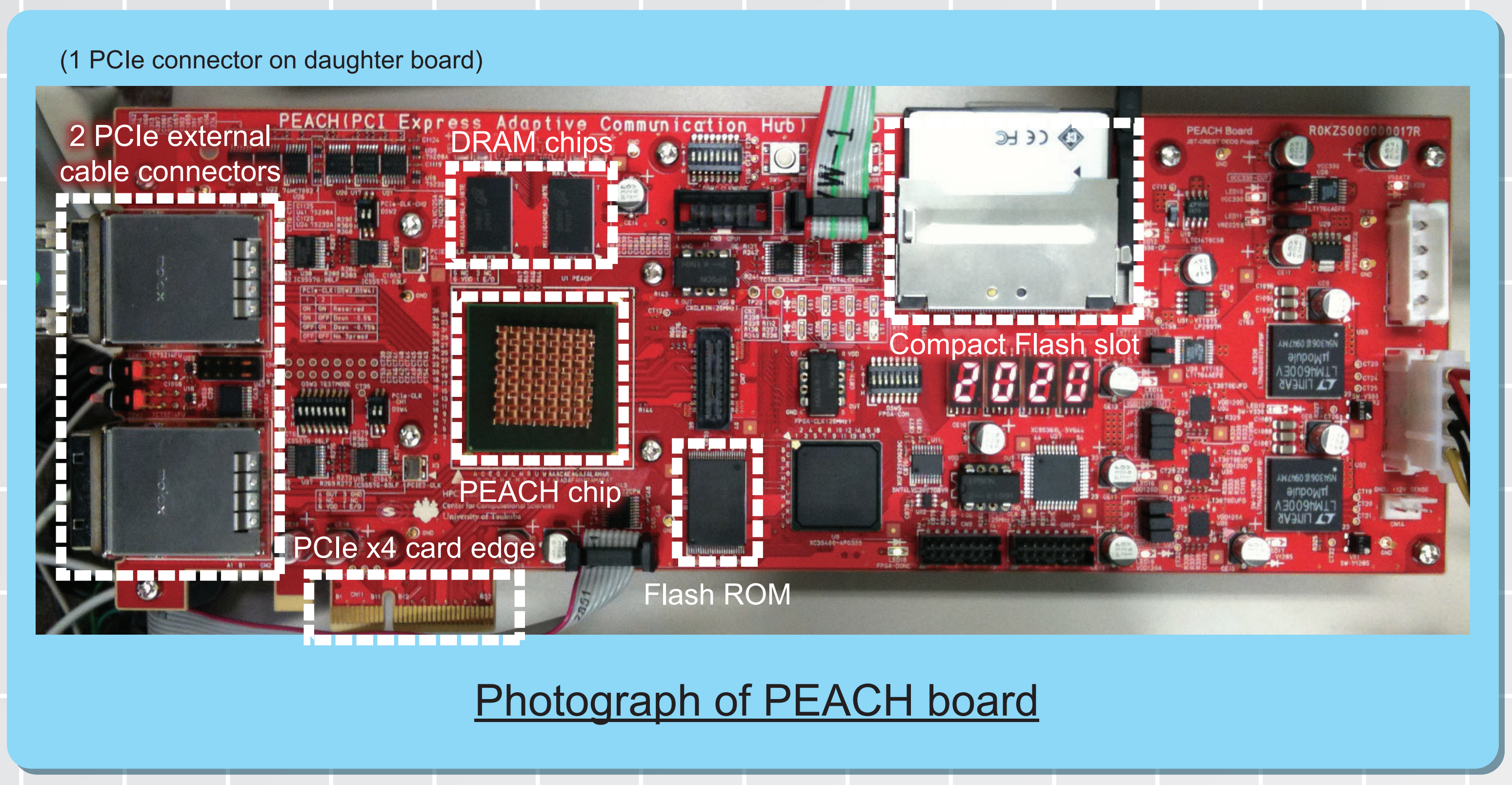
Overview of PEACH Chip

- Embedded CPU: M32R (Renesas Electronics, 4 cores, SMP)
- PCI Express Gen 2, x4 lanes (20 Gbps) 4 ports
- Max Payload Size: 1 Kbytes

Overview of PEACH Board

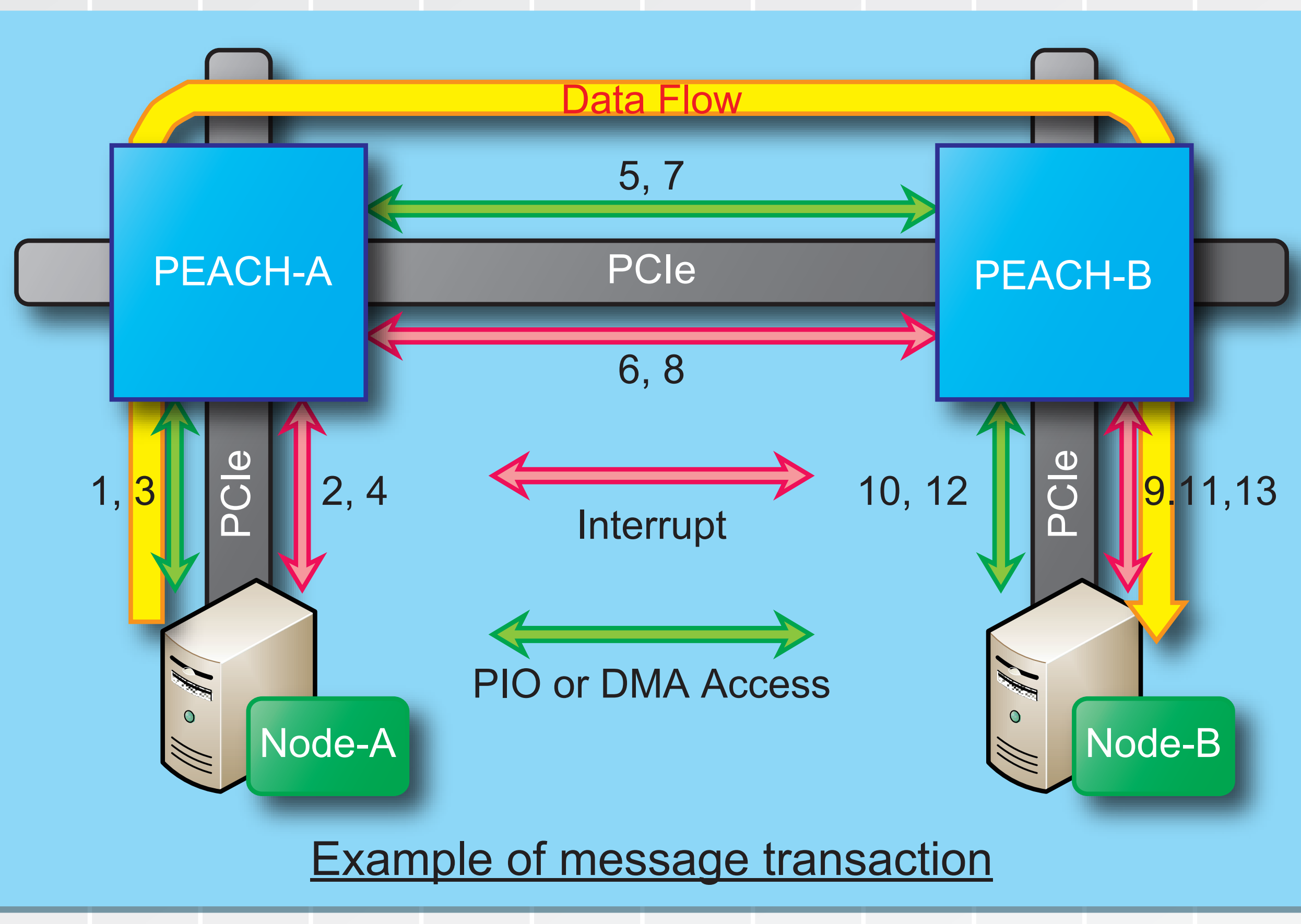
- PCI Express x4 host adapter board
- 3 PCI Express external cable ports
- Operates independently of host system.

(1 PCIe connector on daughter board)



Photograph of PEACH board

Communication Example



Example of message transaction

Node-A
1. Write header to PEACH-A
2. Interrupt for “DMA read ready” to PEACH-A

PEACH-A
3. Copy Payload from Node-A via DMA
4. Interrupt for “DMA read done” to Node-A
5. Write header to PEACH-B
6. Interrupt for “DMA read ready” to PEACH-B

PEACH-B
7. Copy payload from PEACH-A via DMA
8. Interrupt for “DMA read done” to PEACH-A
9. Interrupt for “Read data ready” to Node-B

Node-B
10. Read Header from PEACH-B
11. Interrupt for “DMA write ready” to PEACH-B

PEACH-B
12. Copy payload to Node-B via DMA
13. Interrupt for “DMA write done” to Node-B

Acknowledgements

This project is sponsored by the Japan Science and Technology Core Research for Evolutional Science and Technology(JST/CREST) under the "Computational Platform for Power-aware and Reliable Embedded Parallel Processing Systems" program in the area of "Dependable Embedded Operating Systems for Practical Use"(Oct. 2006 - Nov. 2011)