

# Configurable Crossbar Switch for Deterministic, Low-latency Inter-blade Communications in a MicroTCA Platform

*Machine Protection—Using a MicroTCA-based platform to protect the Spallation Neutron Source's particle accelerator from its own high energy beam*

By Eric Breeding, Oak Ridge National Laboratory, Saeed Karamooz, VadaTech, Inc., and Alan Justice, Oak Ridge National Laboratory



Eric Breeding



Saeed Karamooz



Alan Justice

## INTRODUCTION

The Spallation Neutron Source (SNS) provides the most intense pulsed neutron beams in the world for scientific research and industrial development. The high power generated by the SNS particle accelerator (Figure 1) is inherently accompanied by hazards of uncontrolled energy release (beam loss), uncontrolled power flow, and failure of hardware systems. These hazards pose a dual threat to accelerator components: physical damage and radioactivation. To mitigate risks associated with these hazards, the accelerator utilizes a Machine Protection System (MPS). The MPS monitors more than 1,000 sensors throughout the accelerator complex for potential problems. When such problems are detected, the MPS must terminate beam production within 20 microseconds to protect accelerator components. A highly reliable MPS is critical to maintaining the demanding availability requirements for the SNS facility. The existing MPS is functional and operational; however, its reliability and maintainability are declining due to aging hardware and component obsolescence. Consequently, a new MPS system is being developed to replace the current MPS. This system is based on the MicroTCA standard.

Due to the large physical area covered by the sensors, the new MPS employs a distributed architecture interconnected with high-speed serial communication links. The stringent beam control demands require both inter-shelf and inter-blade links to be low-latency and deterministic. Legacy bus architectures accomplished inter-blade communications using dedicated parallel buses across the backplane. Because of limited fabric resources on its backplane, MicroTCA uses the carrier hub (MCH) for this purpose. Unfortunately, MCH products from commercial vendors are limited to standard bus protocols such as PCI Express, Serial RapidIO, and 10/40 Gigabit Ethernet (GbE). While these protocols have exceptional throughput capability, they are neither deterministic nor necessarily low-latency. The development of an MCH with a user-configurable switch fabric overcomes this limitation, offering the system architect/developer complete

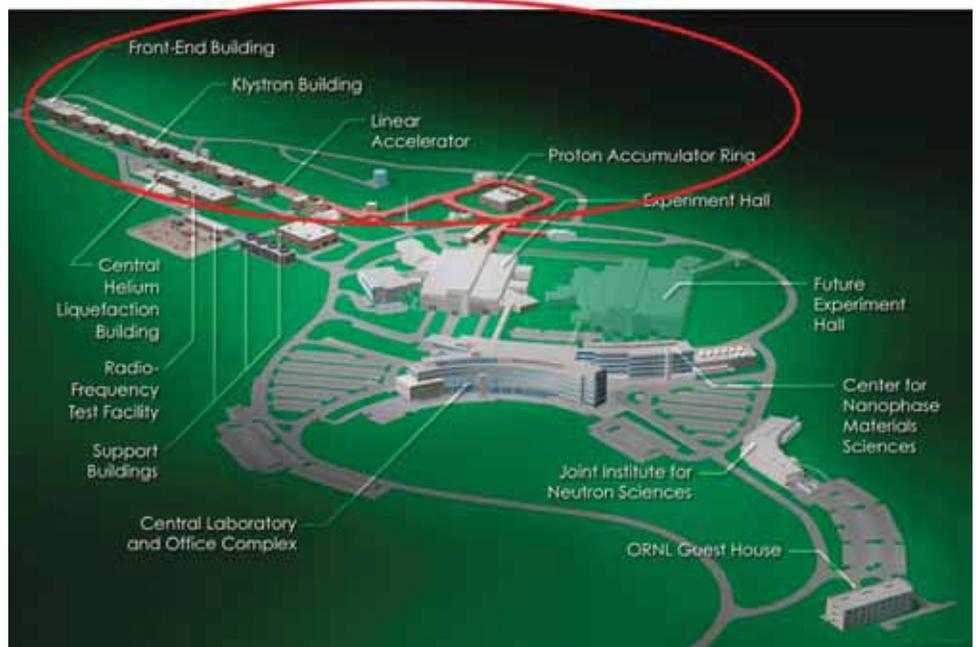


Figure 1: Oak Ridge National Laboratory's Spallation Neutron Source uses a linear accelerator and accumulator ring to generate a 1.4MW pulsed proton beam.

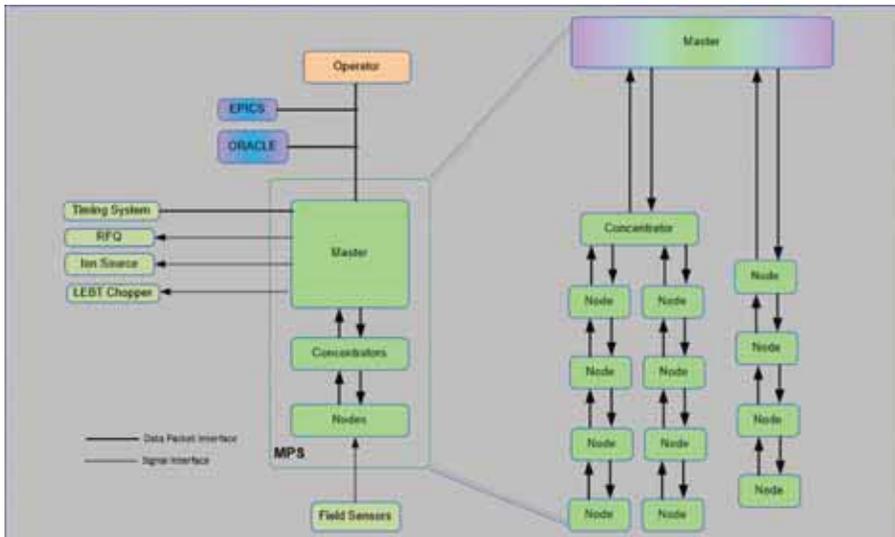


Figure 2: Illustration of Typical MPS Hierarchical Topology

flexibility in both interface protocol and the routing of information between blades.

## ARCHITECTURE

Comprising two distinct subsystems, master controller and field node, the MPS architecture employs a single master controller (Figure 2). The master controller's prime function is to disable beam generation when a qualified fault is reported from a downstream node. The master controller enables and disables beam delivery by controlling the timing pulses to specific front-end systems that generate and accelerate the beam. These front-end systems include the ion source plasma RF generator, the RF generator for the first accelerating structure in the accelerator (RFQ), and the gating pulse for the low-energy beam transport (LEBT) chopper. The MPS con-

sists of multiple field nodes—herein referred to simply as nodes—distributed throughout the accelerator facility. The primary function of each node is to interface with a group of sensors. If a sensor indicates an error, the node immediately reports the fault information to the master controller. Fault information includes the node identifier, the sensor identifier, and a timestamp.

Because the sensors being monitored by the nodes are located across a wide area, it is expedient for the MPS to adhere to a distributed architecture. The subsystems are arranged as a hierarchical topology with the master controller residing at the top. A domain is a fabric of nodes, concentrators, and links associated with one port of the master controller. Concentrators are

interspersed to reduce the overall fault propagation time for a given domain. A full-duplex, high-speed serial link running the Aurora 8B/10B protocol handles communications among subsystems. To ensure quick beam termination, the time required for fault information to propagate to the master controller must be minimal. Thus, in a hierarchical topology, it is critically important that the latency of the serial links be as low as practicable. Using the user flow control (UFC) feature of Aurora, an upper bound on the latency is guaranteed for the Xilinx 7 Series of FPGAs.

## CONFIGURABLE CROSSBAR SWITCH

The master controller is built on a MicroTCA.1 platform with a PCIe switch fabric. Specifically, the base-implementation consists of an MCH, a 12x12 full-duplex (x4) crossbar switch, multiple (up to 11) node processor (NP) blades and a single beam control (BC) blade (Figure 3). Any one of the NP blades must communicate a “terminate-

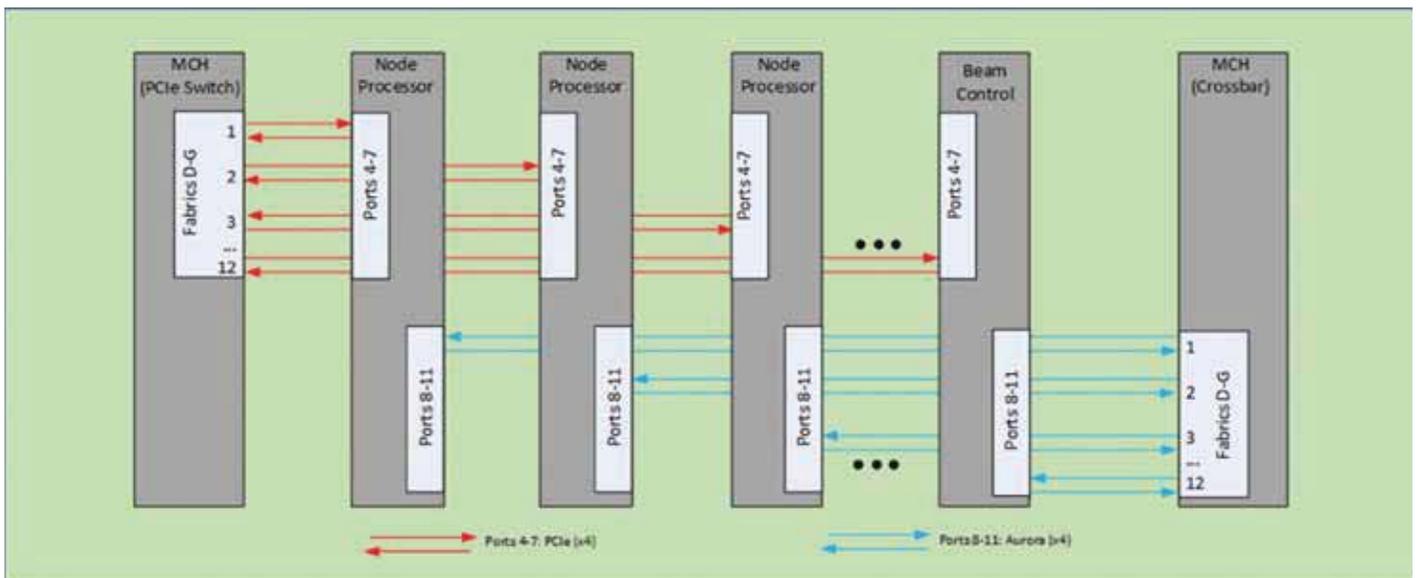


Figure 3: MPS Master Controller's Utilization of the MicroTCA Backplane

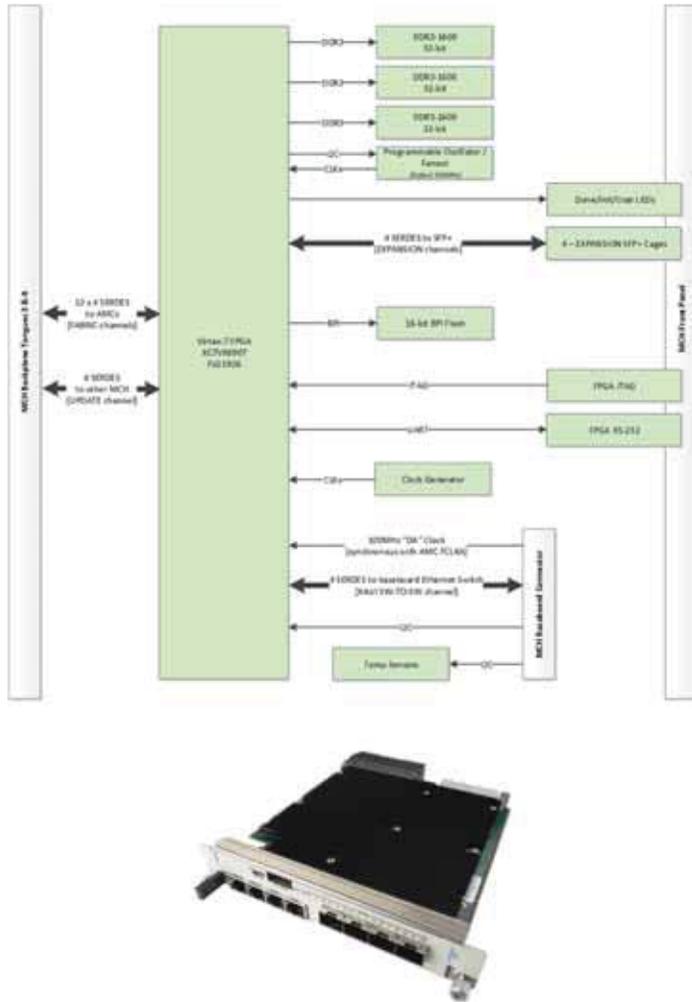


Figure 4: Vadatech UTC006 Configurable MCH w/ FPGA Switch Fabric

beam” message to the BC blade immediately upon receipt of a fault packet from a down-stream field node. Inter-blade communication is handled by a 12x12 full-duplex crossbar switch. Each communication port within the switch is comprised of four high-speed serial links. These links are implemented on Ports 8-11 of the MicroTCA switch fabric. By using the switch fabric, the crossbar eliminates the need for rear-panel transition modules and front-panel fiber-optics to handle module-to-module communications. From a purely hardware perspective, the crossbar switch is connected to the MicroTCA backplane as if it were a standard MCH module.

The crossbar switch is implemented on the VadaTech UTC006 (Figure 4). The UTC006 is a double module MCH (MTCA.4) with a user-configurable switch fabric. This switch fabric is based on the Xilinx Virtex-7 690T FPGA and consists of 12x4 full-duplex serial lanes to MCH Tongues 3 and 4. In addition, the MCH has four fiber-optic ports on the front panel connected directly to the FPGA multi-gigabit transceivers (MGTs), thus allowing direct external communication using any standard or propriety protocol up to 12 Gb/s. Direct communication between the Virtex-7 and a processing blade is easily achieved

by including the Xilinx PCIe IP core in the FPGA design. Other MCH features include three banks of 1GB DDR3-1600 memory and a 128 MB flash memory for FPGA configuration. The MCH has a managed-layer 3-port GbE switch base fabric. The base fabric connects to the FPGA via a 10GbE interface, allowing the FPGA to be fully monitored by an external source using IP. The MCH has additional GbE ports on the front panel which can be used as egress ports.

Using all four lanes at a line rate of 6.25 Gbps, the crossbar has demonstrated consistent 320 ns transfers between blades using a 16-byte UFC message. For high throughput applications not requiring determinism, the Aurora AXI4-Stream user interface yields an aggregate bandwidth approaching 60 GB/s.

*Eric Breeding received the B.S. and M.S. degrees in electrical engineering from the University of Tennessee, Knoxville, in 1987 and 1994, respectively. In 2013, he joined UT-Battelle as a Senior Staff Engineer at the Oak Ridge National Laboratory, Oak Ridge TN. He currently serves as a technical lead for hardware and firmware development at the Spallation Neutron Source. He holds six U.S. patents related to data acquisition for positron emission tomography and is also the co-author of several publications.*

*Saeed Karamooz received his B.S. in Chemical Engineering, B.A. in Mathematics, B.S. in Computer Science, M.S. in Electrical Engineering and M.S. in Computer Science all from the University of Nebraska at Lincoln. In 2004 he started VadaTech Inc. in pursuit of introducing high performing systems into embedded applications. Before that he worked as the CTO for GE/Fanuc Embedded division. He holds several U.S. patents related to embedded systems.*

*Alan Justice received a B.S. degree in electrical engineering from Tennessee Technological University in 2003. In 2006, he joined UT-Battelle as a Staff Engineer at the Oak Ridge National Laboratory, Oak Ridge TN. He currently serves as a Hardware Engineer for the Research Accelerator Division at the Spallation Neutron Source. For more information, contact the authors at: breedingj@ornl.gov and Saeed.Karamooz@vadatech.com.*

**ACKNOWLEDGMENTS**

ORNL is managed by UT-Battelle, LLC, under contract DE-AC05-00OR22725 for the U.S. Department of Energy. This material is based upon work supported by the U.S. Department of Energy, Office of Science under contract number DE-AC05-00OR22725.

Notice: This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains, and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).