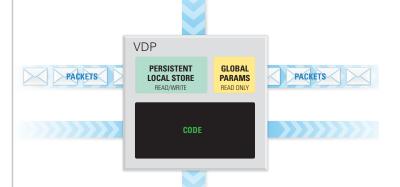
PULSAR (Parallel Ultra Light Systolic Array Runtime) offers a simple programming model for large-scale distributed-memory machines with multicore processors and hardware accelerators. PULSAR automates multithreading, message-passing, and multi-stream multi-GPU programming.

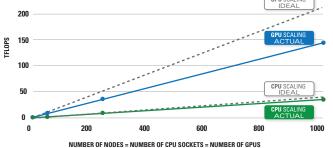
# PROGRAMMING MODEL



PULSAR offers a simple programming model, where the user defines the computation in the form of a Virtual Systolic Array (VSA), which is a set of Virtual Data Processors (VDPs), connected with data channels. The VDP is assigned a function, which defines its operation. Within that function, the VDP has access to a set of global parameters, its private, persistent local storage, and its channels. The runtime invokes that function when there are packets in all of the VDP's input channels. This is called firing. When the VDP fires, it can fetch packets from its input channels, call computational kernels, and push packets to its output channels.

#### **WEAK SCALING OF THE CANNON'S ALGORITHM USING 8KX8K** (A, B AND C) MATRICES PER NODE

Titan@ORNL, CPUs: 16-core AMD Interlagos, GPUs: NVIDIA Tesla K20X. 200



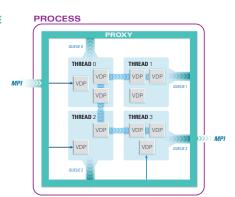
## BENEFITS OF PULSAR

- Simple Programming Model
- Lightweight Runtime System
- Multithreading & Message-passing
- Multi-stream Multi-GPU execution
- Minimum Overhead / Maximum Scalability

## **RUNTIME SYSTEM**

This programming model is accessible to the user through a very small and simple Application Programming Interface (API). All the complexity of executing the workload on a large-scale system is hidden in the runtime implementation. While the user invokes simple push and pull channel operations, the runtime takes appropriate actions, depending on the boundaries crossed by the channel, i.e., uses shared memory for VDPs residing in the same node, uses message-passing for VDPs residing in different nodes, uses Direct Memory Access (DMA) transfers between CPU VDPs and GPU VDPs.

#### **MPI MULTICORE**



### **MPI MULTICORE ACCELERATORS**

