

# Open MPI on the Cray XT

Presented by

Richard L. Graham

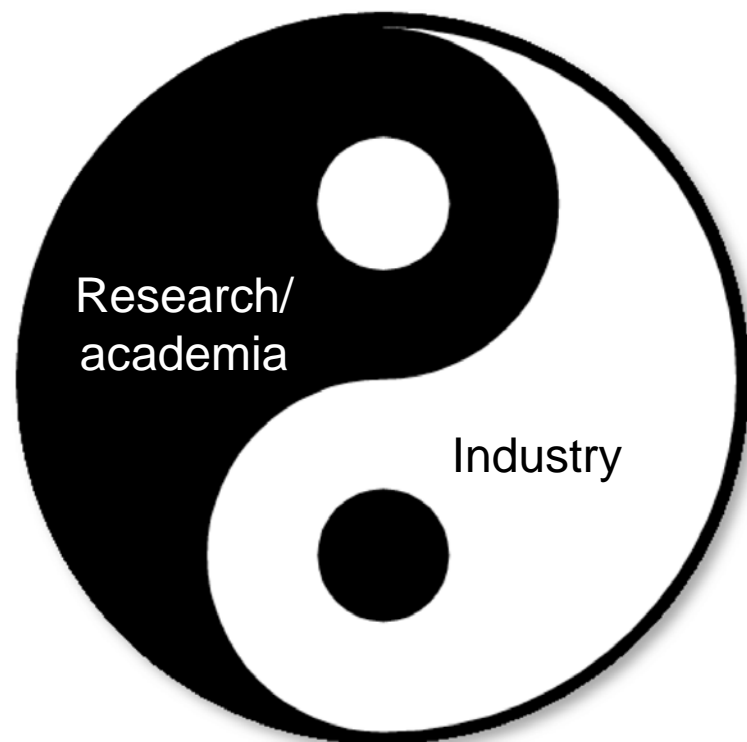
Tech Integration

National Center for Computational Sciences



# Why does Open MPI exist?

- **Maximize all MPI expertise:**
  - research/academia,
  - industry,
  - ...elsewhere.
- **Capitalize on (literally) years of MPI research and implementation experience.**
- **The sum is greater than the parts.**



# Current membership

14 members, 6 contributors

4 US DOE labs

7 vendors

8 universities

1 individual



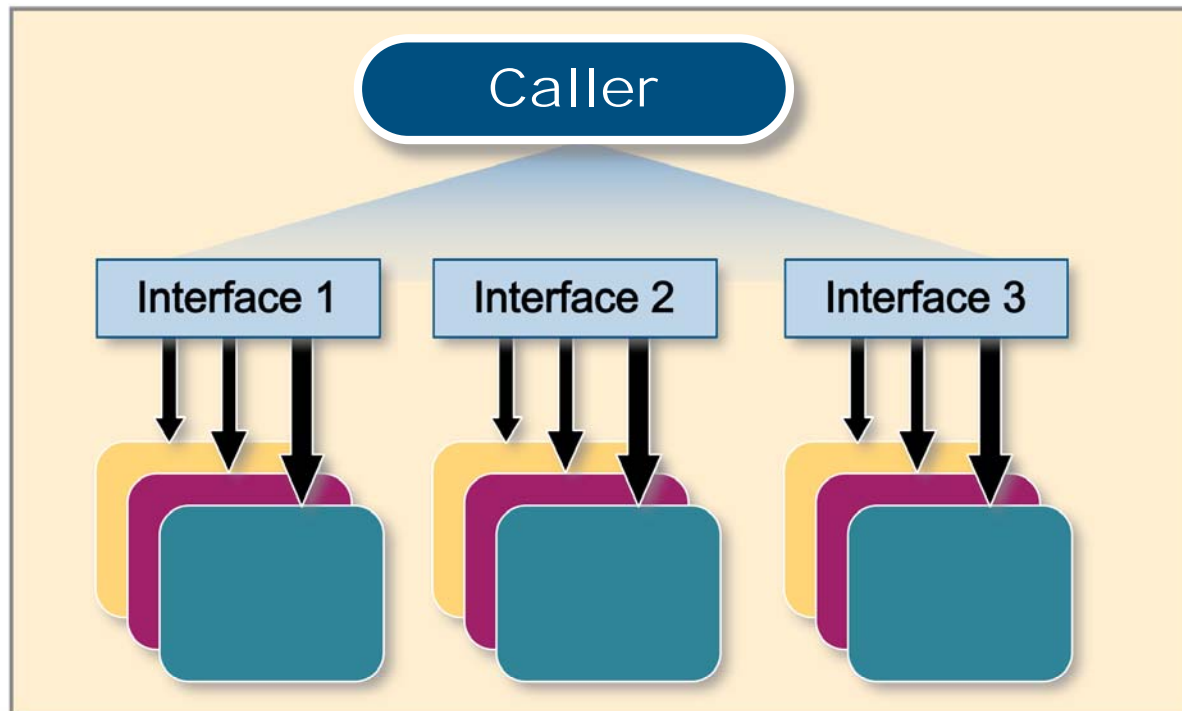
pervasive **technology** labs  
AT INDIANA UNIVERSITY



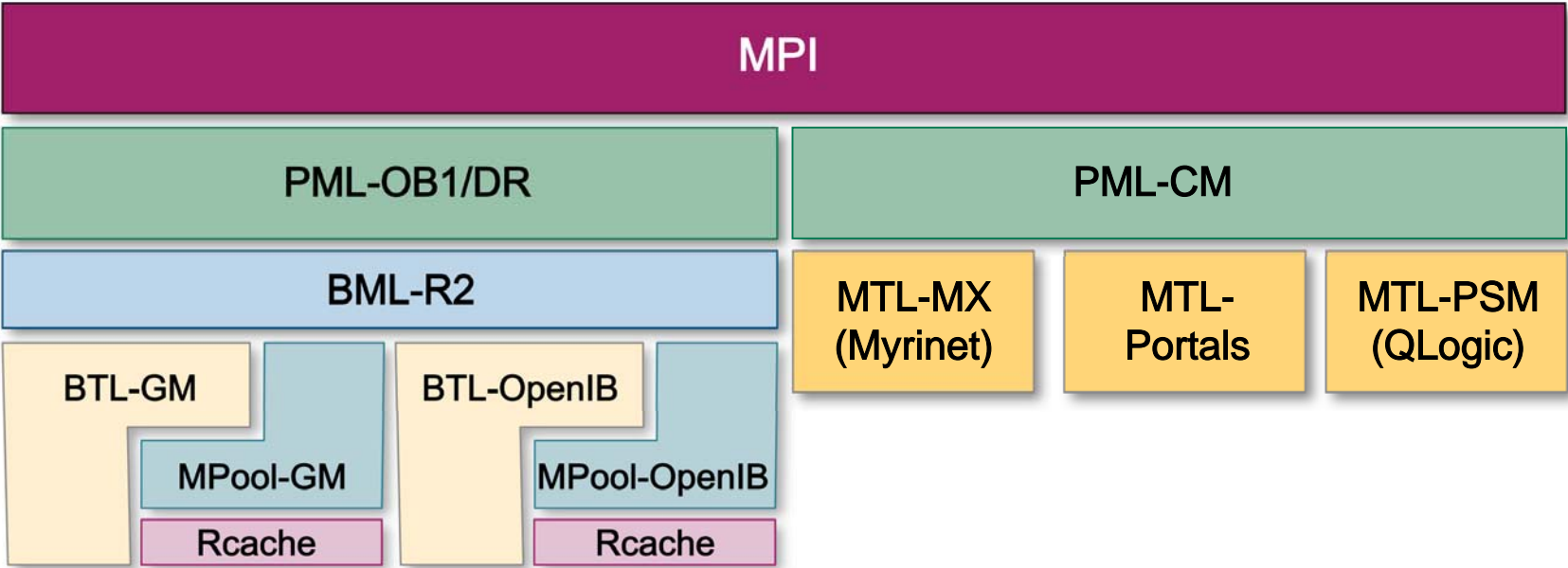
# Key design feature: Components

## Formalized interfaces

- Specifies “black box” implementation
- Different implementations available at run-time
- Can compose different systems on the fly



# Point-to-point architecture



# Portals port: OB1 vs. CM

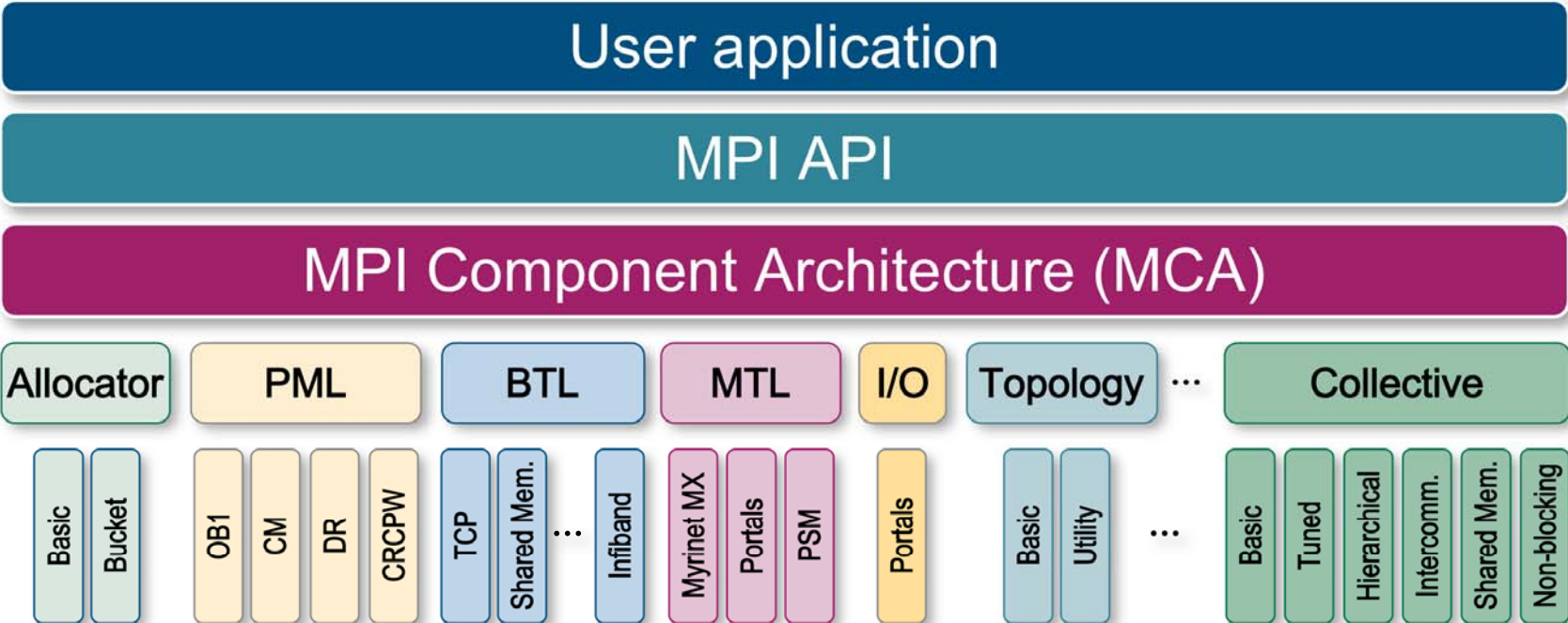
## OB1

- Matching in main-memory
- Short message: eager, buffer on receive
- Long message: rendezvous
  - Rendezvous packet: 0 byte payload
  - Get message after match

## CM

- Matching maybe on NIC
- Short message: eager, buffer on receive
- Long message: eager
  - Send all data
    - If Match: deliver directly to user buffer
    - No Match: discard payload, and get() user data after match

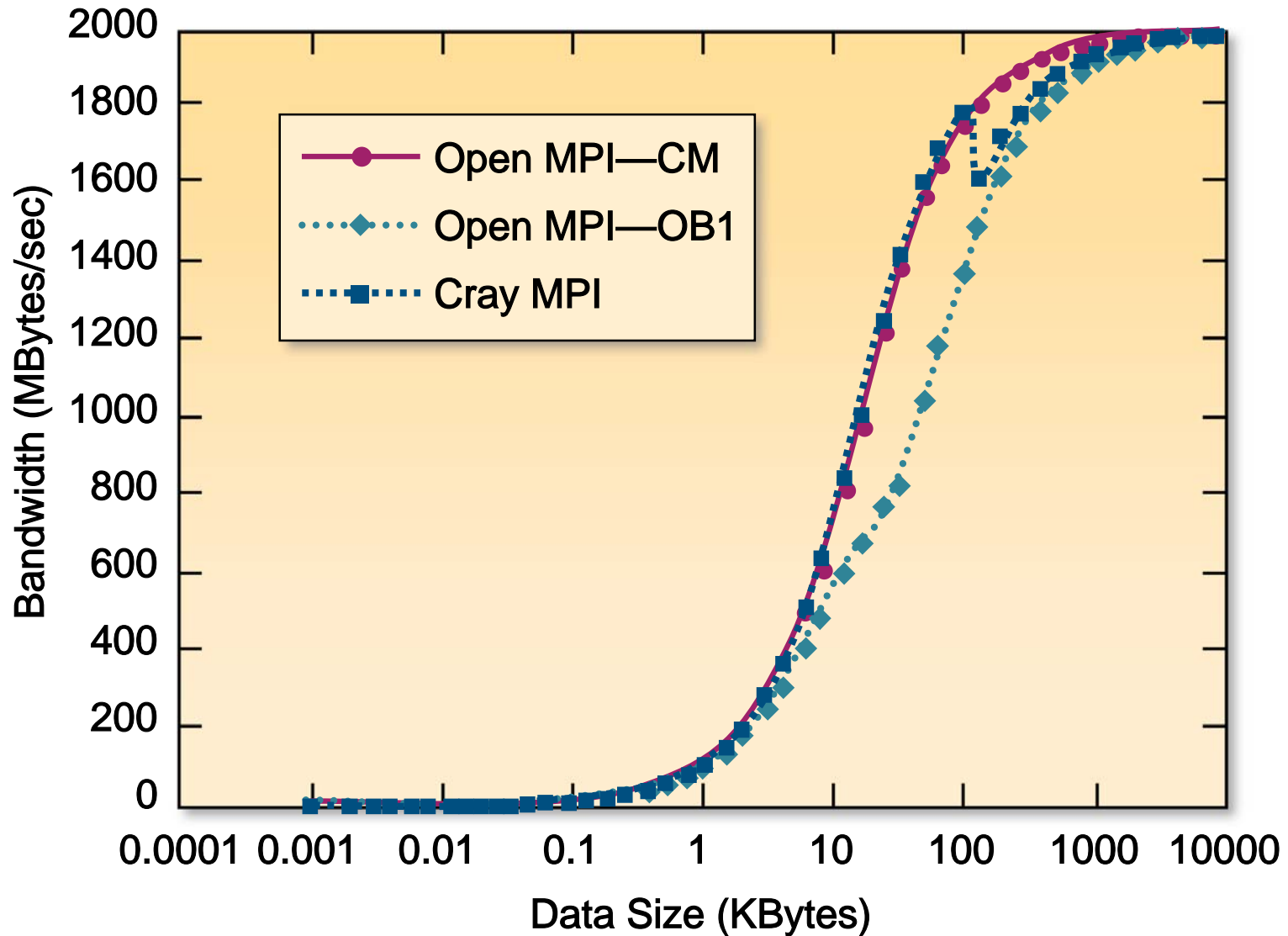
# Collective communications component structure



# Benchmark Results



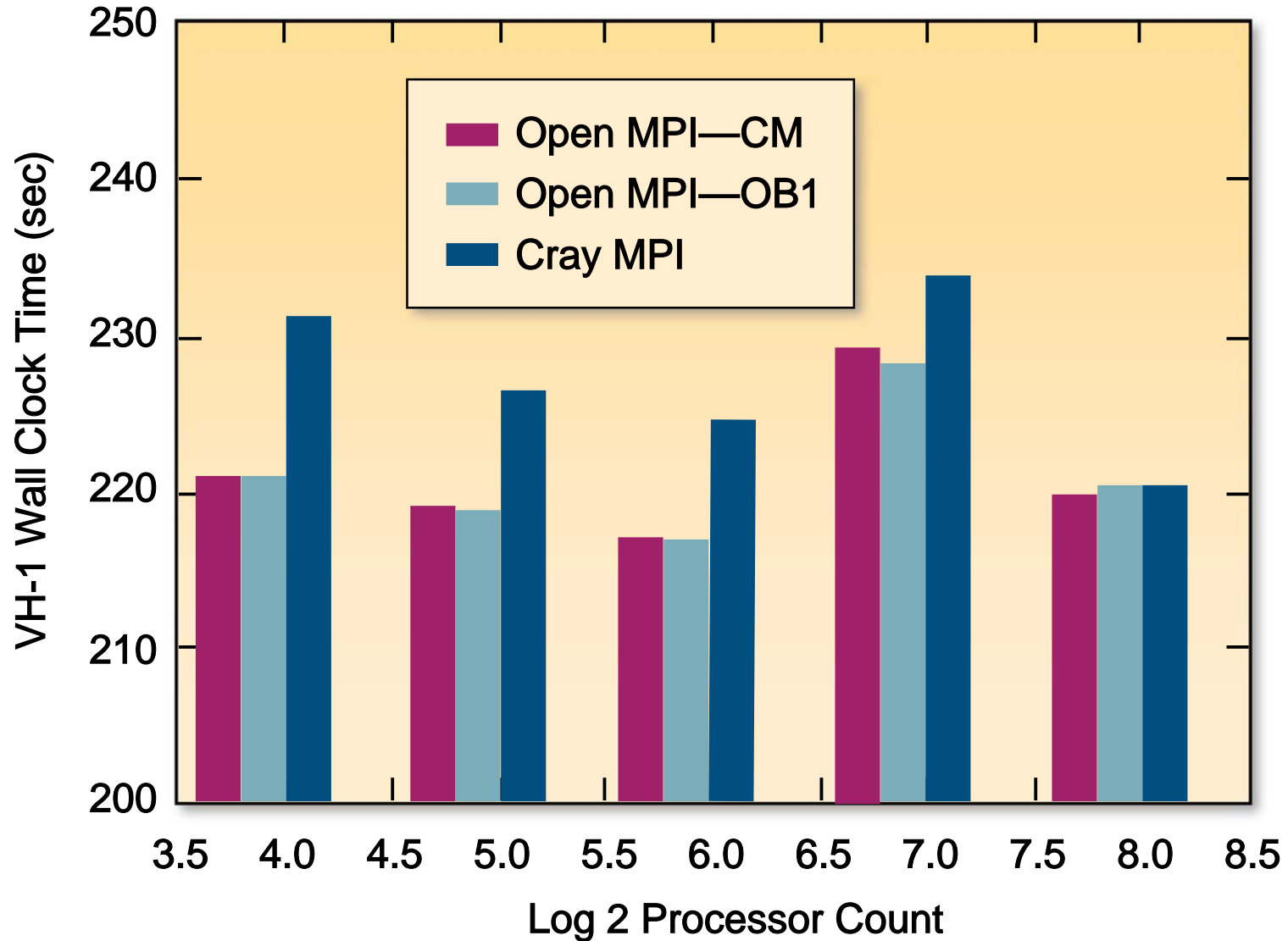
# NetPipe bandwidth data (MB/sec)



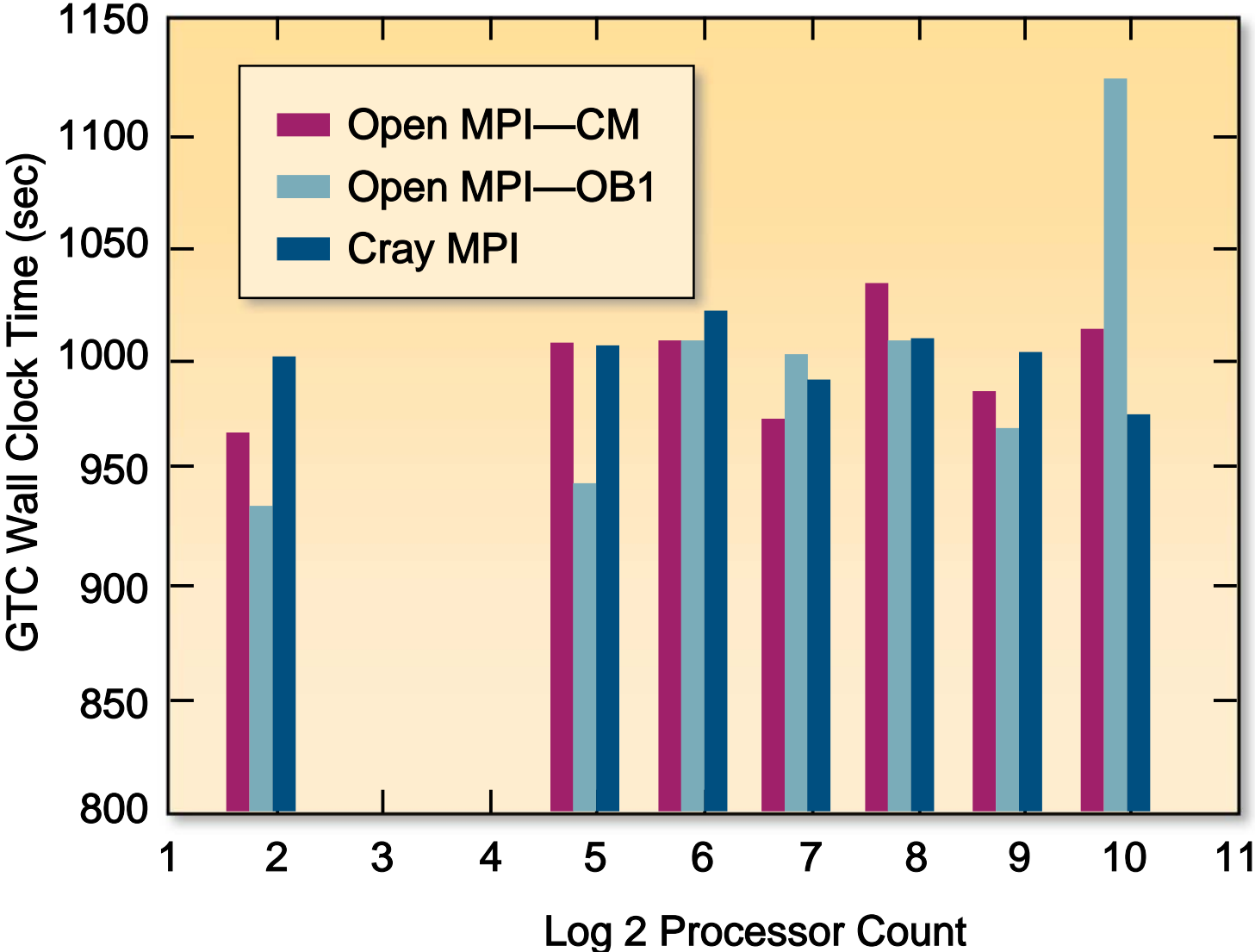
# Zero byte ping-pong latency

Open MPI—CM	4.91 $\mu$ sec
Open—OB1	6.16 $\mu$ sec
Cray MPI	4.78 $\mu$ sec

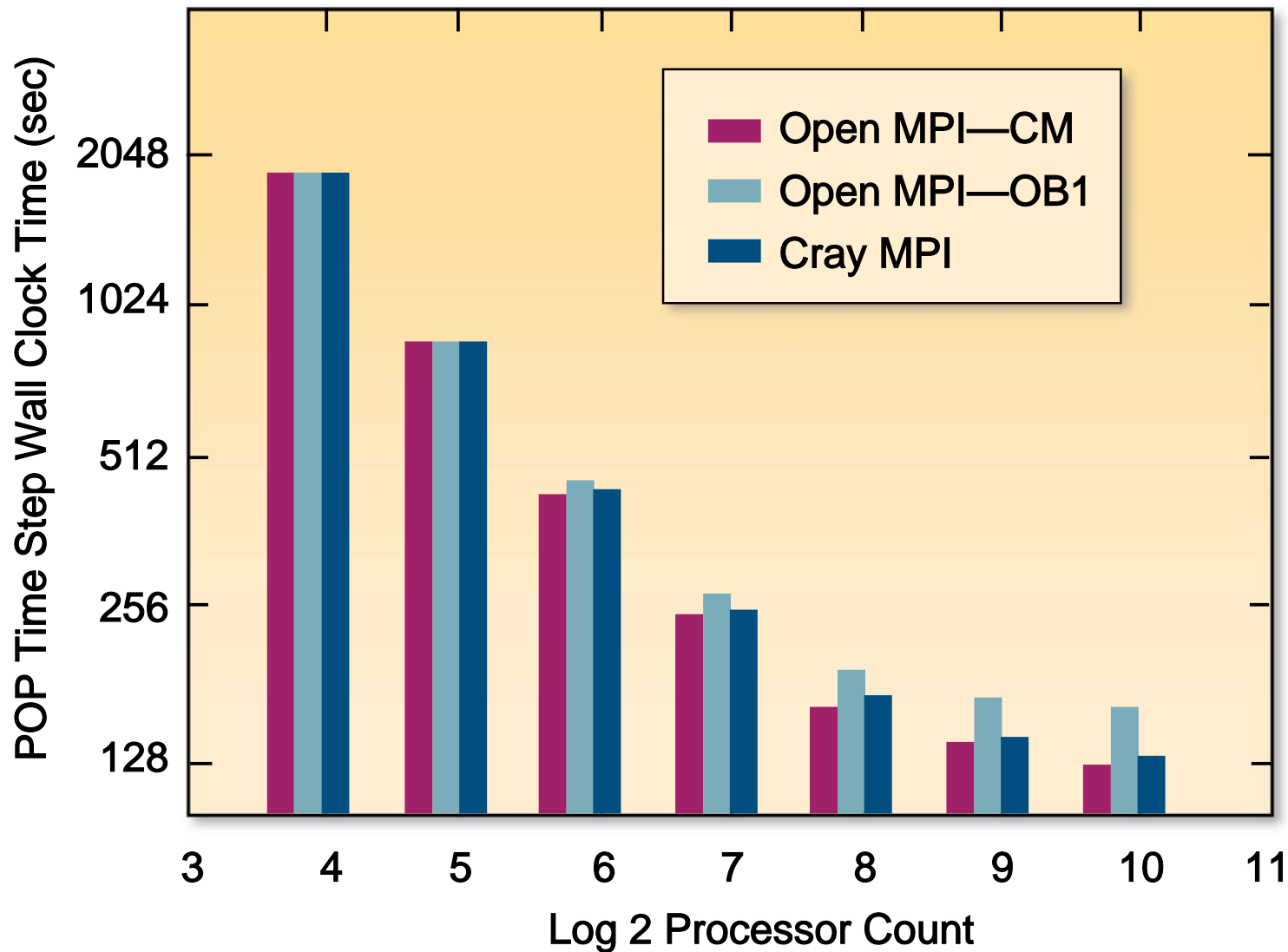
# VH1—Total runtime



# GTC—Total runtime



# POP—Step runtime



# Summary and future directions

- Support for XT (Catamount and Compute Node Linux) within standard distribution
- Performance (application and micro-benchmarks) comparable to that of Cray MPI
- Support for recovery from process failure is being added

# Contact

Richard L. Graham

Tech Integration

National Center for Computational Sciences

(865) 356-3469

rlgraham@ornl.gov



[www.open-mpi.org](http://www.open-mpi.org)