

Simulation of Advanced Large-Scale HPC Architectures

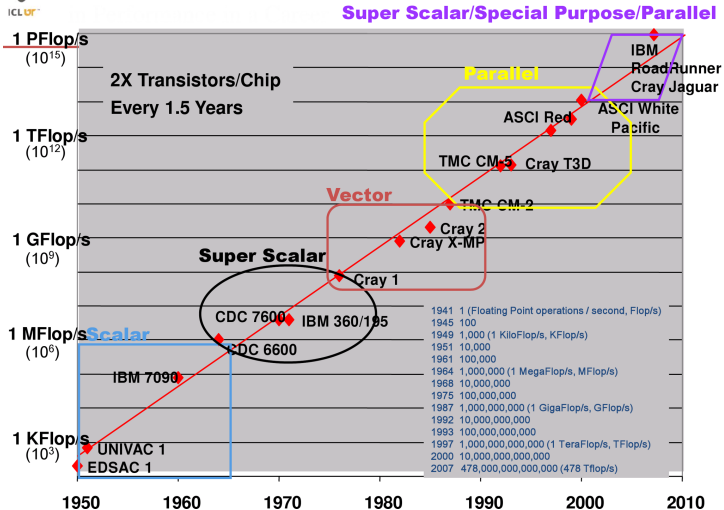
Frank Lauer

28. April 2010

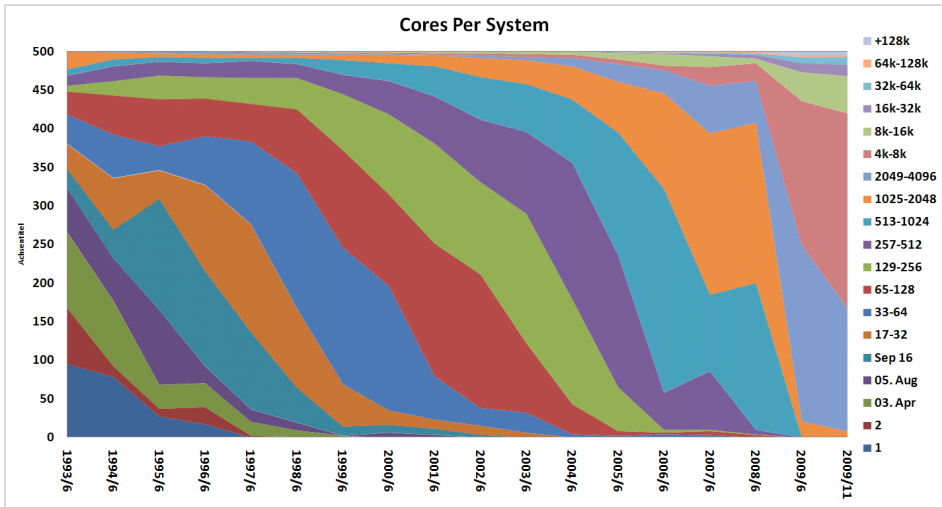
Overview

- Motivation
- Design
- Implementation / Results
- Future Aspects

History HPC



History HPC



Project AIM

The project aim is to provide a transparent platform for analysing MPI based applications in advanced large HPC architectures.

Project AIM

The project aim is to provide a transparent platform for analysing MPI based applications in advanced large HPC architectures.

- Run unmodified MPI code

Project AIM

The project aim is to provide a transparent platform for analysing MPI based applications in advanced large HPC architectures.

- Run unmodified MPI code
- Scale applications at least to 10^7 entities

Project AIM

The project aim is to provide a transparent platform for analysing MPI based applications in advanced large HPC architectures.

- Run unmodified MPI code
- Scale applications at least to 10^7 entities
- Fault injection into the program flow

Project AIM

The project aim is to provide a transparent platform for analysing MPI based applications in advanced large HPC architectures.

- Run unmodified MPI code
- Scale applications at least to 10^7 entities
- Fault injection into the program flow
- Collecting metric data during a run

Related Work

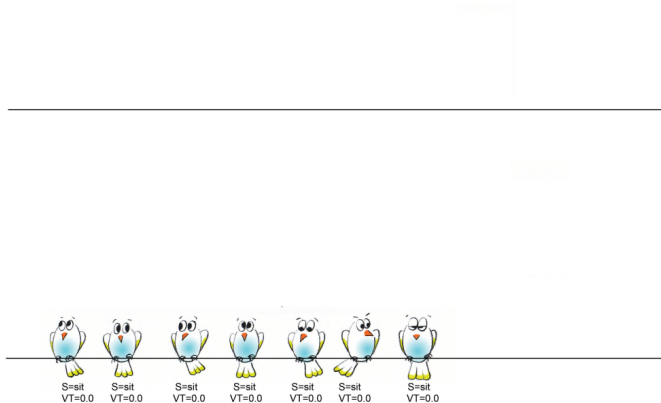
- BIGSIM
- JCAS
- SST
- $\mu\pi$
- AMPI CHAMP₊₊
- MPI-Sim

Design

- Project based on parallel discrete event simulation (PDES)

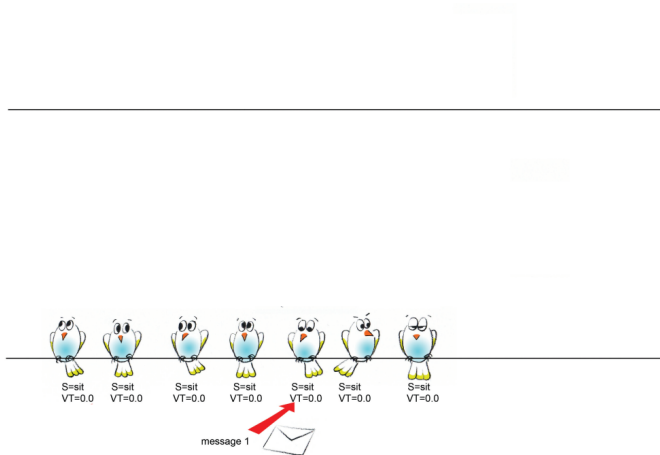
Design

- Project based on parallel discrete event simulation (PDES)



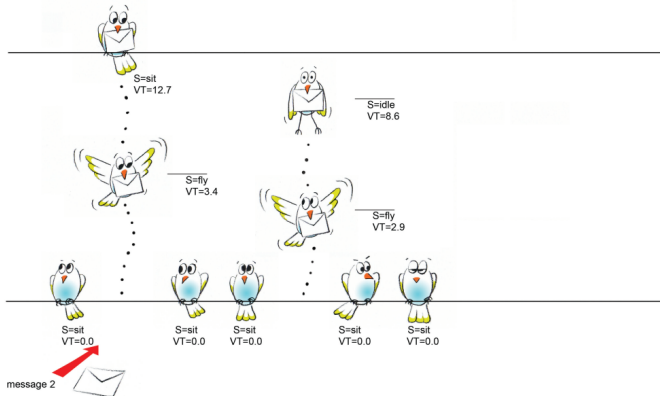
Design

- Project based on parallel discrete event simulation (PDES)



Design

- Project based on parallel discrete event simulation (PDES)



Design

- Project based on parallel discrete event simulation (PDES)
 - Logical process $\rightarrow$ Virtual compute node
 - Event $\rightarrow$ MPI message
 - Virtual time $\rightarrow$ CPU utilisation time + network delay

Design

- Project based on parallel discrete event simulation (PDES)
 - Logical process $\rightarrow$ Virtual compute node
 - Event $\rightarrow$ MPI message
 - Virtual time $\rightarrow$ CPU utilisation time + network delay
 - Global virtual time

Design

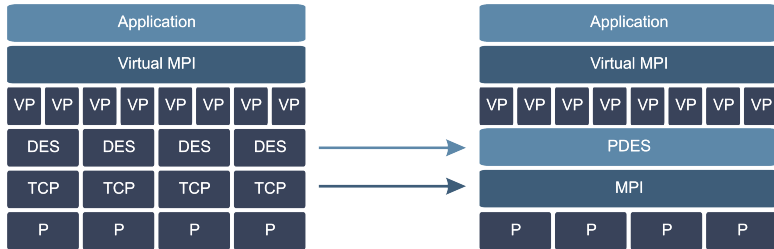
- Project based on parallel discrete event simulation (PDES)
 - Logical process $\rightarrow$ Virtual compute node
 - Event $\rightarrow$ MPI message
 - Virtual time $\rightarrow$ CPU utilisation time + network delay
 - Global virtual time
 - Synchronisations mechanisms

Design

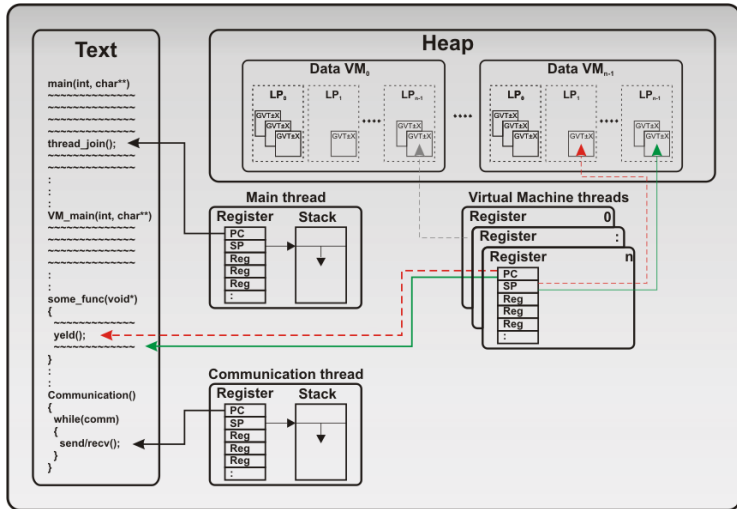
- Project based on parallel discrete event simulation (PDES)
 - Logical process $\rightarrow$ Virtual compute node
 - Event $\rightarrow$ MPI message
 - Virtual time $\rightarrow$ CPU utilisation time + network delay
 - Global virtual time
 - Synchronisations mechanisms
- Reuse of the JCAS architecture

Design

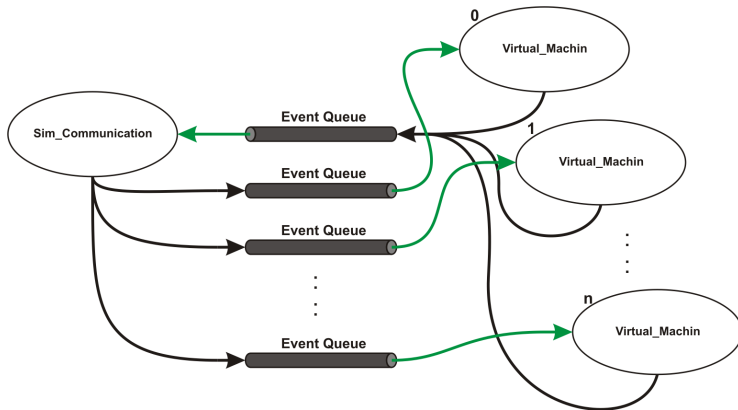
- Project based on parallel discrete event simulation (PDES)
 - Logical process $\rightarrow$ Virtual compute node
 - Event $\rightarrow$ MPI message
 - Virtual time $\rightarrow$ CPU utilisation time + network delay
 - Global virtual time
 - Synchronisations mechanisms
- Reuse of the JCAS architecture



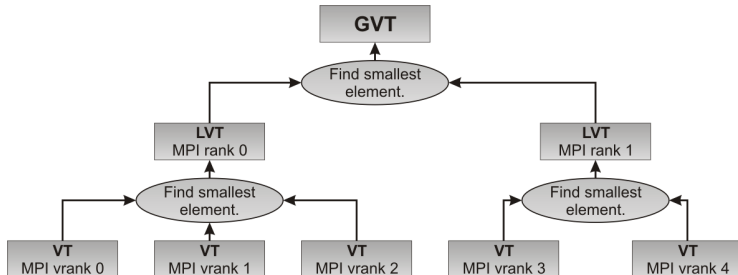
Design - Simulator Instance



Design - Internal Communication



Design - GVT Synchronisation



Implementation

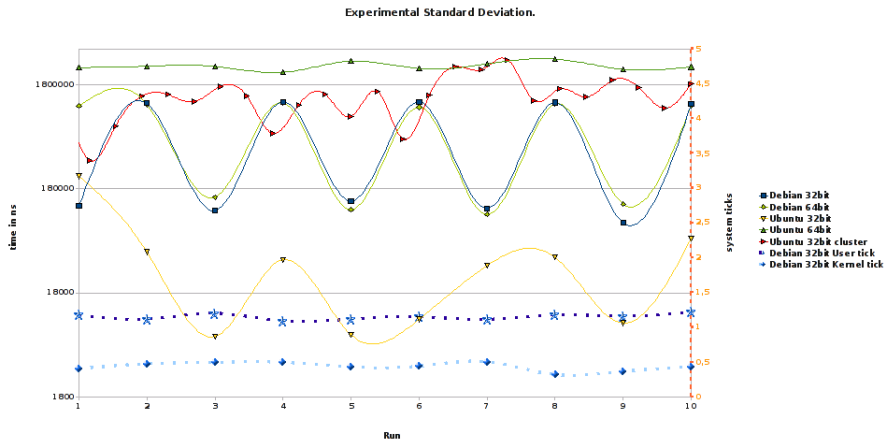
- Unmodified code
 - Add header file
 - Compile to object file
 - Link against simulator
- Supported MPI-calls
 - MPI_Init(...)
 - MPI_Finalise()
 - MPI_Comm_rank(...)
 - MPI_Comm_size(...)
 - MPI_Isend(...)
 - MPI_Recv(...)
- Simulation size only restricted by memory
- Variable stack size

Implementation - Known Restrictions/Issues

- Wrapper function calls only in `main(...)`
- Memory Segmentation fault
- Memory out of Resources
- `printf(...)` floating point values in 64bit

Test Result - Record Virtual Time

- Time accounting for the VT and the GVT synchronisation

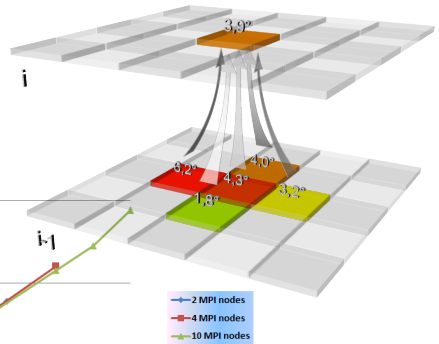
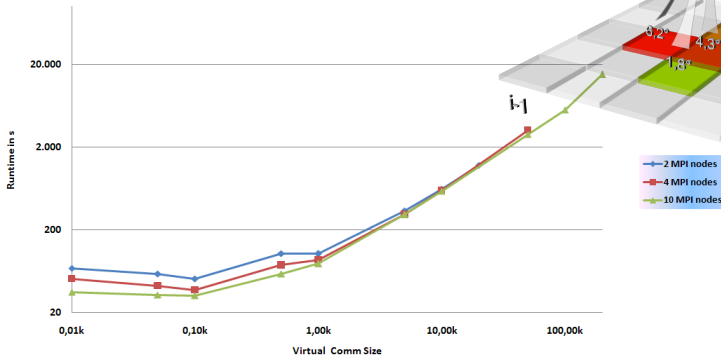


Simulation Test App. one: Heat Transfer

Determine the heat distribution on a two dimensional area

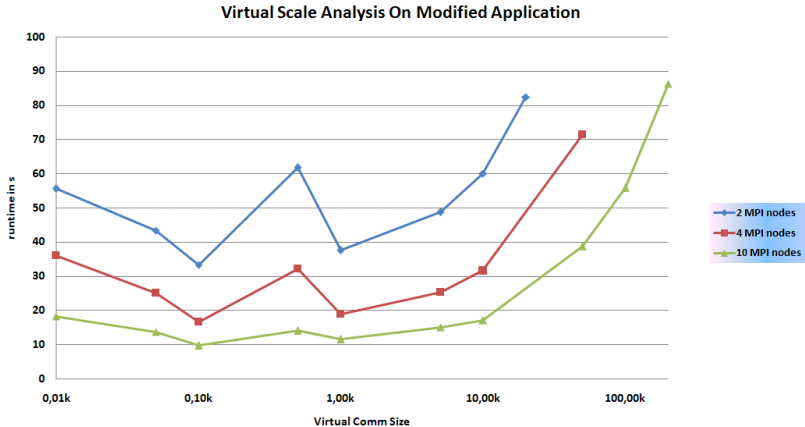
$$T_i = \frac{T_{i-1} + T_{i-1}^{up} + T_{i-1}^{down} + T_{i-1}^{right} + T_{i-1}^{left}}{5}$$

Virtual Scale Analysis



Simulation Test App. one: Heat Transfer

Modified to non blocking finalization.

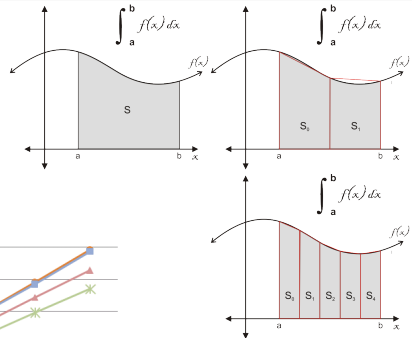
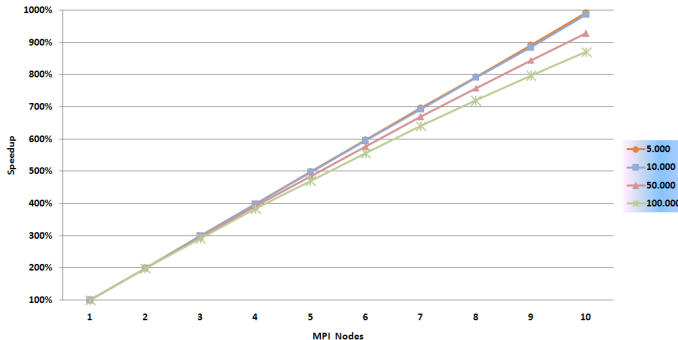


Simulation Test App. two: Numerical Quadrature

Solve a integral numerical, by dividing it into subintegrals and using approximation:

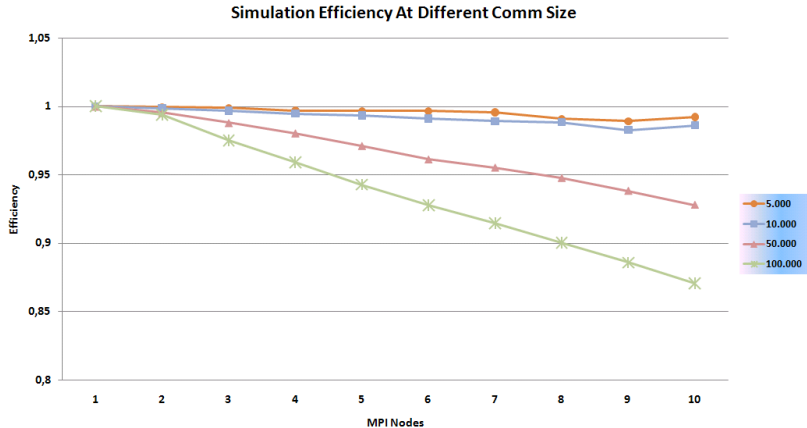
$$S = \sum_{i=0}^{n-1} A_i, \quad A_i = \frac{h_i(f(x_i) + f(x_{i+1}))}{2}$$

Simulation Speedup At Diffrent Comm Size



Simulation Test App. two: Numerical Quadrature

Modified, where no data is gathered.



Future Aspects

- Embed more complex MPI applications

Future Aspects

- Embed more complex MPI applications
- Virtual time for LP

Future Aspects

- Embed more complex MPI applications
- Virtual time for LP
- Enhance MPI instruction set

Future Aspects

- Embed more complex MPI applications
- Virtual time for LP
- Enhance MPI instruction set
- Fault injection

Questions ?

Thank you for your attention!