

Coupling DDT and Marmot for Debugging of MPI Applications

Bettina Krammer, Valentin Himmler
University of Stuttgart
High-Performance Computing-Center Stuttgart (HLRS)

www.hlrs.de

David Lecomber

Allinea Software

www.allinea.com



Hochleistungsrechenzentrum Stuttgart

ParMA

Parco2007, Jülich

HLRTS

allinea
SCALE TO NEW HEIGHTS

Outline

- Motivation
- Introduction to DDT and Marmot
- Coupling DDT and Marmot
- Conclusions



Slide 2

Hochleistungsrechenzentrum Stuttgart

ParMA

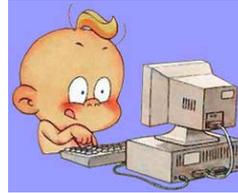
Parco2007, Jülich

HLRTS

allinea
SCALE TO NEW HEIGHTS

Motivation

- Parallel programming is complex and error-prone
- Couple stand-alone tools to cover different levels of debugging and correctness checking of MPI applications
- Need for a user-friendly environment



Slide 3

Hochleistungsrechenzentrum Stuttgart

ParMA

Parco2007, Jülich

H L R T S

allinea
SCALE TO NEW HEIGHTS

Introduction to DDT and Marmot



Slide 4

Hochleistungsrechenzentrum Stuttgart

ParMA

Parco2007, Jülich

H L R T S

allinea
SCALE TO NEW HEIGHTS

DDT – Distributed Debugging Tool

- Source-level debugger for scalar, multi-threaded and large-scale parallel C, C++ and Fortran codes
- Support for all MPIs
- Easy-to-use graphical interface
- www.allinea.com



Slide 5

Hochleistungsrechenzentrum Stuttgart

ParMA

Parco2007, Jülich

H L R T S

allinea
SCALE TO NEW HEIGHTS

What is Marmot?



Slide 6

Hochleistungsrechenzentrum Stuttgart

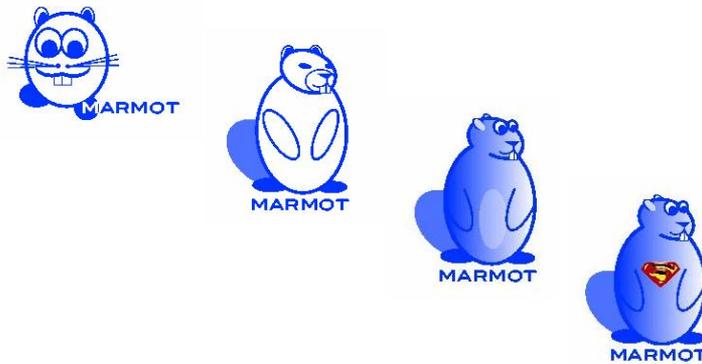
ParMA

Parco2007, Jülich

H L R T S

allinea
SCALE TO NEW HEIGHTS

History of Marmot



Slide 7

Hochleistungsrechenzentrum Stuttgart

ParMA

Parco2007, Jülich

H L R T S

allinea
SCALE TO NEW HEIGHTS

Now – what Marmot REALLY is

- Tool for the development of MPI applications
- Automatic correctness checking at runtime:
 - Detect incorrect use of MPI
 - Detect non-portable constructs
 - Detect possible race conditions and deadlocks
- no source code modification, just relinking and run with 1 additional process
- Different output formats (txt, html, xml)
- C and Fortran binding of MPI -1.2 is supported, also C++ and mixed C/Fortran code
- Development is still ongoing (not every possible functionality is implemented yet...)
- Tool makes use of the so-called *profiling interface*



Slide 8

Hochleistungsrechenzentrum Stuttgart

ParMA

Parco2007, Jülich

H L R T S

allinea
SCALE TO NEW HEIGHTS

output html

export MARMOT_LOGFILE_TYPE=1

```
0 0 None Test performing  
Call: MPI_Isend  
Test performing  
Call: MPI_Isend  
Test: ERROR: MPI_Isend Request is still in use !!  
Argument request  
0 0 None Information for Reuse of type MPI_Request  
instead of request-reuse.c line 55  
and get freed.  
Call: MPI_Isend  
Test: ERROR: MPI_Isend Request is still in use !!  
Argument request  
0 0 None Information for Reuse of type MPI_Request  
instead of request-reuse.c line 55  
and get freed.  
Call: MPI_Isend  
Test performing  
Call: MPI_Wait  
Test performing  
Call: MPI_Wait  
Test performing  
Call: MPI_Finalize  
Test: NOTE: do not need a maximum of 1 request!  
Test: WARNING: MPI_Finalize: There are still pending messages!  
Call: MPI_Finalize  
Test: WARNING: all clients are pending!
```



Slide 9

Hochleistungsrechenzentrum Stuttgart

ParMA

Parco2007, Jülich

HLRZS

allinea
SCALE TO NEW HEIGHTS

output cube

export MARMOT_LOGFILE_TYPE=2

```
0 Messages  
28 Infos  
0 Warnings  
1 WARNING - Pending messages  
14 Notes  
0 Errors  
2 ERROR - Request is still in use  
2 WARNING - A Deadlock might f
```

```
0 request-reuse2.c  
0 MPI_Init @line: 47  
0 MPI_Comm_size @line: 48  
0 MPI_Comm_rank @line: 49  
0 MPI_Isend @line: 55  
0 MPI_Isend @line: 69  
1 MPI_Recv @line: 59  
1 MPI_Isend @line: 73  
0 MPI_Wait @line: 61  
0 MPI_Wait @line: 75  
0 MPI_Finalize @line: 78
```

```
0 MPI-Environment  
0 MPI-Processes  
0 rank 0  
0 rank 1
```



Slide 10

Hochleistungsrechenzentrum Stuttgart

ParMA

Parco2007, Jülich

HLRZS

allinea
SCALE TO NEW HEIGHTS

Marmot

- Portable tool , tested on many platforms (ia32/ia64, Opteron/Xeon, IBM, SX6, SX8,...)



- Different compilers (Intel, GNU, NEC,...)
- Different MPIs (mpich, Open MPI, Lam-MPI, NEC MPI, intel MPI, Voltaire MPI,...)
- Marmot works basically anywhere, main challenges:
 - Guess all configure options right
 - **not possible to find everything automatically, e.g. --with-mpi-dir=... etc.**
 - Link examples correctly
- www.hlr.de/organization/amt/services/tools/debugger/marmot



Slide 11

Höchstleistungsrechenzentrum Stuttgart

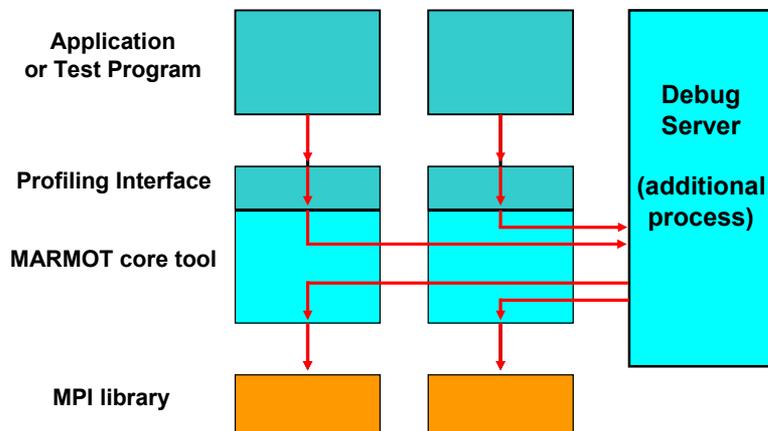
ParMA

Parco2007, Jülich

HLRZS

allinea
SCALE TO NEW HEIGHTS

Design of MARMOT (1)



Slide 12

Höchstleistungsrechenzentrum Stuttgart

ParMA

Parco2007, Jülich

HLRZS

allinea
SCALE TO NEW HEIGHTS

DDT Marmot



Slide 13 Höchstleistungsrechenzentrum Stuttgart

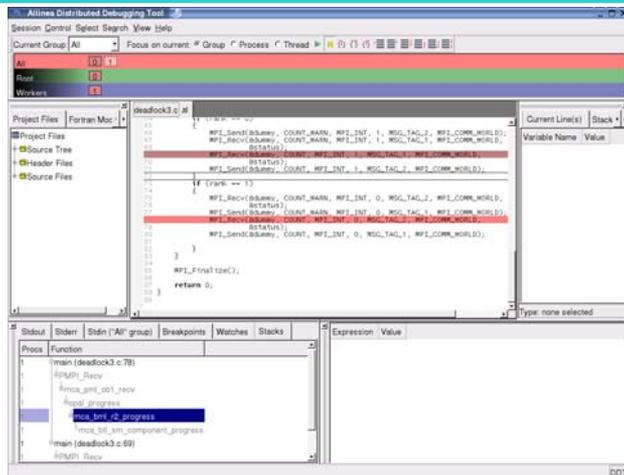
ParMA

Parco2007, Jülich

H L R T S 

allinea
SCALE TO NEW HEIGHTS

Simple example – with DDT, without Marmot



```
1 // MPI_Send
2 MPI_Send(&array, COUNT_MAIN, MPI_INT, 1, MSG_TAG_2, MPI_COMM_WORLD);
3 MPI_Recv(&array, COUNT_MAIN, MPI_INT, 1, MSG_TAG_1, MPI_COMM_WORLD);
4 MPI_Status(status);
5 MPI_Send(&array, COUNT_MAIN, MPI_INT, 1, MSG_TAG_1, MPI_COMM_WORLD);
6 MPI_Recv(&array, COUNT_MAIN, MPI_INT, 0, MSG_TAG_2, MPI_COMM_WORLD);
7 MPI_Status(status);
8 MPI_Send(&array, COUNT_MAIN, MPI_INT, 0, MSG_TAG_1, MPI_COMM_WORLD);
9 MPI_Recv(&array, COUNT_MAIN, MPI_INT, 0, MSG_TAG_1, MPI_COMM_WORLD);
10 MPI_Status(status);
11 MPI_Send(&array, COUNT_MAIN, MPI_INT, 0, MSG_TAG_1, MPI_COMM_WORLD);
12 MPI_Finalize();
13 return 0;
```



Slide 14 Höchstleistungsrechenzentrum Stuttgart

ParMA

Parco2007, Jülich

H L R T S 

allinea
SCALE TO NEW HEIGHTS

Design of MARMOT (2)

Client (Application)

```
...
MPI_Init{
  PMPI_Init();
  initComm;
  MapResources;
  ...
}
...
MPI_Recv{
  doSomeChecks;
  ...
  PMPI_Recv;
}
...
```

Debug Server

```
...
MPI_Init{
  PMPI_Init();
  initComm;
  MapResources;
  ...
  //don't leave MPI_Init
  DebugServerCode;
  logging;
  ...
}
```



Slide 15

Hochleistungsrechenzentrum Stuttgart

ParMA

Parco2007, Jülich

H L R T S

allinea
SCALE TO NEW HEIGHTS

Integration with DDT

- Marmot's special MPI_Init (last process):
export DDT_MPI_Init=PMPI_Init
- Compile everything with **-g** (including marmot code)



Slide 16

Hochleistungsrechenzentrum Stuttgart

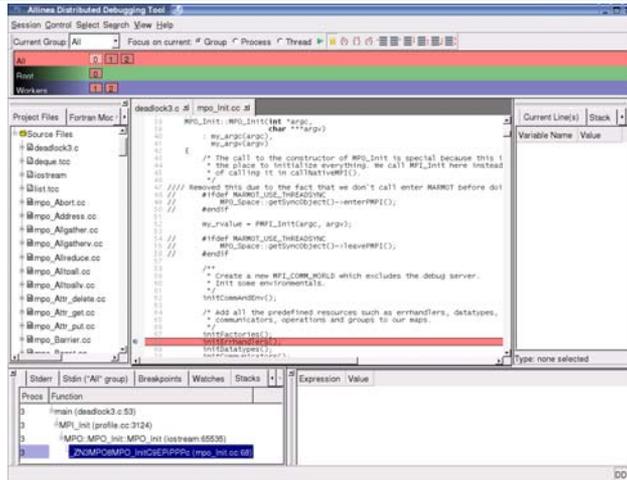
ParMA

Parco2007, Jülich

H L R T S

allinea
SCALE TO NEW HEIGHTS

Simple example – with DDT & Marmot (scary, isn't it)



Design of MARMOT (3)

Client (Application)

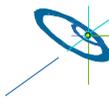
```

...
MPI_Init{
  PMPI_Init();
  initComm;
  MapResources;
  ...
}
...
MPI_Recv{
  doSomeChecks;
  ...
  PMPI_Recv;
  ...
}
...
    
```

Debug Server

```

...
MPI_Init{
  PMPI_Init();
  initComm;
  MapResources;
  ...
  //don't leave MPI_Init
  DebugServerCode;
  logging;
  ...
  if (error)
    insertBreakpointError;
  if (warning)
    insertBreakpointWarning;
  ...
}
    
```



DDT & Marmot

- Insert breakpoint calls in our debug server (highest rank) when error/warning is detected (mpo-breakpoints.cc ~ 30 lines of code)
- Just compile Marmot's mpo-breakpoints.cc source file with `-g`
- Display Marmot warnings (stderr/variables windows, further ideas)



Slide 19

Hochleistungsrechenzentrum Stuttgart

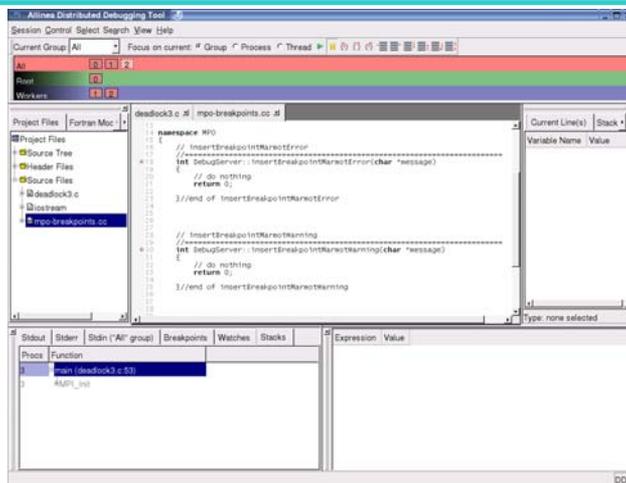
ParMA

Parco2007, Jülich

H L R T S

allinea
SCALE TO NEW HEIGHTS

Simple example – with DDT & Marmot (breakpoints)



Slide 20

Hochleistungsrechenzentrum Stuttgart

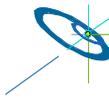
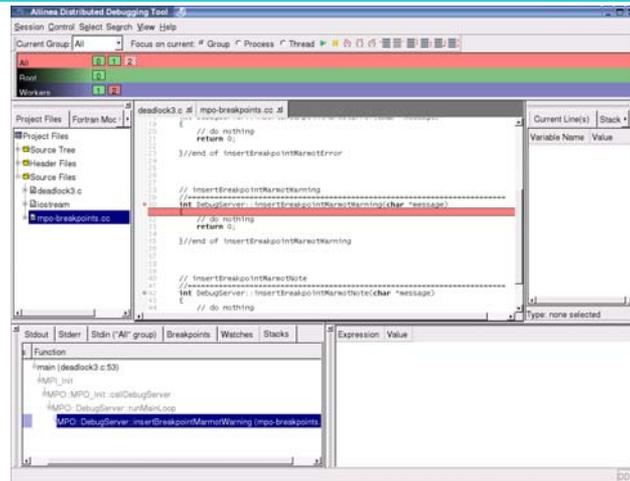
ParMA

Parco2007, Jülich

H L R T S

allinea
SCALE TO NEW HEIGHTS

Simple example – with DDT & Marmot (warning detected)



Slide 21 Höchstleistungsrechenzentrum Stuttgart

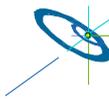
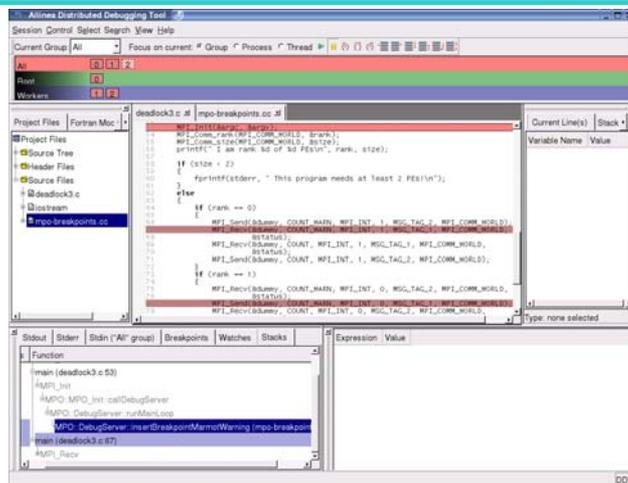
ParMA

Parco2007, Jülich

HLRS

allinea
SCALE TO NEW HEIGHTS

Simple example – with DDT & Marmot (warning detected)



Slide 22 Höchstleistungsrechenzentrum Stuttgart

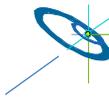
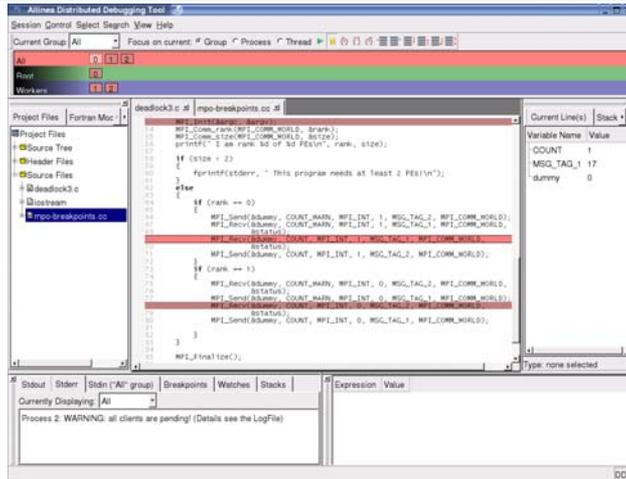
ParMA

Parco2007, Jülich

HLRS

allinea
SCALE TO NEW HEIGHTS

Simple example – with DDT & Marmot (error display)



Slide 23 Höchstleistungsrechenzentrum Stuttgart

ParMA

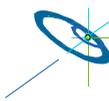
Parco2007, Jülich

HLRS

allinea
SCALE TO NEW HEIGHTS

Conclusions and Future Work

- DDT & Marmot – it works!
- It's not rocket science... but takes some considerations, e.g.
 - how to handle this last Marmot process (hide it completely eventually? DDT to add breakpoints automatically)
 - how to handle error dialogues? Add information? Message queues?
 - adapt Marmot's build process (shared libraries) to be able to switch Marmot on/off through DDT GUI (LD_PRELOAD)



Slide 24 Höchstleistungsrechenzentrum Stuttgart

ParMA

Parco2007, Jülich

HLRS

allinea
SCALE TO NEW HEIGHTS

Conclusions and Future Work

- Tested with simple examples on various platforms
 - Tests with real applications to be performed
 - Shared lib approach tested on few platforms



Slide 25

Hochleistungsrechenzentrum Stuttgart

ParMA

Parco2007, Jülich

HLRTS

allinea
SCALE TO NEW HEIGHTS

Conclusions and Future Work

- Marmot is alive and kickin'
- It's under active development by HLRS and ZIH
- Also within the ParMA project – Parallel Programming for Multi-core Architectures
www.parma-itea2.org
(collaboration with Vampir, Kojak, DDT, OPT,...)



allinea ddt
the distributed debugging tool

allinea opt
the optimization and profiling tool



Slide 26

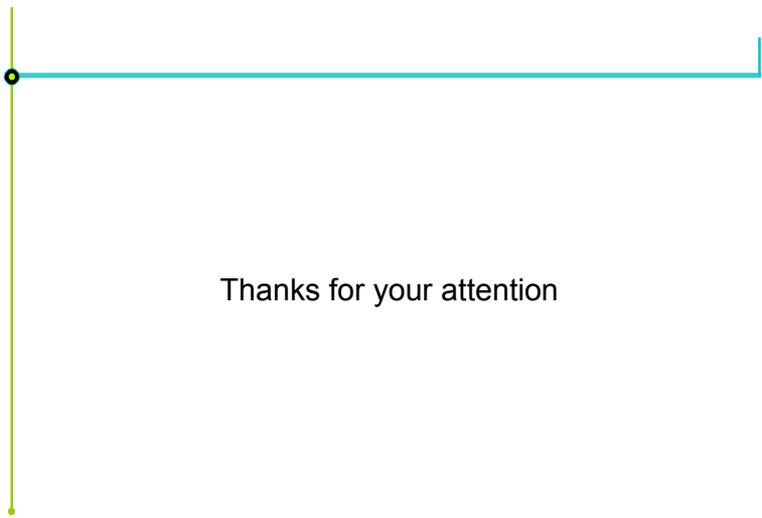
Hochleistungsrechenzentrum Stuttgart

ParMA

Parco2007, Jülich

HLRTS

allinea
SCALE TO NEW HEIGHTS



Thanks for your attention



Slide 27

Hochleistungsrechenzentrum Stuttgart

ParMA

Parco2007, Jülich

H L R T S 

allinea
SCALE TO NEW HEIGHTS