

Matthias Lieber
Center for Information Services and High Performance Computing (ZIH)

Parallel Debugging with DDT

Parallel Programming with MPI, OpenMP, and Tools
Dresden, 8-12 February 2021

Why using a Debugger?

Your program shows incomprehensible behavior, e.g.

- Program terminates abnormally

```
% icc myprog.c -o myprog  
% ./myprog  
Segmentation fault
```

- Program produces wrong results

```
% ./myprog  
Pi = 3.573
```

You want to know what your program is (really) doing

What can a Debugger do?

Observe a running program:

- Print variables (scalars, arrays, structures / derived types, classes)
- Inform about current source code line and function (function call stack)

Control program execution:

- Stop the program at a specific source code line (**Breakpoints**)
- Stop the program when certain conditions are true (**Conditional Breakpoints** and Watchpoints)
- Stop the program before terminating abnormally
- Execute the program line-by-line (**Stepping**)

Typical Usage of a Debugger

Development workflow

- Compile the program with **-g**

```
mpif90 -g myprog.c -o myprog
```
- Run the program under control of the debugger

```
dtt <mpirun command> ./myprog
```
- Use the Debugger to locate the position of the problem and examine variables
- Understand the cause of the problem and correct the source code
- Repeat until problem is solved



Hints:

Always compile your application with the **-g flag**, especially during developing and testing. It adds **symbolic debug info** to the binary and has no performance impact.

Optimizations often interfere with debugging (e.g. functions or variables of interest are “optimized away”). If necessary, compile with the **-O0 flag** to disable optimizations.

Debugger Operation Modes

Start program under debugger control

- Most common way to use a debugger
- Not useful if you want to observe what the program does after a long runtime or you do not expect problems

Attach to an already running program

- Program was not started under debugger
- Useful if program has been running for a long time

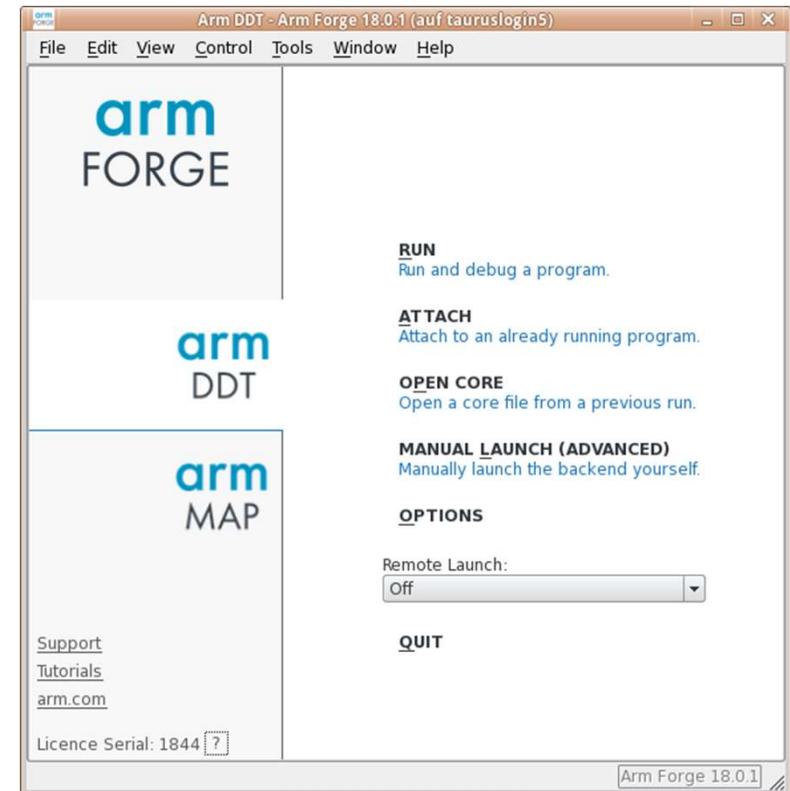
Core files / core dumps

- Core files are memory state of a crashed program written to file
- Only static analysis of program's data after termination
- Useful if you don't expect a crash or don't want to wait until a crash happens (probably after long runtime)

Arm DDT

Distributed Debugging Tool

- Commercial debugging tool by Arm (Arm acquired Allinea in 2016)
- Languages: C, C++, Fortran
- Parallel Support: Pthreads, OpenMP, MPI, PGAS languages, CUDA, OpenACC
- Available for all common HPC platforms
- Intuitive graphical user interface



- More info:
<https://developer.arm.com/tools-and-software/server-and-hpc/debug-and-profile/arm-forge/arm-ddt>

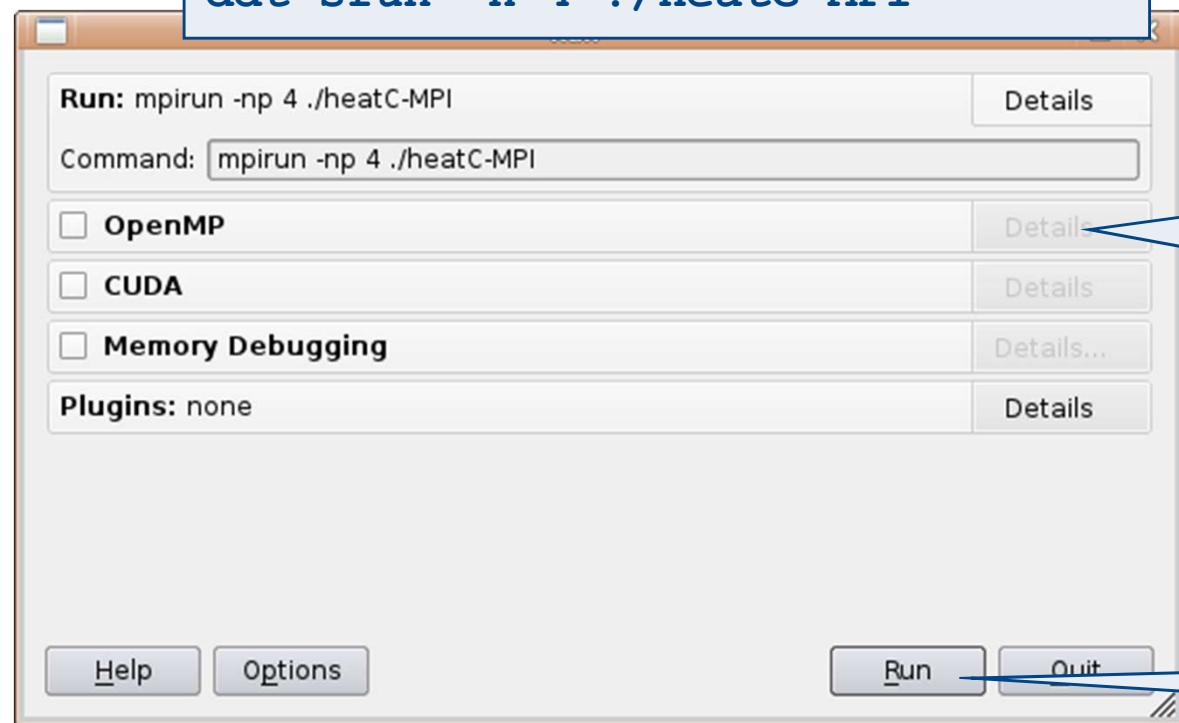
DDT: Program Start

```
% mpicc -g -O0 heatC-MPI.c -o heatC-MPI  
% ddt mpirun -np 4 ./heatC-MPI
```

Compile with Debugging

On Taurus we use srun instead of mpirun:
`ddt srun -n 4 ./heatC-MPI`

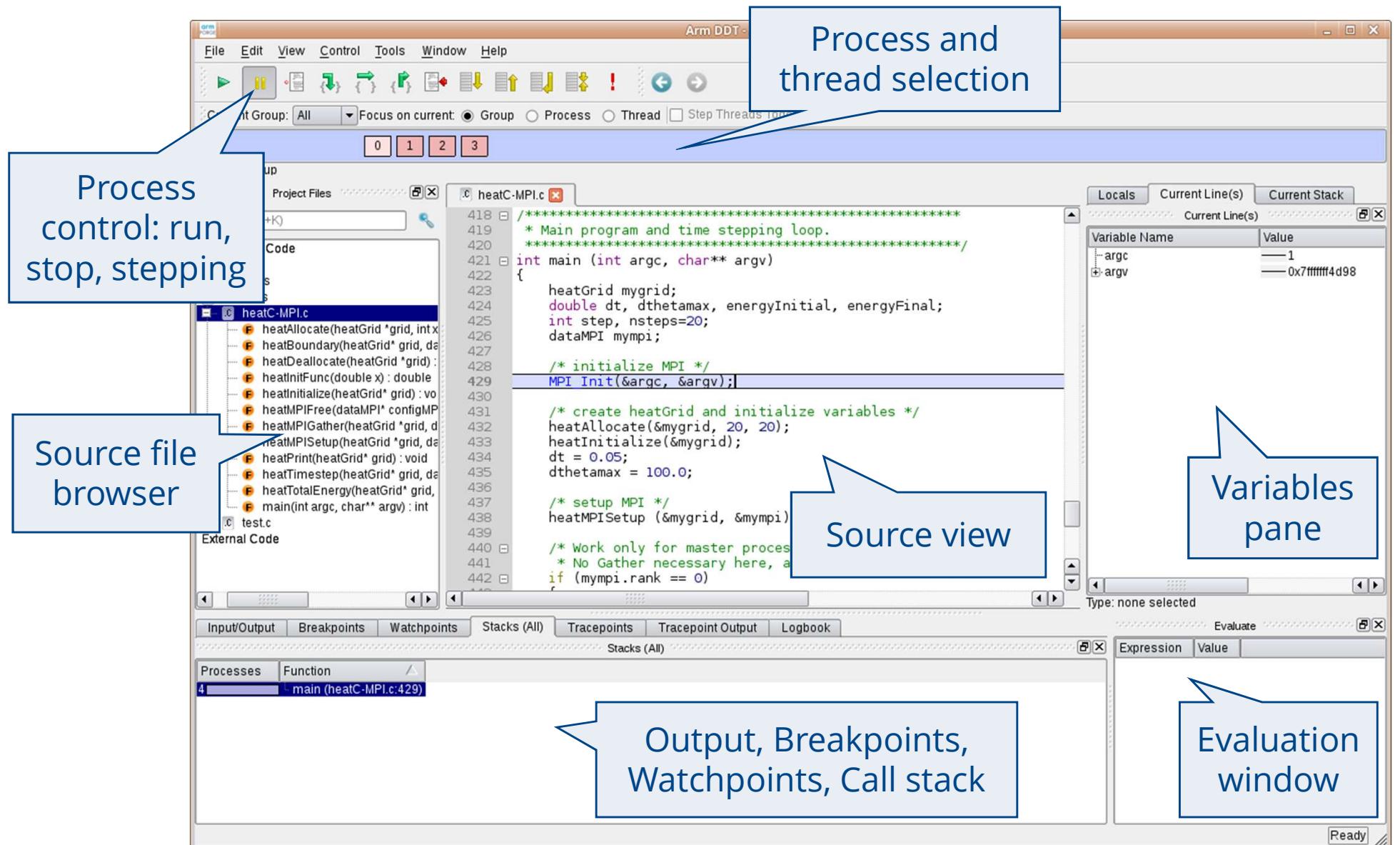
Start DDT: prepend `ddt` to mpirun command line



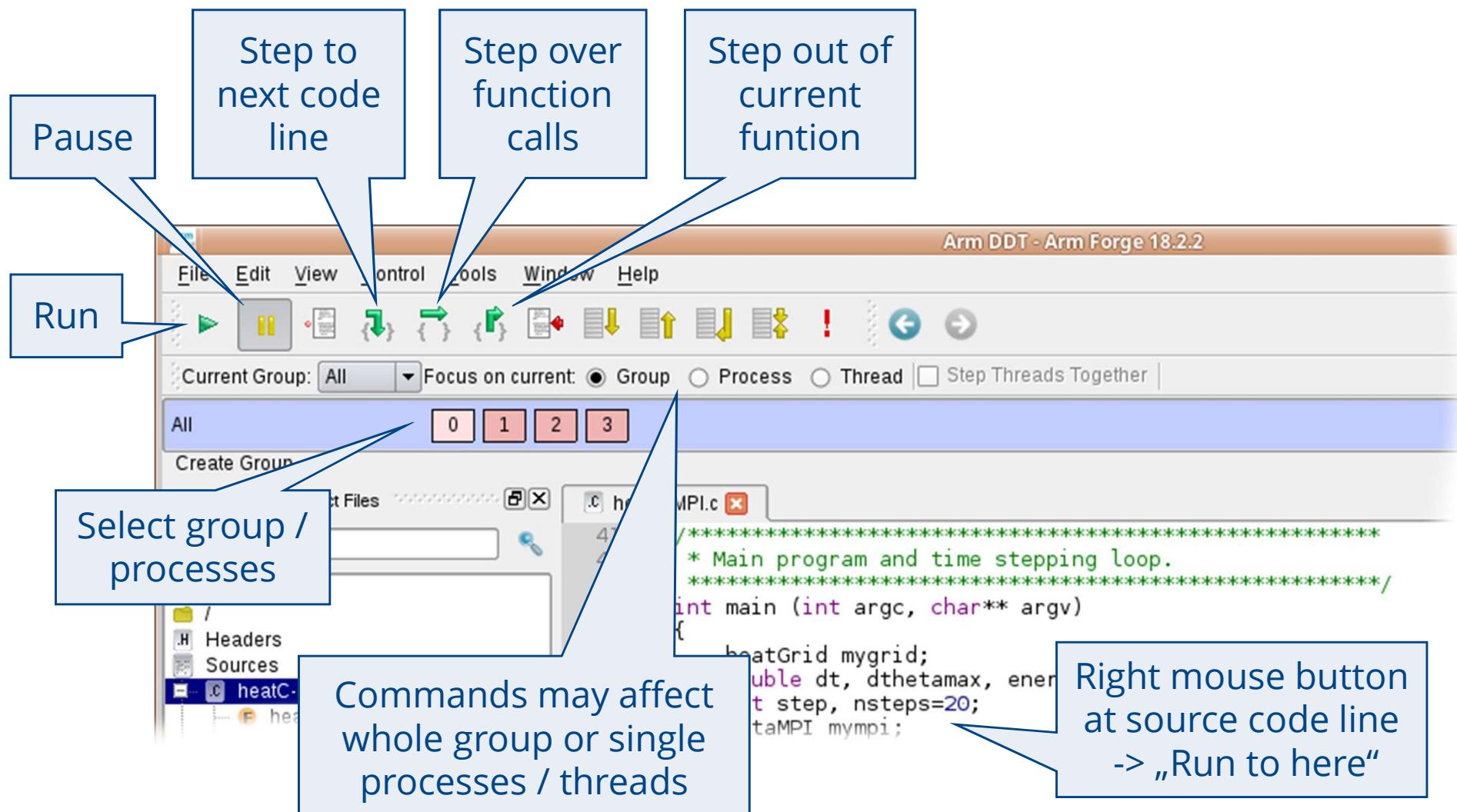
Enable/disable OpenMP and set number of threads if necessary

Start Program

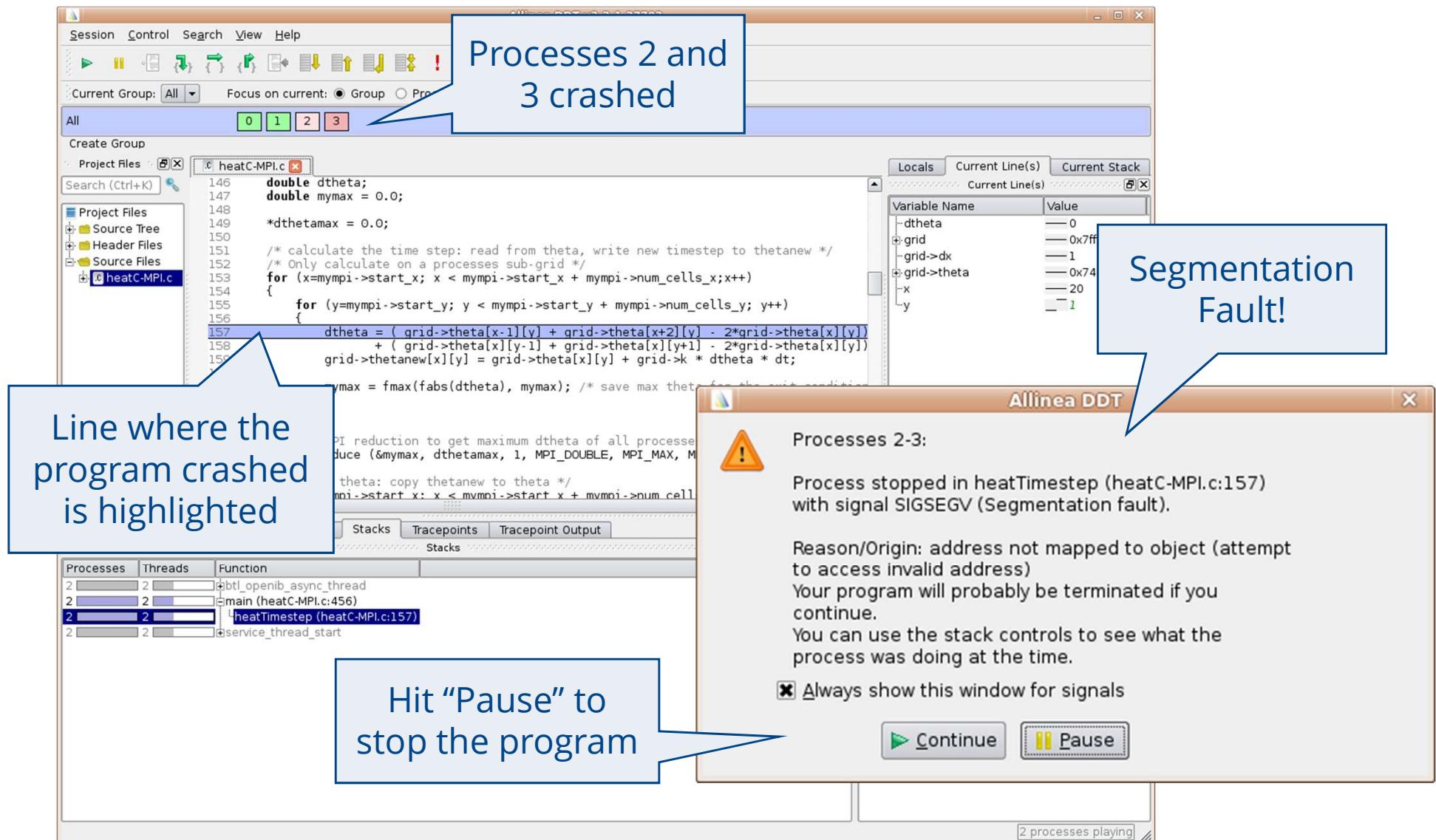
DDT: Main Window



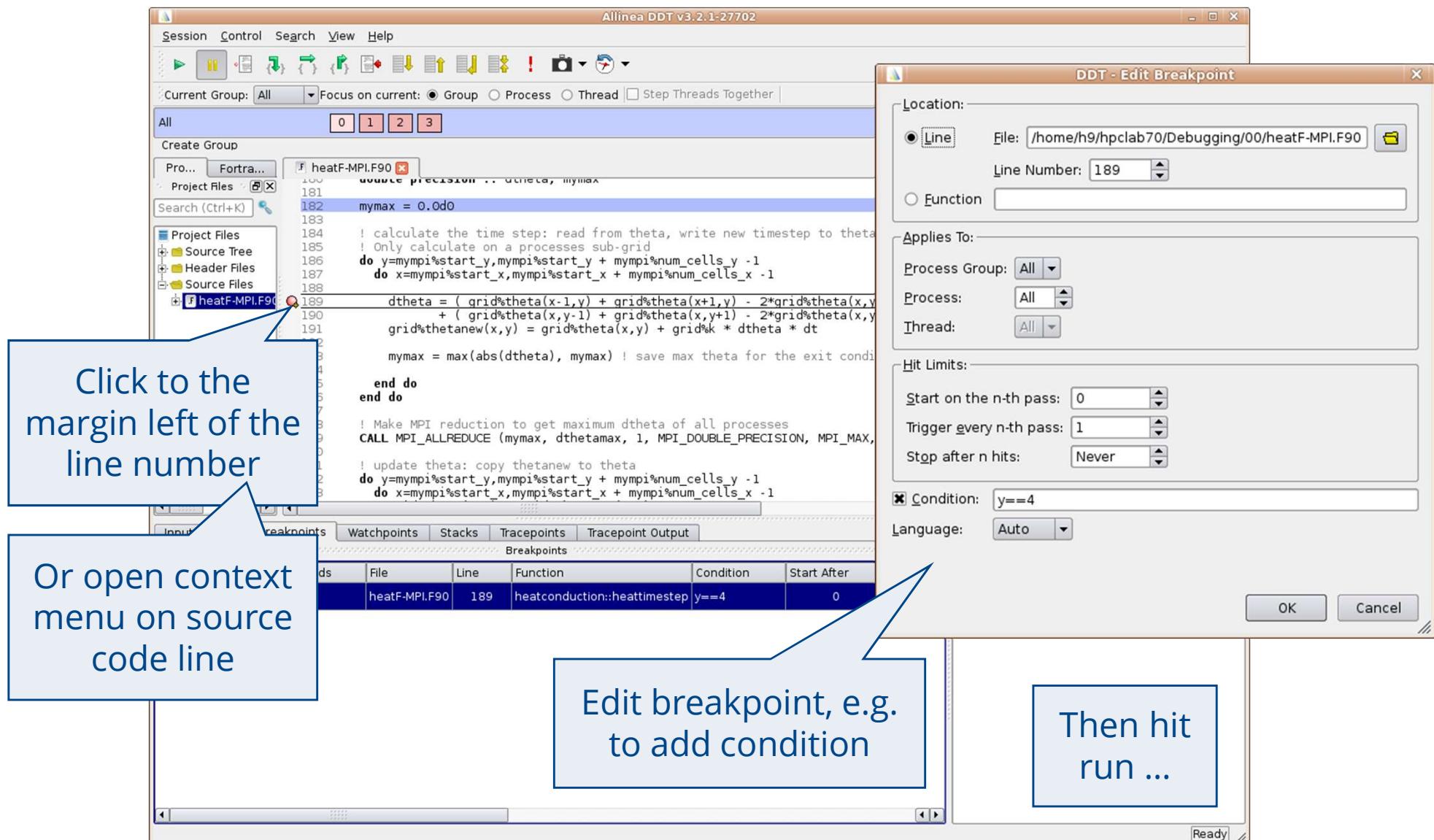
DDT: Process Control & Stepping



DDT: Segmentation Fault



DDT: Breakpoints (1/2)



DDT: Breakpoints (2/2)

The screenshot shows the Allinea DDT interface version 3.2.1-27702. The main window displays a Fortran source file named `heatF-MPI.F90`. A conditional breakpoint is set at line 189. The code implements a finite difference scheme for heat conduction, calculating a new timestep and updating the grid. The variable `dtheta` is calculated as the average of the differences between adjacent grid points. The `grid%thetanew` variable is updated by adding `dtheta` scaled by `dt`.

The Locals pane shows the current values of variables:

Variable Name	Value
<code>dt</code>	0.0500000000
<code>dtheta</code>	0
<code>dthetamax</code>	100
<code>err</code>	0
<code>grid</code>	(rank = 0, call)
<code>mymax</code>	0
<code>mypi</code>	1
<code>x</code>	1
<code>y</code>	4

A modal dialog box titled "Allinea DDT" is displayed, stating "Processes 0,2: Process stopped at breakpoint in heatconduction::heattimestep (heatF-MPI.F90:189)." It includes a checkbox "Always show this window for user-defined breakpoints" and buttons for "Continue" and "Pause".

A callout box in the foreground states: "Processes 0 and 2 stopped at conditional breakpoint".

DDT Practical 1: Conditional Breakpoints

C:

```
% cd ~/Debugging/c  
% mpicc -g -O0 heatC-MPI.c -o heatC-MPI  
% ddt srun -n 4 ./heatC-MPI
```

Fortran 90:

```
% cd ~/Debugging/f90  
% mpif90 -g -O0 heatF-MPI.F90 -o heatF-MPI  
% ddt srun -n 4 ./heatF-MPI
```

In the DDT run window:
unchecked OpenMP, CUDA,
Mem. debugging and hit run

Task A:

- Find out the value of dthetamax after step 10 has been computed.
- Hint: Use a conditional breakpoint in the time stepping loop (main program)

Task B (optional):

- Which process contributed the maximum to dthetamax at the MPI_Allreduce in heatTimestep after step 10 has been computed?
- Hint: use an additional breakpoint at the MPI_Allreduce, then right click on the variable mymax in the variables pane and select “Compare Across Processes”

DDT Practical 1: Task A Solution

Breakpoint in time stepping loop, condition: step==11

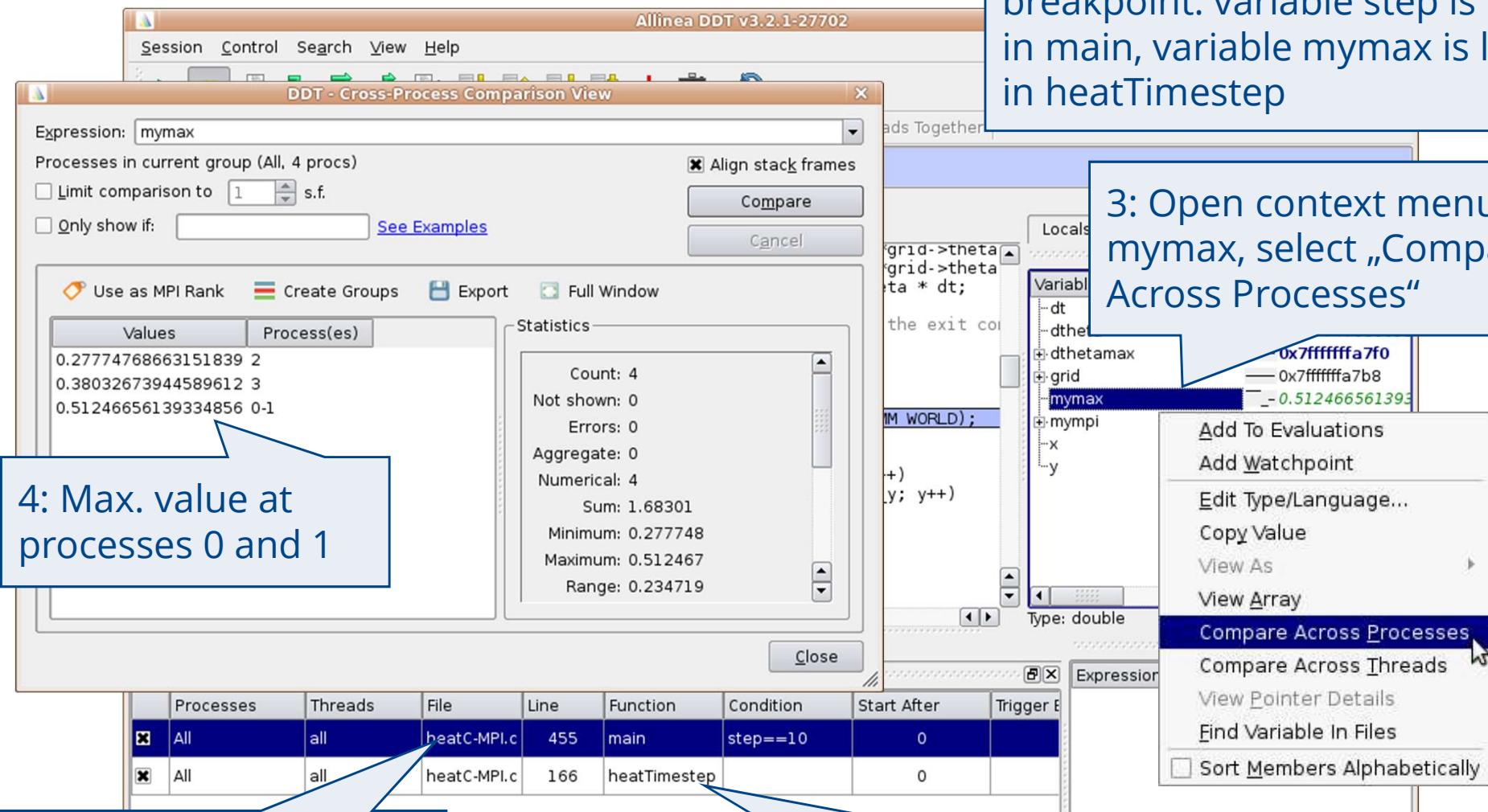
dthetamax = 0,512

```
447      /* energy of initial grid */
448      heatTotalEnergy(&mygrid, &energyInitial);
449  }
450
451  /* time stepping loop */
452  for( step=1 ; step<=nsteps ; step++)
453  {
454      heatBoundary(&mygrid, &mympi);
455      heatTimestep(&mygrid, &mympi, dt, &dthetamax );
456
457
458
459  /* Gather data on process 0 for output*/
heatMPIGather (&mygrid, &mympi);

/* Work only for master process*/
if (mympi.rank == 0)
{
    /* output of final grid */
}
```

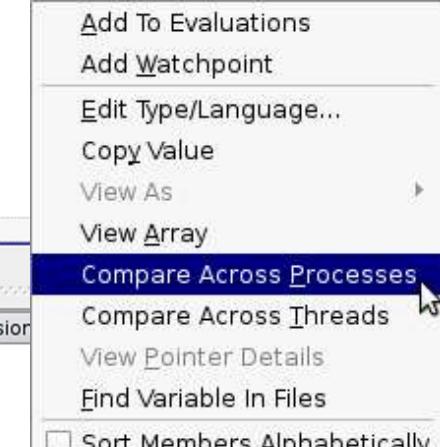
Processes	Threads	File	Line	Function	Condition	Start After	Trigger E
All	all	heatC-MPI.c	455	main	step==11	0	

DDT Practical 1: Task B Solution



Not possible with a single
breakpoint: variable step is local
in main, variable mymax is local
in heatTimestep

3: Open context menu for
mymax, select „Compare
Across Processes“



1: Run to breakpoint in
time stepping loop,
condition: step==10

2: Then add breakpoint at
MPI_Allreduce in
heatTimestep

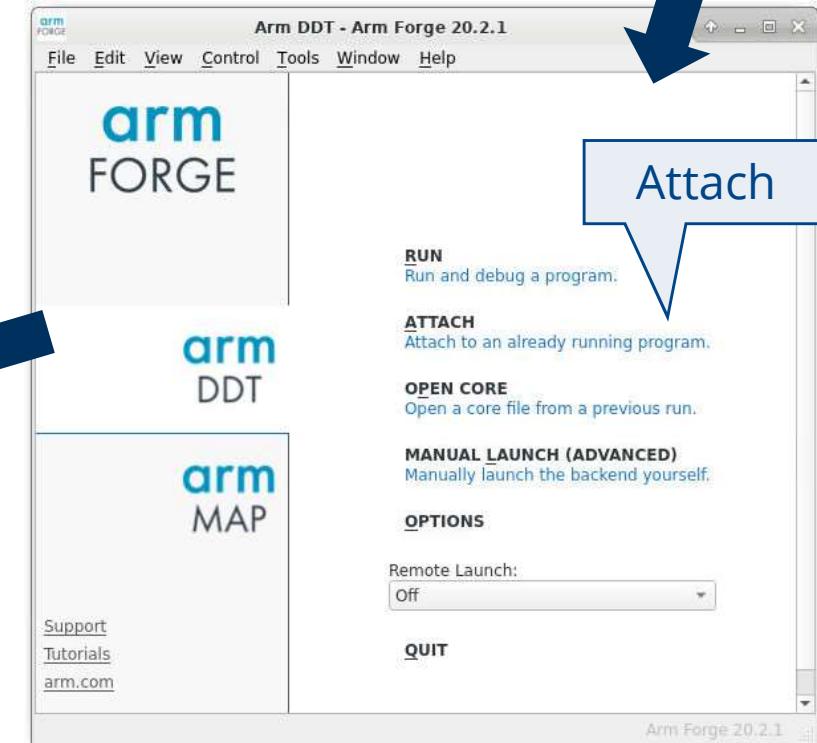
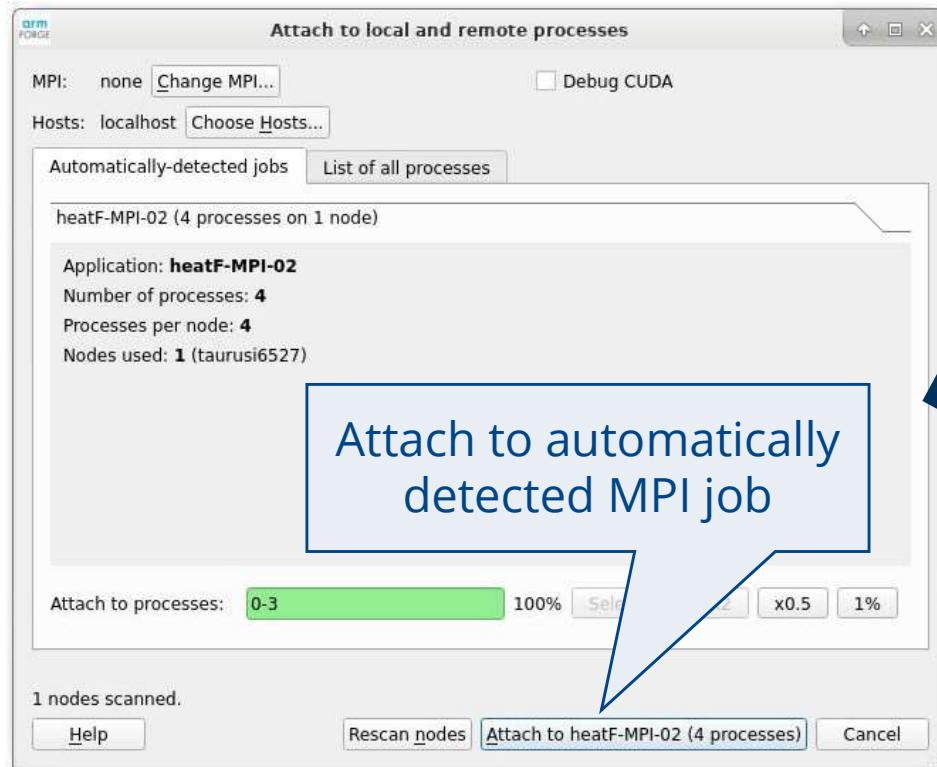
DDT: Attach to running program

```
% mpif90 -g heatF-MPI-02.F90 -o heatF-MPI-02  
% srun -n 4 ./heatF-MPI-02  
. . .
```

Program runs – you want to know what it is doing?

Start DDT in a 2nd terminal

```
% ddt
```



DDT: Core Files (1/2)

```
% mpif90 -g -O0 heatF-MPI-01.F90 -o heatF-MPI-01  
% ulimit -c  
0  
% ulimit -Sc 100000  
% export FOR_DUMP_CORE_FILE=yes  
% srun -n 2 ./heatF-MPI-01
```

Check core file size limit (reports kB) and increase if required (sets to 100 MB)

```
...  
  
forrtl: severe (174): SIGSEGV, segmentation fault occurred
```

Intel Fortran only

Segmentation Fault

```
...  
  
srun: error: taurusi6595: tasks 0-1: Exited with exit code 174
```

srun realizes crash

```
...  
  
% ls -lh core*  
-rw----- 1 gpu59 1111111 42M Jan 27 13:21 core.taurusi6595.taurus.hrsk.tu-dresden.de.3371  
-rw----- 1 gpu59 1111111 42M Jan 27 13:21 core.taurusi6595.taurus.hrsk.tu-dresden.de.3372  
% ddt
```

Per-process core files

Analyze with DDT

DDT: Core Files (2/2)

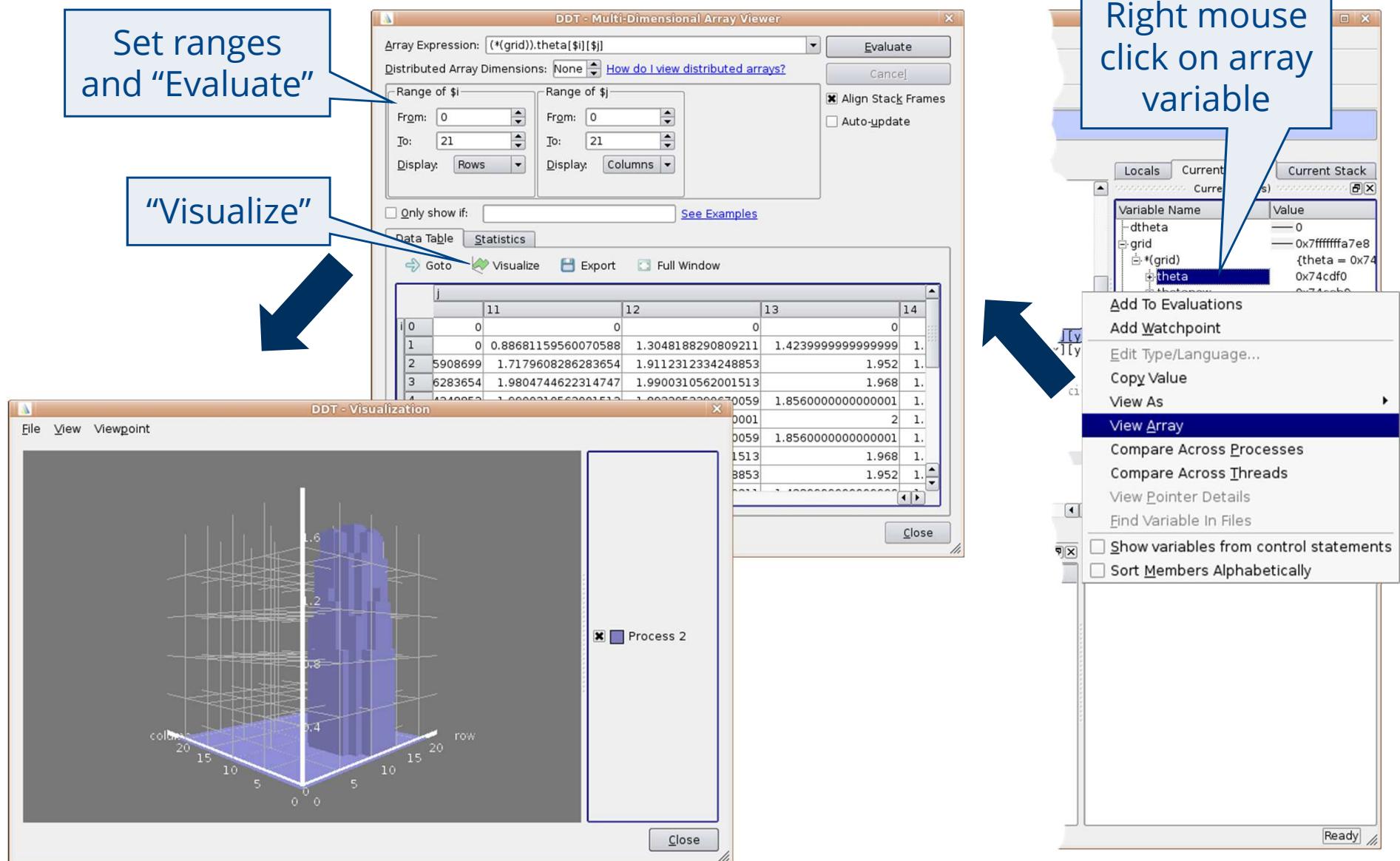
The screenshot illustrates the workflow for using DDT with core files:

- Initial DDT Window:** Shows the main menu (File, Edit, View, Control, Tools, Window, Help) and several options:
 - Open Core**: A large blue callout points to this option.
 - RUN**: Run and debug a program.
 - ATTACH**: Attach to an already running program.
 - OPEN CORE**: Open a core file from a previous run.
 - MANUAL LAUNCH (ADVANCED)**: Manually launch the backend yourself.
 - OPTIONS**: Remote Launch dropdown set to Off.
 - QUIT**.
- Open Core Files Dialog:** A modal window titled "Open Core Files" is shown, containing:
 - Executable:** /home/h9/hpclab70/Debugging/16/f90/heatF-MPI-01
 - Core files:** /home/h9/hpclab70/Debugging/16/f90/core.22292
/home/h9/hpclab70/Debugging/16/f90/core.22293
 - Add...** and **Remove** buttons.
 - OK** and **Cancel** buttons.A blue arrow points from the "Open Core" button in the main window to this dialog.
- DDT Debugging Environment:** The main DDT interface shows:
 - The project **heatF-MPI-01.F90** is open in the Project Files tab.
 - The code editor displays the source code for **heatF-MPI-01.F90**, specifically the **heatBoundary** subroutine.
 - The **Locals** pane shows variable values at the time of the crash, including **dt**, **dtheta**, **dthetamax**, **err**, **grid**, **mymax**, **mympi**, **x**, and **y**.
 - The **Stacks** pane shows the call stack with entries for **heatexample** and **heatconduction_heat timestep**.
 - The **Expression** pane is currently empty.A blue arrow points from the "Open Core Files" dialog to this main DDT window.

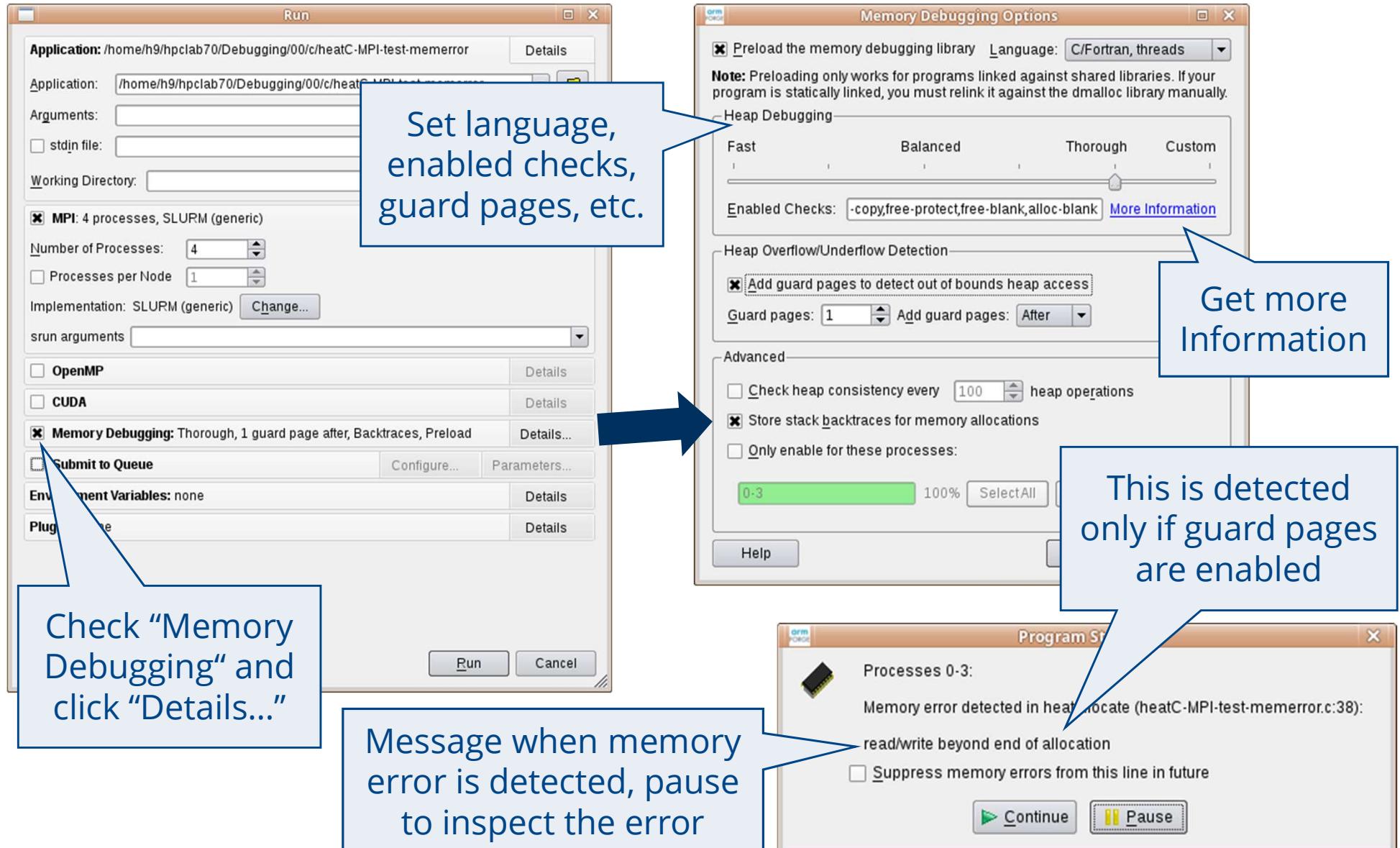
Text Box Summary:

DDT shows position of the crash in the source code and values of variables at the time of the crash.
But no running or stepping possible!

DDT: Multidimensional Array Visualization



DDT: Memory Debugging



DDT Practical 2: Find the Bugs!

Find the bug in each of the three programs!

- Compile and first run normally (4 processes) to observe the behavior, then use DDT to find the bug
- If the program stops in MPI, DDT may complain about missing source files: ignore and select an application function in the call stack view

heatC-MPI-01 / heatF-MPI-01

- Produce core dumps (with up to 4 MPI processes) and open with DDT
- You already know this example from the Intro's practical

heatC-MPI-02 / heatF-MPI-02

- Run without DDT and then attach DDT (use a second terminal session)
- In case of trouble when attaching: ensure that Options - System – Debugger is set to GNU 7.6.2

heatC-MPI-03 / heatF-MPI-03

- Hint: compare arguments of send and receive call

Optional OpenMP task on next slide

DDT Practical 3: Breakpoints with OpenMP (optional)

C:

```
% cd ~/Debugging/c  
% icc -g -O0 -fopenmp heatC-omp.c -o heatC-omp  
% ddt ./heatC-omp
```

Fortran 90:

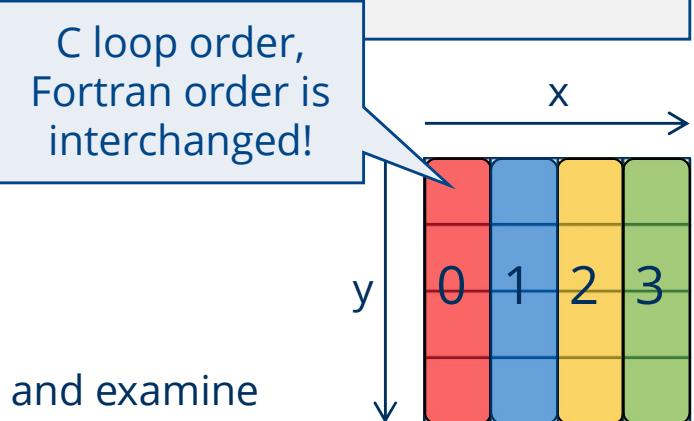
```
% cd ~/Debugging/f90  
% ifort -g -O0 -fopenmp heatF-omp.F90 -o heatF-omp  
% ddt ./heatF-omp
```

Task:

- Run with 4 threads in DDT and find out which thread computes which part of the 20 x 20 grid

Hints:

- Use a breakpoint in the inner compute loop in heatTimestep and examine loop variable x, use condition y==1 to jump to the next iteration of the x-loop
- Fortran: x and y are interchanged compared to C, y-loop is parallelized: need to examine y
- Breakpoints in OpenMP parallel regions sometimes behave unexpected - it helps to run a single thread, while pausing all others: select “Focus on current thread” and select the thread below



DDT Practical 2 Task 01: Solution (C)

Segmentation fault writing grid->thetanew[0] in heatAllocate
Check if the array has been allocated

The screenshot shows the Arm DDT interface. The code editor displays a C file named 'heatC-MPI-01.c' with the following relevant code:

```
25     grid->xsize = xsize;
26     grid->ysize = ysize;
27     grid->theta  = (double**) malloc (sizeof(double)*
28                                     (xsize+2)*(ysize+2));
29     grid->theta [0] = (double*) malloc (sizeof(double)*
30                                     (xsize+2)*(ysize+2));
31
32     for (i = 0; i < xsize+2; i++)
33     {
34         grid->theta [i] = grid->theta [0]+i*(ysize+2);
35         grid->thetanew[i] = grid->thetanew[0]+i*(ysize+2);
36
37         for (j = 0; j < ysize+2; j++)
38         {
39             grid->theta [i][j] = 0.0;
40             grid->thetanew[i][j] = 0.0;
41         }
42
43         grid->dx = 1.0;
44         grid->dy = 1.0;
45         grid->k = 1.0;
```

The line `grid->thetanew[i] = grid->thetanew[0]+i*(ysize+2);` is highlighted in blue. A callout box points from this line to the variable `grid->thetanew` in the Locals window.

The Locals window shows the following variable values:

Variable Name	Value
grid	0xfffffff7858
*(grid)	{theta = 0x6164b0, thetanew = 0x6164b0}
theta	0x6164b0
thetanew	0x0
thetanew	<Cannot access memory at address 0x6164b0>
xsize	20
ysize	20
dx	6.9533558061358138e-310
dy	2.079947693866876e-317
k	1.2414507464325953e-312
i	0
j	1
xsize	20
ysize	20

The Processes window shows two threads:

- main (heatC-MPI-01.c:432)
- heatAllocate (heatC-MPI-01.c:34)

DDT Practical 2 Task 01: Solution (Fortran)

The screenshot shows the Arm DDT - Arm Forge 18.0.1 interface. The left pane displays the project structure under 'Application Code' with 'heatF-MPI-01.F90' selected. The right pane shows the Fortran code for heatTimestep, where line 192 is highlighted. A callout box points to this line with the text: "Segmentation fault writing grid%thetanew(1,1) in heatTimestep Check if the array has been allocated". The variable viewer on the right shows 'grid%thetanew' as <not associated>. Another callout box points to the variable viewer with the text: "Fortran version crashes later than C version".

```
182 mymax = 0.0d0
183
184 ! calculate the time step: read from theta, write to thetanew
185 ! Only calculate on a processes sub-grid
186 do y=mypi%start_y,mypi%start_y + mympi%num_ce
187   do x=mypi%start_x,mypi%start_x + mympi%num_cells_x - 1
188     dtheta = ( grid%theta(x-1,y) + grid%theta(x+1,y) - 2*grid%theta(x,y) ) / (grid%
189       + ( grid%theta(x,y-1) + grid%theta(x,y+1) - 2*grid%theta(x,y) ) / (grid%
190       grid%thetanew(x,y) = grid%theta(x,y) + grid%k * dtheta * dt
191
192     mymax = max(abs(dtheta), mymax) ! save max theta for the exit condition
193
194   end do
195 end do
196
197 ! Make MPI reduction to get maximum dtheta of all processes
198 CALL MPI_ALLREDUCE (mymax, dthetamax, 1, MPI_DOUBLE_PRECISION, MPI_MAX, MPI_COMM_WORLD)
199
200 ! update theta: copy thetanew to theta
201 do x=mypi%start_x,mypi%start_x + mympi%num_cells_x - 1
202   do y=mypi%start_y,mypi%start_y + mympi%num_ce - 1
203     grid%theta(x,y) = grid%thetanew(x,y)
204   end do
205 end do
```

Variable Name	Value
dt	0.0500000000000000
dtheta	0
dthetamax	100
err	0
grid	
grid%theta	
grid%thetanew	<not associated>
xsize	20
ysize	20
dx	1
dy	1
k	1
mymax	0
mypi	
x	-1
y	-1

DDT Practical 2 Task 02: Solution

The screenshot shows the Arm DDT - Arm Forge 18.2.2 interface. The left pane displays the project structure for 'heatC-MPI-02.c' and its source code. The source code contains MPI_Bsend and MPI_Recv calls for peer-to-peer communication between processes. The right pane shows the 'Locals' window with variable values for process 0. A callout box highlights a specific MPI_Recv call at line 260, which is waiting because the received tag (122) does not match the expected tag (123). The variable values are:

Variable Name	Value
grid	0x7fffffff2a68
grid->theta	0x618660
mympi	0x7fffffff2a08
mympi->num_cells_y	10
mympi->start_x	1
mympi->start_y	1

All processes are waiting at this MPI_Recv in heatBoundary
Reason: Tags are not matching

DDT Practical 2 Task 03: Solution

The screenshot shows the Arm DDT - Arm Forge 18.2.2 interface. The left pane displays the project structure and source code for `heatC-MPI-03.c`. The right pane shows the debugger's state. A callout box highlights a bug where ranks 1-3 send 4 MPI_DOUBLE values, which exceeds the receive buffer of rank 0. Another callout box indicates that Rank 0 receives 4 MPI_INT values. A third callout box provides a caution about the error.

Ranks 1-3 send 4 MPI_DOUBLE, which is wrong and exceeds the receive buffer of rank 0.

Rank 0 receives 4 MPI_INT

Caution: If ranks 1-3 would send 4 MPI_FLOAT, MPI would not abort because the buffer size fits! Only MUST could detect this error.

```
386 MPI_DOUBLE, /* old type */
387 &blocktype /* new type */ );
388 MPI_Type_commit (&blocktype);
389
390 MPI_Send (block_size, 4, MPI_DOUBLE, 0, 50, MPI_COMM_WORLD);
391 MPI_Send (&grid->theta[mympi->start_x][mympi->start_y], 1, blocktype, 0, 51, MPI_COMM_WORLD);
392
393 MPI_Type_free (&blocktype);
394 }
395 else
396 /*Master Receives data*/
397 {
398 MPI_Comm_size (MPI_COMM_WORLD, &size);
399 for (i = 1; i < size; i++)
400 {
401 /*Receive Block Info*/
402 MPI_Recv (block_size, 4, MPI_INT, i, 50, MPI_COMM_WORLD, &status);
403
404 /* Create datatype to communicate one block*/
405 MPI_Type_vector (
```

Program Stopped

Process 0:

Program stopped at MPID_Abort.

Always show this window for default breakpoints

Continue Pause

DDT Practical 3: Solution

The screenshot shows the Arm DDT - Arm Forge 18.0.1 interface. The code editor displays a C program with OpenMP parallel for loops. A callout box labeled "1: Run to a breakpoint in OpenMP parallel for loop" points to the code. The variable viewer shows local variables like dt, dtheta, grid, etc. A callout box labeled "3: Compare x across threads (via context menu)" points to the variable viewer. The call stack comparison view shows thread assignments for values 1, 6, 11, and 16. A callout box labeled "2: Select bottom call stacks" points to this view. The bottom right corner features the ZIH logo.

1: Run to a breakpoint in OpenMP parallel for loop

```
/* calculate the time step: n */  
/* OpenMP 3.1: New reduction */  
#pragma omp parallel for private(dtheta)  
for (x=1; x <= grid->xsize; x++)  
{  
    for (y=1; y <= grid->ysize; y++)  
    {  
        dtheta = ( grid->theta[x-1][y] + grid->theta[x+1][y] - 2*grid  
                  + ( grid->theta[x][y-1] + grid->theta[x][y+1] - 2*grid  
                  grid->thetanew[x][y] = grid->theta[x][y] + grid->k * dtheta *  
                  dthetamax_reduction = fmax(fabs(dtheta), dthetamax_reduction)
```

3: Compare x across threads (via context menu)

Thread 0 computes x=1,..., x=5
Thread 1 computes x=6,..., x=10
etc.

2: Select bottom call stacks