

High performance tools to debug, profile, and analyze your applications

Debugging and Profiling your HPC Applications

Ryan Hulguin, Application and Support Analyst

rhulguin@allinea.com

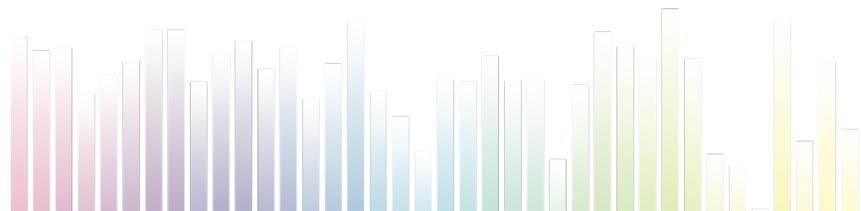


About this talk



allinea
FORGE
DDT + MAP

- Learn how to debug and profile your code
 - Techniques to take home
- Tools we will use: Allinea Forge
 - Debugging with Allinea DDT
 - Profiling with Allinea MAP
 - NB. Allinea MAP is not supported on BG/Q
- Where to find Allinea's tools
 - > 70% of Top 500 have at least one Allinea tool



allinea

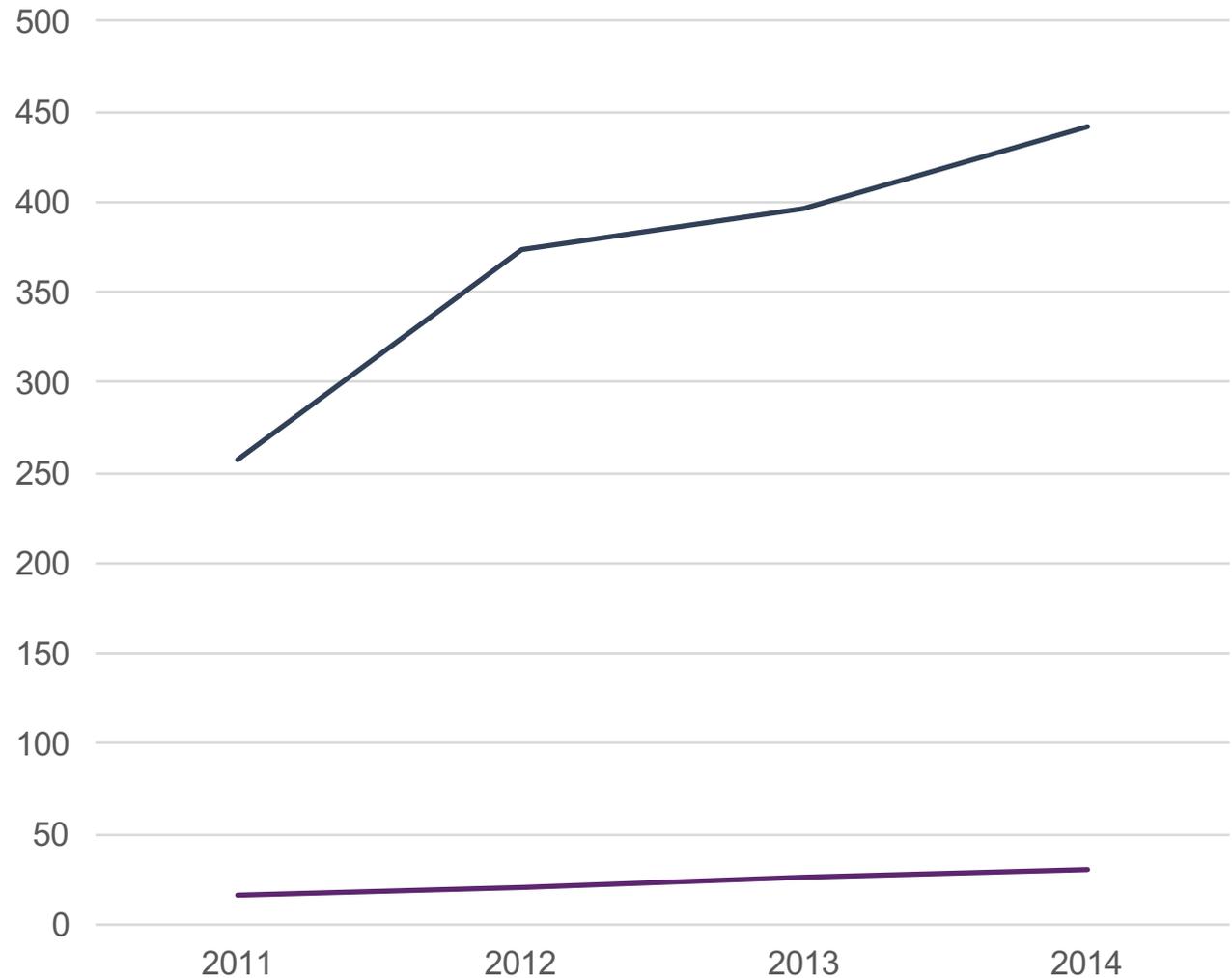
Motivation

- “Without capable highly parallel software, large supercomputers are less useful”
 - Council on Competitiveness

- “1% of HPC application codes can exploit 10,000 cores”

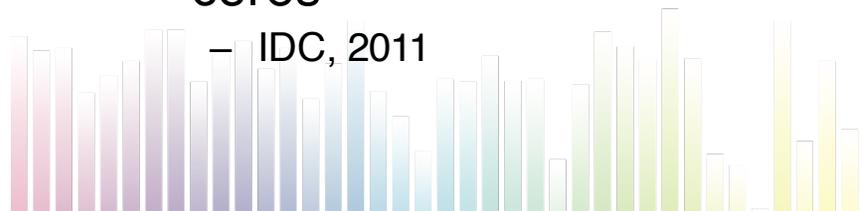
– IDC, 2011

Number of 10,000 core systems

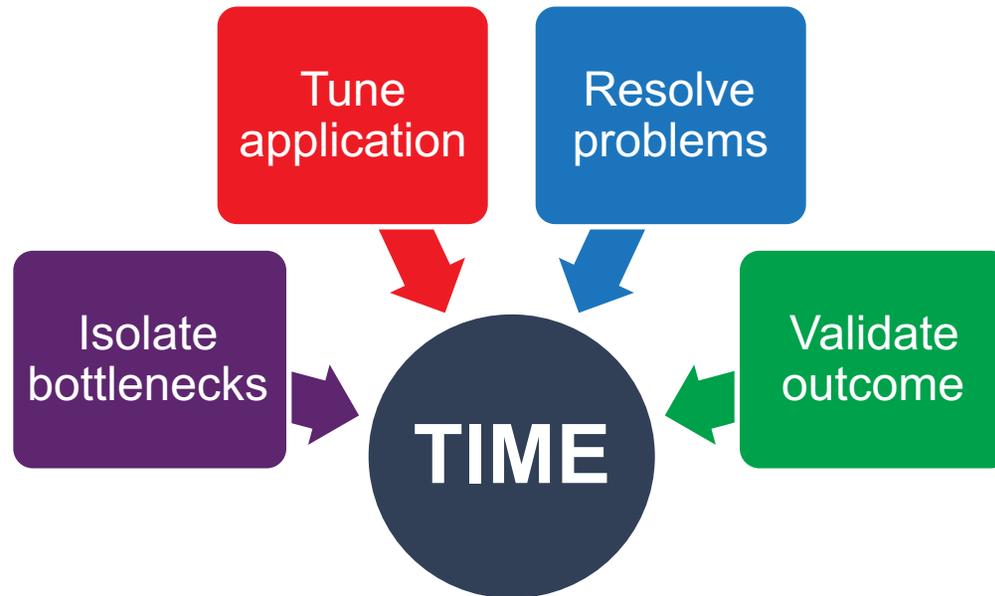


— > 10,000 cores — > 100,000 cores

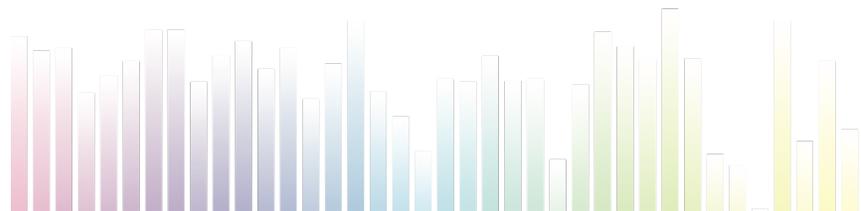
allinea



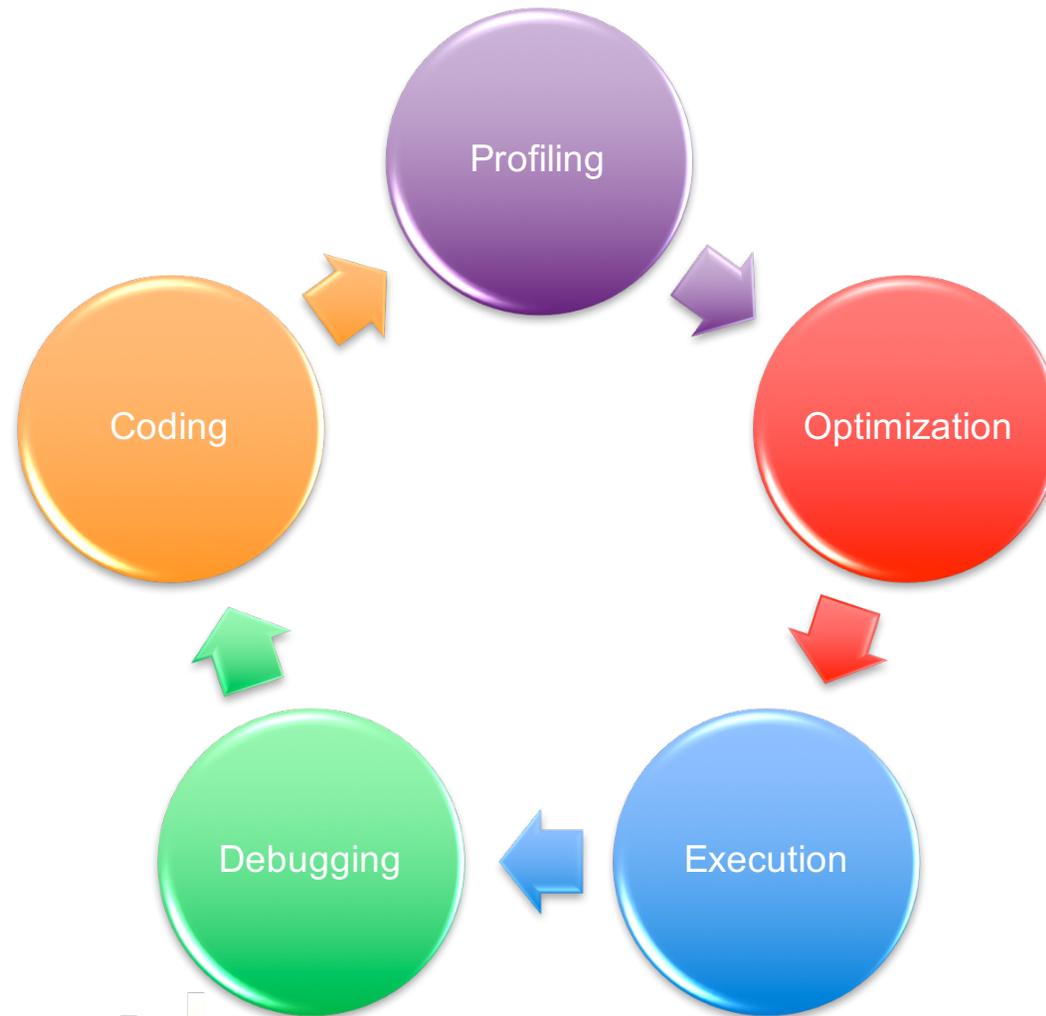
Use the right software tool to be faster



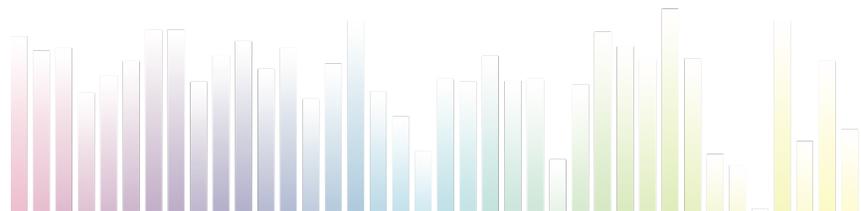
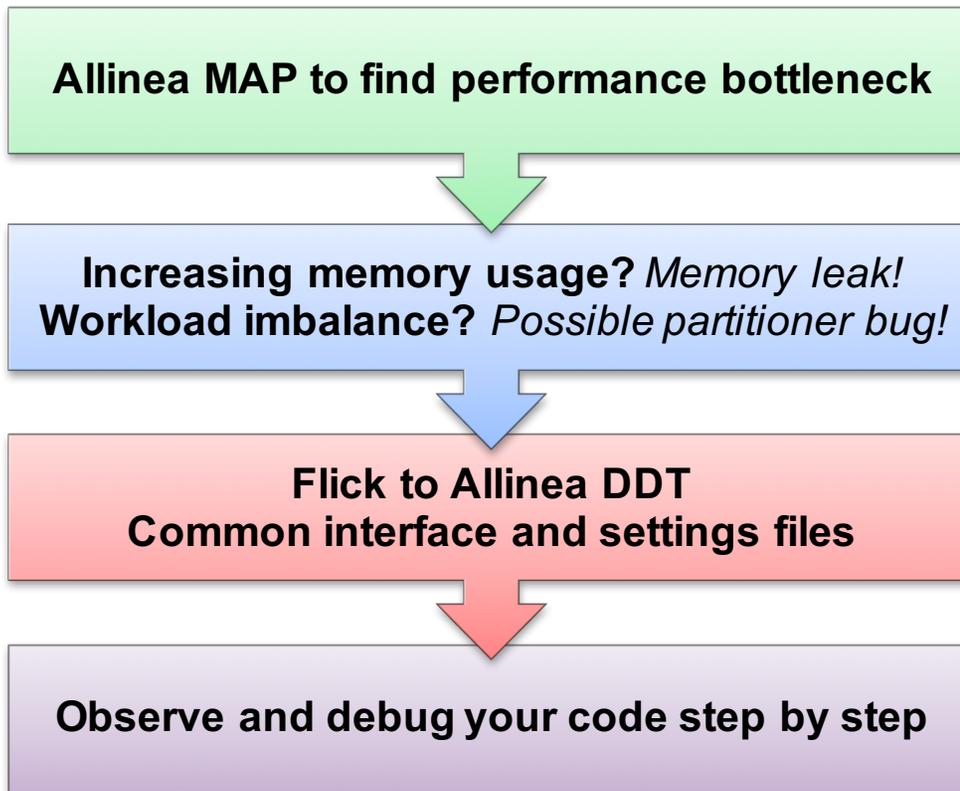
- What parts of the code would benefit most from being rewritten?
- How should I modify a code to make it better (or work at all)?



Application Development Workflow



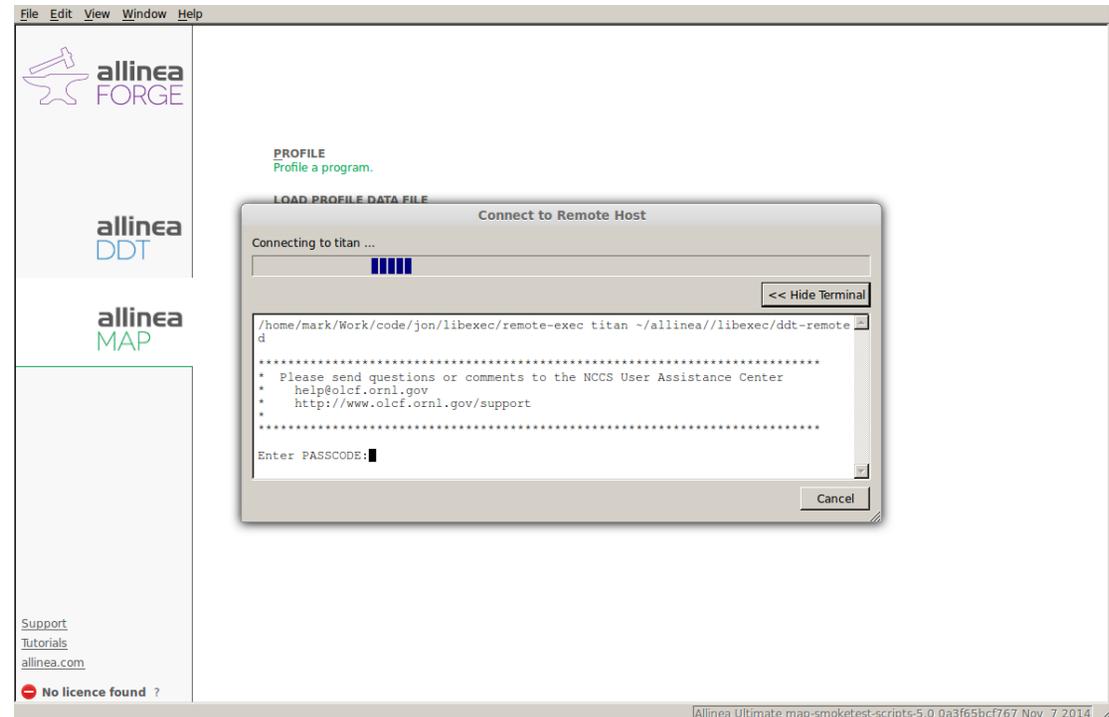
Hello Allinea Forge!



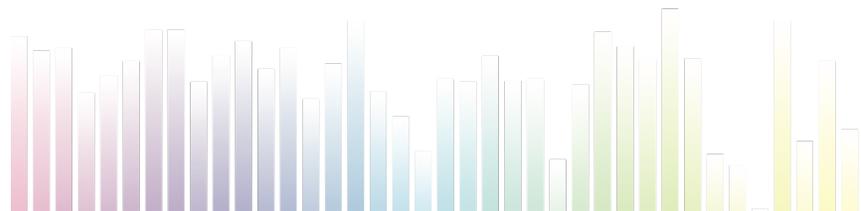
HPC means being productive on remote machines



- ✓ Linux
- ✓ OS/X
- ✓ Windows
- ✓ Multiple hop SSH
- ✓ RSA + Cryptocard
- ✓ Uses server license



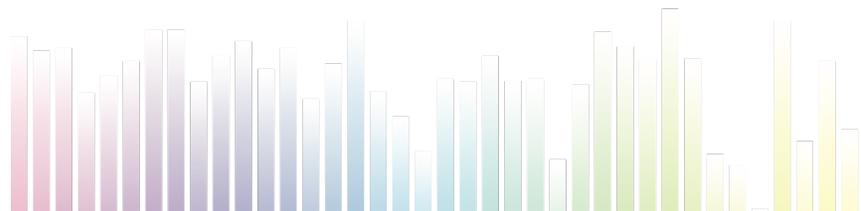
allinea



- Code optimisation can be time-consuming...

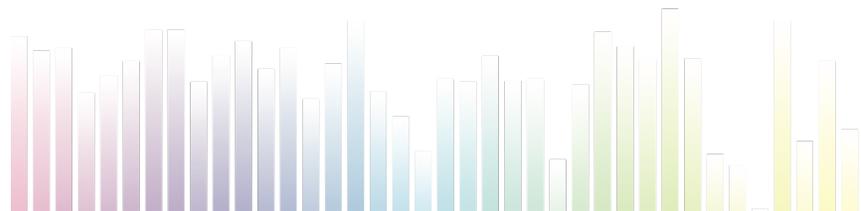
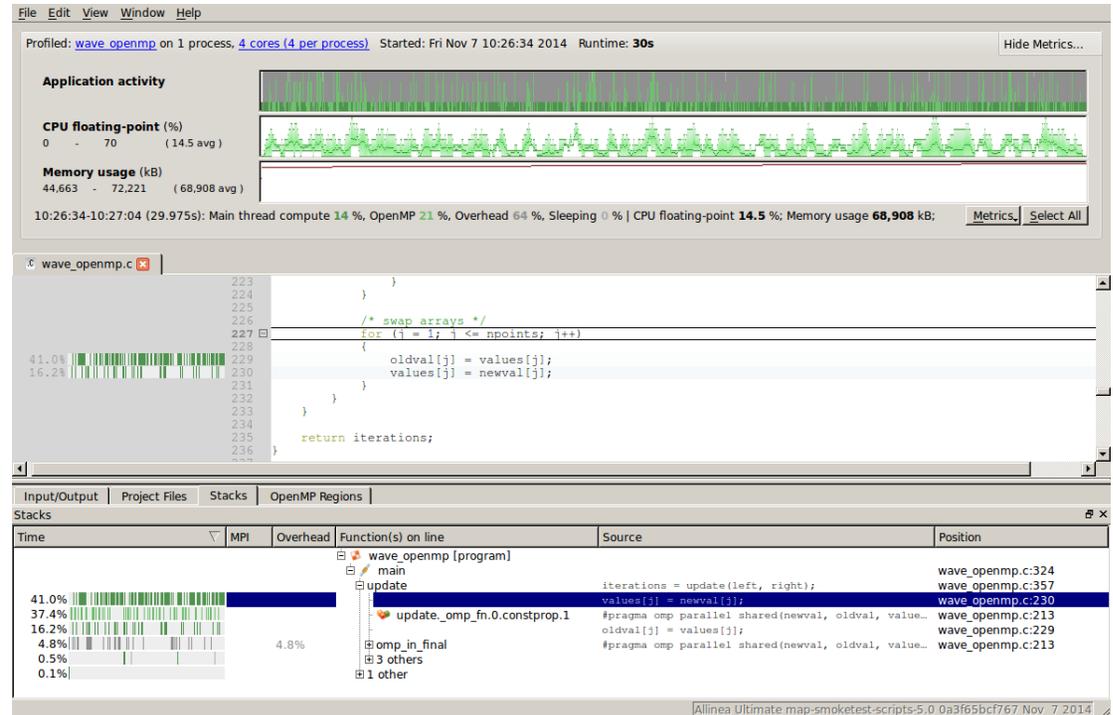


– (image courtesy of xkcd.com)

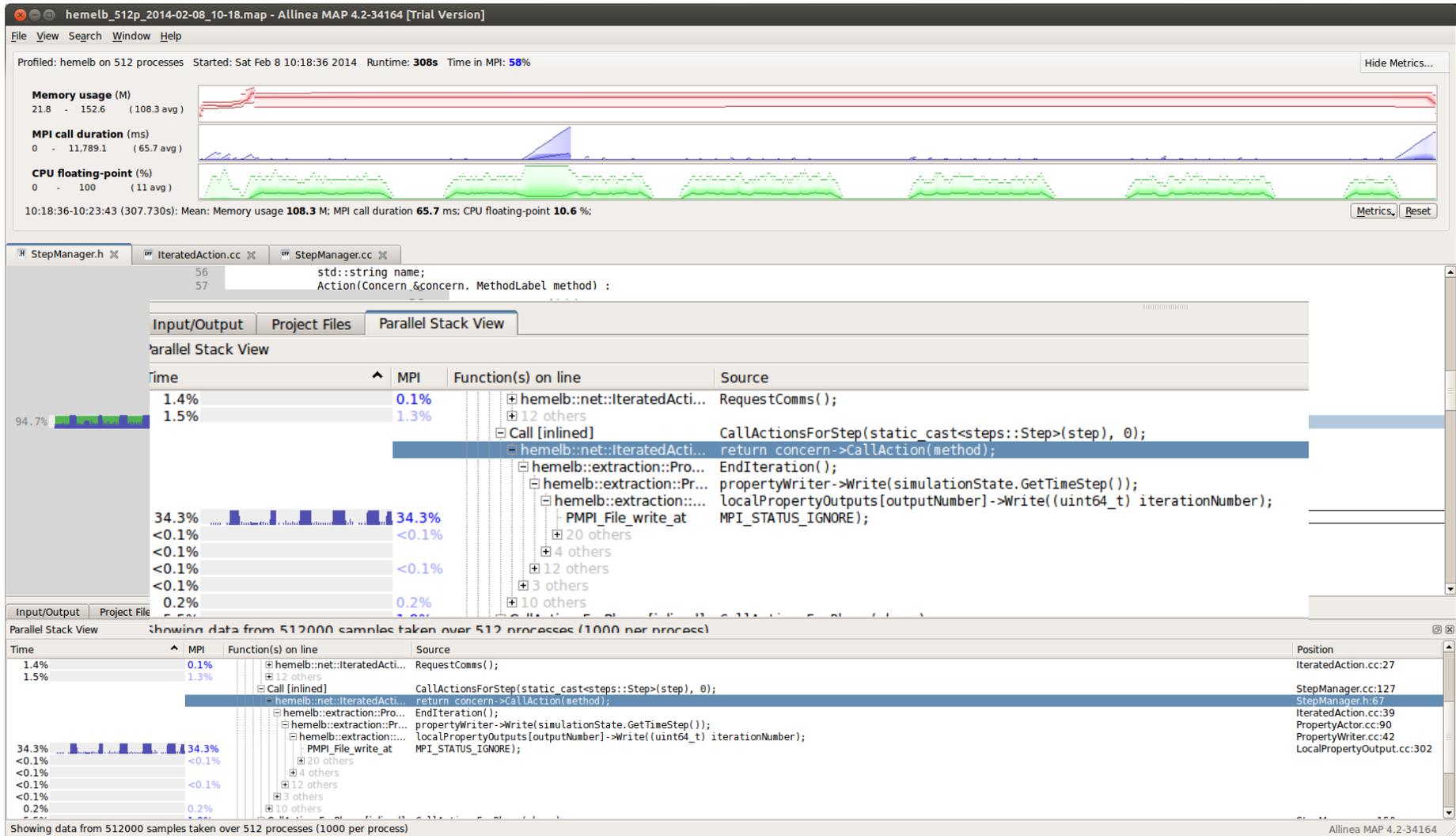


MAP in a nutshell

-  Small data files
-  <5% slowdown
-  No instrumentation
-  No recompilation



Scaling issue – 512 processes



Simple fix... reduce periodicity of output

Some types of bug

Bohrbug

Steady, dependable bug

Heisenbug

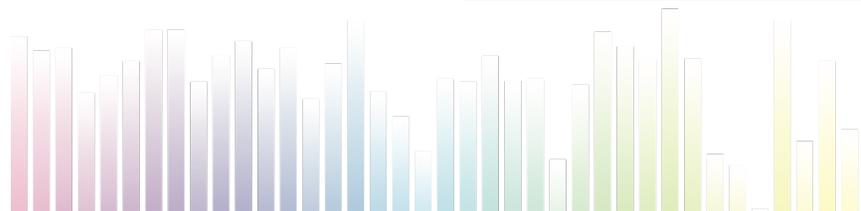
Vanishes when you try to debug (observe)

Mandelbug

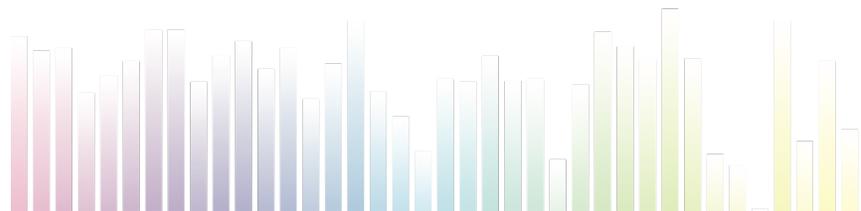
Complexity and obscurity of the cause is so great that it appears chaotic

Schroedinbug

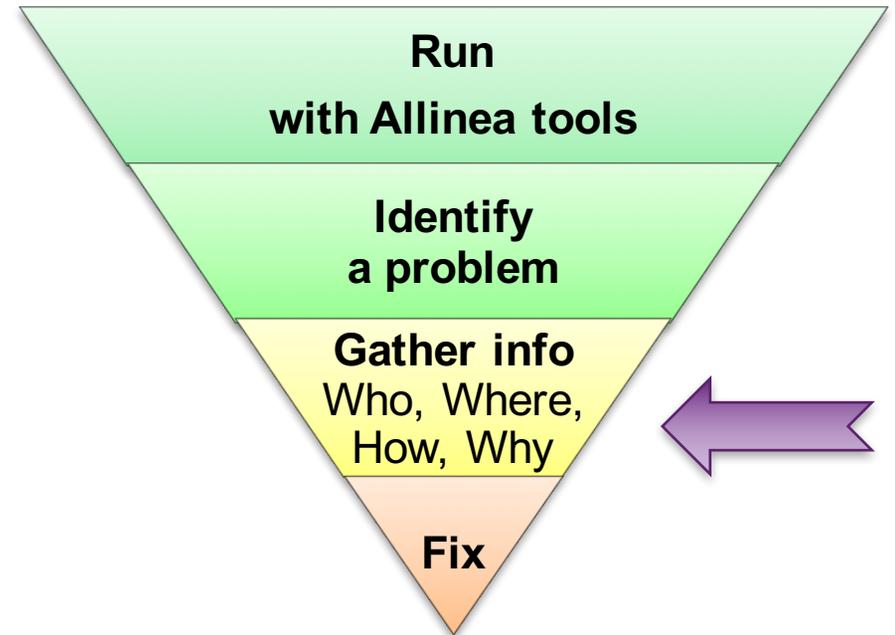
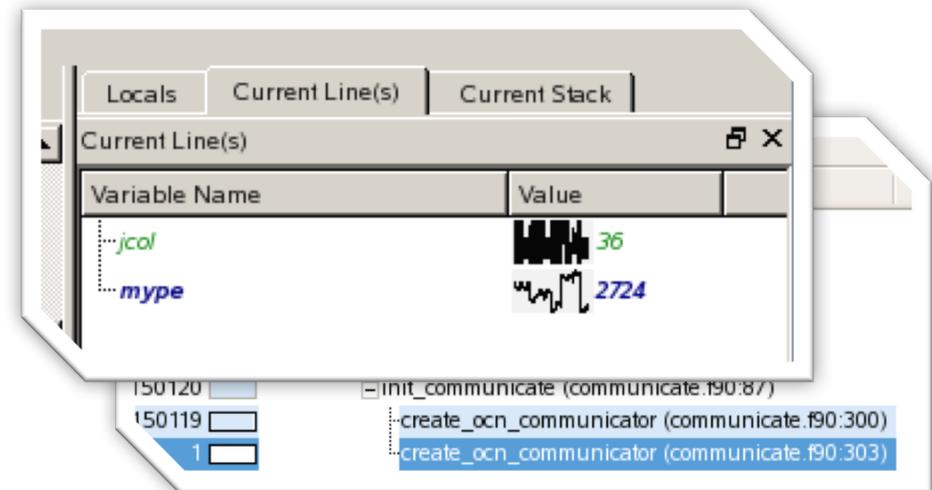
First occurs after someone reads the source file and deduces that it never worked, after which the program ceases to work



- Transforming a broken program to a working one
- How? TRAFFIC!
 - Track the problem
 - Reproduce
 - Automate - (and simplify) the test case
 - Find origins – where could the “infection” be from?
 - Focus – examine the origins
 - Isolate – narrow down the origins
 - Correct – fix and verify the testcase is successful
- Suggested Reading:
 - Zeller A., “Why Programs Fail”, 2nd Edition, 2009
 - Zen and the Art of Motorcycle Maintenance, Robert M. Pirsig



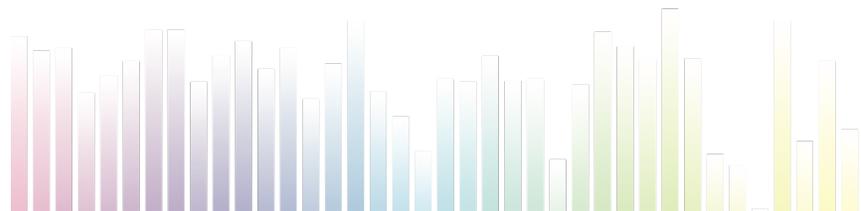
- Who had a rogue behavior ?
 - Merges stacks from processes and threads
- Where did it happen?
 - leaps to source
- How did it happen?
 - Diagnostic messages
 - Some faults evident instantly from source
- Why did it happen?
 - Unique “Smart Highlighting”
 - Sparklines comparing data across processes

Variable Name	Value
<i>icol</i>	36
<i>mype</i>	2724

Stack Trace:

- 150120 [] - init_communicate (communicate.f90:87)
- 150119 [] - create_ocn_communicator (communicate.f90:300)
- 1 [] - create_ocn_communicator (communicate.f90:303)



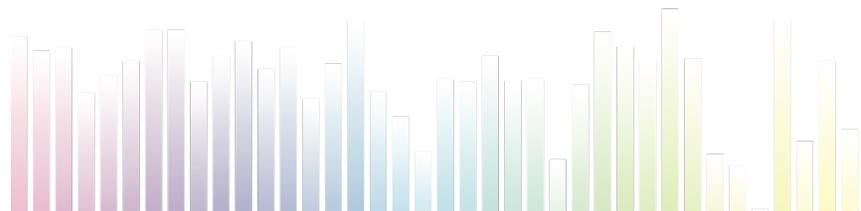
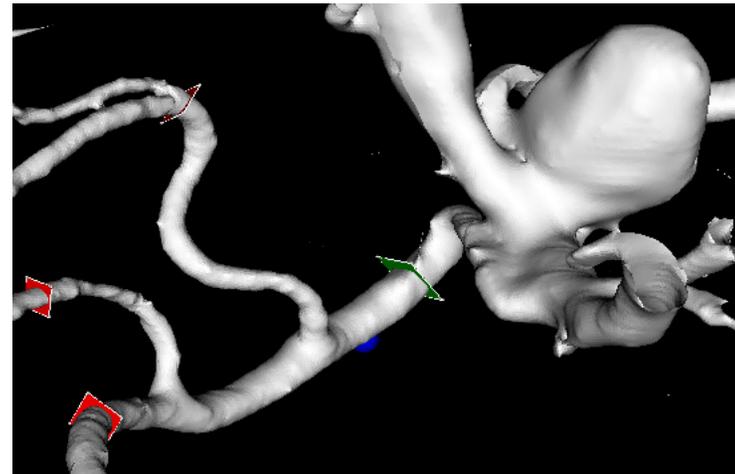
HPC could be brain surgery

- Brain aneurysms
 - 2-5% of population – most are undiagnosed
 - 30,000 rupture in US each year – 40% fatal
 - Early discovery and treatment increases survival rates
- Neurosurgery as HPC
 - MRI provides the blood vessel structure
 - Intra-cranial blood flow and pressures is just complex CFD
 - Full brain 3D model is 2-10GB geometry
- Individualized HPC
 - Patient’s MRI scan enables surgical decision: whether to operate, how to operate, ...
 - Circle of Willis requires super-Petascale ~~machine~~ software
 - Need answer in minutes or hours
- ... but it crashes at 49152 cores

Saccular



Fusiform



Allinea DDT 4.2.1-36484

File View Control Search Tools Window Help

Current Group: All Focus on current: Group Process Thread Step Threads Together

All 24576 processes (0-24575) Paused: 17223 Playing: 7353 Finished: 0
 Currently selected: 260 (on nid00194, pid 9481, main thread IWP 9481)

Create Group

Project Files

Search (Ctrl+K)

VolumeTrav
 wave.c
 weird.c
 WholeGeom
 Writer.cc
 wspace.c
 XdrFileWrite
 XdrMemRea
 XdrMemWrit
 XdrReader.c
 XdrWriter.cc
 XmiAbstract
 xyzpart.c
 External Code

MpiEnvironment.cc xyzpart.c

```

551 ikvsortii(ntsamples, allpicks);
552
553
554 /* Select the final splitters. Set the boundaries to
555 for (i=1; i<nps; i++)
556     mypicks[i] = allpicks[i*ntsamples/nps];
557 mypicks[0].key = IDX_MIN;
558 mypicks[nps].key = IDX_MAX;
559
560
561 WCOREPOP; /* free allpicks */
562
563 STOPTIMER(ctrl, ctrl->AuxTmr2);
564 STARTTIMER(ctrl, ctrl->AuxTmr3);
565
  
```

Locals Current Line(s) Current Stack

Current Line(s)

Variable Name	Value
allpicks	0x2aab8055e010
i	2245
mypicks	0x2a6f8f0
nps	24575
ntsamples	1818550

Type: none selected

Input/Output Breakpoints Watchpoints Stacks Tracepoints Tracepoint Output Logbook

Stacks

Processes	Threads	Function
17223	17223	main (main.cc:37)
17223	17223	SimulationMaster::SimulationMaster (SimulationMaster.cc:63)
17223	17223	SimulationMaster::Initialise (SimulationMaster.cc:154)
17223	17223	hemelb::geometry::GeometryReader::LoadAndDecompose (GeometryReader.cc:188)
17223	17223	hemelb::geometry::GeometryReader::OptimiseDomainDecomposition (GeometryReader.c
17223	17223	hemelb::geometry::decomposition::OptimisedDecomposition::OptimisedDecomposition
17223	17223	hemelb::geometry::decomposition::OptimisedDecomposition::CallParmetis (Optimise
17223	17223	ParMETIS_V3_PartGeomKway (gkmetis.c:90)
17223	17223	libparmetis_Coordinate_Partition (xyzpart.c:58)
17223	17223	libparmetis_PseudoSampleSort (xyzpart.c:556)

Evaluate

Expression	Value
i * ntsamples	-212322546

Type: int
 Range: from -2147259746 to -12282046
 49/17223 processes equal

7353 processes playing

Computer titan-ext7 Allinea DDT 4.2.1-36484 Sun Aug 10, 7:50 PM

Favorite Allinea DDT Features for Scale

Stacks (All)

Processes	Function
150120	start
150120	__libc_start_main
150120	main
150120	pop (POP:190:81)
150120	init_communicate (initial:190:119)
150120	init_communicate (communicate:190:87)
150119	create_ocn_communicator (communicate:190:300)
150119	create_ocn_communicator (communicate:190:303)

Parallel

Parallel stack view

Current Line(s)

Variable Name	Value
mycol	36
mypr	2724

Parallel

Automated data comparison: sparklines

Array Expression: bigArray[S:I]

Distributed Array Dimensions: 1

Range of Sx (Distributed): From: 0 To: 9999

Range of Si: From: 0 To: 9999

Display: Rows Columns

Only show if: Svalue == 1

	2444	2733	3011	3185	4704	5343	6795	7881	9108	9467
x 0										
1										
2										
3										
4										
5										

Parallel

Parallel array searching

All

0	1	2
1	2	
0		

ddt.bin

licen.reserver

Current Group: All

Focus on current: Group Process Thread

200004 processes (0-200003) Paused: 200004 Running: 0

Currently selected: 0

Parallel

Step, play, and breakpoints

Process stopped in sched_yield (syscall-template.S:82) with signal SIGSEGV (Segmentation fault). Reason: Origin: kill, sigsend or raise. Your program will probably be terminated if you continue. You can use the stack controls to see what the process was doing at the time.

Stacks

Processes	Threads	Function
1-15	15	main (tu: f:118)
1-15	15	error: f:121
1-15	15	error: f:193
1-3-5-7-9-11-13-15-17		exchange_1 (exchange_1: f:37)
1-3-5-7-9-11-13-15-17		popl_recv
1-3-5-7-9-11-13-15-17		popl_recv
1-3-5-7-9-11-13-15-17		opnl_progress
1-3-5-7-9-11-13-15-17		sched_yield (syscall-template.S:82)
4-8-12-14	4	exchange_1 (exchange_1: f:37)
4-8-12-14	4	popl_recv
4-8-12-14	4	popl_recv
4-8-12-14	4	opnl_progress
4-8-12-14	4	sched_yield (syscall-template.S:82)

Stack for process 1

Every process in your program has terminated.

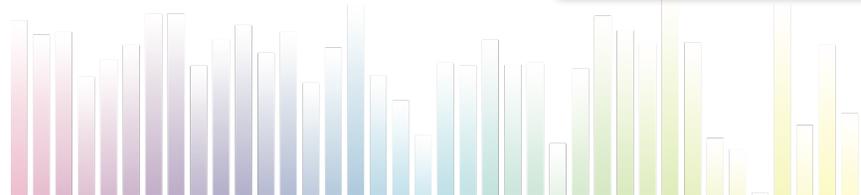
Messages Tracepoints Output

Tracepoints

#	Time	Tracepoint	Processes	Values
1	00:13:45.1	subdomain F9E0		
2	00:13			
3	00:13			

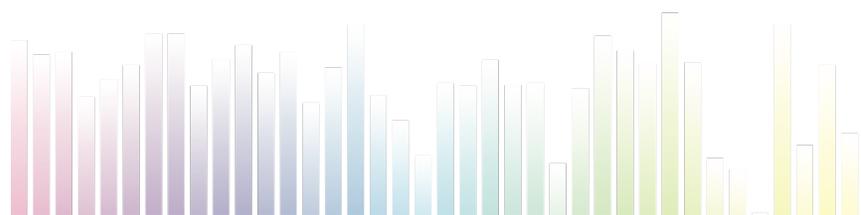
Parallel

Offline debugging



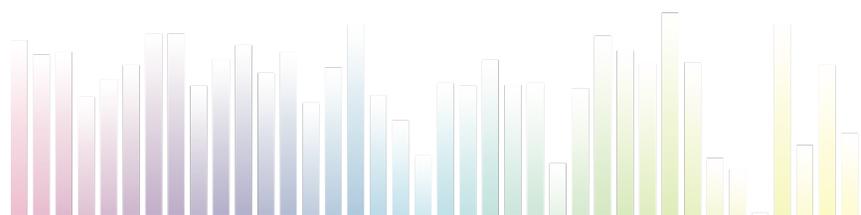
Today's Status on Scalability

- Debugging and profiling
 - Active users at 100,000+ cores debugging
 - 50,000 cores is largest profiling tried to date (and was Very successful)
 - ... and active users with just 1 process too
- Deployed on
 - ORNL's Titan, NCSA Blue Waters, ANL Mira etc.
 - Hundreds of much smaller systems – academic, research, oil and gas, genomics, etc.
- Tools help the full range of programmer ambition
 - Very small slow down with either tool (< 5%)



Getting started on Mira/Cooley

- Install local client on your laptop
 - www.allinea.com/products/forge/downloads
 - Linux – installs full set of tools
 - Windows, Mac – just a remote client to the remote system
 - Run the installation and software
 - “Connect to remote host”
 - Hostname:
 - username@cetus.alcf.anl.gov
 - username@cooley.alcf.anl.gov
 - Remote installation directory: /soft/debuggers/ddt
 - Click Test
- Congratulations you are now ready to debug on Mira/Vesta/Cetus – or debug and profile on Cooley.



Five great things to try with Allinea DDT

Tracepoint	Processes	Values logged
vhone f90 85	916, ranks 12, 14-17, 22-23, 12...	mype 2178-3527 jcol 2-83 mod pey
vhone f90 81	900, ranks 12, 14-17, 22-23, 12...	ks 1 kmax pez
vhone f90 85	942, ranks 12, 14-17, 22-23, 12...	mype 2178-3527 jcol 2-83 mod pey
vhone f90 81	939, ranks 12, 14-17, 22-23, 12...	ks 1 kmax pez
vhone f90 85	919, ranks 12, 14-17, 22-23, 12...	mype 2178-3527 jcol 2-83 mod pey
vhone f90 81	898, ranks 12, 14-17, 22-23, 12...	ks 1 kmax pez
vhone f90 85	884, ranks 12, 14-17, 22-23, 12...	mype 2178-3527 jcol 2-83 mod pey
vhone f90 81	880, ranks 12, 14-17, 22-23, 12...	ks 1 kmax pez

The scalable print alternative

```

for (i = 0 ; i < SIZE M; i++)
  for (j = 0 ; j < SIZE O; j++)
    C[i][j] = 0;

for (i = 0 ; i < SIZE M; i++)
  for (j = 0 ; j < SIZE N; j++)
    for (k = 0 ; k < SIZE O; k++)
      C[i][j] += A[i][k] * B[k][j];
  
```

Program Stopped

Process 0:
 Process stopped at watchpoint "rank" in main (watchmatrix.c:45).
 Old value: 0
 New value: 1074790400
 Always show this window for watchpoints

Continue Pause Pause All

Stop on variable change

```

hello.c
43 else
44     test=-1;
45 }
46
47 void func3()
48 {
49     void* i = (void*) 1;
50     while(i++ || !i)
51         free((void*)i);
  
```

portability 'i' is of type 'void *'. When using void pointers in calculations, the pointer must be cast to the appropriate type.

Static analysis warnings on code errors

```

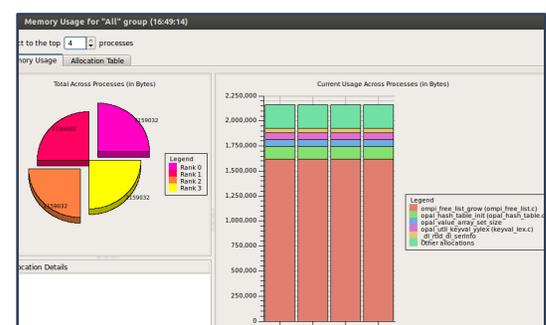
&& !strcmp(argv[i], "crash")) {
  0;
  s", *(char **)argv[i]);
  ll se
  
```

Program Stopped

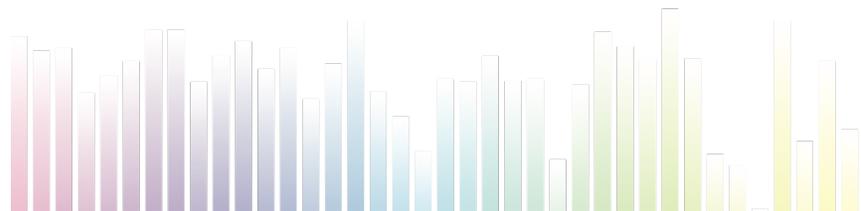
Processes 0-3:
 Memory error detected in main (hello.c:118):
 null pointer dereference or unaligned memory access

Note: the latter may sometimes occur spuriously if guard pages are not enabled
 Tip: Use the stack list and the local variables to explore current state and identify the source of the error.

Detect read/write beyond array bounds



Detect stale memory allocations



Six Great Things to Try with Allinea MAP



compute 76 % MPI 24 % File

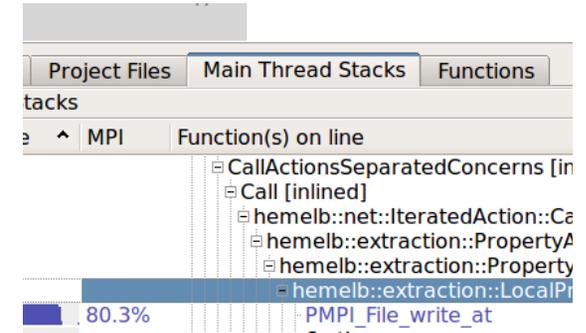
Find the peak memory use

```

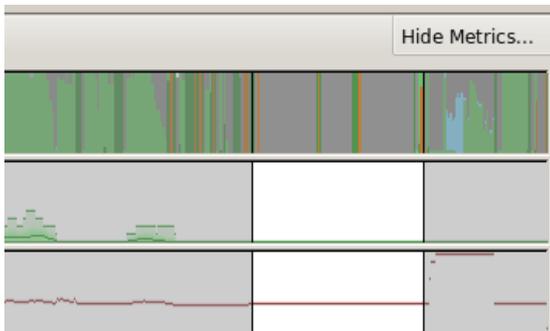
30 ! 'late to the party
31 do j=1,20*nprocs; a
32 end if
33
34 if (pe /= 0) then
35 call MPI_SEND(a, si
36 else
37 do from=1,nprocs-1
38 call MPI_RECV(b,
39 do j=1,50; b=sqrt
40 print *, "Answer f
41 end do
42 end if
43 end do
44 call MPI_BARRIER(MPI CO

```

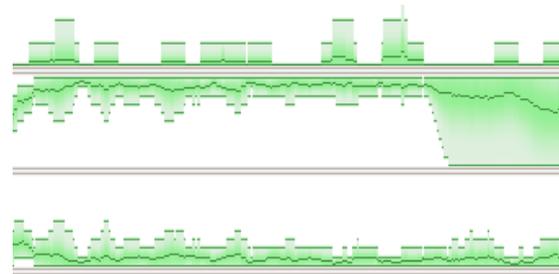
Fix an MPI imbalance



Remove I/O bottleneck



Make sure OpenMP regions make sense



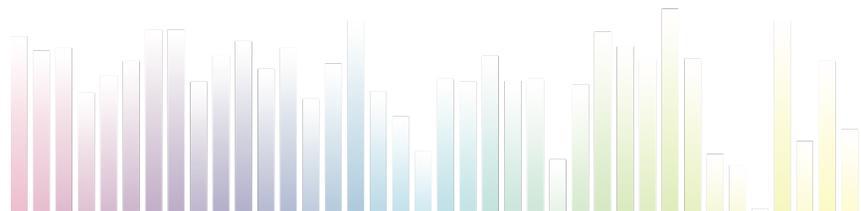
Improve memory access

```

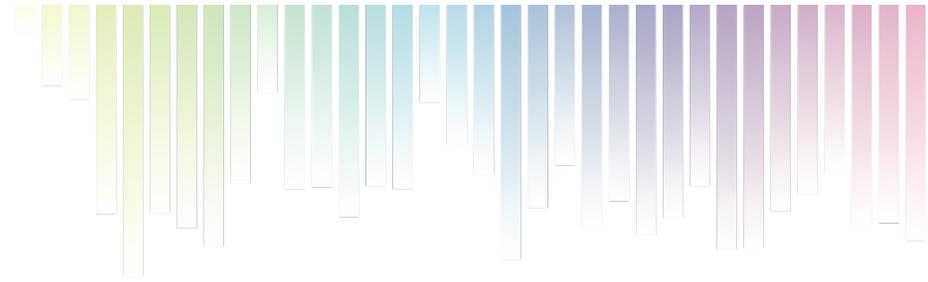
size, nproc, mat_a
A[i*size+k]*B[k*s

```

Restructure for vectorization



allinea



High performance tools to debug, profile, and analyze your applications

Thank you!

Ryan Hulguin, Application and Support Analyst
rhulguin@allinea.com

