# MPI Programming Tools

## John Urbanic

Parallel Computing Scientist Pittsburgh Supercomputing Center

Distinguished Service Professor Carnegie Mellon University

Copyright 2024

# **Tools In Parallel Programming**

Given that the philosophy of these workshops is to teach portable programming techniques and avoid platform or version specific information, I have always reluctantly avoided talking about tools. The best ones are licensed, and in parallel programming there are no open source debuggers worth speaking of.

However, we always get feedback requesting this, and there are two types of tools that are wildly useful in this field:

- Debuggers
- Profilers

So we will carve out a few minutes to expose you to two of the best examples.

# Debuggers

A good debugger is an invaluable time saver. It is shameful how under-exploited they are in general serial programming. In parallel programming this is no less true. However there is one excuse: there are only two decent parallel debuggers (Totalview and DDT) and they are both proprietary.

### They are very easy to use

Just compile with the –g option and your executable will include everything needed to interact with the debugger at runtime. Note that this does impact performance substantially, so you wouldn't want to do this all the time.

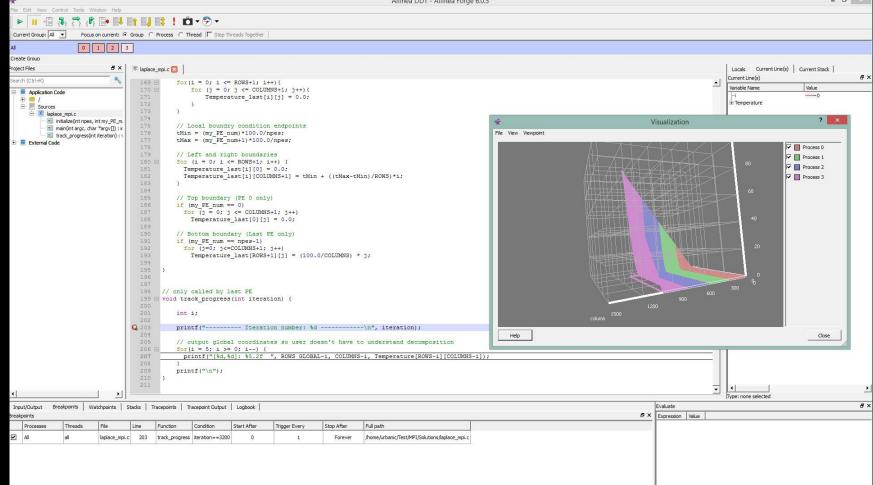
## DDT allows us to skip the X Windows complication

Most unix-land tools with nice GUIs require an X server driven display. DDT now has a desktop client that avoids what have been tricky configuration and maddening performance issues.

Note that remote usage, especially with a queuing system such as Slurm, requires the debugger to negotiate details with the remote platform (Bridges). It will do all of this automatically for you *once you configure the system the first time*. Our web pages will explicitly step you through this process.

*								Allinea DDT - Allinea	orge 6.0.5					- 1	0 ×
File Edit View Control Tools Wind				5											
] ▶ <b>    </b> ·⊡ ( <b>ᠯ</b> ) ( <b>下</b> ) ( <b>下</b> )				• 💬 •											
		Group C Proc	.cess C Thread							 					
	1 2 3			📕 Thi	S IS W	hat mak	it a	parallel db.							
Create Group		_						· · · · · · · · · · · · · · · · · · ·					1		
Project Files		C laplace_mpi.	и.с 🔯										Locals Current Line(s)	Current Stack	ē×
Search (Ctrl+K)	8	118 119	11			from below in						2	Variable Name	Value	
Application Code     Application Code     Sources     Sources     Modes_mpic     Iminipatient rose, and     manifinit arcs, and     manifinit arcs, and     manifinit arcs, and     Trads_progress(int it     External Code	*argv[]): ir	110 0 120 1 121 1 122 1 123 1 124 1 126 1 127 1 128 1 129 1 130 1 131 1 132 1 133 1 134 1 135 1 136 1 137 1 138 1 141 1 142 1 143 1 144 1 145 1 145 1 145 1 155 1 156 1 157 1 158 1 157 1 158 1 157 1 158 1 157 1 157 1 158 1 157 1 15	<pre>if if dt for } // MPI NPI NPI NPI NPI NPI NPI NPI NPI NPI N</pre>	<pre>(my PE num MPI_Reov ) ) ) ) ) ) ) ) ) ) ) ) ) ) ) ) ) ) )</pre>	<pre>n != npes-1 v(sTemperat i &lt;= ROWS; 1; j &lt;= CO = fmax(fab.perature la beal dt dt, sdt_gldbal, llly print on % 100) == EE num == n ck_progress ; s accurate COMM_WORLD tim(g and i)(v(stop_tim(s)) tim(s) error s to error st ero</pre>	<pre>){ ure_last[ROWS+ i++){ i++){ i++){ idouble_last[ROWS+ ist[i][j] = Tem obal, 1, MPI_I ist[i][j] = Tem obal, 1, MPI_DUBLE, test values - = 0) { pres-1){ idouble_last idou</pre>	//unless til[], Colt il[]]. Colt []]. Colt []]	<pre>we are bottom PE DRNS, MPI_DOUBLE, my_PE_num+ erature_last[i][j]), dt); [[j]; [[j]; MAX, 0, MPI_COMM_WORLD); M_WORLD); E in lower corner</pre>		itatus);			- Temperature De Temperature De Temperature Last		
Input/Output Breakpoints Wate	chooints   {	Stacks Tracer	points Trace	epoint Output	Logbook							Evaluate	Type: none selected		đ×
Breakpoints	inpenter 1 -	Autor   near	pointer   11223	point output 1	collecon 1			87		 	e×				-
	File	Line Fun	unction Co	Condition :	Start After	Trigger Every	Stop After	Full path							
Al all	laplace_mpi.c	c 139 mair	in ite	eration==3200	0	1	Forever	/home/urbanic/Test/MPI/Solutions/laplace_	mpi.c						
	2		ja.	Î	7		0								

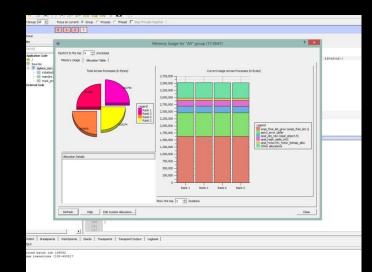
Ready Connected to: urbanic@bridges.psc.edu



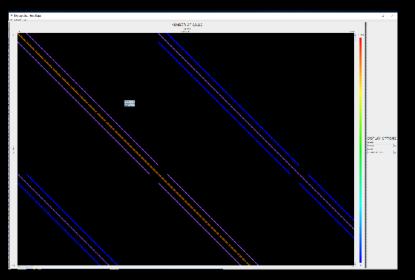
Ready Connected to: urbanic@bridges.psc.edu

# **Other Nifty Features**

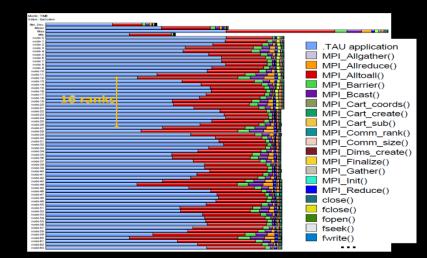
- Can deal with very large numbers of PEs
  - Has ways of coalescing PEs into groups, etc.
  - Has ways of comparing and contrasting across PEs to find anomalies
- Can connect to running jobs
- Also great for OpenMP and CUDA/OpenACC
- Also good for memory debugging
- and MPI tracing



## **Profilers** Great Open Source Option: TAU

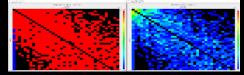


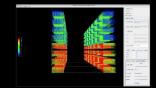
Nearest neighbor communications in 3D.



Comp/comm ratio and other hotspots.

### Lots more...





# **Profilers** Intel: Vtune, Advisor & Trace Analyzer

•

•

•

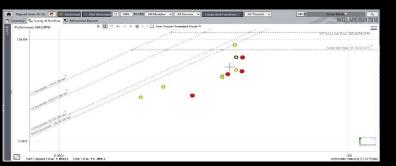
•

•

•

Intel® Trace Analyzer - [1: Dr/PSC/PRAC 2017/psdns/tac/bridges_1024_512_02/trace_data/DNS	52c_mpi_p8xstf)		- 0
file Options Project Windows Help			
🕎 🚍 🛣 34.319.926 - 31.867.771 - 6.547.875 Seconds 🔷 😪 All,Processes 😡	HPI seconded in (Majir Function Croups) 🕢 🍸 🔝 🖉 🔝 🖉		
			tins xy partoen
the test of the bard as bard as being the set of the se	and the many states of the last of the second states of the second state	which is a second provide the second second second	and the second states of the
Aschoslik AppMPL Allow Aspital AppR Aug APL Atmatical to Aspital AppR App	station Application Application MPLA	In Asil Asplane Mos Min - Macol Kostiumon A	ApApplication Applied Di Linguna
A BORRETINA PORT A THAT A DESIGNATION AND A DESI	READO ADDICATION ADDIC	ALANDAR DEPOSITION NO. A DALARD ALD ALANDAR	
And Acone all And	interior Application Application MP dealers	Appendix MITACIAA DOA DOMINION MITA DAD DIRECTOR	
A Rest & Dest & A part & a part of the Apartic Apart & State Apart & State & S	Realister Application Application Mill Ap/Application	atter MILL A privatigation Mill Aphagination Appl	1 (A.10 + 4)
NEUTRALINATION AND AREAD TO COMPANY A ROCK OF THE TAXABLE PARTY OF THE	Roter Application Application application and application applicat	PT ADARCARCHICASON MPLACARDISONON A CONCAPT ATTACK	
AREAAUMA DESAGARE A RADORD LL PLUT COMPANY AND ARE ARE	Institut-Application Appl/AppApphiation PETA VALApplicationAppApApplica	HER AT A PERSON ADD - A THE PERS	A CONTRACTOR OF
	it is policity of a policy ment of it politic is policy and a policy of the second second second second second		
	REAR A SOLOMADOA A SERIEM SCAASCING ASOLOGIUSTA		y Altopli on
P Reprint Reprint (SAM) - a first again (SAM) - A local sector (SAM)	items Application application and an instance in Application of Application and Applications		
AND A DESCRIPTION OF A DESCRIPTION AND A DESCRIPTION OF A			Jappanois
2 ROOM ADDRESS A TRANSPORTATION OF A POLICY DESCRIPTION ADDRESS OF A POLICY OF	nestor Applement Applyshis policybos 1 801 Applaced	COLORDON HIP & CALL OF A DOMESTICAL STRATE OF A DATA OF	Sector Alterall
	Asaber Rooksabar Rooksabar Rooksabar Rooksabar	Asti-Asticution MELAOSAAcc-Institution MARPADELD	MART ARTICLE
<ol> <li>Artenisk polici Artenis and an artenisk policy in activation of Artenis Control of Artenistic Artenisti Artenistic Artenistic Artenisti Artenistic Artenistic Arten</li></ol>	Anders Application And extending the strend	Antheory and an anti-Antheory and an Antheory and Antheory antheo	Application of Designation of the Designation
S Applikacie Meter Altheide - D Application (altheid attract Modification) Application Application (Application)	Australia Australia Australia Australia and Australia	A STATISTICS OF A STATISTICS O	AND ADDRESSON ADDRESS ADDRESSON
A ADARDER STORES ATTACK REPORT ADDRESS AND A TAXAGE A DRAW A ADARDER AND A DRAW A ADDRESS ADDRES	Branse Apationsie ApatiesApoApptiverie (***) Attiviet	THE DESCRIPTION OF A REPORT OF	AND ADDIENT ADDIT ADDIED
7 alsa besa landere i a fotori a la farin nisaren besaret a itmat ia di trianica a vieren azartinea polaren	incettor applicate particition life a brok	DELAPOROLOGICATION NET LIPSCH COLORADORNOON NET LIP	ou paleeton Appener altrait amount
2 Advantage of the second s	nutur Assumption Assumption of a state of a state		And And Attent
	Anna and a second	and south a second data with the second data and the second data and the second data and the second data and the	110000
	a Nas Managementan Apply and a second s		2000AU
		a and a state of the	1 ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) (
مراجع معالمة المتعالية المنابعة المالية المستعد التعانية من بها التجارية المتعانية المن تكريمات من 	all and a sublement of the second state of the		
A			
Ret Profile Level Belerco Coll Tree Cel Graph		and a second second second	
K_Pricesses *	CONTRACTOR NO.	Reported max time	T Internet and
ame TSelf TSelf TTotal #Calls TSelf /Call	. Start of an	for this xy-Alltoall	End of th
All races All races All races and All races	inks xy-Alltoall	tor the ny thread	xy-Alltoal
Group Application 162,053 200,343 0 m.a. MP1_Alteall 87,3522 1 8257 26,8191e-3 1 SVNChTC	xy-Alitoali		xy-Alitoa
MPI Witne 56,152e-3 s 56,152e-3 s 56,057e-9 s			
MPI_Alfreduce 31.0123 = 31.0123 400 64.7527e-33 after All	reduce		

#### Message tracking a tricky All-To-All based code.



#### **Tune OpenMP Scalability**

Hotspot (Statistical call tree), Call counts

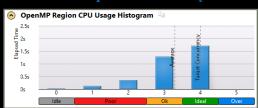
GPU Offload and OpenCL<sup>™</sup> Kernel Tracing View Results on the Source / Assembly

Visualize Thread & Task Activity on the Timeline

Cache miss, Bandwidth analysis...<sup>1</sup>

OpenMP Scalability Analysis

Thread Profiling – Concurrency, Lock & Waits Analysis



#### **Quickly Find Tuning Opportunities**

	CPU Time+ 🛪 📧					
Function / Call Stack	Effective Time by Utilization	Spin Time	Overhead Time	1		
	🔲 idle 📕 Poor 📒 Ok 📕 ideal 📒 Over					
FireObject::checkCollision	4.507s	Os	05	i		
IFireObject::ProcessFireCollisionsRange	3.444s	Os	05			
NtWaitForSingleObject	0s	3.406s	i 0s	5		
std::basic_ifstream <char,struct std::char_traits<="" td=""><td>3.359s</td><td>0s</td><td>: Os</td><td></td></char,struct>	3.359s	0s	: Os			
Ogre::FileSystemArchive::open	3.359s	Os	0s			
CBaseDevice::Present	2.335s	0.671s				
Selected 1 row(s):	1.151s	0.728s	: Os	V		

#### Visualize & Filter Data

06040-0 <b>4</b>			Frame Rate
Frame Rate	the second s	-	Auf Frame Rate
wWinMainCRTStartup		~	Thread
_endthreadex (TID: 91 endthreadex (TID: 91			Running
endthreadex (TID: 91			Waits
CBatchFilterl::LHBatc			CPU Time
CPU Usage	in the second		Spin and Over     CPU Sample
Thread Concurrency			P P Tasks
	5		Transitions

### Roofline Analysis.