# ParaTools ThreadSpotter ParaTools, Inc.

John C. Linford et al. jlinford@paratools.com

Army HPC User Group Review 18 May 2017, ARL

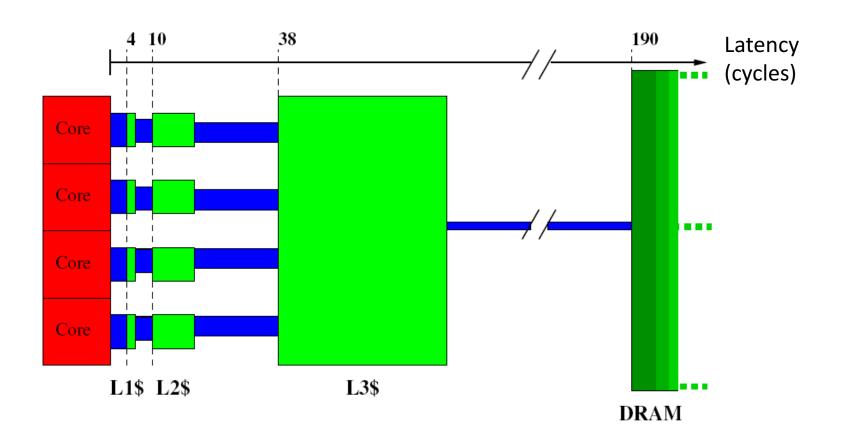


### ParaTools ThreadSpotter

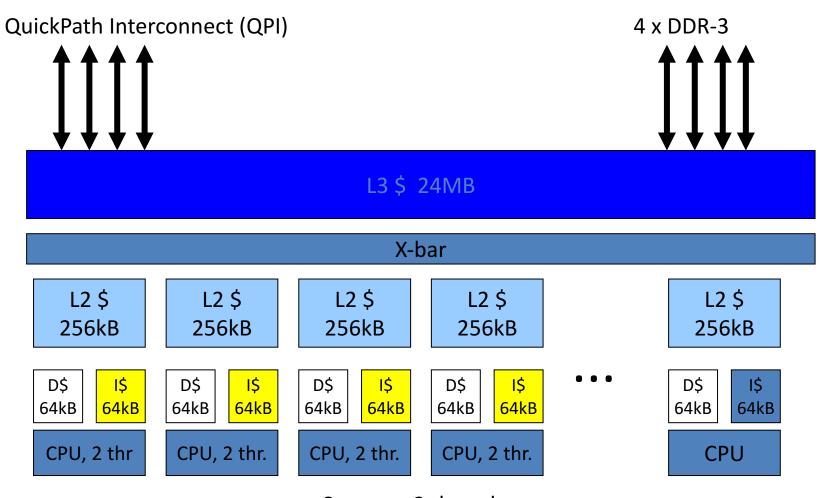
 Shrinking cache memory per core is causing some applications to spend over half their runtime waiting for data to arrive in cache.

- ThreadSpotter can:
  - Analyze memory bandwidth and latency, data locality and thread communications.
  - Identify specific issues and pinpoints troublesome areas in source code.
  - Provide guidance towards a resolution.
- Supports mixed language distributed memory applications like CREATE-AV HELIOS and KESTREL.
  - Integrates with the TAU Performance System<sup>®</sup>.

# Cache Capacity/Latency/BW



### Intel i7



8 cores x 2 threads

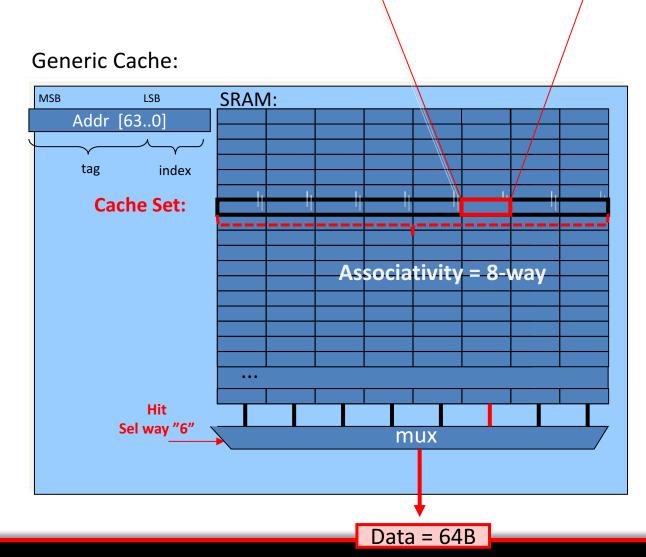
# Typical Cache Implementation

Cacheline, here 64B:

**AT S Data** = 64B

L3 \$ 24MB

L2 \$ 256kB



**Army HPC User Group Review** 

# **THREADSPOTTER REPORTS**



# Report Index for 2040 MPI Ranks

### ThreadSpotter™

ThreadSpotter™ is a tool to quickly analyze an application for a range of performance problems, particularly related to multicore optimization

### Read more... Manual

A modern processor has two types of resources that are both limited and highly responsible for application performance:

- . The memory bus, which transports data between the memory and the processor, and
- The caches, which store often used data so each access doesn't have to go all the way to memory.

A third resource, the hardware prefetcher, is also important for application performance, as it anticipates the application's memory access patterns, and tries to fetch data ahead of time. If it is successful, this hides some of the time it takes to access memory.

It is of great importance for application performance and scalability on multicore processors that these resources are used efficiently.

ThreadSpotter™ looks for abuse of these resources and points out the responsible slowspots.

Node	Memory Bandwidth	Memory Latency	Data Locality	Thread Communication / Interaction
<u>o</u>			$\bigcirc$	$\triangle$
1			$\triangle$	$\triangle$
<u>2</u>			$\triangle$	4
<u>3</u>			$\triangle$	4
4			<u>~</u>	4
<u>5</u>	<u>⟨</u>		$\bigcirc$	
<u>6</u>			$\bigcirc$	4
<u>7</u>			$\bigcirc$	4
<u>8</u>			$\bigcirc$	4
<u>9</u>			$\triangle$	
<u>10</u>			$\triangle$	$\triangle$
<u>11</u>	<u> </u>		$\sim$	4
<u>12</u>			$\triangle$	
<u>13</u>	<u>₩</u>		$\triangle$	<u> </u>
<u>14</u>			$\bigcirc$	4
<u>15</u>			$\bigcirc$	4
<u>16</u>	<u>⟨</u>		$\bigcirc$	$\triangle$
<u>17</u>			$\bigcirc$	$\triangle$
<u>18</u>				$\triangle$
<u>19</u>	· A		$\triangle$	$\triangle$
<u>20</u>				$\triangle$
<u>21</u>			$\bigcirc$	$\bigcirc$
<u>22</u>	<u>~</u>		$\bigcirc$	4
<u>23</u>			$\bigcirc$	$\bigcirc$

### **Para**Tools

### **Next Steps**

The prepared report is divided into sections.

- Select the tab Summary to see global statistics for the entire application.
- Select the tabs Bandwidth Issues, Latency Issues and MT Issues to browse through the detected problems.
- Select the tab Loops to browse through statistics and detected problems loop by loop.

The Issue and Source windows contain details and annotated source code for the detected problems.



### Resources

Manual

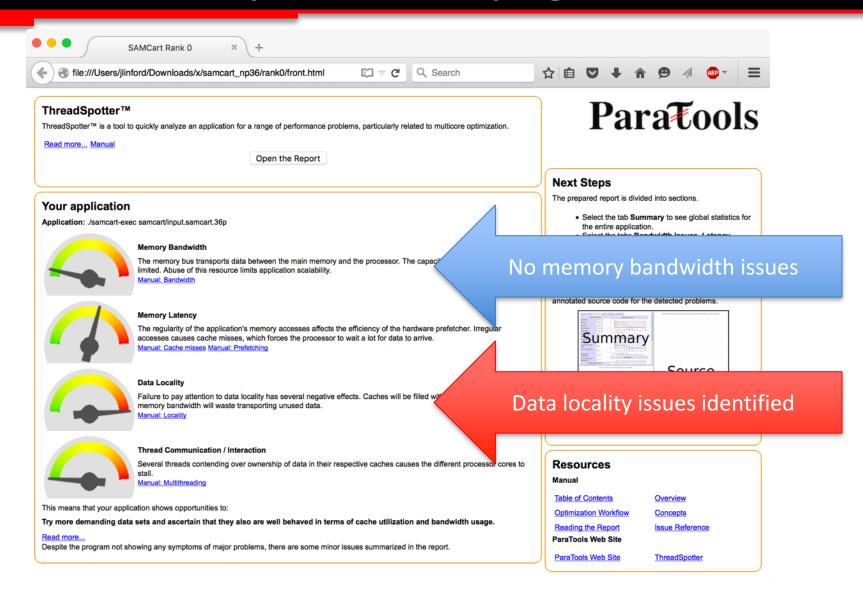
Table of Contents
Optimization Workflow
Reading the Report

Overview
Concepts
Issue Reference

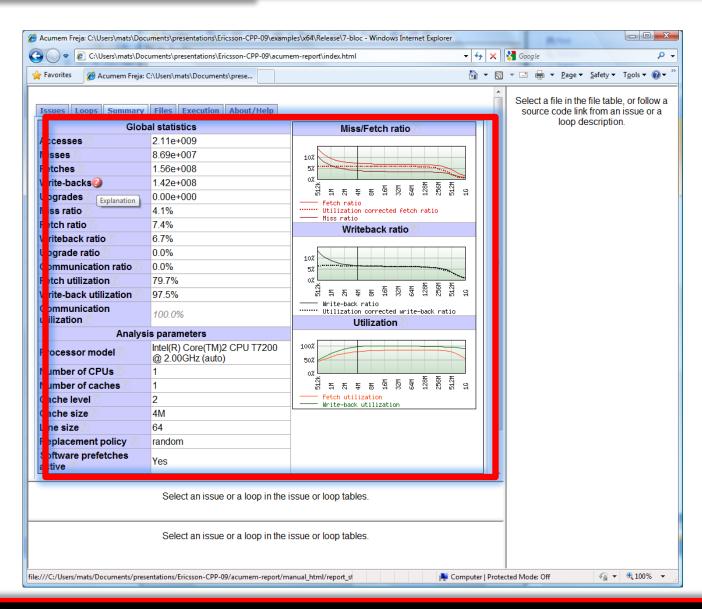
ParaTools Web Site

ThreadSpotter

### Report Cover page



### **Global statistics**



# ThreadSpotter Concepts

### Issue

- A problem or opportunity to improve performance
- Presented with backing statistics and source navigation
- Online help to explain details

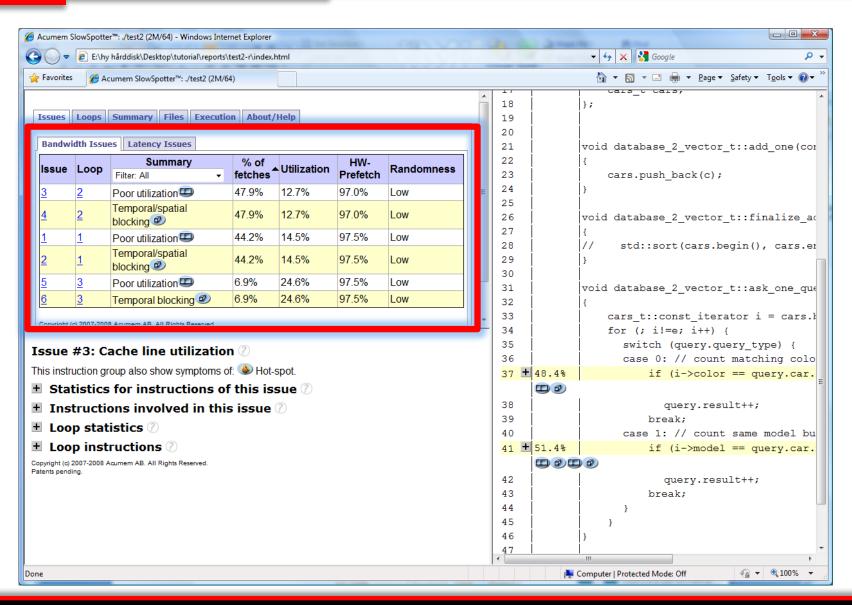
### Loop

An instruction cycle usually corresponding to a code loop

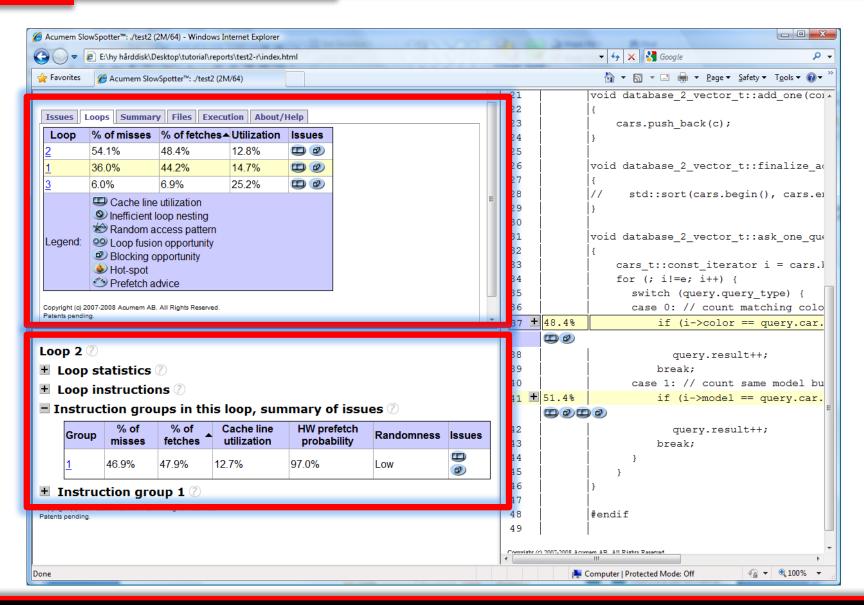
### Instruction group

A collection of instructions touching related data

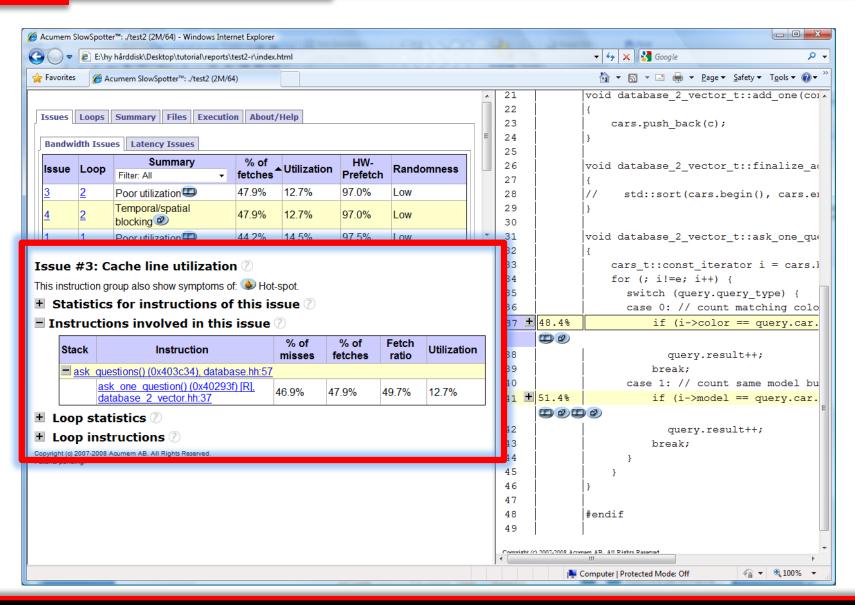
# ParaTools ThreadSpotter report: Issues



# ParaTools ThreadSpotter report: Loops



# ParaTools ThreadSpotter report: Details

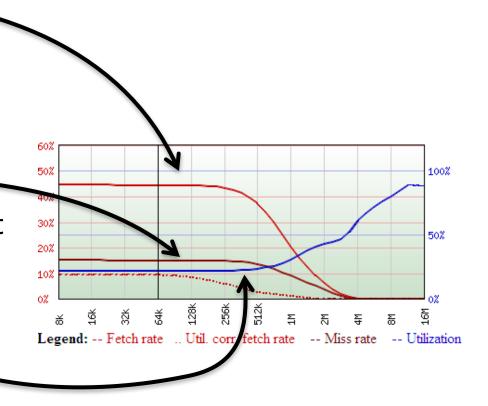


### Metrics as a function of cache size

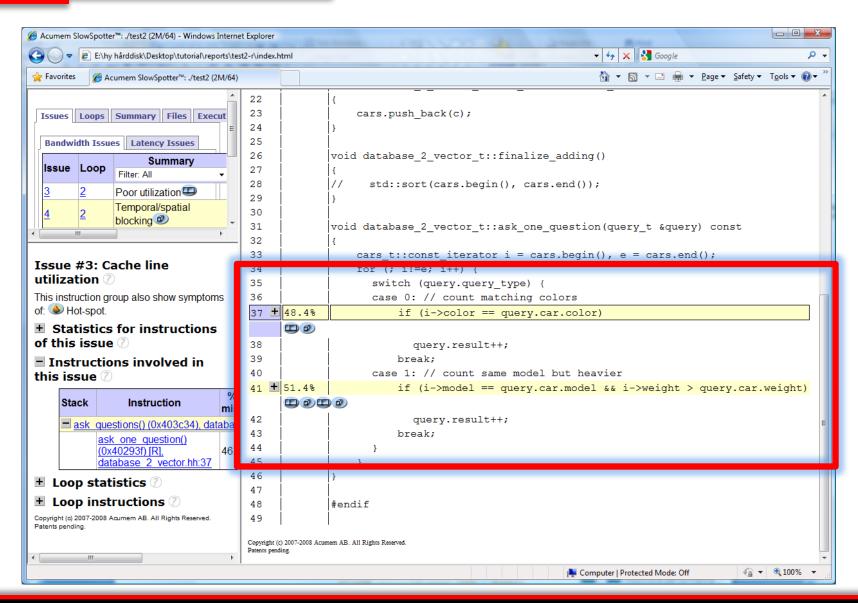
Fetch ratio

Memory operations
 that cause a data
 transfer to/from RAM

- Miss ratio
  - Memory operations that stall due to cache misses.
- Fetch utilization
  - Fraction of the data loaded into the cache that are actually used



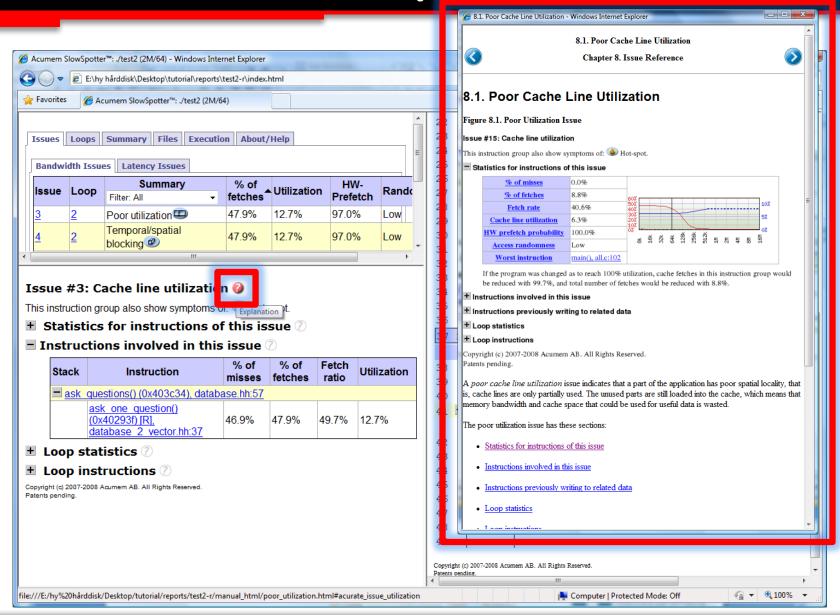
### ParaTools ThreadSpotter report: Source annotation



### Report Vocabulary

- Miss ratio: What is the likelihood that a memory access will miss in a cache?
- **Miss rate**: Misses per unit, e.g. per-second or per-1000-instructions
- Fetch ratio/rate: What is the likelihood that a memory access will cause a fetch to the cache (including HW prefetching)
- Fetch utilization: What fraction of a cacheline was used before it got evicted
- Writeback utilization: What fraction of a cacheline written back to memory contains dirty data
- Communication utilization: What fraction of a communicated cacheline is ever used?

ParaTools ThreadSpotter report: help



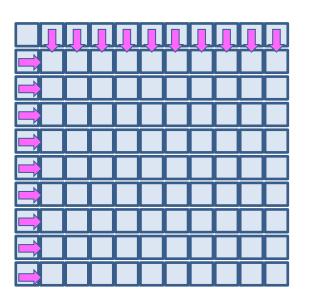
**Army HPC User Group Review** 

# **EXAMPLE: TEMPORAL BLOCKING**



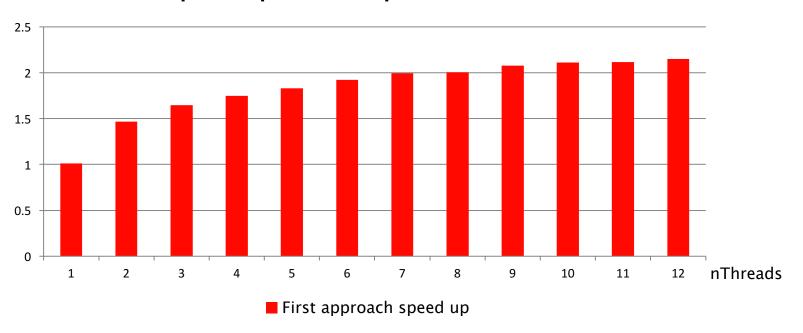
### Overview of the forward elimination stage

```
for i=1 to n-1
🛶 find pivotPos in column i
   if pivotPos ≠ i
     exchange rows(pivotPos,i)
   end if
   for j=i+1 to n
     A(i,j) = A(i,j)/A(i,i)
   end for j
  !$omp parallel do private ( i ,j )
   for j=i+1 to n+1
     for k=i+1 to n
         A(k,j)=A(k,j)-A(k,i)\times A(i,j)
     end for k
   end for j
end for i
```



# First approach speed up

### Speed up w.r.t. sequential version

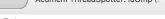


ParaTools ThreadSpotter report



Acumem ThreadSpotter: luOmp (₺ ¾

+





file:///Users/jlinford/Documents/ParaTools/ARL:APG/AHUGR17/luOmpGE/front.html



Q Search









### **Acumem ThreadSpotter**

Acumem ThreadSpotter is a tool to quickly analyze an application for a range of performance problems, particularly related to multicore optimization.

Read more... Manual

Open the Report

### Your application

Application: ./luOmp



### **Memory Bandwidth**

The memory bus transports data between the main memory and the processor. The capacity of the memory bus is limited. Abuse of this resource limits application scalability.

Manual: Bandwidth



### Memory Latency

The regularity of the application's memory accesses affects the efficiency of the hardware prefetcher. Irregular accesses causes cache misses, which forces the processor to wait a lot for data to arrive.

Manual: Cache misses Manual: Prefetching



### Data Locality

Failure to pay attention to data locality has several negative effects. Caches will be filled with unused data, and the memory bandwidth will waste transporting unused data.

Manual: Locality



### Thread Communication / Interaction

Several threads contending over ownership of data in their respective caches causes the different processor cores to stall.

Manual: Multithreading

This means that your application shows opportunities to:

Try more demanding data sets and ascertain that they also are well behaved in terms of cache utilization and bandwidth usage.

### Read more.

Despite the program not showing any symptoms of major problems, there are some minor issues summarized in the report.

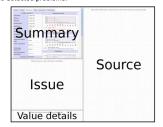
### **ParaTools**

### **Next Steps**

The prepared report is divided into sections.

- Select the tab Summary to see global statistics for the entire application.
- Select the tabs Bandwidth Issues, Latency Issues and MT Issues to browse through the detected problems.
- Select the tab Loops to browse through statistics and detected problems loop by loop.

The Issue and Source windows contain details and annotated source code for the detected problems.



### Resources

### Manual

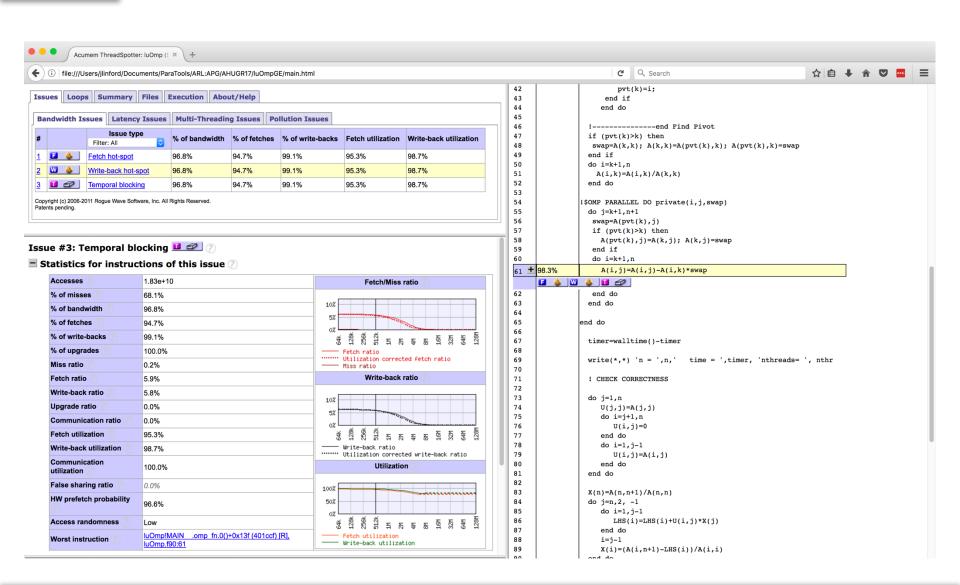
Table of Contents Overview

Optimization Workflow Concepts

Reading the Report Issue Reference

Rogue Wave Software Web Site

Rogue Wave Web Site Tutorials



### What went wrong!

- For each prepared pivot, the whole matrix is accessed. The algorithm requires pivots to be calculated in order.
- Repeated eviction of the matrix' cache lines.

### What went wrong!

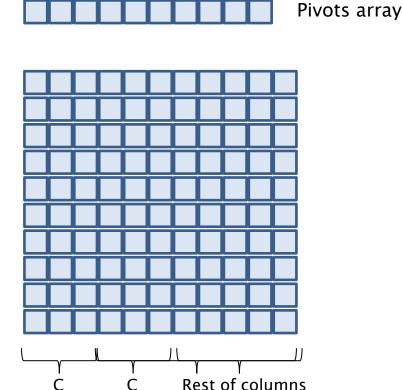
- For each prepared pivot, the whole matrix is accessed. The algorithm requires pivots to be calculated in order.

  Making things right!
- Repeated eviction of the matrix' cache lines.
  - Observation: Each column is an accumulation of eliminations using previous columns!
  - Temporal Blocking Advice says:
     Use each column many times before it gets evicted
    - → Arrange code to make more pivots available!

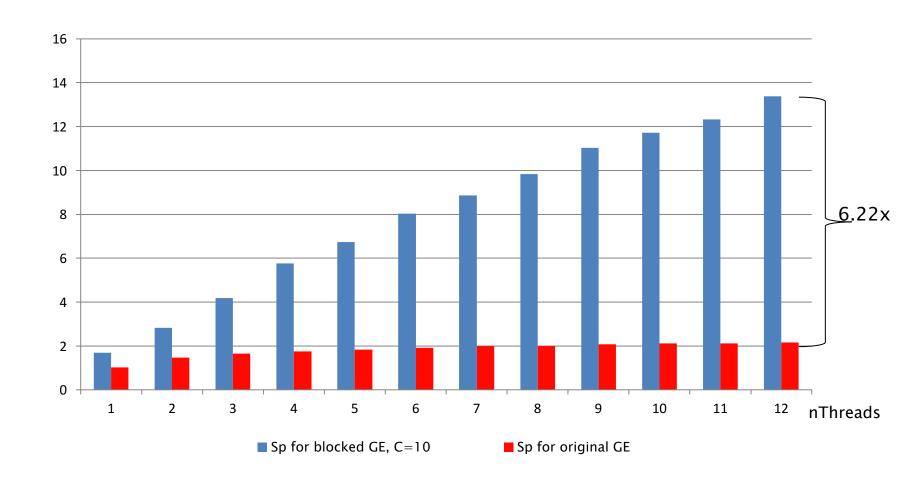
### **Blocking GE**

```
for k=1 to n-1, step C
 BlockEnd=min(k+C-1,n)
 <sup>1</sup> GE on A(k:n,k:BlockEnd) &
 - Store C pivots' positions
  !$omp parallel do private ( i ,j )
  for each column j after BlockEnd
     for i=k to BlockEnd
     swap using pivots(i)
          elimination i on j
     end for i
  end for each i
End for k
```

The row exchange turned into a two-element swap before column elimination



# Speed up w.r.t sequential time



Army HPC User Group Review

# COMMAND LINE USAGE



### PTTS Integration with TAU

- tau\_exec --ptts
  - Sampling (sample\_ts)
    - Generates sample files named ptts/sample.#.smp
  - Issue Identification (report\_ts)
    - Generates report files named ptts/report.#.tsr
  - Reporting (view-static\_ts )
    - Generates node\_# directories with html reports
- Each rank produces log files for each step
  - sample\_ts.#.log, report\_ts.#.log, and viewstatic\_ts.#.log

# PTTS Integration with TAU

Flag	Description		
-ptts	Use PTTS to generate performance data.		
-ptts-post	Skip application sampling and post-process existing sample files. Useful for analyzing performance at each cache layer, or predicting performance on other architectures.		
-ptts-num= <n></n>	Indicate number of MPI ranks if the size of the MPI communicator cannot be automatically detected, e.g. on Cray.		
-ptts-sample-flags= <flags></flags>	Additional flags to pass to the sample_ts command. Can also be specified in the TAU_TS_SAMPLE_FLAGS environment variable.		
-ptts-report-flags= <flags></flags>	Additional flags to pass to the report_ts command. Can also be specified in the TAU_TS_REPORT_FLAGS environment variable.		

## Sampler settings: Start & Stop

- Controlling when to start and stop sampling
- Start conditions
  - delay, -d <seconds>
  - --start-at-function <function name>
  - --start-at-address <0x1234123>
  - --start-at-ignore <pass count>
- Stop conditions
  - Duration, -t <seconds>
  - --stop-at-function <function name>
  - ---stop-at-address <0x123123>
  - --stop-at-ignore <pass count>

# Report settings: Issue selection

### depth

- How long call stack to consider while differentiating instructions
- --depth 0 just consider the PC
- --depth 1 consider the PC and the site which called this function

### percentage

- cutoff. Suppress uninteresting issues
- --percentage 3 (% of fetches)

## Report settings: Source/debug

- source directory look for source in other places
  - Useful if the directories are not recorded in the debug information
  - -s directory
- binary look for binaries (containing debug information in other places)
  - Useful if the binary has moved since sampling
  - -b path-to-actual-binary
- debug directory (for debug information stored external to the elf file)
  - -D directory
- debug level (varying degree of information)
  - --debug-level 0-3