

Multi-Core Processors – Are they Here Yet?

My shopping basket at Fry's electronics on BlackFriday

Item	Cost
Motherboard+Intel® Quad core 2.4Ghz	200
4GB Memory	70
0.5TB Disk	80
Case	10
Graphics	10
CD/DVD	10
Total	380



Outline

Parallelism in Intel® processors

Intel® tools for parallel programming

Some thoughts



A New Era...

THE NEW

THE OLD

Performance Equals IPC Multi-Core

Power Efficiency

Microarchitecture Advancements

Performance Equals Frequency

Unconstrained Power

Voltage Scaling



Intel® Wide Dynamic Execution

EACH CORE

EFFICIENT 14 STAGE PIPELINE

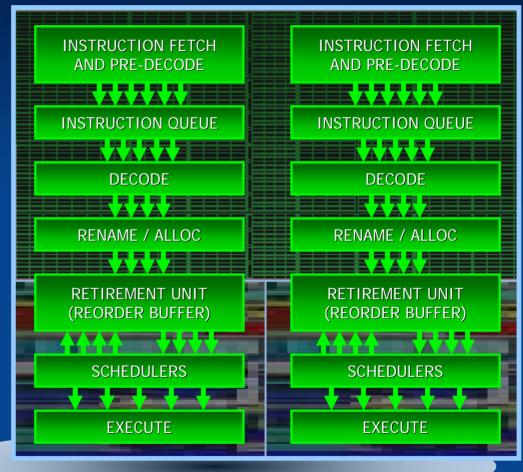
DEEPER BUFFERS

> 4 WIDE -DECODE TO EXECUTE

4 WIDE -MICRO-OP EXECUTE

> MICRO and MACRO FUSION

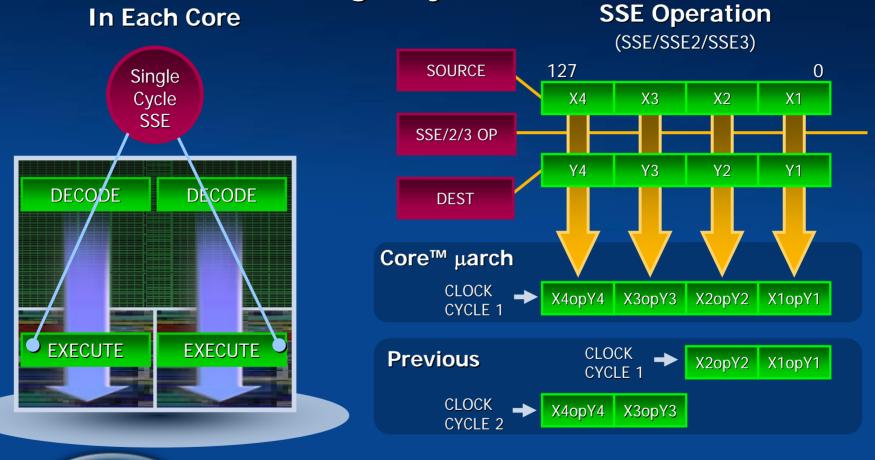
ENHANCED ALUs CORE 1 CORE 2





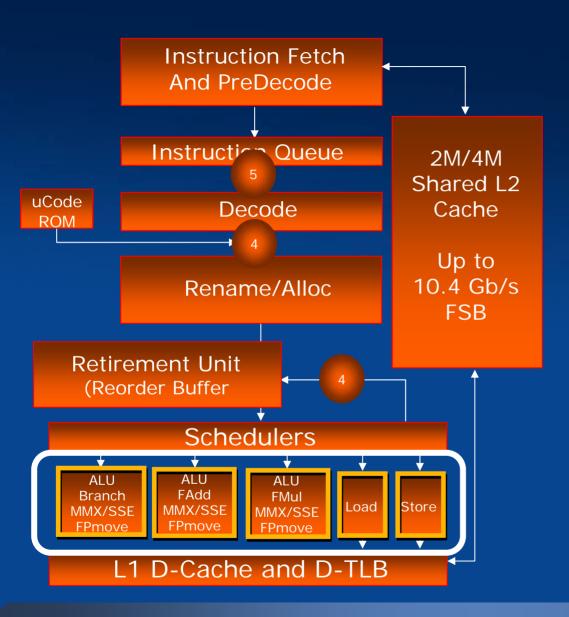
Multiple cores. Multiple instruction issue.

Intel® Advanced Digital Media Boost Single Cycle SSE





SIMD instructions compute multiple operations per instruction



Advanced Digital Media Boost

128-bit Packed Multiply

plus 128-bit Packed Add

plus 128-bit packed Load

plus 128-bit packed Store

plus
Compare and jump

SIMD Operations Combined with Multiple Instruction Issue

Many Levels of Parallelism

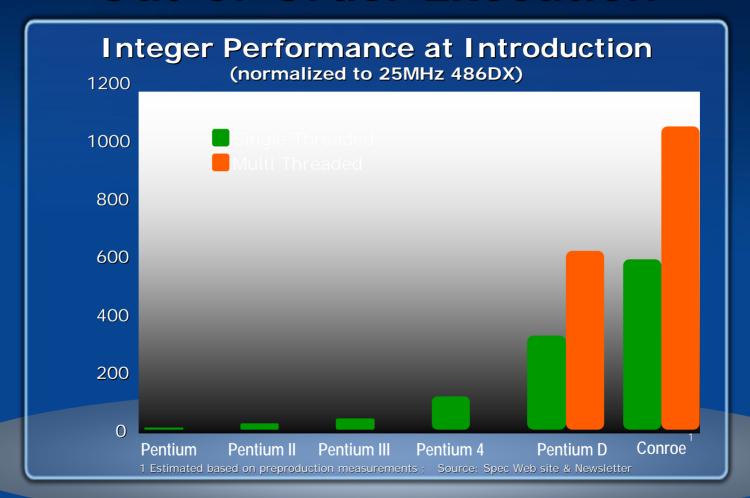
Instruction-level parallelism

SIMD (vector) parallelism

Multi-core parallelism



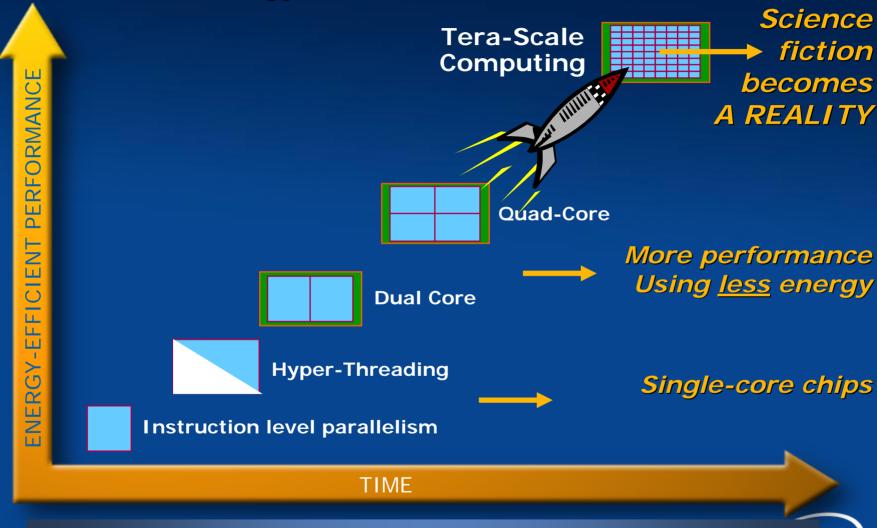
Biggest Performance Leap Since Out-of-Order Execution



Parallel Processor: Instruction-level Parallelism, SIMD Parallelism, Multi-core Parallelism

Tera-Leap to Parallelism:

Energy Efficient Performance



Many core requires a high degree of scaling in an application.

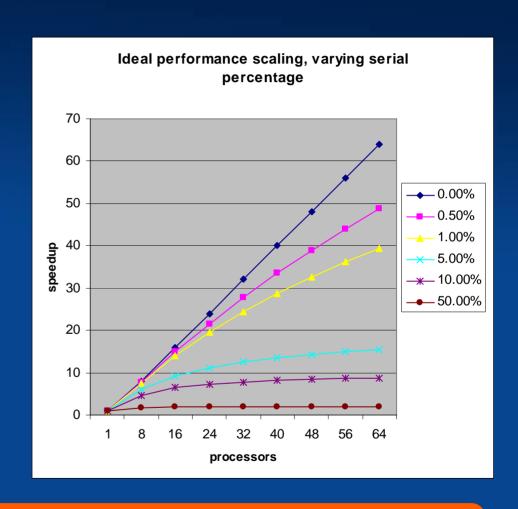


Amdahl's Law

Runtime = serial + parallel

With perfect scaling: Runtime = serial + parallel/P

To get 50x speedup on 65 processors, serial time must be less than 0.5%



Emerging class of highly parallel workloads: games, media processing, personal data search, and more

Application Example

Sports Video Indexing



Recognize players, objects and events

Mine the video for target scenes

Synthesize a summary of what happened



Outline

Parallelism in Intel[®] processors

Intel® tools for parallel programming

Some thoughts



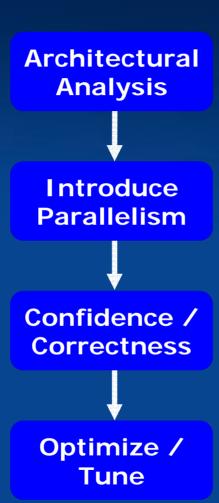
Parallel Programming Deserves Great Tools

Visualization of applications and the system

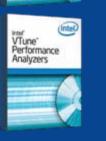
Highly optimizing compilers delivering scalable solutions

Detect latent Programming bugs

Tune for performance and scalability













Example: Prime Number Generation

Outer loop steps through numbers that may be prime

Inner loop tests factors of number

```
#include <stdio.h>
const long N = 18;
long primes[N], number of primes = 0;
main()
 printf( "Determining primes from 1-%d \n", N );
 primes[ number of primes++ ] = 2; // special case
  for (long number = 3; number <= N; number += 2)
   long factor = 3;
   while ( number % factor ) factor += 2;
    if (factor == number)
     primes[ number of primes++ ] = number;
 printf( "Found %d primes\n", number_of_primes );
```



Example: Prime Number Generation

```
#include <stdio.h>
const long N = 18;
long primes[N], number of primes = 0;
main()
 printf( "Determining primes from 1-%d \n", N );
 primes[ number of primes++ ] = 2; // special case
  for (long number = 3; number <= N; number += 2)
   long factor = 3;
    while ( number % factor ) factor += 2;
    if ( factor == number )
     primes[ number of primes++ ] = number;
  printf( "Found %d primes\n", number_of_primes );
```

```
11 3,5,7,9,11 1 3 3,5,7,9,11,13 15 3
```

3 3 5 3,5

7 3,5,**7**

9 3

```
17 3,5,7,9,11,13,15,17
```

```
C:\Intel(R)_Threading_Tools\build\Release\Primes.exe"

Determining primes from 1-18

Found 7 primes

Press any key to continue_
```

Using the Intel Programming Tools

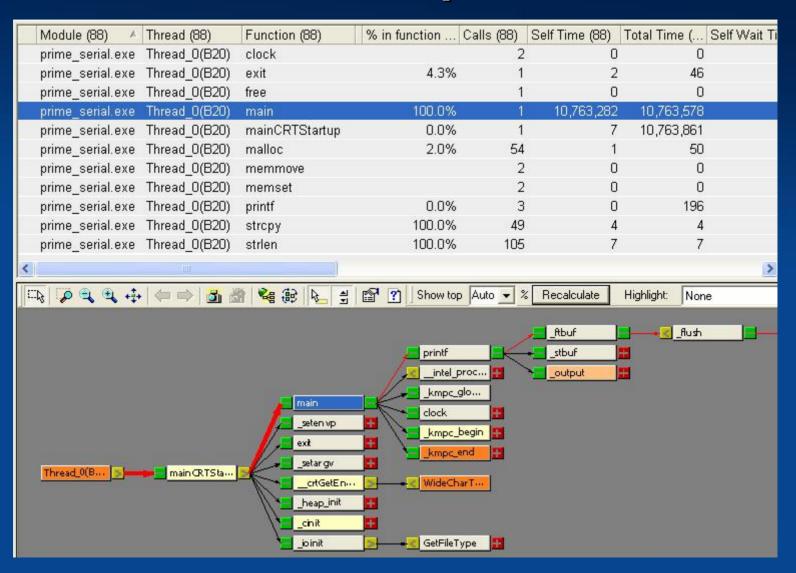
Architectural Analysis Introducing **Threads** Debugging **Performance Tuning**

Intel® VTune™ Performance Analyzer

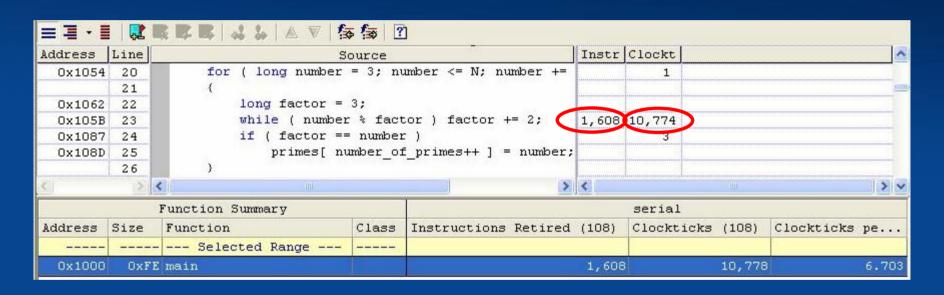
- Call Graph
 - Functional Structure
 - Execution Times
 - Counts



Vtune™ Call Graph Profile



Vtune™ Sampling Profile



Sampled Cycles: 10,774

Sampled Instructions: 1,608

CPI: 6.7



Hardware Support for Sampling



Event counters count events that occur in pipeline.

Key events:

- Clock_cycle
- Instruction_retired

Counters count down to zero.

- At zero, interrupt and capture instruction pointer.
- Reset counter and continue execution.

Event2

Event1

Statistically associate events with instructions

Example: Prime Number Kernel

```
for (long number = 3; number <= N; number += 2)
                       long factor = 3;
   3,5
                       while ( number % factor ) factor += 2;
                       if ( factor == number )
   3,5,7
                        primes[ number_of_primes++ ] = number;
11 3,5,7,9,11
13 3,5,7,9,11,13
15 3
17 3,5,7,9,11,13,15,17
```



Using the Intel Programming Tools

Architectural Analysis Introducing **Threads Debugging Performance Tuning**



- OpenMP Loop Construct
 - Creates one thread per core
 - Assigns iterations to threads



Example: Add OpenMP Directive

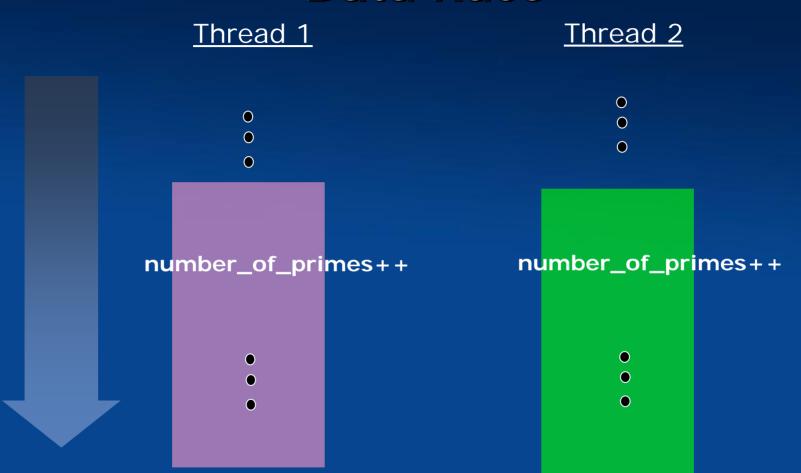
```
#include <stdio.h>
const long N = 100000;
long primes[N], number_of_primes = 0;
main()
  printf( "Determining primes from 1-%d \n", N );
  primes[ number of primes++ ] = 2; // special case
#pragma omp parallel for
  for (long number = 3; number <= N; number += 2)
    long factor = 3;
    while ( number % factor ) factor += 2;
    if (factor == number)
     primes[ number_of_primes++ ] = number;
  printf( "Found %d primes \n", number of primes );
```

Example: Add OpenMP Directive

```
#include <stdio.h>
const long N = 100000;
long primes[N], number_of_primes = 0;
main()
                    Thread 1
  print
  prime
              Iterations 1..50000
#pragma
  for (long number = 3; number <= N; number +=
     long factor = 3;
                   Thread 2
          Iterations 50000..100000
  printf( "Found %d primes \n", number_of_primes );
```

Example: Add OpenMP Directive

```
#include <stdio.h>
const long N = 100000;
                              Command Prompt
long primes[N], number of p
                               C:\Primes\Release>Primes.exe
main()
                               Determining primes from 1-100000
                               Found 9592 primes
  printf( "Determining prin C:\Primes\Release>Primes.exe
                              Determining primes from 1-100000
  primes[ number of primes+
                              Found 9589 primes
#pragma omp parallel for
                              C:\Primes\Release\Primes.exe
   for (long number = 3; nu
                              Determining primes from 1-100000
                               Found 9590 primes
     long factor = 3;
                               C:\Primes\Release>Primes.exe
                              Determining primes from 1-100000
     while ( number % factor
                              Found 9588 primes
     if (factor == number
                               C:\Primes\Release>Primes.exe
       primes[ number of pri
                              Determining primes from 1-100000
                              Found 9591 primes
  printf( "Found %d primes C:\Primes \Release >_
```



The updates to number_of_primes are not atomic!



Thread 1

Thread 2

$$t1 = LOAD n_of_p$$

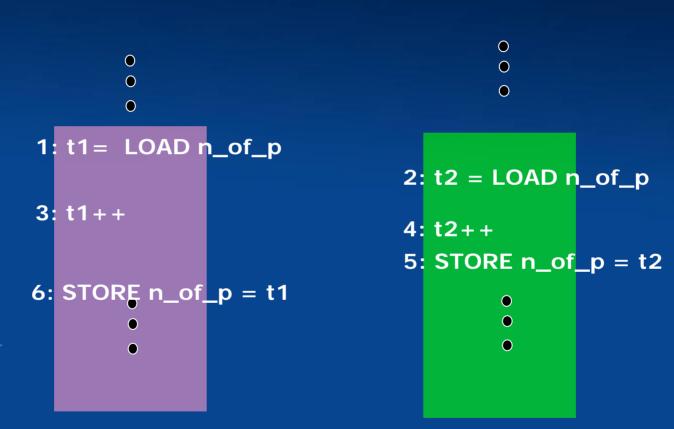
STORE
$$n_{of} = t1$$

$$t2 = LOAD n_of_p$$

The updates to number_of_primes are not atomic!



Thread 1 Thread 2



Lose one increment to number of primes!



Thread 1 Thread 2

```
CRITICAL {
    t1 = LOAD n_of_p
    t1++
    STORE n_of_p = t1
}
```

```
CRITICAL {
    t2 = LOAD n_of_p
    t2++
    STORE n_of_p = t2
}
```

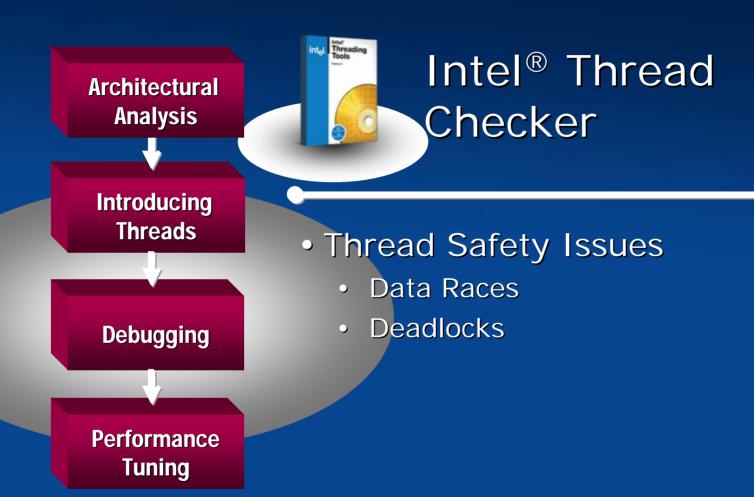
Add synchronization to remove data race



Example: Add Synchronization Directive

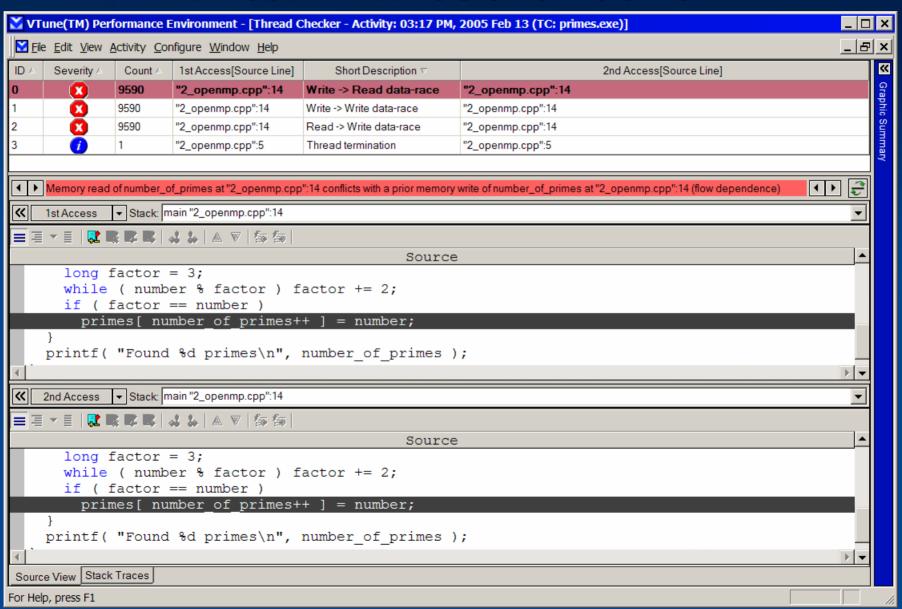
```
#include <stdio.h>
const long N = 100000;
                               Command Prompt
long primes[N], number of pr
                               C:\Primes\Release>Primes
main()
                               Determining primes from 1-100000
                               Found 9592 primes
  printf( "Determining prime C:\Primes\Release>Primes
  primes[ number_of_primes++ Determining primes from 1-100000
                               Found (9592) primes
#pragma omp parallel for
                               C:\Primes\Release>Primes
  for (long number = 3; num
                               Determining primes from 1-100000
                               Found 9592 primes
    long factor = 3;
                               C:\Primes\Release>Primes
                               Determining primes from 1-100000
    while ( number % factor
                               Found 9592 primes
    if (factor == number)
                               C:\Primes\Release>Primes
#pragma omp critical
                               Determining primes from 1-100000
      primes number of prim Found (9592) primes
                               C:\Primes\Release>_
  printf( "Found %d primes \n", number of primes );
```

Using the Intel Programming Tools





Intel® Thread Checker



Thread 1 \odot Lock(L); n_of_p++ Unlock(L); Thread 2

0

Lock(L);

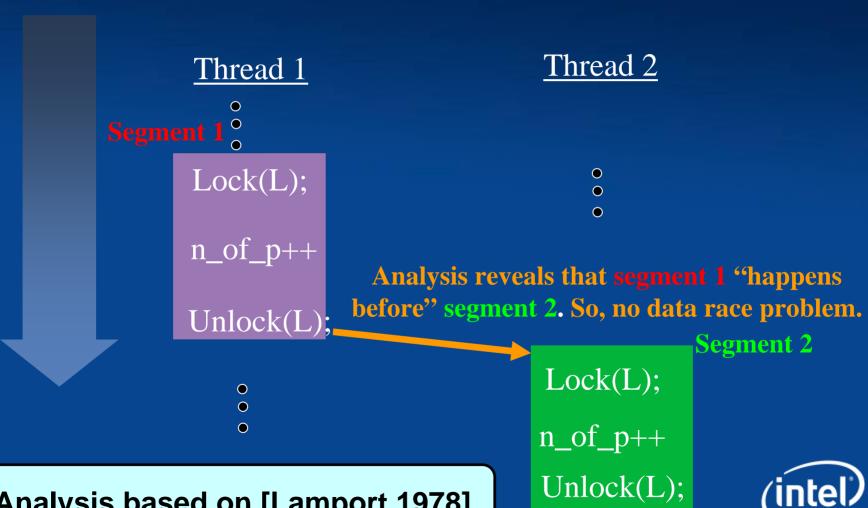
n_of_p++;

Unlock(L);

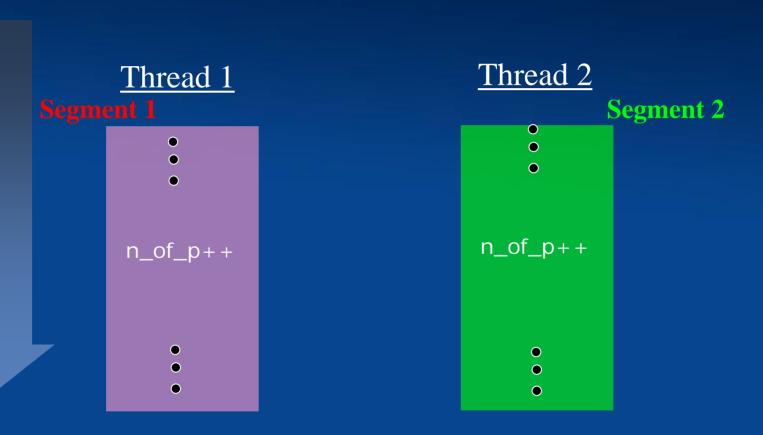


Use binary instrumentation

```
Thread 2
Thread 1
      record lock(L)
Lock(L); record read(n_of_p)
      record write(n_of_p)
n_of_p++
      record unlock(L)
                                    record lock(L)
Unlock(L);
                             Lock(L);
                                    record read(n_of_p)
                            record write(n_of_p)
n_of_p++;
                                     record unlock(L)
                             Unlock(L);
```



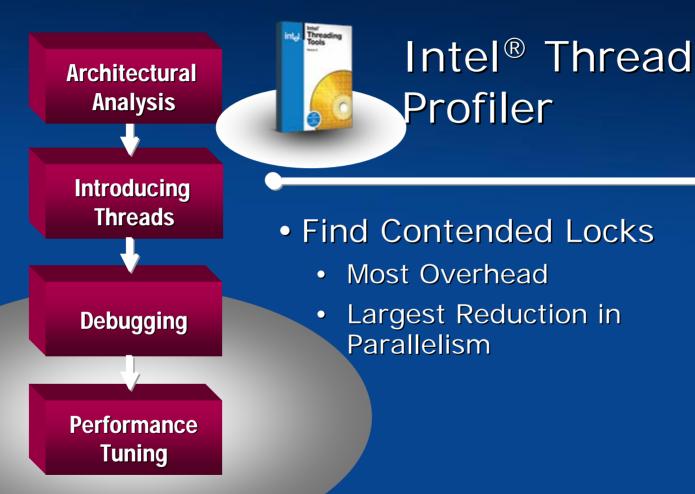
Analysis based on [Lamport 1978]



With no locks, neither segment "happens before" the other. So there is a data race!



Using the Intel Programming Ttools





Profiling the Two Threads in Primes



Work is not balanced between threads



```
for ( long number = 3; number <= N; number += 2 )
{
    long factor = 3;
    while ( number % factor ) factor += 2;
    if ( factor == number )
        primes[ number_of_primes++ ] = number;
}</pre>
```

Thread 1
Iterations 1..50000

Thread 2 Iterations 50000..100000

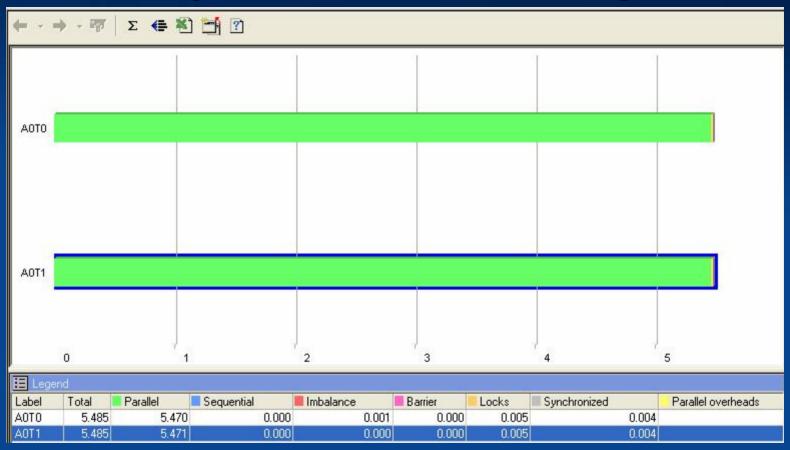
Thread 2 does 3x more work than thread 1



Example: Change to Dynamic Scheduling

```
• #include <stdio.h>
onst long N = 100000;
 long primes[N], number of primes = 0;
main()
   printf( "Determining primes from 1-%d \n", N );
   primes[ number_of_primes++ ] = 2; // special case
 #pragma omp parallel for schedule(dynamic)
   for (long number = 3; number <= N; number += 2)
     long factor = 3;
     while ( number % factor ) factor += 2;
     if (factor == number)
 #pragma omp critical
       primes[ number of primes++ ] = number;
   printf( "Found %d primes\n", number of primes );
```

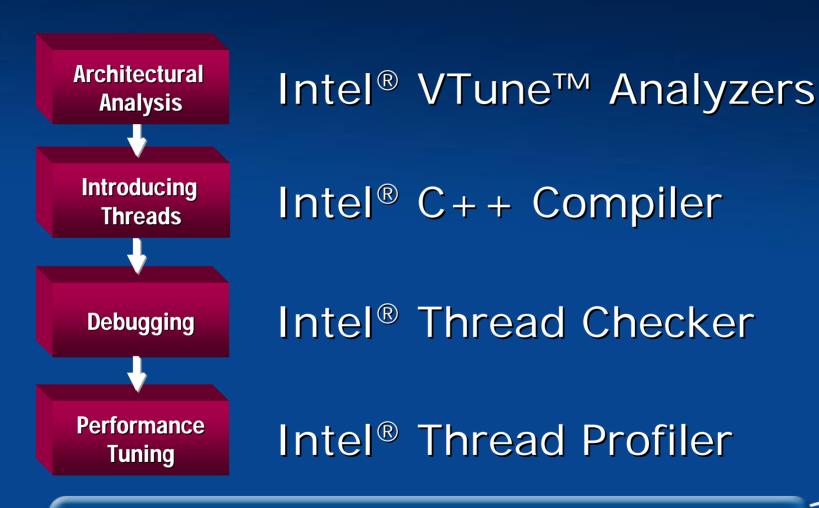
Profiling the Two Threads with Dynamic Scheduling



Work balanced between the threads



Using the Intel® Programming Tools



All these tools are available for free at Intel® website for academic use



Outline

Parallelism in Intel[®] processors

Intel® tools for parallel programming

Some thoughts



The landscape of parallel tools

- Tools need to just work out of the box
 - Provide noise less information
 - No false positives
 - Information must be relevant and accurate
- Must be scalable and work in a variety of environments
 - Language, Architecture and OS independent
 - Must work in the presence of VMMs
- Must work on existing binaries
 - No recompilation, special preparation of binaries for tools
- Performance Degradation must be in acceptable range
- Need more capable tools to handle future multi-core processors
 - Non-Uniform memory access
 - Multiple layers of memory hierarchy visible to programmer
 - Very large number of cores



Other Software development tools in our lab

- Data Mining and machine learning techniques for utilization of PMU counters to improve program performance
- Pin a dynamic instrumentation system
 - Computer Architecture Research
 - For emulating new instructions
 - Modeling micro-architectural features
 - Branch Predictor Models:
 - Cache Models:
 - Simple Timing Models:
 - Performance Analysis Tools
 - Dynamic/Static Instruction Counting Tools
 - CallGraph/CallCount Tools
 - Tool to produce Annotated CFG
 - Program Analysis Tools
 - Code Coverage Tools
 - Dynamic Optimization Tools
 - Memory Checking Tools
 - Race Detection Tools
 - Replay Tools
- Dynamic optimization of programs
- Parallel libraries/domain specific languages



Summary

- Parallel processors
 - Instruction level parallelism
 - SIMD parallelism
 - Multi-core parallelism
- Intel[®] tools for parallel programming
 - Intel® Fortran/C++ compiler
 - Intel[®] VTune[™]
 - Intel® Thread Checker
 - Intel® Thread Profiler
- Parallel applications
 - New emerging class of parallel applications
 - Tune for all levels of parallelism
 - Need more research into parallel tools and methodologies



