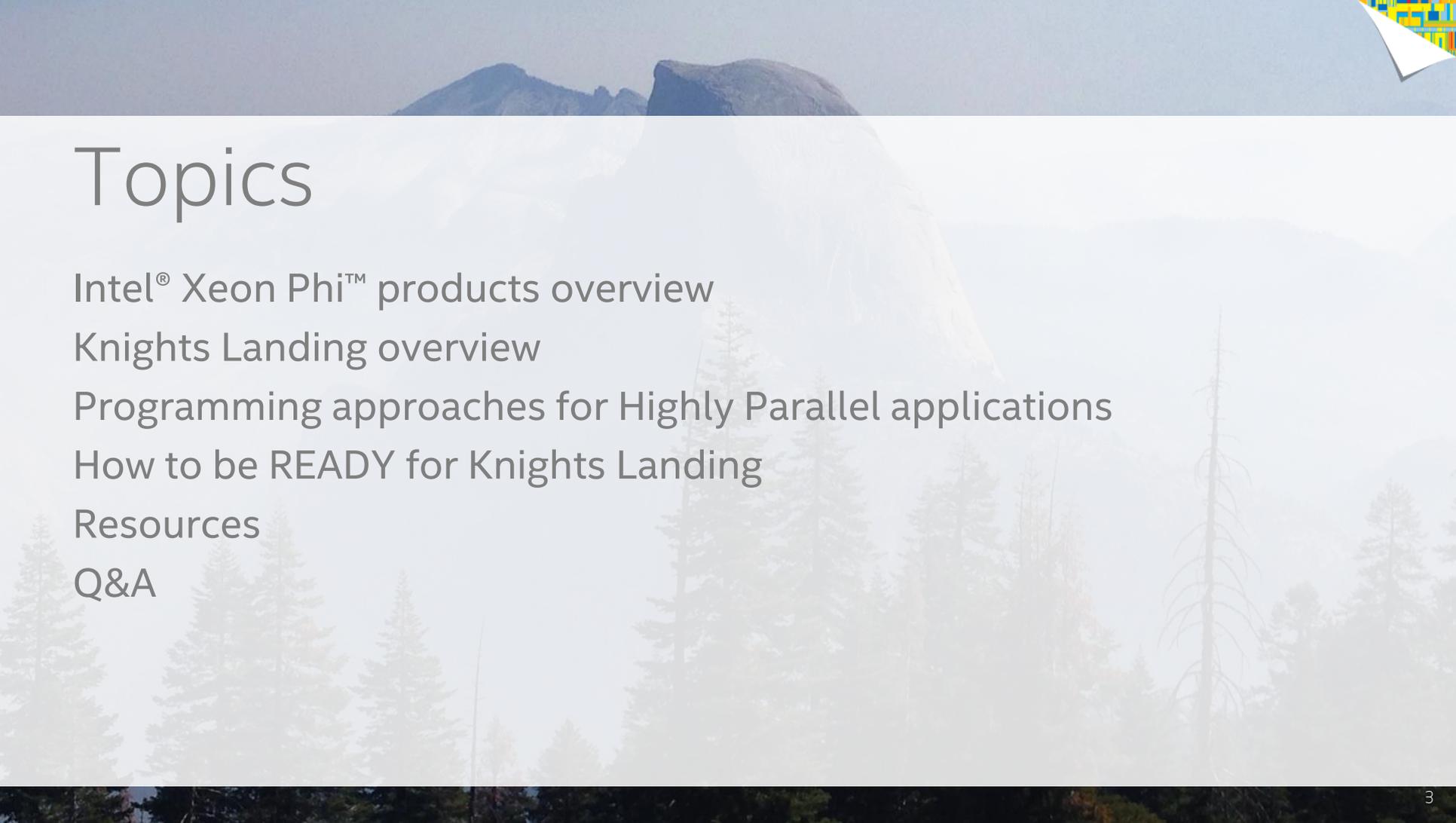


# Knights Corner: Your Path to Knights Landing

James Reinders, Intel  
Wednesday, September 17, 2014; 9-10am PDT



Photo (c) 2014, James Reinders; used with permission; Yosemite Half Dome rising through forest fire smoke 11am on September 10, 2014



# Topics

Intel® Xeon Phi™ products overview

Knights Landing overview

Programming approaches for Highly Parallel applications

How to be READY for Knights Landing

Resources

Q&A

# Topics

Intel® Xeon Phi™ products overview

Knights Landing overview

Programming approaches for Highly Parallel applications

How to be READY for Knights Landing

Resources

Q&A

# Intel® Xeon Phi™ Coprocessors

Up to 61 cores, 1.1 GHz, 244 threads.

Up to 16GB memory.

Up to 352 GB/s bandwidth.

Runs Linux OS.

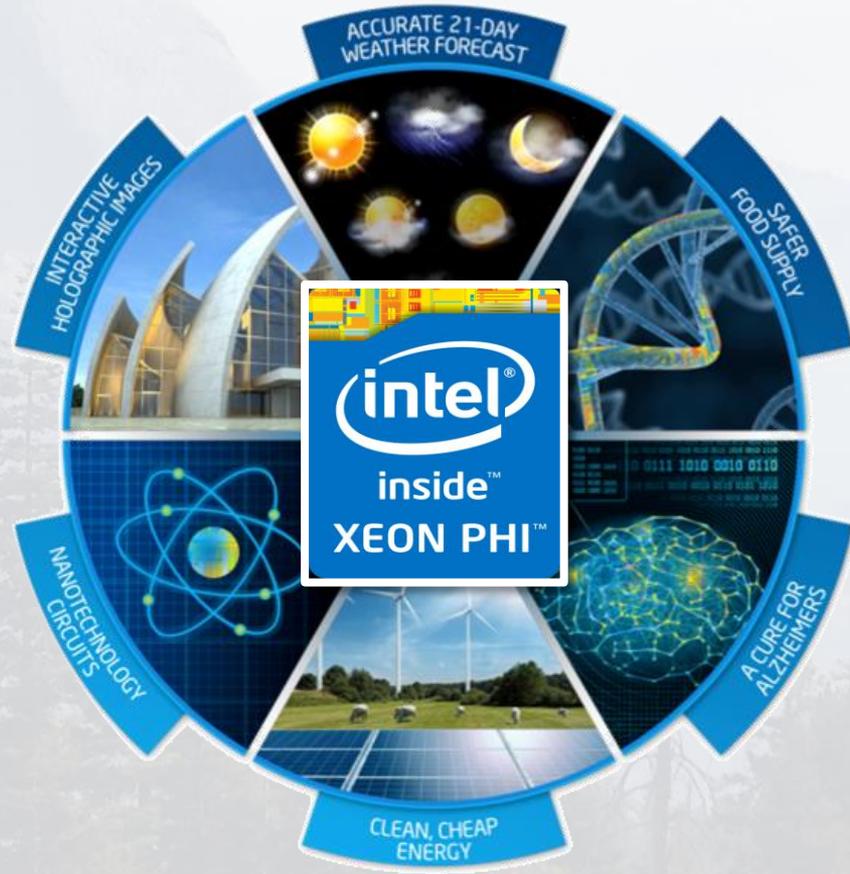
Standard tools, models, languages.

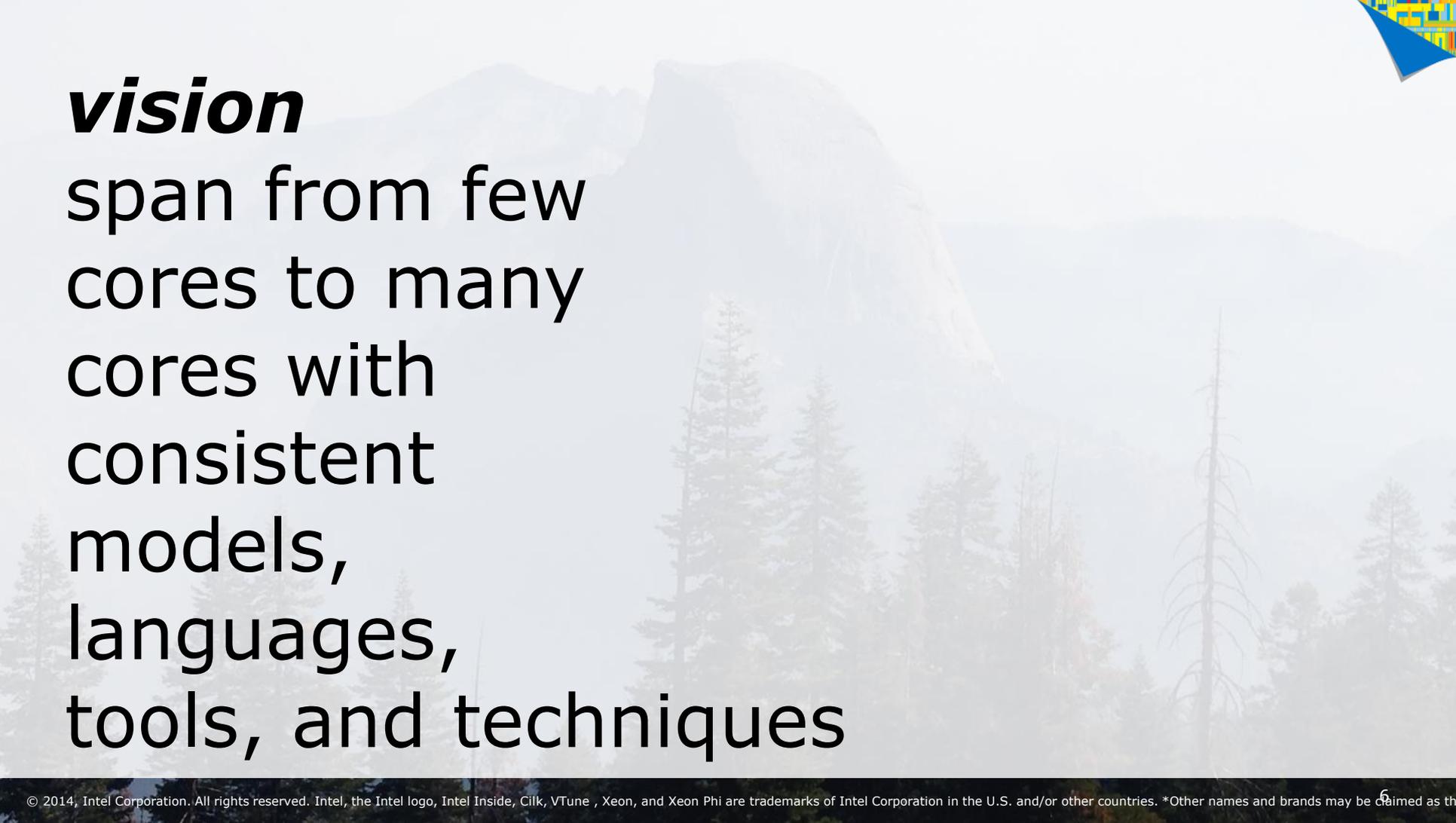
1 TFLOP/s DP FP peak.

Better for parallelism than processor...

Up to 2.2X performance

Up to 4X more power efficient

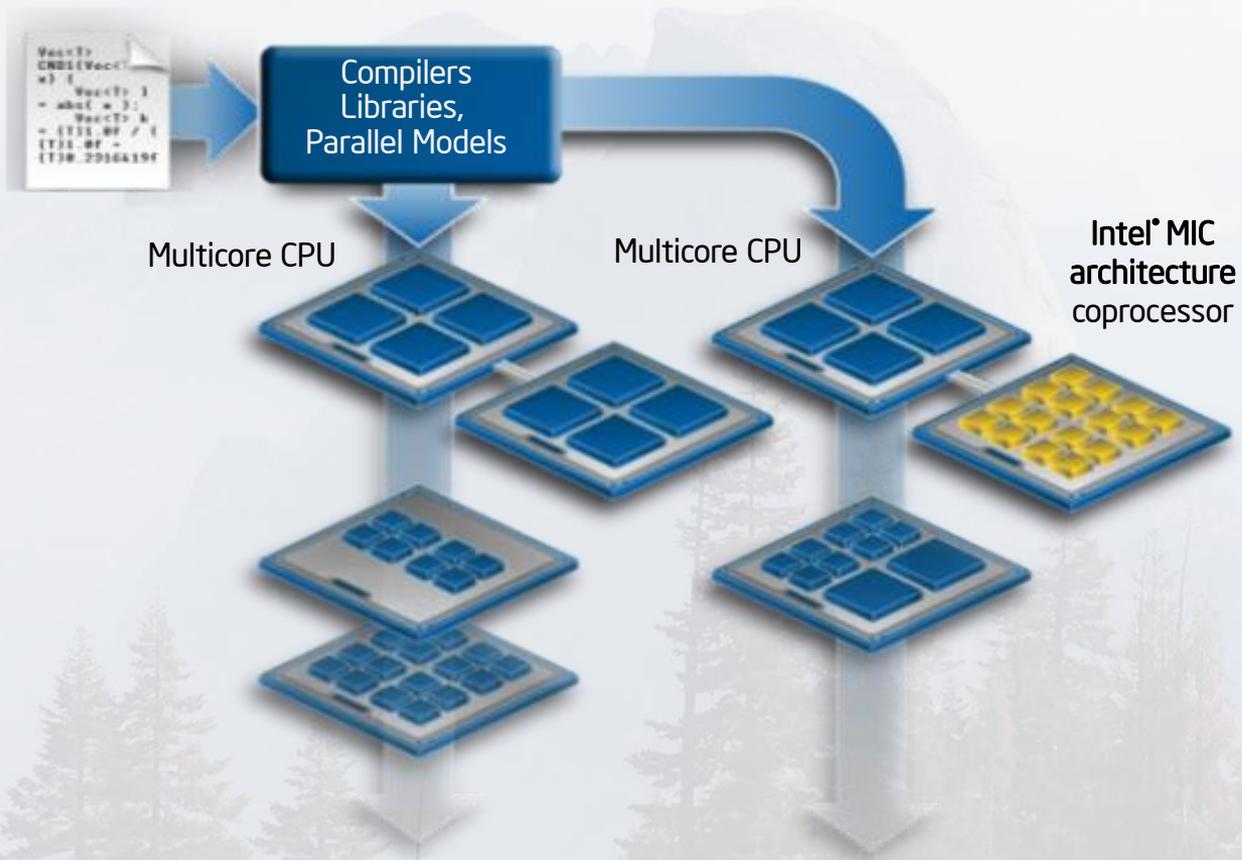




# ***vision***

span from few  
cores to many  
cores with  
consistent  
models,  
languages,  
tools, and techniques

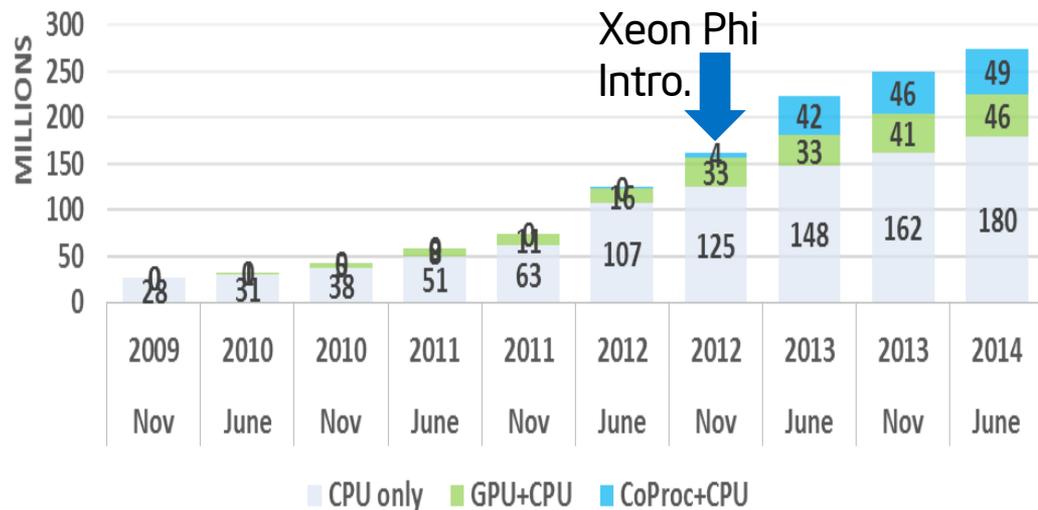
# Source



# Standards

- ✓ **OpenMP**
- ✓ **MPI**
- ✓ **Fortran**
- ✓ **TBB**
- ✓ **C++**

## TOP500 GFLOPS CO-PROCESSOR / ACCELERATORS



Source: June 2014 "Top 500" - [www.top500.org](http://www.top500.org)

Top 500 (June 2014):

Again... the

**#1** system

(third time)

is a

**Neo-heterogeneous**

system

(Common

Programming Model)

(Intel® Xeon® Processors +

Intel® Xeon Phi™ Coprocessor)

# Intel® Xeon Phi™ Coprocessors

Up to 61 cores, 1.1 GHz, 244 threads.

Up to 16GB memory.

Up to 352 GB/s bandwidth.

Runs Linux OS.

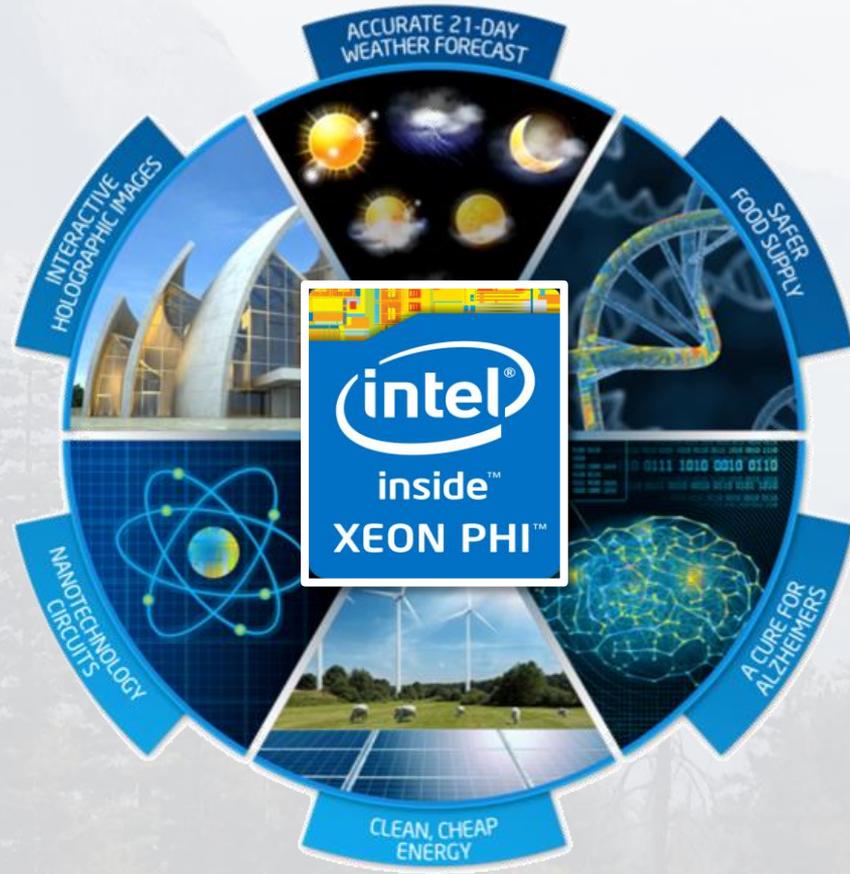
Standard tools, models, languages.

1 TFLOP/s DP FP peak.

Better for parallelism than processor...

Up to 2.2X performance

Up to 4X more power efficient



# Topics

Intel® Xeon Phi™ products overview

**Knights Landing overview**

Programming approaches for Highly Parallel applications

How to be READY for Knights Landing

Resources

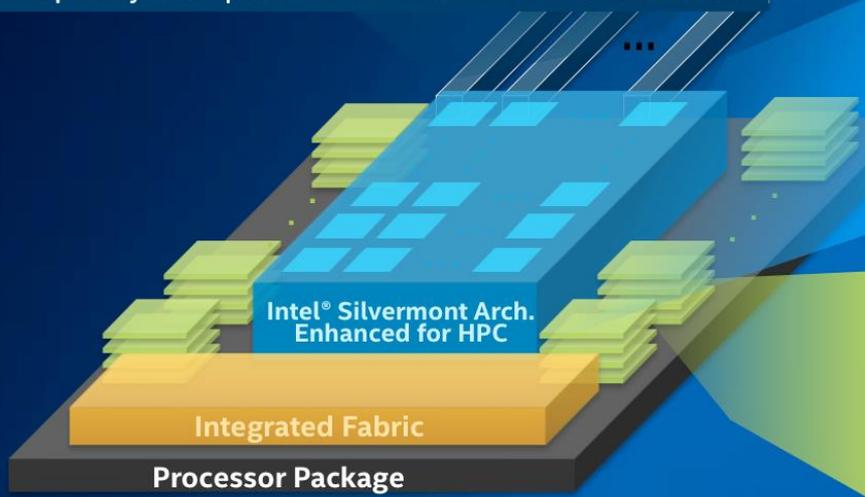
Q&A

# Next Intel Xeon Phi product: Knights Landing

## Knights Landing

(Next Generation Intel® Xeon Phi™ Products)

**Platform Memory:** DDR4 Bandwidth and Capacity Comparable to Intel® Xeon® Processors



Conceptual—Not Actual Package Layout

**3+ TFLOPS**  
In One Package  
Parallel Performance & Density

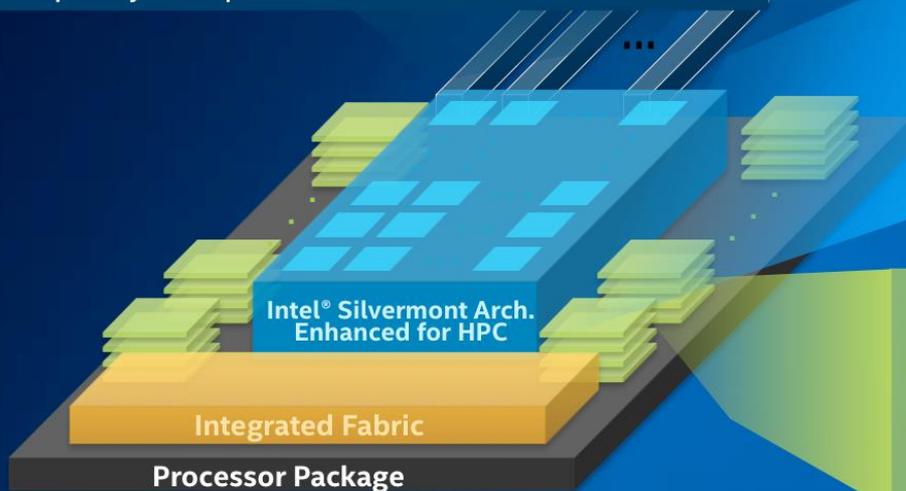
Source: June 2014 Intel @ ISC'14

- Processor  
*(no host required)*
- Out-of-order cores
- High bandwidth memory on-package
- Integrated fabric

# Knights Landing

(Next Generation Intel® Xeon Phi™ Products)

**Platform Memory:** DDR4 Bandwidth and Capacity Comparable to Intel® Xeon® Processors



Conceptual—Not Actual Package Layout

Continued programming model advantage  
Add Intel® AVX-512 instructions  
gcc work well underway

**Compute:** Energy-efficient IA cores

- Microarchitecture enhanced for HPC
- **3X** Single Thread Performance vs Knights Corner
- Intel Xeon Processor Binary Compatible

**On-Package Memory:**

- up to **16GB** at launch
- **1/3X** the Space
- **5X** Bandwidth vs DDR4
- **5X** Power Efficiency

*Jointly Developed with Micron Technology*

★ **3+ TFLOPS**  
In One Package  
Parallel Performance & Density

★ **2<sup>nd</sup> half '15**  
1<sup>st</sup> commercial systems

Source: June 2014 Intel @ ISC'14



## PERFORMANCE

3+ TeraFLOPS of double-precision peak theoretical performance per single socket node<sup>0</sup>

## INTEGRATION

Intel® Omni Scale™ fabric integration

High-performance on-package memory (MCDRAM)

Over 5x STREAM vs. DDR4<sup>1</sup>

Up to 16GB at launch

NUMA support

Over 5x Energy Efficiency vs. GDDR5<sup>2</sup>

Over 3x Density vs. GDDR5<sup>2</sup>

In partnership with Micron Technology

Flexible memory modes including cache and flat

## SERVER PROCESSOR

Standalone bootable processor (running host OS) and a PCIe coprocessor (PCIe end-point device)

Platform memory capacity comparable to Intel® Xeon® Processors

Reliability (“Intel server-class reliability”)

Power Efficiency (Over 25% better than discrete coprocessor)<sup>4</sup>

Density (3+ KNL with fabric in 1U)<sup>5</sup>

## PROGRAMMABILITY

Continues to deliver highly parallel capabilities using standard non-restrictive programming models consistent with CPUs; recompile KNC code for processor, Xeon binaries work.

## MICROARCHITECTURE

Based on Intel's 14 nanometer manufacturing technology

Binary compatible with Intel® Xeon® Processors

Support for Intel® Advanced Vector Extensions 512 (Intel® AVX-512)

3x Single-Thread Performance compared to Knights Corner<sup>6</sup>

Cache-coherency

60+ cores in a 2D Mesh architecture

“Based on Intel® Atom™ core (based on Silvermont microarchitecture) with many HPC enhancements”

4 Threads / Core

Deep Out-of-Order Buffers

Gather/scatter in hardware

Advanced Branch Prediction

High cache bandwidth

## AVAILABILITY

First commercial HPC systems in 2H'15

## ALREADY ANNOUNCED SYSTEMS (FUTURE)

Cori Supercomputer at NERSC (National Energy Research Scientific Computing Center at LBNL/DOE) was the first publically announced Knights Landing based system.

“Trinity” Supercomputer at NNSA (National Nuclear Security Administration) is a \$174 million deal awarded to Cray that will feature Haswell and Knights Landing.

All products, computer systems, dates and figures specified are preliminary based on current expectations, and are subject to change without notice.

All projections are provided for informational purposes only. Any difference in system hardware or software design or configuration may affect actual performance.

<sup>0</sup> Over 3 Teraflops of peak theoretical double-precision performance is preliminary and based on current expectations of cores, clock frequency and floating point operations per second.

<sup>1</sup> Projected result based on internal Intel analysis of STREAM benchmark using a Knights Landing processor with 16GB of ultra high-bandwidth versus DDR4 memory with all channels populated.

<sup>2</sup> Projected result based on internal Intel analysis comparison of 16GB of ultra high-bandwidth memory to 16GB of GDDR5 memory used in the Intel® Xeon Phi™ 7200P.

<sup>3</sup> Compared to 1<sup>st</sup> Generation Intel® Xeon Phi™ 7120P Coprocessor (formerly codenamed Knights Corner)

<sup>4</sup> Projected result based on internal Intel analysis using estimated performance and power consumption of a rack sized deployment of Intel® Xeon® processors and Knights Landing coprocessors as compared to a rack with KNL processors only

<sup>5</sup> Projected result based on internal Intel analysis comparing a discrete Knights Landing processor with integrated fabric to a discrete Intel fabric component card.

<sup>6</sup> Projected peak theoretical single-thread performance relative to

1<sup>st</sup> Generation Intel® Xeon Phi™ Coprocessor 7120P

# Topics

Intel® Xeon Phi™ products overview

Knights Landing overview

**Programming approaches for Highly Parallel applications**

How to be READY for Knights Landing

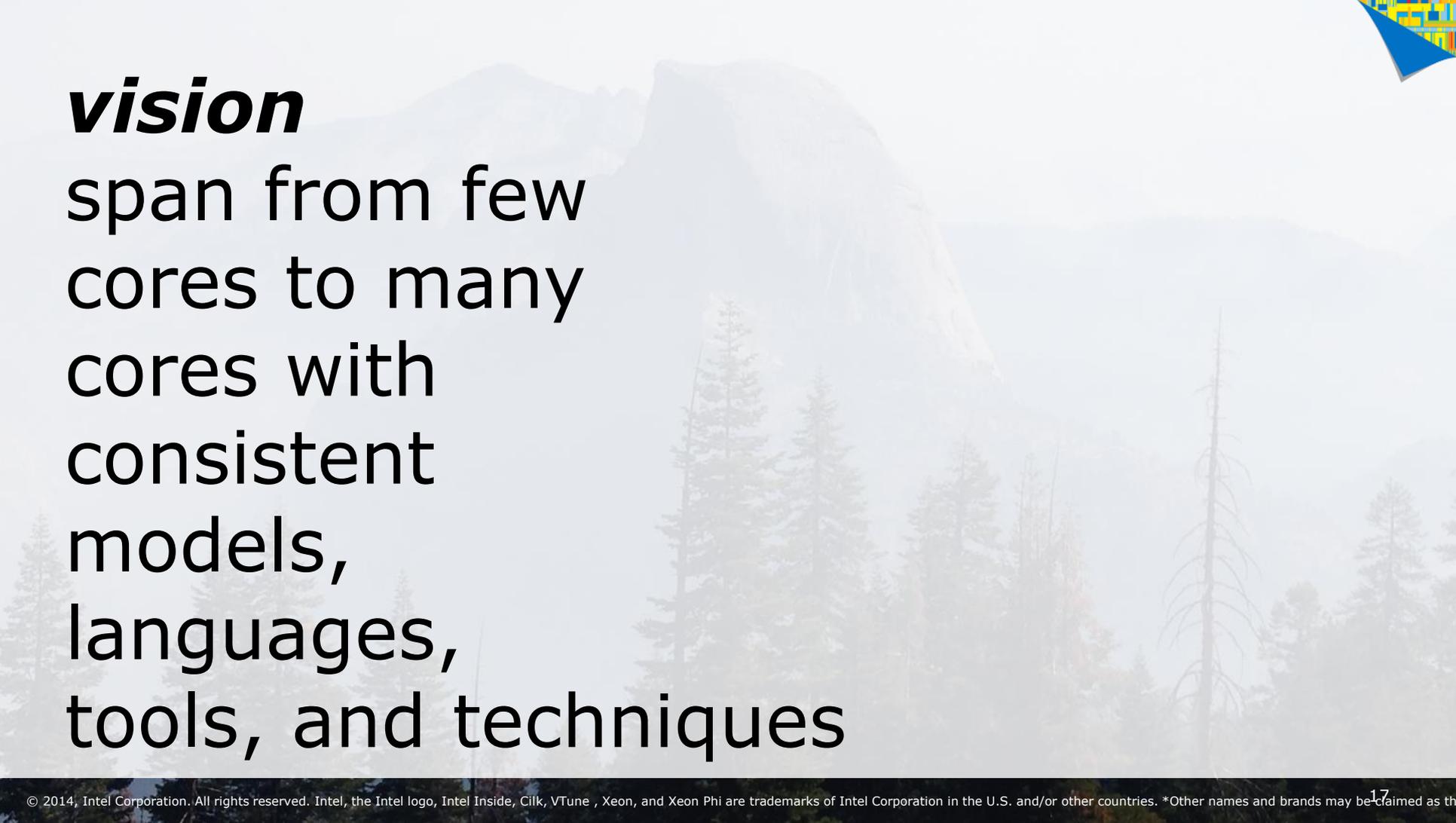
Resources

Q&A



***Intel® Xeon Phi™ products***  
are NOT GPU accelerators.

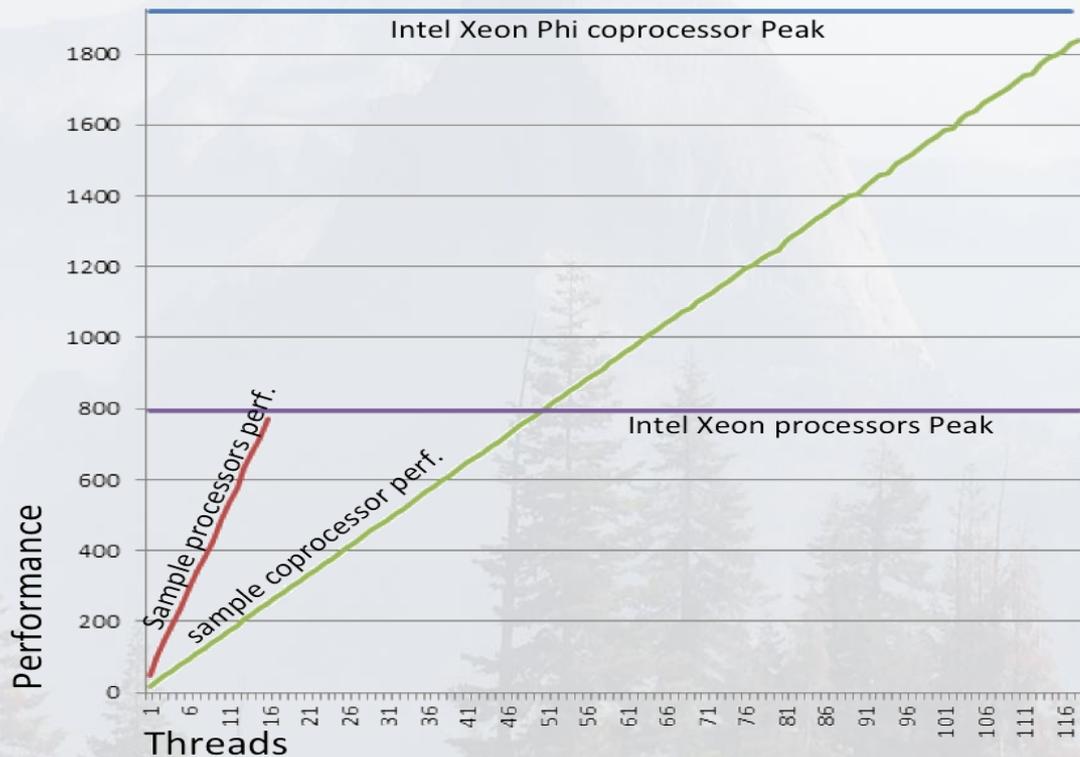
This is VERY GOOD NEWS  
for programmers, users, buyers,  
owners and system administrators.



# ***vision***

span from few  
cores to many  
cores with  
consistent  
models,  
languages,  
tools, and techniques

# Picture worth many words



© 2013, James Reinders & Jim Jeffers, diagram used with permission

# Concurrency required

Intel® Xeon Phi™ based platforms **require** workloads to have great concurrency across multiple dimensions to realize their value proposition.

Many supercomputer workloads today lack this.

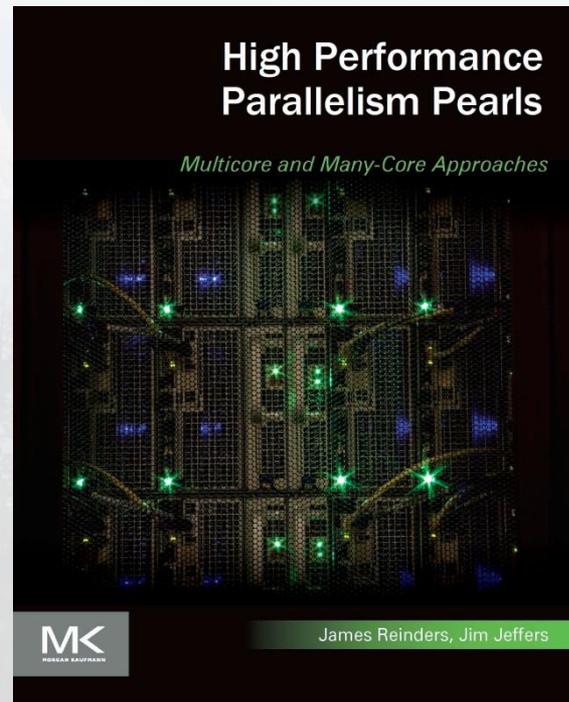
There is a *huge* opportunity to help contemporary workloads reveal their concurrency readiness .

# “Inspired by 61 cores”

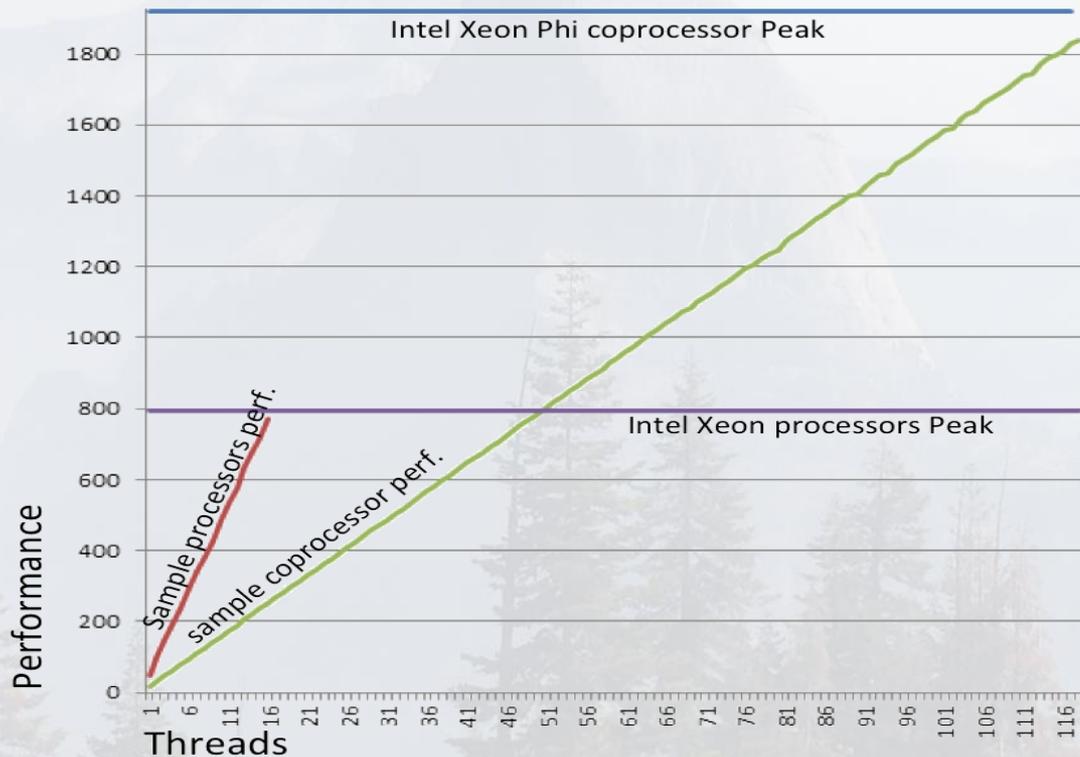
A key realization while preparing our new book.

Over and over and over again...

“Inspired by 61 cores” was the biggest reason for people to work on scaling and vectorization... while benefiting processors and Intel Xeon Phi coprocessors BOTH!



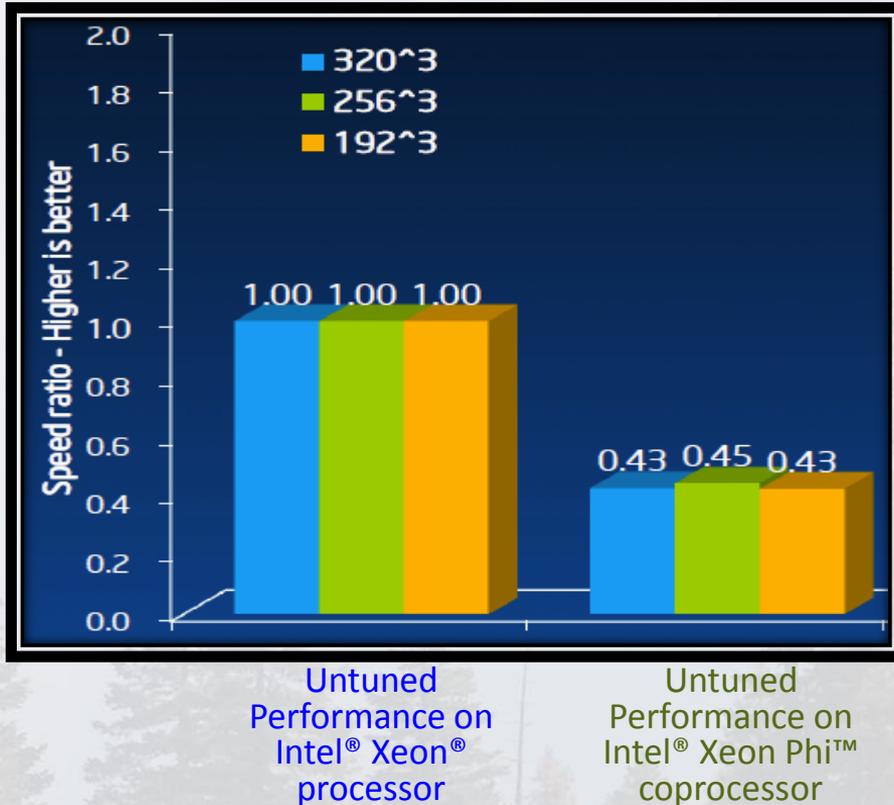
# Picture worth many words



© 2013, James Reinders & Jim Jeffers, diagram used with permission

# Illustrative example

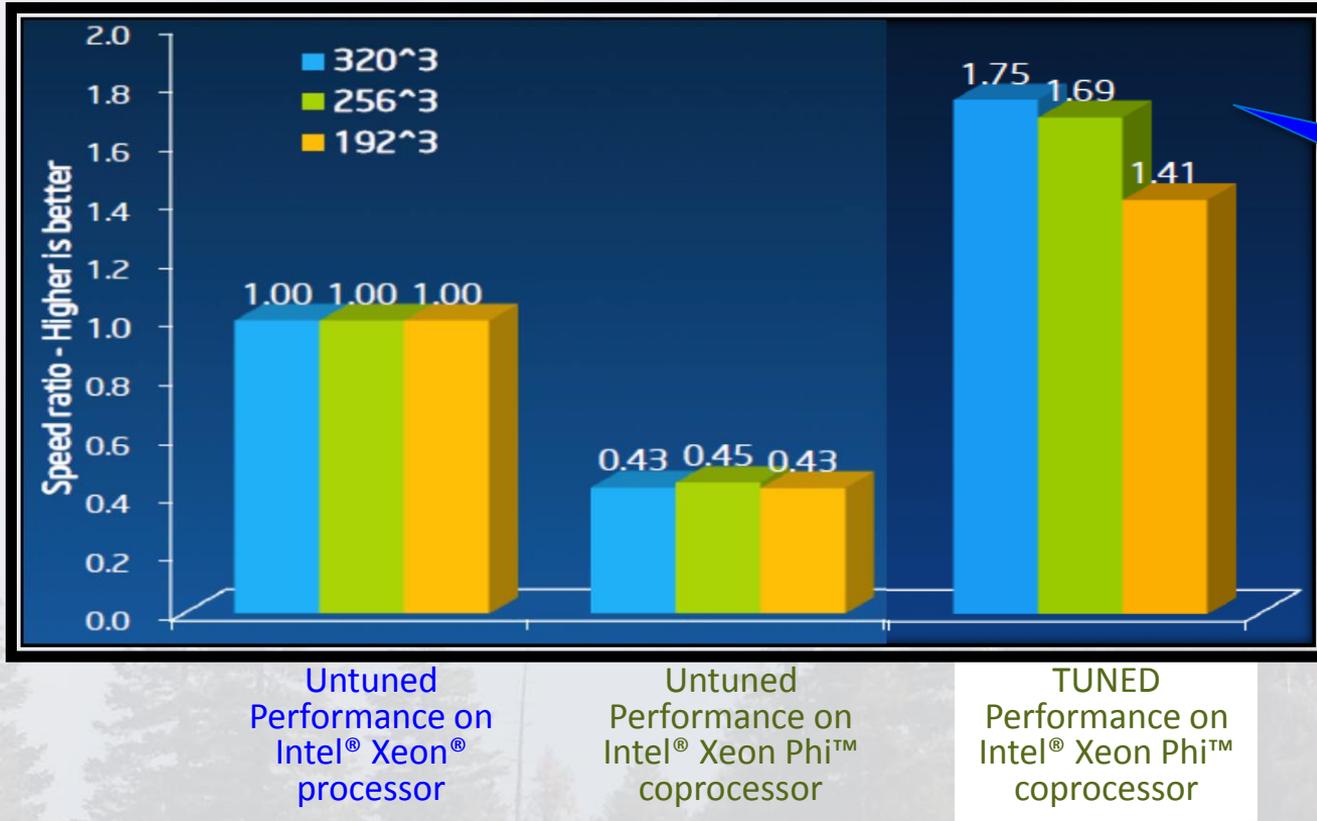
Fortran code using MPI, single threaded originally.  
Run on Intel® Xeon Phi™ coprocessor natively (no offload).



Based on an actual customer example.  
Shown to illustrate a point about common techniques.  
Your results may vary!

# Illustrative example

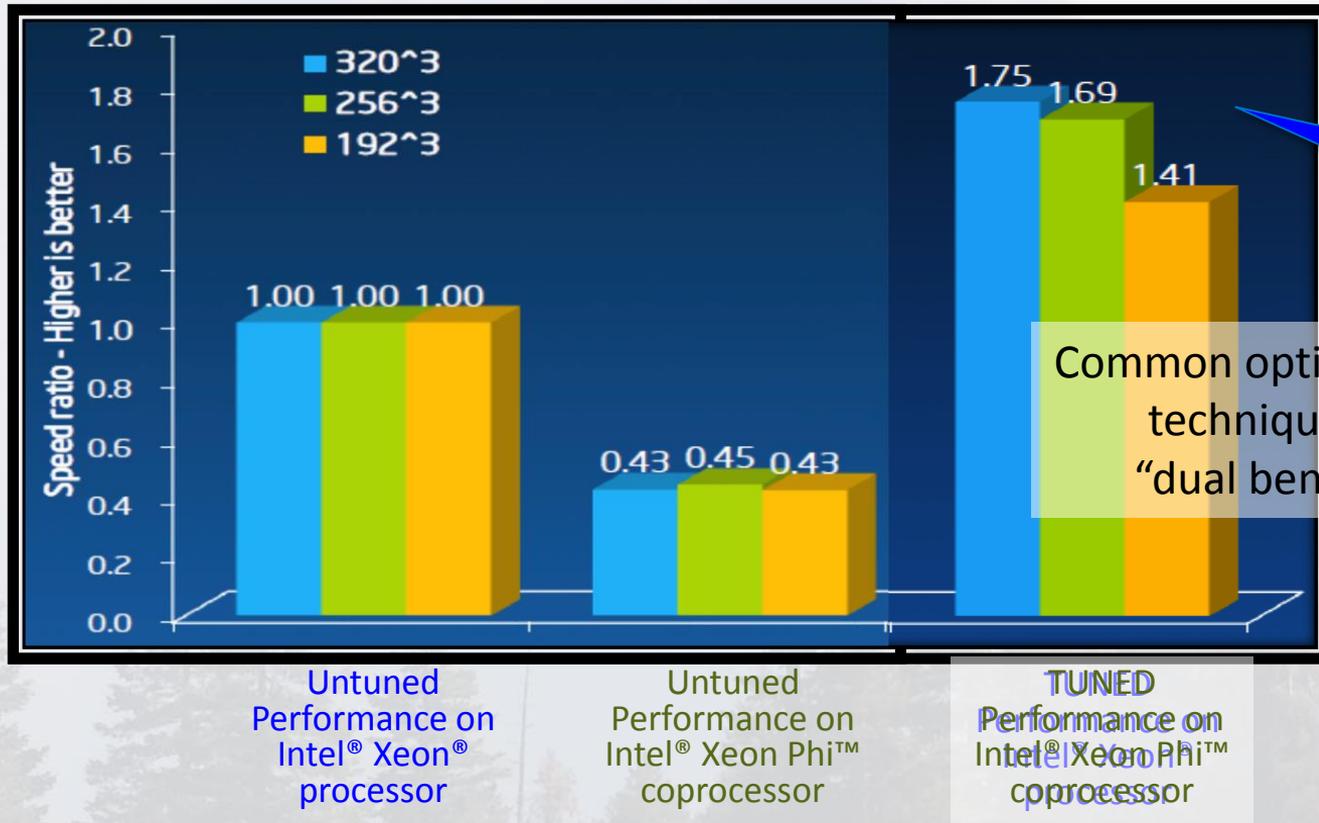
Fortran code using MPI, single threaded originally.  
Run on Intel® Xeon Phi™ coprocessor natively (no offload).



Yeah!

# Illustrative example

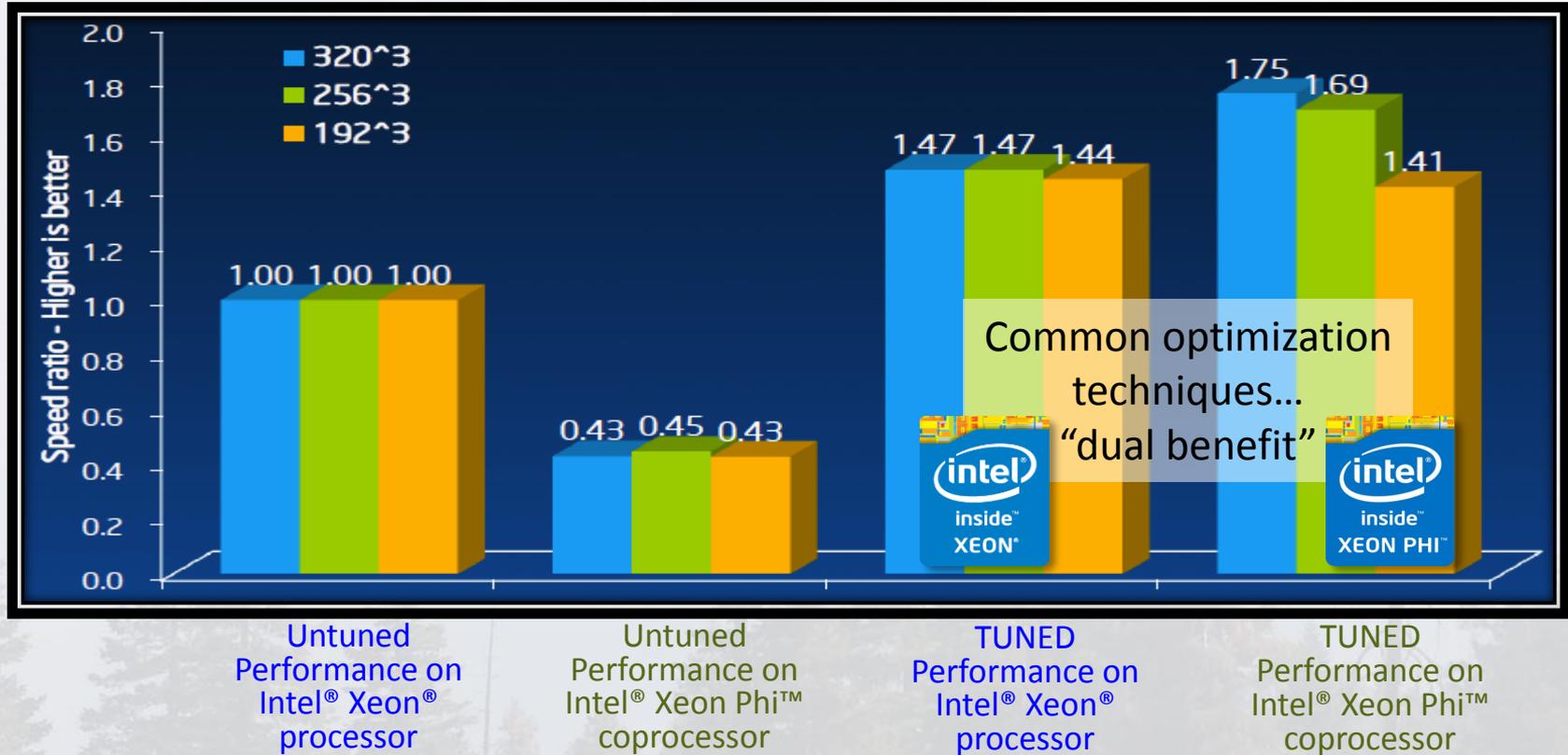
Fortran code using MPI, single threaded originally.  
Run on Intel® Xeon Phi™ coprocessor natively (no offload).



Yeah!

# Illustrative example

Fortran code using MPI, single threaded originally.  
Run on Intel® Xeon Phi™ coprocessor natively (no offload).

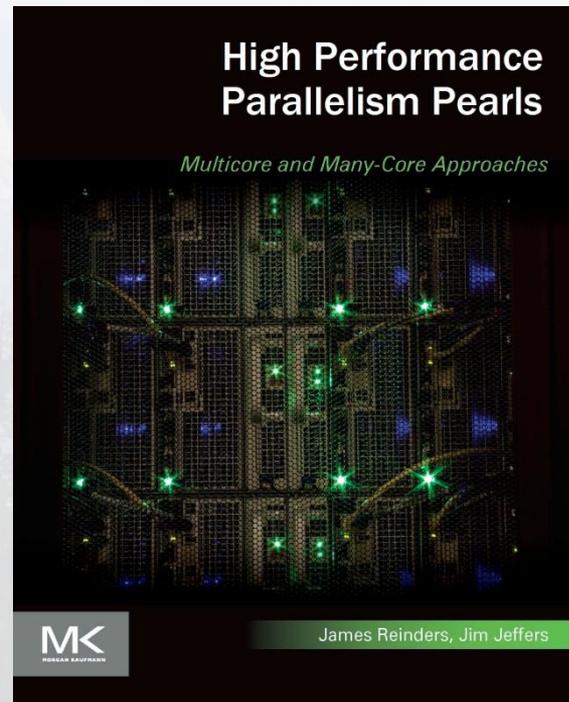


# “Inspired by 61 cores”

A key realization while preparing our new book.

Over and over and over again...

“Inspired by 61 cores” was the biggest reason for people to work on scaling and vectorization... while benefiting processors and Intel Xeon Phi coprocessors BOTH!



# Topics

Intel® Xeon Phi™ products overview

Knights Landing overview

Programming approaches for Highly Parallel applications

**How to be READY for Knights Landing**

Resources

Q&A

# KNL ready?

YES: when Highly parallel (**thread scaling**) and **vectorization profitability**.  
Applications for cluster also need **fabric scaling**.

How can we know? Plot them on a KNC or a high-core count processor based system (Intel® Xeon® processor).

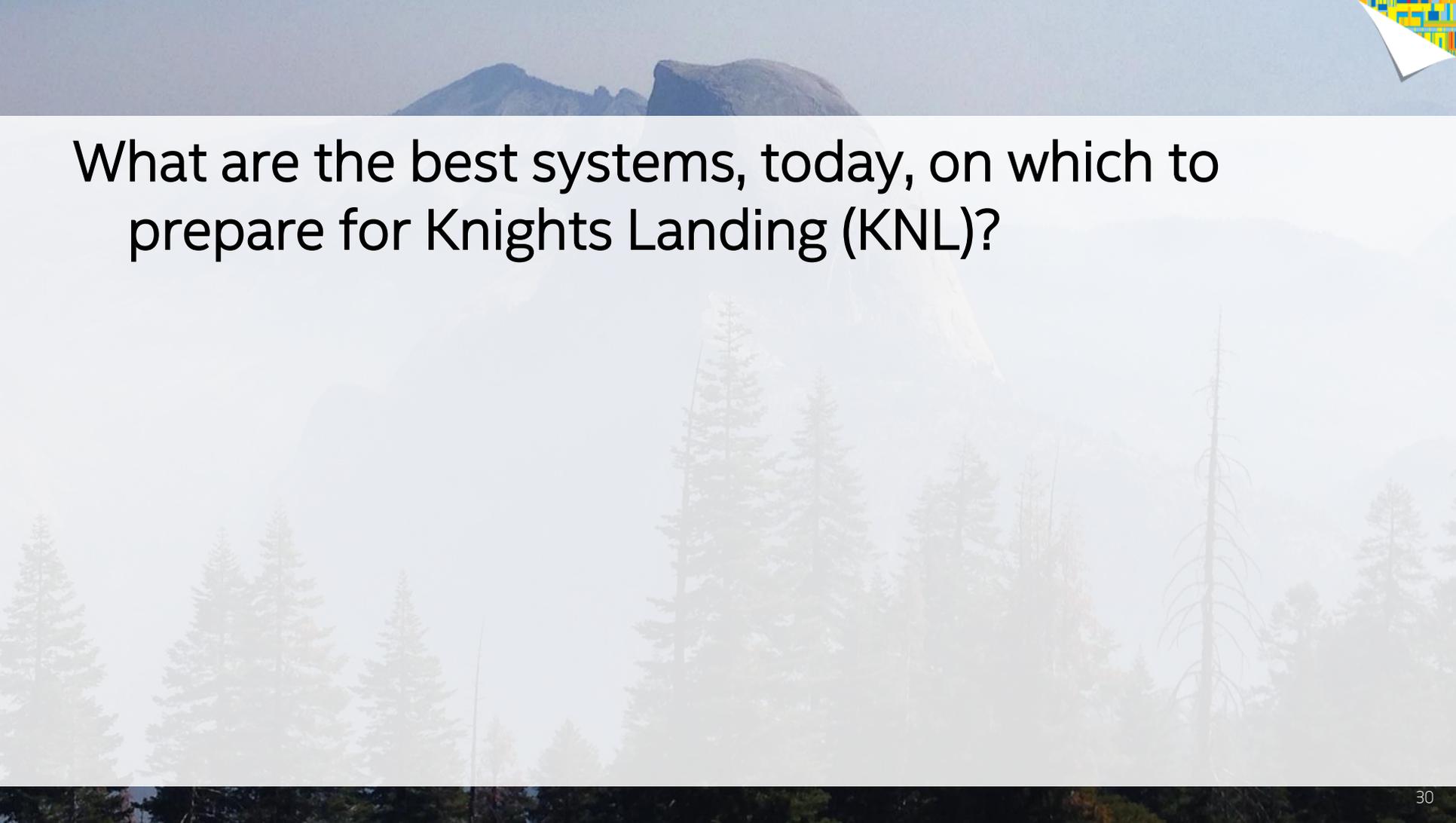
# What to Plot: Concurrency Survey

Establish a performance and configuration discipline and cluster baseline for your workload at scale.

**Fabric:** Plot performance vs #ranks/node.

**Vector:** Plot vectorization profitability (min vs. max vectorization flags).

**Thread Scalability:** Plot OpenMP (or TBB) thread scaling, with a minimal ranks/node.



What are the best systems, today, on which to prepare for Knights Landing (KNL)?

# What are the best systems, today, on which to prepare for Knights Landing (KNL)?

- Knight Corner (KNC), without a doubt, “if” KNC fit your application.
- Key reasons KNL might be excellent, but KNC is less of a fit:
  - Limited memory
  - Offloading across PCI
  - Low serial performance of KNC
  - High I/O or communication requirements
- If KNC is not the right choice for your application, then a high count Xeon (EX) system is needed.

# Examples of KNC specific optimizations

- Limiting code running on KNC because of:
  - more limited memory than host processor
  - very limited serial code performance on KNC
  - “low level vector programming” - Direct coding in KNC vector intrinsics
  - “static parameters” - Exquisite load balancing for offload and symmetric workloads based upon current standard test configurations (lacks scalability across all skus, current and future)

# Examples of KNC specific optimizations

- Limiting code running on KNC because of:
  - more limited memory than host processor
  - very limited serial code performance on KNC
  - “low level vector programming” - Direct coding in KNC vector intrinsics
  - “static parameters” - Exquisite load balancing for offload and symmetric workloads based upon current standard test configurations (lacks scalability across all skus, current and future)

## Knights Landing improvements help!

- Additional memory size enables many more applications and use models
- Much higher serial code perf.
- High source level compatibility but improvements in KNL encourage change. High level options BEST.
- Retune on any system tweak... including KNL
- Integrated fabric offers new scaling optimizations

# What are the best systems, today, on which to prepare for Knights Landing (KNL)?

- Knight Corner (KNC), without a doubt, “if” KNC fit your application.
- Key reasons KNL might be excellent, but KNC is less of a fit:
  - Limited memory
  - Offloading across PCI
  - Low serial performance of KNC
  - High I/O or communication requirements
- If KNC is not the right choice for your application, then a high count Xeon (EX) system is needed.

# What are the best systems, today, on which to prepare for Knights Landing (KNL)?

- Knight Corner (KNC), without a doubt, “if” KNC fit your application.
- Key reasons KNL might be excellent, but KNC is less of a fit:
  - Limited memory
  - Offloading across PCI
  - Low serial performance of KNC
  - High I/O or communication requirements
- If KNC is not the right choice for your application, then a high count Xeon (EX) system is needed.

# What are the best systems, today, on which to prepare for Knights Landing (KNL)?

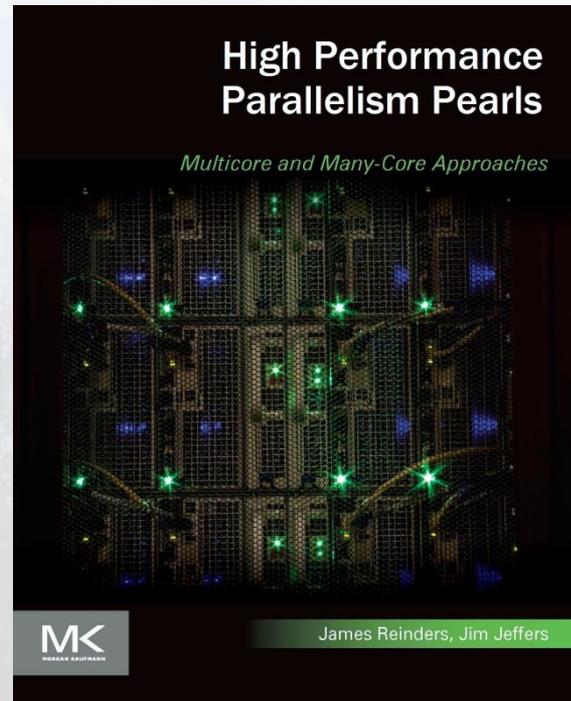
- Knight Corner (KNC), without a doubt, “if” KNC fit your application.
- Key reasons KNL might be excellent, but KNC is less of a fit:
  - Limited memory
  - Offloading across PCI
  - Low serial performance of KNC
  - High I/O or communication requirements
- If KNC is not the right choice for your application, then a high count Xeon (EX) system is needed.

# “Inspired by 61 cores”

A key realization while preparing our new book.

Over and over and over again...

“Inspired by 61 cores” was the biggest reason for people to work on scaling and vectorization... while benefiting processors and Intel Xeon Phi coprocessors BOTH!



# ***vision***

span from few  
cores to many  
cores with  
consistent  
models,  
languages,  
tools, and techniques

I never get tired of saying this.  
You should be glad.

Because... "getting ready" for  
KNL means investing generically  
on the exposure of concurrency.

This has LASTING VALUE.

# *vision*

span from few  
cores to many  
cores with  
consistent  
models,  
languages,  
tools, and techniques

I never get tired of saying this.  
You should be glad.

Because... “getting ready” for  
KNL means investing generically  
on the exposure of concurrency.

This has LASTING VALUE.

# Topics

Intel® Xeon Phi™ products overview

Knights Landing overview

Programming approaches for Highly Parallel applications

How to be READY for Knights Landing

**Resources**

Q&A

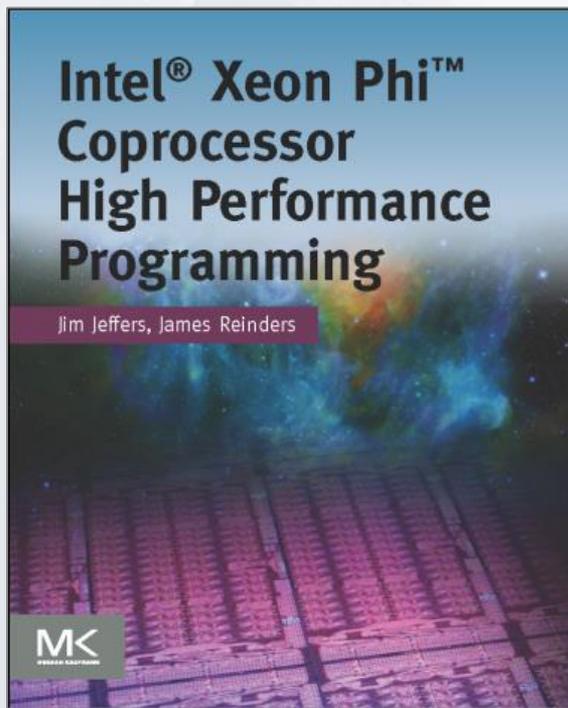
# Intel® Xeon Phi™ Coprocessor High Performance Programming

It all comes down to  
PARALLEL PROGRAMMING!  
(applicable to processors  
and Intel® Xeon Phi™  
coprocessors both)

Forward, Preface

Chapters:

1. Introduction
  2. High Performance Closed Track  
Test Drive!
  3. A Friendly Country Road Race
  4. Driving Around Town:  
Optimizing A Real-World  
Code Example
  5. Lots of Data (Vectors)
  6. Lots of Tasks (not Threads)
  7. Offload
  8. Coprocessor Architecture
  9. Coprocessor System SW
  10. Linux on the Coprocessor
  11. Math Library
  12. MPI
  13. Profiling and Timing
  14. Summary
- Glossary, Index



Intel® Xeon Phi™ Coprocessor High Performance Programming,  
Jim Jeffers, James Reinders, (c) 2013, publisher: Morgan Kaufmann

Book cover designed used with permission of publisher.

[www.lotsofcores.com](http://www.lotsofcores.com)

**This book belongs on the bookshelf of every HPC professional. Not only does it successfully and accessibly teach us how to use and obtain high performance on the Intel MIC architecture, it is about much more than that. It takes us back to the universal fundamentals of high-performance computing including how to think and reason about the performance of algorithms mapped to modern architectures, and it puts into your hands powerful tools that will be useful for years to come.**

**—Robert J. Harrison  
Institute for Advanced  
Computational Science,  
Stony Brook University**

**All figures, diagrams and code  
freely downloadable.**

Over 25 chapters,  
69 expert contributors.

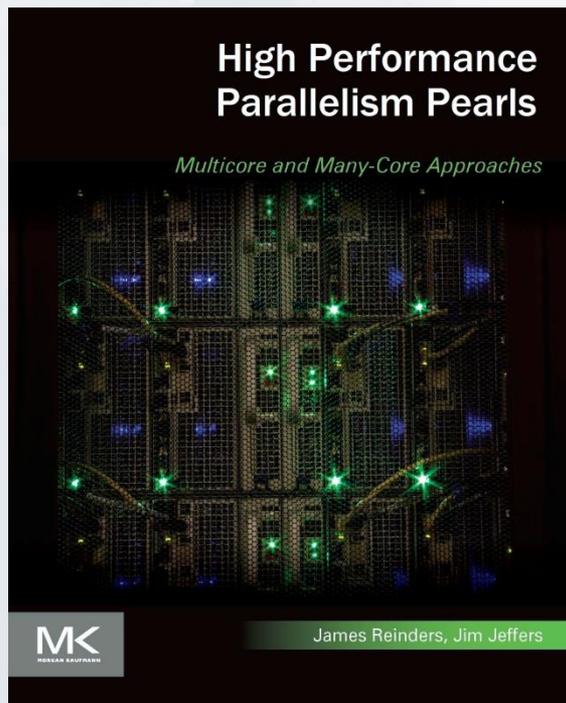
Numerous “Real World” Code  
“Recipes” and examples using  
OpenMP, MPI, OpenCL, C, C++,  
Fortran.

Successful techniques, tips for  
vectorization, scalable parallel coding,  
load balancing, data structure and  
memory tuning, applicable to both  
processors and coprocessors!

Domains include Molecular  
Dynamics, CFD, Financial Services,  
Visualization, System Administration,  
and more...

Learn the methods and reap the  
rewards of modernizing  
code for parallelism!

available  
November 2014



High Performance Programming Pearls

Editors Jim Jeffers, James Reinders, (c) 2014, publisher: Morgan Kaufmann

Book cover designed used with permission of publisher.

[www.lotsofcores.com](http://www.lotsofcores.com)

***This book will make it much easier  
in general to exploit high levels of  
parallelism including programming  
optimally for the Intel Xeon Phi  
products. The common  
programming methodology  
between the Xeon and Xeon Phi  
families is good news for the entire  
scientific and engineering  
community; the same programming  
can realize parallel scaling and  
vectorization for both multicore and  
many-core.***

**—Sverre Jarp  
CERN Honorary Staff Member  
(CTO CERN openlab emeritus)**

All figures, diagrams and code  
freely downloadable. (Nov'14)

# Structured Parallel Programming

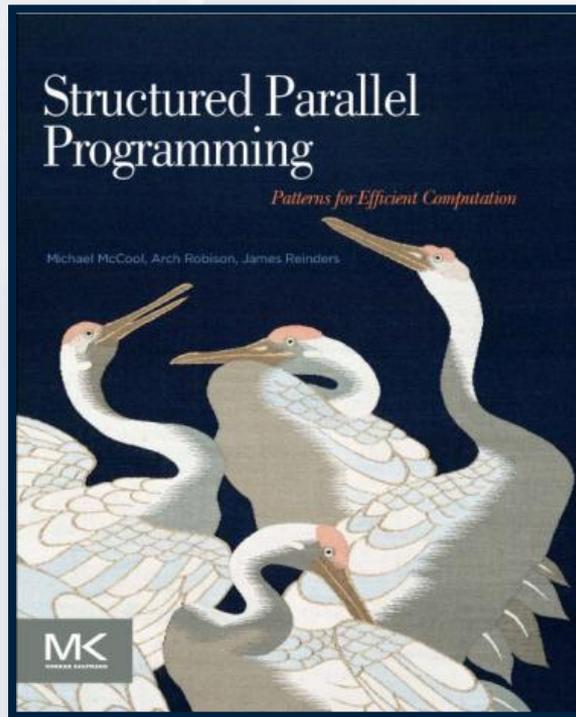
Teaches parallel programming using a new pattern-based approach.

Extensive examples in Cilk Plus and TBB.

Not about any specific hardware, but relevant to all.

It's about effective parallel programming.

Great for teaching!



Structured Parallel Programming, Michael McCool, Arch Robison, James Reinders

(c) 2012, publisher: Morgan Kaufmann

Book cover designed used with permission of publisher.

[www.parallelbook.com](http://www.parallelbook.com)

**This is a really great book...**

**I've been dreaming for a while of a modern accessible book that I could recommend to my threading-deprived colleagues and assorted enquirers to get them up to speed with the core concepts of multithreading as well as something that covers all the major current interesting implementations.**

**Finally I have that book.**

**—Martin Watt,  
Principal Engineer,  
Dreamworks Animation**

**All figures, diagrams and code  
freely downloadable.**

# Online

- <http://software.intel.com/mic-developer>
  - The Training tab has Beginner and Advanced workshop videos, and links to past/future webinars
  - Tools & Downloads tab has useful links
- Book figures, diagrams, code examples:
  - [parallelbook.com](http://parallelbook.com)
  - [lotsofcores.com](http://lotsofcores.com)

# Topics

Intel® Xeon Phi™ products overview

Knights Landing overview

Programming approaches for Highly Parallel applications

How to be READY for Knights Landing

Resources

Q&A



Thank-you  
*Q&A time now*

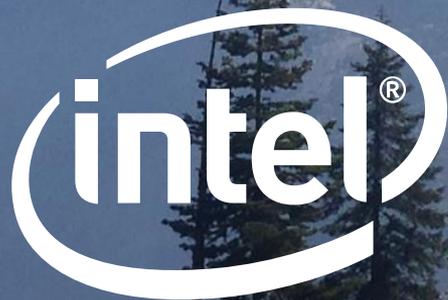


Photo (c) 2014, James Reinders; used with permission; Yosemite Half Dome rising through forest fire smoke 11am on September 10, 2014

# Thank-you



Photo (c) 2014, James Reinders; used with permission; Yosemite Half Dome rising through forest fire smoke 11 am on September 10, 2014

# Legal Disclaimer & Optimization Notice

INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS". NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO THIS INFORMATION INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

Copyright © 2014, Intel Corporation. All rights reserved. Intel, Pentium, Xeon, Xeon Phi, Core, VTune, Cilk, and the Intel logo are trademarks of Intel Corporation in the U.S. and other countries.

## Optimization Notice

Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

Notice revision #20110804