Programming and Optimization with Intel Xeon Phi Coprocessors

Colfax Developer Training
Four-day Workshop

CDT 401

COLFAX
Customized Solutions

Abstract: Colfax Developer Training (CDT) is an in-depth intensive course on efficient parallel programming of Intel Xeon family processors and Intel Xeon Phi coprocessors.

The 4-day workshop features presentations and hands-on exercises on the available programming models and best optimization practices for the Intel many-core platform, and on the usage of the Intel software development and diagnostic tools.

Day One − 9 am to 12 noon: Lecture session.	Day One − 1 pm to 5 pm: Hands-on session.
Introduction, native & offload programming, application porting.	Orientation, Chapter 2 labs: "Hello World" to advanced, "Open MIC".
Day Two − 9 am to 12 noon: Lecture session.	**Day Two − 1 pm to 5 pm: Hands-on session.**
Beginning optimization: automatic vectorization, profiling, scalar and general considerations.	Chapter 3 and 4 labs: VTune, ITAC, automatic vectorization, scalar tuning, "Open MIC".
Day Three − 9 am to 12 noon: Lecture session.	**Day Three − 1 pm to 5 pm: Hands-on session.**
Multi-threading: synchronization, false sharing, load balancing, affinity control, memory traffic.	Chapter 4 labs: exposing parallelism, reduction, synchronization, affinity, loop tiling, "Open MIC".
Day Four − 9 am to 12 noon: Lecture session.	**Day Four − 1 pm to 5 pm: Hands-on session.**
Offload persistence, MPI fabrics. Inter-operation of MPI and OpenMP, using Intel MKL.	Chapter 4: data persistence, fabrics, OpenMP with Intel MPI, modes of MKL utilization, "Open MIC".

Intel Xeon Phi coprocessors, featuring the Intel Many Integrated Core (MIC) architecture, are novel many-core computing accelerators for highly parallel applications, capable of delivering greater performance per system and per watt than general-purpose CPUs. Unlike GPGPUs, they support traditional HPC programming frameworks, including OpenMP and MPI, and require the same optimization methods as multi-core CPUs.

Schedule

Day 1, Lecture Session (9 am – 12 noon)

9:00–10:00   Introduction to the Intel Many Integrated Core (MIC) architecture
  – MIC architecture from programmer's perspective
  – Software tools for Intel Xeon Phi coprocessors
  – Will application X benefit from the MIC architecture?

10:15–11:00   Programming models for Intel Xeon Phi coprocessors
  – Native and offload approaches
  – Using multiple coprocessors
  – MPI applications on clusters with coprocessors
  – Future-proofing: intrinsics vs a high level language

11:15–12:00   Porting applications to the MIC architecture
  – Choosing the programming model
  – Clustering vs offload
  – Compiling libraries
  – Performance expectations

— Lunch break —

Day 1, Practical Session (1 pm – 5 pm)

1:00–1:45   Remote Access Configuration, Lab Orientation
2:00–4:00   Programming Models
  – Native Programming (Lab 2.1: "Hello World" and resolving dependencies)
  – Explicit Offload (Lab 2.2: basic blocking offload; 2.3: data persistence, 2.4: complex example)
  – Virtual-shared memory (Lab 2.5: sharing complex objects)
  – OpenMP 4.0 Target model (new lab: pragma target)
  – Multiple coprocessors (Lab 2.6: concurrent offloads; 2.7: asynchronous offload; 2.8: MPI)

4:00–5:00   "Open MIC" for Application Porting
  – Students can experiment with porting larger applications provided by the instructor or their own codes.

Day 2, Lecture Session (9 am – 12 noon)

9:00–10:00   Beginning optimization for the MIC architecture
- Optimization checklist
- OpenMP support in heterogeneous programming
- Automatic vectorization
- Array notation and elemental functions
- Tuning pragmas

10:15–11:00   Profiling-guided optimization
- Finding bottlenecks using Intel VTune Amplifier
- MPI diagnostics using Intel Trace Analyzer and Collector

11:15–12:00   Optimization Essentials
- Scalar optimization considerations
- Enforcing automatic vectorization
- Data structures for optimal vector math

— Lunch break —

Day 2, Practical Session (1 pm – 5 pm)

1:00–1:30   Optimization Tools
- Node-level profiling with VTune (Lab 4.1)
- Cluster-level profiling with ITAC (Lab 4.2)

1:30–3:00   Vectorization Control
- Assisting automatic vectorization (Lab 3.1: alignment, array notation, programming with intrinsics)
- Controlling automatic vectorization (Lab 4.4: data structures, 4.5: compiler hints)
- SIMD constructs in OpenMP 4.0 (new lab)

3:00–4:00   Scalar Math Tuning
- Compiler settings (Lab 4.3)
- Intel Math Library specifics (Lab 4.3, continued)

4:00–5:00   "Open MIC" for Vectorization
- Students can work on an open-ended example provided by the instructor (N-body calculation or Monte Carlo financial simulation code) or their own codes

Day 3, Lecture Session (9 am − 12 noon)

9:00–10:15   Common thread parallelism pitfalls
- Dealing with insufficient parallelism
- Avoiding synchronization
- Making vectorization co-exist with multi-threading
- Avoiding false sharing
- Load balancing and scheduling overhead

10:30–12:00   Memory traffic optimization
- Thread affinity control
- Memory access pattern in parallel algorithms

Day 3, Practical Session (1 pm − 5 pm)

1:00–2:00   Multi-threading
- Synchronization and reduction (Lab 4.7: OpenMP critical section, atomics, reducers; new lab: user-defined reducers in OpenMP 4.0)
- Load balancing (Lab 4.8: diagnostics and treatment of load imbalance)
- Expanding iteration space (Lab 4.9: strip-mining)

2:00–4:00   Memory optimization
- Affinity control (Lab 4.a: bandwidth-bound and compute-bound workload and thread affinity)
- Memory traffic optimization (Lab 4.b: tiling, Lab 4.c: recursive cache-oblivious algorithm, Lab 4.d: inter-procedural fusion)

4:00–5:00   "Open MIC" for multi-threading and memory optimization
- Students can work on an larger example provided by the instructor (in-place matrix transposition, quicksort or DGEMM in high-level language) or their own codes

Day 4, Lecture Session (9 am – 11 am)

9:00–9:30   Offload optimization
– Data persistence between offloads
– Memory retention between offloads
– Load balancing in heterogeneous workloads

9:30–10:30   MPI application optimization
– MPI and InfiniBand with Intel Xeon Phi coprocessors
– Load balancing in heterogeneous MPI applications
– Inter-operation of OpenMP and MPI

10:30–11:00 Using the Intel Math Kernel Library
– Intel MKL functionality
– MKL usage modes with Intel Xeon Phi coprocessors

Day 4, Practical Session (11 am – 5 pm + lunch)

11:00–12:00 Optimization with explicit offload model
– Data and memory persistence (Lab 4.e: leaving data and allocated buffers on coprocessor)
– Heterogeneous load balancing (new lab: various methods to share a workload between the host and the coprocessor)

— Lunch break —

1:00–2:00   Optimization with MPI
– Controlling MPI fabrics (new lab: virtual Ethernet and InfiniBand with Intel Xeon Phi coprocessors)
– Load balancing (Lab 4.f: diagnostics and relief of load imbalance with static and dynamic scheduling)
– Inter-operation with OpenMP (new lab: shifting parallelism between multiple threads and multiple processes)

2:00–2:30   Using Intel MKL
– DGEMM with MKL (new lab: linking with MKL, automatic offload, compiler-assisted offload, native execution)

2:30–5:00   "Open MIC" for MPI and other topics
– Students work on an larger example provided by instructor (computational fluid dynamics) or their own codes.

Instructor: Vadim Karpusenko, Ph. D., is Principal HPC Research Engineer at Colfax International involved in training and consultancy projects on data mining, software development and statistical analysis of complex systems. His research interests are in the area of physical modeling with HPC clusters, highly parallel architectures, and code optimization. Vadim holds a PhD from North Carolina State University for his computational biophysics research on the free energy and stability of helical secondary structures of proteins. He is a co-author of the book "Parallel Programming and Optimization with Intel Xeon Phi Coprocessors"[1], and a regular contributor to the online resource Colfax Research[2].

Instructor: Andrey Vladimirov, Ph. D., is Head of HPC Research at Colfax International. His primary interest is the application of modern computing technologies to computationally demanding scientific problems. Prior to joining Colfax, A. Vladimirov was involved in computational astrophysics research at Stanford University, North Carolina State University, and the Ioffe Institute (Russia), where he studied cosmic rays, collisionless plasmas and the interstellar medium using computer simulations. He is a co-author of the book "Parallel Programming and Optimization with Intel Xeon Phi Coprocessors", a regular contributor to the online resource Colfax Research, and an author or co-author of over 10 peer-reviewed publications in the fields of theoretical astrophysics and scientific computing.

Instructor: Ryo Asai is a Researcher at Colfax International. Ryo holds a B. A. degree in Physics from University of California, Berkeley. He develops optimization methods for scientific applications targeting emerging parallel computing platforms, computing accelerators and interconnect technologies. Having joined Colfax's research team early on, Ryo has acquired deep domain expertise in programming the Intel MIC architecture. He has committed a great deal of work to the Colfax Developer Training materials, and his peer-reviewed work is among the most widely read publications of Colfax Research.

---

[1]March 2013, ISBN-10: 0-9885234-1-8, ISBN-13: 978-0-9885234-1-8, more details available at http://www.colfax-intl.com/nd/xeonphi/book.aspx

[2]http://research.colfaxinternational.com/

Notes

Presentations

Video and audio recording and still photography during Colfax Developer Training (CDT) is permitted only for private or institutional use by the attendees and their direct collaborators. No recorded materials shall be publicly disseminated without explicit written authorization from Colfax International.

Materials

The slides of all presentations will be made available to all attendees in electronic form. Attendees are free to use these materials privately and share them with direct collaborators. However, no materials shall be publicly disseminated without explicit written authorization from Colfax International.
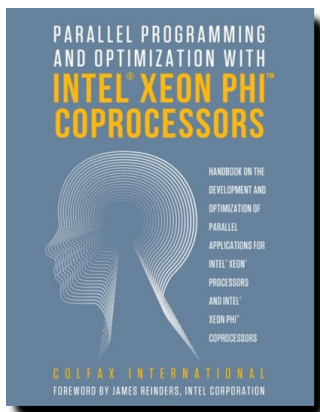
The book on which the CDT is based, "Parallel Programming and Optimization with Intel Xeon Phi Coprocessors", is available in the electronic format and as a hard copy at http://www.colfax-intl.com/nd/xeonphi/book.aspx. An electronic copy of the book and enclosed codes of exercises is included in the training price.

Contacts and Resources

The instructors of this CDT can be contacted via email at vadim@colfax-intl.com, andrey@colfax-intl.com and ryo@colfax-intl.com.

You may also find useful our online resource research.colfaxinternational.com, where explanatory and research publications can be found.

General inquiries regarding Colfax's business can be sent to phi@colfax-intl.com. Colfax's business Web site www.colfax-intl.com contains information about the company's hardware solutions, education and consulting offerings.