INTEL® HPC CODE MODERNIZATION WORKSHOP
FOR LIFE SCIENCES – CAMBRIDGE 2016

2-DAY WORKSHOP IN CAMBRIDGE, ENGLAND – JULY 6TH & 7TH 2016

Wellcome Genome Campus Conference Centre
Hinxton, Cambridge, England
CB10 1RQ

Free to attend! Places are limited. Register early to reserve a seat: http://www.inteldevconference.com/

DAY 1		AGENDA — WEDNESDAY, JULY 6TH
08:30	10:00	Registration and Breakfast
10:00	10:45	**Introduction: Intel Technology Platform & Processor Update** *Ralph de Wargny, Intel Software & Janko Strassburg, Bayncore* Learn about the parallel architecture, technical advances and features of the latest and future Intel processors, especially Intel® Xeon™ and Intel® Xeon Phi™.
10:45	11:30	**Trends in Efficient Parallel Computing and Performance** *Martin Hilgeman, Technical lead HPC benchmarking group, Dell* With all the advances in massively parallel and multi-core computing with CPUs and accelerators it is often overlooked whether the computational work is being done in an efficient manner. This presentation shows the well-known laws of parallel performance from the perspective of a system builder. It also covers through the use of real case studies, examples of how to program for energy efficient parallel application performance.
11:30	11:45	Short Coffee Break
11:45	12:30	**Developing High Performance Parallel Code for Intel® Xeon™ and Intel® Xeon Phi™** *Janko Strassburg, Bayncore*
12:30	13:15	**Guest Session: Using Task-Based Parallelism to Accelerate Rational Drug Design** *Dr. Christopher Woods, EPSRC Research Software Engineering (RSE) Fellow, Univ. Bristol* Adapting software for the multicore present, and massively multicore future, forces researchers to look beyond traditional parallelisation strategies, such as message passing or parallelising loops. Task based parallelism provides an intuitively clear abstraction. This talk will show how task based programming using threading building blocks has been used to parallelise the program "ligandswap". Ligandswap is used for predicting relative binding affinities of molecules to proteins. The program is divided into three levels of nested tasks, with threading building blocks handling the dynamic allocation of compute resources to those tasks throughout the calculation.
13:15	14:15	Lunch
14:15	15:00	**Vectorization – Parallelism at Core Level (SIMD)** *Janko Strassburg, Bayncore* Vectorization is one of the critical elements to maximize parallel performance, especially when it comes to the Intel Advanced Vector Extensions 512 (AVX512). This session will show how to get started with vectorization and avoid common pitfalls.
15:00	15:45	**Guest Session: Open Computational Biology Platform (OCBP) on Intel® Xeon™ and AVX** *Ignacio Medina Castello, Head of Computational Biology Lab, High Performance Computing Service, University of Cambridge*
15:45	16:15	Coffee Break & Networking
16:15	17:00	**Advanced Vectorization** *Janko Strassburg, Bayncore* Vectorization is one of the critical elements to maximize parallel performance. In this session we will review common cases where automatic vectorization fails, providing tips and best known methods for effective vectorization. This session is also illustrated with a couple of real-world case examples by using Intel® Parallel Studio XE suite.
17:00	17:45	**Guest Session: Developing High Performance Software Solutions for Life Sciences** *Enric Gibert, CEO Pharmacelera* Pharmacelera provides hardware and software solutions for computer-aided drug design, with products like PharmScreen and PharmQSAR. Enric will talk about his key learnings from developing high performance drug discovery software solutions (Computer-Aided Drug Design) and detail how Pharmacelera reduced the execution time of PELE, a state-of-the-art molecular docking software from the Barcelona Supercomputing Center.
17:45	18:00	**Q&A**
18:00	20:00	**Networking with drinks & finger food**

In partnership with DELL

DAY 2		AGENDA	THURSDAY, JULY 7TH
08:30	10:00	Breakfast	
10:00	10:45	**Pushing Machine Learning to a New Level with Intel® Xeon™ and Intel® Xeon Phi™** *Ralph de Wargny, Intel Software*	
10:45	11:30	**Deep Learning on Intel® Xeon Phi™ (KNL – Knights Landing) and OPA** *Gaurav Kaul, Intel*	
11:30	11:45	Short Coffee Break	
11:45	12:30	**Guest Session: A Bioinformatic Facility for Research** *Paolo Aretini, PhD, Senior Bioinformatics Analyst, Fond. Pisana per la Scienza ONLUS* In this presentation we describe the activity of the Bioinformatic Lab of Pisa Science Foundation, a new and small research institute. The intensive calculation power needed to perform genomic mapping, identification of SNPs, InDel, structural variations and differential expression analysis from RNAseq experiments, was provided by the collaboration of Information Technology Center of University of Pisa. The main application fields are molecular and genetic oncology (breast and brain cancer) and medical genetics. We discuss here the issues raised from NGS data analysis with the special regard to variant call algorithm and differential expression analysis. Paolo Aretini joined Pisa Science Foundation in 2013. In collaboration with other lab members, he conducts bioinformatics research, focusing on the development of computational and integrative methods, databases and tools to carry out the analysis of big amount of "omics" data: whole genome and whole exome analysis and differential expression analysis from RNA-seq experiments. His main research interest is the bioinformatics and biostatistics applied to medical genetics.	
12:30	13:30	Lunch	
13:30	14:15	**Tutorial - Real World Examples for Vectorization** *Francois Fayard, Bayncore*	
14:15	15:00	**Using Intel Lustre for Large-Scale High Bandwidth Storage in Life Sciences Applications** *Gabriele Paciucci, Intel* I/O needs in Life Science are challenging developers and architects. Genomics workloads needs superfast "IOPS oriented" file system when they are collecting data from last generation of sequencers, but also need high throughput super cheaper and very large parallel file system for back processing and long term archival. In this presentation, developers and architects can understand how the Scalable System Framework holistic design can help to design solutions that meet so different requirements. Gabriele Paciucci is a Solution Architect in the Enterprise & HPC Platform Group at Intel. In this role, Gabriele provides technical consultations to partners and customers and evangelizes the Lustre technology worldwide. Gabriele is author of several researches presented at major Lustre and HPC events.	
15:00	15:30	Coffee Break	
15:30	16:15	**Guest Session: Profiling genome pipeline I/O patterns** *Dr. Rosemary Francis, PhD, Founder and CEO, Ellexus* Genome pipelines are typically IO bound: some optimization can be made by exploiting parallelization of the sequence analysis algorithms, but unless the storage can deliver the increased work loads you are only going to waste compute cycles and network bandwidth. Ellexus make tools to profile file IO and to identify IO hotspots. Ellexus analysed the IO patterns of well known genomics applications at the Sanger Institute and Cancer Research UK Cambridge. Rosemary will present the results of the IO analysis and she will look at the impact those patterns have on the Lustre storage and the efficiency of those applications on an HPC cluster. She will highlight how to find and prevent bad IO patterns using the IO profiling tools from Ellexus. Developers can adopt these techniques in order to protect the storage and improve the quality of service.	
16:15	17:00	**Case Study: The n-body problem on Intel® Xeon Phi™ (KNL – Knights Landing)** *Francois Fayard, Bayncore*	
17:00	17:30	**Q&A**	

Consulting Partner: Bayncore
www.bayncore.com

BAYNCORE

In partnership with DELL

DELL