Part II Courses

Module 2.1:	Data Structures and Sorting & Searching
Lecturer	Axel Kohlmeyer (Temple U. & ICTP)
Mandatory	
Module Description	Introduction to fundamental data structures and their impact on performance and memory consumption. Study parallelization issues.
Main Topics	• Implementation of data structures like arrays, linked lists, hash tables, trees, maps in different programming languages • Comparison of performance for data access, data insert, data removal, data append, scaling with problem size, and need for auxiliary storage • Accessing and updating data structures in parallel (using locks and lock-free) • Data structures for numerical problems (example: sparse matrix storage) • Some Considerations on Data Structure Design • Implementation of selected popular sort algorithms • Comparison of performance for unsorted and presorted data • Searching in unsorted or presorted data
Objectives	On successful completion of this module students should be able to determine which data structure is best suited for a given task.

Module 2.2:	**High Performance Computing Technology**
Coordinator	IBM Italia
Mandatory	
Module Description	This module introduces state-of-the-art technologies and innovation in High Performance Computing. Main components of computing infrastructure are analyzed and discussed. Students will install and configure a HPC Linux Cluster and will also be exposed to the use of Cloud and Grid Infrastructures.
Main Topics	• HPC system deployment • Software Provisioning (modules) • Managing hardware diversity • Scheduling and resource management • Usage accounting • Data management (quotas, purging, archival) • Sustainable HPC computing infrastructure • Green computing • Grid and Cloud Computing
Objectives	On successful completion of this module students should be able to understand common problems related to the installation and maintenance of a sustainable HPC infrastructure.

Module 2.3:	**Scientific data management**
Coordinator	Stefano Cozzini (CNR) Rossella Aversa (CNR)
Mandatory	
Module Description	The module introduces modern techniques to deal with the large amount of data in scientific and technical computing.
Main Topics	• Introduction to Big data issues • Parallel file systems and parallel I/O • Scientific data formats and libraries (NetCDF, HDF5) • MPI-IO • Data intensive computing (distributed file systems and MapReduce), Hadoop • Web interface and protocols for data exchange (i.e. opendap) • Workflows for data processing • Benchmarking and profiling data intensive calculation
Objectives	On successful completion of this module students have an overview of the main techniques and tools to tackle data-intensive computational problems.

Module 2.4:	**Electronic structure: from blackboard to source code**
Lecturer	Stefano de Gironcoli (SISSA)
CFU	2
Module Description	Material science and condensed matter theorists extensively employ in their research ab initio atomistic simulations as implemented in a number of widely available software codes. Most often these tools are used as 'black boxes' with little or only partial knowledge of the practical implementation of the general theoretical ideas they are based on. This is particularly severe drawback when it hinders the development of new analysis tools or computational experiments due to the lack of insight on the internal structure of the employed research software.
Main Topics	● The basic approximations: Born Oppenheimer adiabatic approximation and the effective potential concept. Density Functional Theory: Density as the basic quantity, Hohenberg-Kohn theorem and variational principle, Kohn-Sham auxiliary equations. ● General description of a plane-wave pseudopotential code. Self-consistent cycle vs global minimization and Car-Parrinello method. Block diagram of a SCF-type and a CP-type code. The basic modules: diagonalization/minimization (needs H*psi), building the density (needs BZ sampling), building the potential (needs Poisson's solver and exchange-correlation functionals). Initialization (the external potential) and termination (forces/stress and ionic evolution). The KS hamiltonian and wave functions in a PW basis set. Fast Fourier transform and the dual space formalism.

	How things scale with system dimensions ? Why do we need pseudopotentals ? ● Pseudopotentials. Empirical pseudopotentials. Transferability. First principles pseudopotential (unscreening of the reference atomic configuration). Norm Conserving Pseudopotentials: PP in the semilocal form, PP in the fully non-local Kleinman-Bylander form. Ghost States. Ultra Soft Pseudopotentials. Generalized eigenvalue problem and orthogonality. Projection Augmented Wave datasets. Total energy break up in grid and atomic contributions. ● Solving the KS equation. Iterative diagonalization drivers: Davidson diagonalization; Conjugate Gradient; DIIS. Hamiltonian preconditioning. Efficient evaluation of the KS hamiltonian and overlap matrix by dual space formalism and Kleinman-Bylander pseudopotential decomposition. ● Building the new charge density. Brillouin zone sampling. Symmetry. Charge density mixing. Building the new potential. ● Parallelization. Parallelization tools and strategies: MPI and OpenMP; data and workload distribution; bandwidth and latency. Basic parallel operations (checkpoint, broadcast, collect, gather/scatter); communication intensive operations. Amdahl's law. Strong and weak scalability. Hierarchy of parallelization levels. Porting computational intensive modules to GPUs.
Objectives	The aim of this course it to provide the students with a detailed knowledge of the internal design of state of the art electronic structure codes, filling the gap that exists between

	the knowledge of the general principles underlying modern atomistic simulations and their practical implementation in actual codes. Tools and codes available in the Open Source Quantum ESPRESSO software distribution will be used as working examples.

Module 2,5:	**Advanced Computer Architectures & Optimizations**
Coordinator	Chris Dahnken (INTEL) Martin Kronbichler (Technical University of Munich)
CFU	4
Module Description	The course presents advanced topics in optimization techniques needed in HPC environment. In particular it will focus on the use of application accelerators in high-performance and scientific computing and issues that surround it.
Main Topics	● Advanced optimization techniques ● Memory management and optimization ● Introduction to novel accelerator processors, systems, and architectures ● Introduction to GPU computing ● Overview of accelerated architecture ● Programming interfaces for accelerator ○ CUDA ○ OpenCL ○ OpenACC ● Specific libraries with accelerator support
Objectives	On successful completion of this module, students will have an overview of the advanced computational architectures and accelerators and how to use them.

Module 2.6:	**The Finite Element Method Using deal.II**
Lecturer	Luca Heltai (SISSA) Jean-Paul Pelteret (Friedrich-Alexander-Universität)
CFU	4
Module Description	Hands-on module that guides the students to solve a simple poisson problem
Main Topics	• Generation simple meshes • Degrees of Freedom - Matrix Sparsities • A Laplace Solver in 2D • Dimension independent Laplace Solver • Adaptively refined meshes • Hanging nodes and other constraints • A Parallel Laplace Solver in 2D
Objectives	On successful completion of this module students should be able to understand existing codes for the solution of PDEs, and to develop efficient HPC enabled scientific codes dedicated to the solution of PDEs using existing parallel libraries and tools, dealII in particular.

Module 2.7:	**Reduced Basis Method**
Lecturer	Gianluigi Rozza (SISSA)
CFU	4
Module Description	In this course we present reduced basis (RB) approximation and associated a posteriori error estimation for rapid and reliable solution of parametrized partial differential equations (PDEs).
Main Topics	• Introduction to RB methods, offline-online computing, elliptic coercive affine problems • Sampling, greedy algorithm, POD • A posteriori error bounds • Primal-Dual Approximation • Time dependent problems: POD-greedy sampling • Non-coercive problems • Approximation of coercivity and inf-sup parametrized constants • Geometrical parametrization • Reference worked problems • Examples of Applications in CFD
Objectives	On successful completion of this module students should know the basic aspects of numerical approximation and efficient solution of parametrized PDEs for computational mechanics problems (heat and mass transfer, linear elasticity, viscous and potential flows).

Module 2.8:	**Fast Fourier Transforms in Parallel and Multiple Dimensions**
Lecturer	Ralph Gebauer (ICTP) Ivan Girotto (ICTP)
CFU	2
Module Description	Introduction to the Discrete Fourier Transform (DFT) and its application to real problems. From the Discrete to the "Fast" version (FFT). Analysis of a most common algorithm for the solution of a multi-dimensional FFT on parallel systems
Main Topics	● Discrete and Continuous Fourier Transform ● The Cooley-Tukey FFT algorithm ● Implementation of an MPI parallel multi-dimensional FFT based on a 1d FFT ● Example use in a diffusion problem
Objectives	On successful completion of this module students should be able to integrate parallel FFTs into applications.

Module 2.9:	Cluster Analysis
Lecturer	Alex Rodriguez (SISSA)
CFU	2
Module Description	Theory and applications of Clustering algorithms.
Main Topics	• Motivation for Clustering • Similarities and Distances • Flat, fuzzy and Hierarchical clustering methods • Clustering methods examples • External and Internal Validation • Clustering applications
Objectives	The students should be able to implement a clustering algorithm method and to choose the one that fits better the problem that they want to solve.

Module 2.10:	**Monte Carlo method**
Lecturer	Roberto Innocente (SISSA) Sandro Sorella (SISSA) Nicolas Salles (University of Nova Gorica)
CFU	4
Module Description	Theory and applications of the Monte Carlo methods. Hands-on with examples, analysis of simulations and parallelization
Main Topics	• Review of Probability theory • Sampling multi-variate Gaussian deviates • Importance sampling • Stochastic processes • Metropolis Monte Carlo • Langevin dynamics • Introduction to Quantum Monte Carlo • resident-time & Gillepsie algorithms
Objectives	Upon completion students will be able to use and develop simple applications of the Monte Carlo Method on openMP, MPI or hybrid programming paradigms

Module 2.11:	Supervised Machine Learning
Lecturer	Valerio Consorti (Prometeia)
CFU	2
Module Description	Data are becoming the new gold mine in modern companies. The ability to retrieve important information from very large data-sets is more and more requested on the market. This course is focused on teaching how to handle a complete complex data analytics process, by leveraging on supervised machine-learning techniques.
Main Topics	Design: ● Data analytics process; ● Feature extraction to describe data; ● The model choice; ● Performances evaluation; Implementation: ● Python package Scikit-learn; ● Custom transformers and estimators; ● Multi-step analysis
Objectives	You will learn how to implement in python a complete analysis, from data management, to the exploration to the actual implementation of algorithms automatically capable to solve regression and classification problems

Module 2.12:	**Approximation and interpolation of simple and complex functions**
Lecturer	Nicola Seriani (ICTP)
CFU	2
Module Description	Introduction to several techniques for efficient approximation of numerical functions to varying degrees of accuracy
Main Topics	• Interpolation (linear, spline) and errors (number and spacing of interpolation points) • Cost of interpolation vs. explicit function evaluation of complex functions • Approximation (taylor/maclaurin, pade). • Lookup plus newton-raphson • Range reduction as an efficient way to approximate exp(), log(),sin(), cos() and others in combination with a spline table or pade approximation for a small interval. • Floating point math tricks
Objectives	Upon completion of this module students will be able to implement efficient approximations and tabulations of numerical functions and determine the accuracy of the approximations.

Module 2.13:	**Spatial locality algorithms**
Lecturer	Riccardo Valdarnini (SISSA)
CFU	2
Module Description	Theory and applications of algorithms for spatial locality
Main Topics	• Space filling curves-theory • Morton and Peano-Hilbert orders • Tree codes, general framework • Quad tree, R- tree, Kd-tree • Nearest neighbor search using space filling curves.
Objectives	Upon completion students will be able to deal with spatial locality algorithms.

Module 2.14:	Big Data Processing with MapReduce
Lecturer	Alfredo Cuzzocrea (University of Trieste)
CFU	4
Module Description	Course Requirements: Basic notions of data models, database management systems, query languages, distributed processing, data processing algorithms, information retrieval, graph data management. The course provides an introduction to big data and their processing with MapReduce. In particular, the course provides foundations of big data and NoSQL databases and describes real life big data applications along with NoSQL databases and Cloud Computing. References: 1. J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A.H. Byers, “Big Data: The Next Frontier for Innovation, Competition, and Productivity”, McKinsey Global, 2011 2. M. Chen, S. Mao, Y. Liu, “Big Data: A Survey”, Mobile Networks and Applications, Vol.19, No. 2, pp. 171 − 209, 2014 3. T. White, “Hadoop: The Definitive Guide”, 4th Edition, O’Reilly Media, 2015 4. D. Miner, A. Shook, “MapReduce Design Patterns Building Effective Algorithms and Analytics for Hadoop and Other Systems”, O’Reilly Media, 2012 5. J. Lin, C. Dyer, “Data-Intensive Text Processing with MapReduce”, Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers, 2010
Main Topics	● Big Data: Foundations and Applications: big data definitions, big data applications, big data processing,

	big data analytics, real-time analytics, big data technology, big data evolution, cloud computing and big data, service-oriented architectures, cloud computing platforms, real-life big data projects. ● NoSQL Databases for Big Data Representation and Storage: the big data explosion, a case study: Twitter, storage layer alternatives for big data, column-oriented databases, MonetDB, the NoSQL paradigm, classification of NoSQL systems, key-value stores, document stores, column family stores, Apache HBase, Apache Hive, MongoDB, introduction to MapReduce, Apache Cloudera. ● MapReduce Programming and Data Processing: limitations of existing data analytics architectures, divide and conquer paradigm, parallel processing methodologies, distributed and parallel programming models, Apache Hadoop, MapReduce big data processing, MapReduce primitives, runtime execution of MapReduce programs, MapReduce program examples, distributed file systems, GFS, HDFS, MapReduce data flow. ● Database Management with MapReduce: mapping database management operators on MapReduce programs, value-to-key conversion, SQL projection in MapReduce, SQL selection in MapReduce, SQL group-by in MapReduce, SQL aggregation in MapReduce, SQL join in MapReduce, join algorithms in MapReduce, Map-side join, in-memory join, relational data processing with MapReduce, Apache Hive examples on relational data processing, Apache Pig Latin examples on relational data processing. ● Information Retrieval with MapReduce: brief introduction to information retrieval, information retrieval models, indexing data structures for

	supporting information retrieval, MapReduce for information retrieval problems, index construction with MapReduce, mapper and reducer for MapReduce-based information retrieval, improving performance. ● Graph Processing with MapReduce: graph problems and representations, graph analytics, typical graph problems, traditional graph representation alternatives, advanced graph representation alternatives, MapReduce-based graph algorithms, data graph models, state-of-the-art graph algorithms, graphs and MapReduce, MapReduce-based implementations of state-of- the-art graph algorithms (Dijkstra, random walks, PageRank).
Objectives	The course focuses the attention on the MapReduce processing model and its application to several big-data-processing-related application scenarios, such as database management, information retrieval and graph processing. Case studies and examples, as well as state-of-the-art systems and tools, are provided and discussed in details.

Module 2.15:	**Lattice Boltzmann**
Lecturer	Sauro Succi (to add to lecturers' list)
CFU	2
Module Description	
Main Topics	
Objectives	

Module 2.16:	**Molecular Dynamics**
Lecturer	Giovanni Bussi (SISSA)
CFU	2
Module Description	Theory and applications of molecular dynamics simulations."Hands-on exercises on neighbor-list, linked cells, and parallelization strategies
Main Topics	• Introduction to molecular dynamics • Neighbor lists and linked cells • Parallelization using MPI • Implementation of multi-replica algorithms (parallel tempering)
Objectives	Upon completion students will be able to run molecular dynamics simulations of a Lennard-Jones system and to optimize and parallelize a molecular dynamics code. Additionally, they will be able to implement parallel tempering algorithms.