Deliverable D2.3
WP2 Final Report on Exascale Education

CONTRACT NO	EESI2 312478
INSTRUMENT	CSA (Support and Collaborative Action)
THEMATIC	INFRASTRUCTURE

Due date of deliverable:	30/04/2015
Actual submission date:	31/03/2015
Publication date:	31/03/2015

Start date of project: 1 September 2012 Duration: 30 months

Name of lead contractor for this deliverable: TOTAL

Authors: Ulrich Rüde (FA University of Erlangen)

Name of reviewers for this deliverable:

Abstract: This is the final report of EESI2 WP2 Education and Strategic Coordination. The document reports the results of task T2.2 identified in EESI2-WP2 and presents the recommendations.

Revision 0.1

Project co-funded by the European Commission within the Seventh Framework Programme (FP7/2007-2013)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Table of Contents

1. Executive Summary

This document reports the final outcome of EESI2 WP2 “Education and Strategic Coordination”.

It represents an update of the deliverable D2.2.

T2.2 State of the art on education courses and training needs:

Over the past two decades HPC and the field of computational science and engineering (CSE) have penetrated the academic and industrial world, with prominent roles in advancing research and innovation and providing interdisciplinary education. A combination of disruptive developments including challenges in extreme-scale parallel computing and a comprehensive broadening of the application fields of CSE and HPC are changing now also the needs in education and training.

The document presents new strategies and directions for HPC in education and research for the next decade. These include strengthening education in HPC as an important subfield of CSE on three different levels of expertise and academic depth. Incentives must be developed so that suitable interdisciplinary educational structures are being developed and made broadly available. This is essential to leverage the potential that HPC offers in research and industry for all disciplines. European funding for HPC should be focussed on institutions that commit to developing suitable interdisciplinary educational structures and programs. More than in the past, the European research expertise must also be employed for developing innovative educational concepts in the field of HPC and CSE. This is the key to making HPC expertise immediately available for the next generation of researchers and industrial work force and thus to ensure European competitiveness in the decades to come.

2. T2.2 Education Courses and Training Needs

2.1 Introduction

HPC education must be considered in the context of the wider field of Computational Science and Engineering (CSE) which provides the base knowledge for making HPC technology universally relevant in science and engineering. CSE is a multidisciplinary field of research and education, lying at the intersection of applied mathematics, computer science, and core disciplines of science and engineering (Figure 1). CSE is devoted to the development and use of computational methods for scientific discovery in all branches of the sciences and for the advancement of innovation in engineering and technology. It is a broad and vitally important field that is fundamental for all research in high-performance computing (HPC). Additionally, CSE and HPC play a central role in the data revolution. CSE and HPC are affecting every aspect of science and technology and are changing higher education worldwide.

The interdisciplinary constellation of CSE is being adapted only hesitatingly in the European educational framework, since it inevitably requires a deep cross-disciplinary collaboration. Successful education in CSE and HPC is significantly hindered by the traditional disciplinary structures prevalent in universities and research organizations. Therefore, much of the potential of CSE and HPC remains under-used and, as a consequence,many investments in HPC technology cannot be exploited to their full potential.

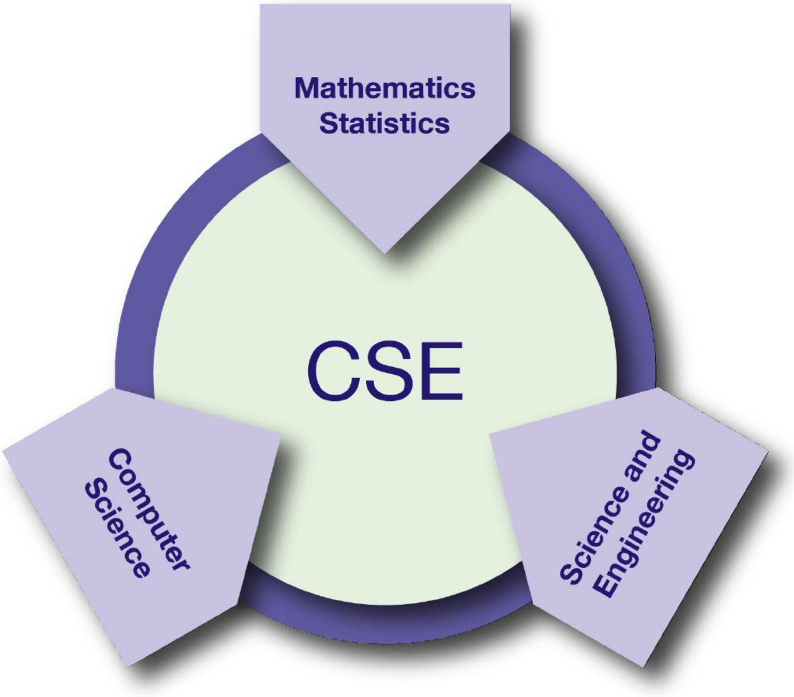

Figure 1: ***CSE at the intersection of mathematics and statistics, computer science, and target disciplines from the sciences and engineering. This combination gives rise to a new field whose character is different from its original constituents. Only in this interdisciplinary constellation the optimal use of HPC technology in science and engineering can be assured.***

CSE is rooted in the mathematics and physical sciences and engineering, where it is now widely recognized as an essential cornerstone that drives scientific and technological progress in conjunction with theory and experiment. The goal of this document is to present the expanding role of CSE and HPC in the 21st-century education landscape.

2.1.1 Importance of CSE and HPC Education

The growing importance of CSE in increasingly many application areas has paralleled the tremendous, exponential growth that has occurred in computing power over the past five decades according to Moore's law. Less well known but crucially important for the success of the CSE paradigm is the fact that these advances in computing power have been matched or exceeded by equivalent advances over the past five decades in the efficiency of the mathematics-based computational algorithms that lie at the heart of CSE. Indeed, the development of efficient new algorithms continues to be a core focus of CSE, which is crucial to the effective use of advanced computing capabilities to address the pressing problems of humankind. HPC education will remain meaningless and irrelevant unless it is set in this wider context.

The emergence and growing importance of massive data sets in many areas of science, technology, and society, in conjunction with the availability of ever-increasing parallel computing power, are transforming the world. Data-driven approaches enable novel ways of scientific discovery and quantifying uncertainties in science and engineering applications. At the same time, relying on new forms of massive data, we can now use the quantitative and model-based approaches that characterize “the scientific method” to drive progress in many areas of society where qualitative forms of analysis, understanding, and decision making were the norm until recently. The CSE paradigm contributes as a cornerstone technology to the data revolution. In these and many other ways, CSE is becoming essential for increasingly broad areas of science, engineering, and technology. It expands human capability beyond its classical limitations, see Figure 2. The figure illustrates that the typical steps in CSE research form a loop which is connected through multiple feedbacks. CSE and HPC education must cover the stages in this loop and the interdependency of their stations as a central paradigm of HPC based development and research.

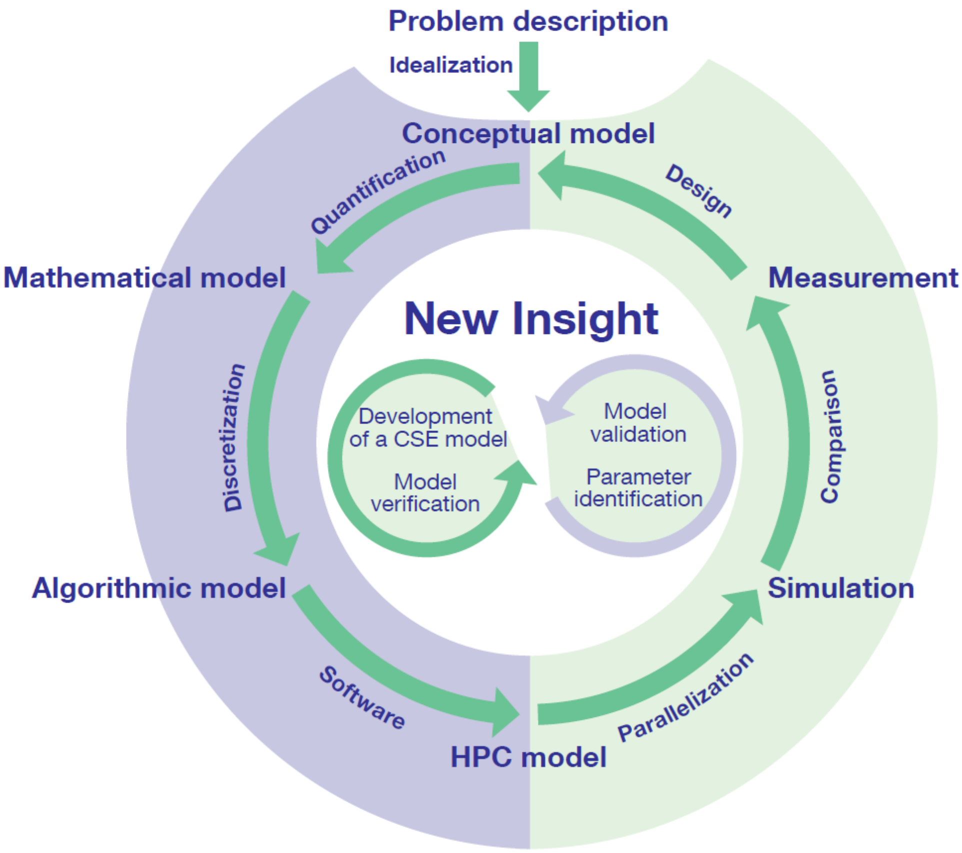

Figure 2: ***CSE pipeline, from application via model and algorithms to efficient implementation in simulation software with verification and validation. The pipeline is actually a loop that requires multiple feedbacks.***

The enormous magnitude of the CSE and HPC challenge can be appreciated when considering the dramatic growth of disciplinary complexity alone. Dominating impediments in many sciences are model complexity and the uncertainty in parameters and data, as well as the huge space and time scales that must be resolved. In computational mathematics, the traditional abstract view of optimality becomes more and more obsolete and must be replaced by new performance-oriented metrics that reflect the true computational cost. In computing, the disruptive transition from modest concurrency to ubiquitous parallelism leaves software development trailing behind. In CSE all these challenges converge and then often pile up to obstructive roadblocks. In research and education these barriers must be overcome by cutting new lanes across the disciplinary borders and by creating a new paradigm of coalescence among the classical disciplines. These barriers must be broken up by novel concepts in training and education that must result in a new generation of HPC-competent computational scientists and engineers.

Nevertheless, high-level CSE and HPC education must differ from courses in mathematics or computer science in that they must be typically oriented to one or more realistic applications of a target discipline in the sciences, engineering, or technology. Such problems often involve complicated three-dimensional geometries, multiple interacting scales, heterogeneities, anisotropies, and multiphysical or biological descriptions; or they may involve complex networks or systems with many components. Thus, the methods developed often thwart rigorous proofs of accuracy or efficiency, and so CSE education must address validation and verification by means other than traditional mathematical analysis. Additionally, CSE and HPC education cannot be limited to any single step of the simulation pipeline. A mathematical simulation algorithm is useful in scientific practice only when it is also implemented on a real computer, and a new parallel processing paradigm becomes relevant for HPC only when it is demonstrated to be applicable for a large-scale numerical computation.

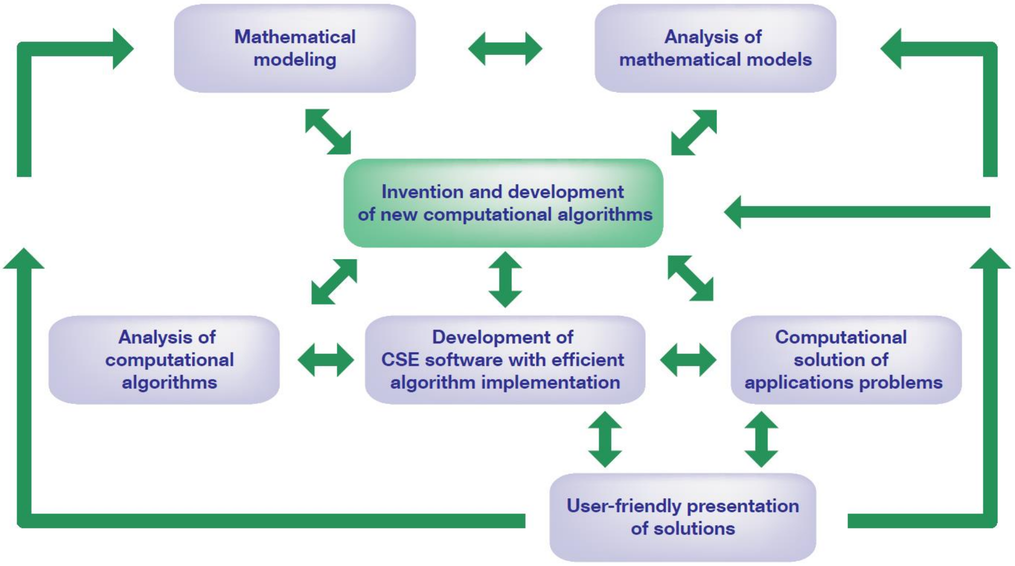

Figure 3: ***Interaction among different CSE components. The development of new algorithms and software is at the core of this view of CSE.***

CSE and HPC often achieve progress through a clever combination of techniques and methods employed for the different stages of the simulation pipeline-loop. In such a case, the innovation may consist in the creativity needed to synthesize a computational solution for a complex problem from the right building blocks. The structure of CSE research and education is illustrated in Figure 3. In essence, CSE research and practice are profoundly interdisciplinary, so that the synthesis achieves a quality that is fundamentally new and unique. CSE researchers and practitioners require a diverse skill set with a focus and drive that is different from what any of the constituent disciplines of mathematics, computer science, and science and engineering can offer by themselves.

2.1.2 The Broad CSE Community

The past two decades have seen tremendous growth in the CSE and HPC community, including a dramatic increase in both size and breadth of intellectual perspectives and interests. As we envision the future of CSE and HPC, and in particular as we consider educational programs, we must keep in mind that such a large and broad intellectual community has a correspondingly broad set of needs. Figure 4 presents one way to view the different aspects of the broad CSE community: *CSE Core Researchers and Developers* - those engaged in the conception, analysis, development, and testing of CSE algorithms and software; and *CSE Domain Scientists and Engineers* - those primarily engaged in applying core CSE algorithms and software in particular science and engineering campaigns. The latter community can usefully be further subcategorized into those who interact with the core technologies at a developer level within their own applications, creating their own implementations and contributing to methodological/algorithmic improvements, and those who are happy to use state-of-the-art CSE technologies as products, combining them with their expert knowledge of an application area to push the boundaries of a particular application. Within the *CSE Core Researchers and Developers* group in Figure 4, we further identify two groups: those focused on broadly applicable methods and algorithms and those focused on methods and algorithms motivated by a specific domain of application. We make this distinction because it is a useful way to cast differences in desired outcomes for different types of CSE educational programs, discussed in detail below. As with any categorization, the dividing lines in Figure 4 are fuzzy, and in fact any single researcher might span multiple categories.

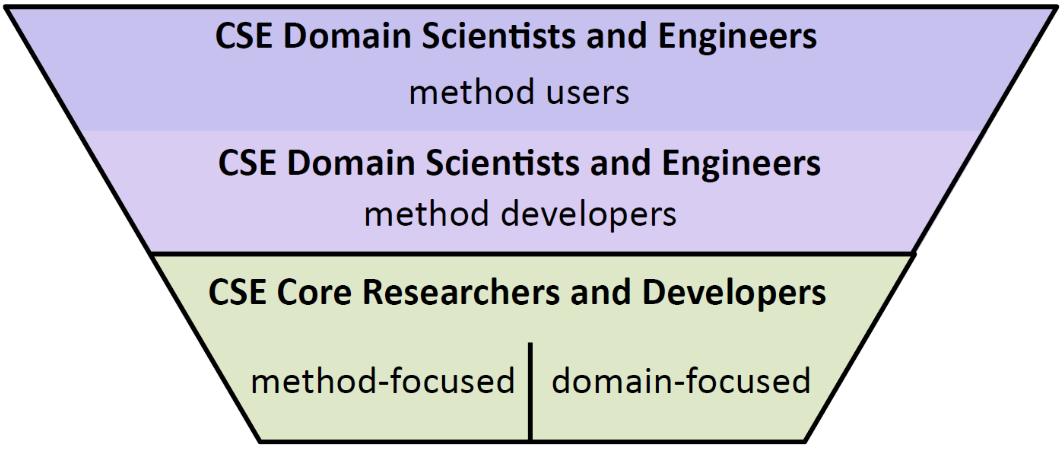

Figure 4: ***One view of the different aspects of the broad CSE community. The part of the CSE community that focuses on developing new methods and algorithms is labeled CSE Core Researchers and Developers. They may be driven by generally applicable methods, or by methods developed for a specific application domain. CSE Domain Scientists and Engineers focus their work primarily in their scientific or engineering domain, but make extensive use of CSE methods in their research or development work. The subcategory of 'method developers' interacts with computational methods and algorithms at a deep level through implementation and method improvements. The subcategory of 'method users'includes end users who often combine the use of advanced CSE methods with a strong focus on applications. HPC technology plays a central role on all levels of this classification.***

2.2 CSE Education and Workforce Development

With all these opportunities on the horizon for the CSE field, there is a growing demand for CSE graduates with strengthened HPC expertise and a need to expand CSE educational offerings in general, and with HPC expertise, in particular. This need includes CSE programs at both the undergraduate and graduate levels, as well as continuing education and professional development programs. In addition, the increased presence of digital educational technologies provides an exciting opportunity to rethink CSE pedagogy and modes of educational delivery.

2.2.1 Growing Demand for CSE Graduates in Industry, National Labs, and Academic Research

Industry, national laboratories, government, and broad areas of academic research are making more use of simulations, high-end computing, and simulation-based decision-making than ever before. This trend is apparent broadly across domains - for example, energy, manufacturing, finance, and transportation are all areas in which CSE is playing an increasingly significant role, with many more examples across science, engineering, business, and government. Research and innovation, both in academia and in the private sector, are increasingly driven by large-scale computational approaches. A National Council on Competitiveness report points out that high-end computing plays a “vital role in driving private-sector competitiveness ... all businesses that adopt HPC consider it indispensable for their ability to compete and survive”. With this significant and increased use comes a demand for a workforce versed in technologies necessary for effective and efficient mathematics-based computational modeling and simulation. There is high demand for graduates with the interdisciplinary expertise needed to develop and/or utilize computational techniques and methods in order to advance the understanding of physical phenomena in a particular scientific, engineering, or business field and to support better decision-making.

As stated in a recent report on workforce development by the US Department of Energy (DOE) Advanced Scientific Computing Advisory Committee “All large DOE national laboratories face workforce recruitment and retention challenges in the fields within Computing Sciences that are relevant to their mission. ... There is a growing national demand for graduates in Advanced Scientific Computing Research-related Computing Sciences that far exceeds the supply from academic institutions. Future projections indicate an increasing workforce gap.” This finding was based on a number of reports, including one from the High End Computing Interagency Working Group stating: “High end computing (HEC) plays an important role in the development and advanced capabilities of many of the products, services, and technologies that are part of our everyday life. The impact of HEC on the agencies of the federal government, the quality of academic research, and on industrial competitiveness is substantial and well-documented. However, adoption of HEC is not uniform, and to fully realize its potential benefits we must address one of the most often cited barriers: lack of HEC skills in the workforce.”

In order to take advantage of the transformation that high-performance and data-centric computing offer to industry, the critical factor is a workforce versed in CSE and capable of developing the algorithms, exploiting the HPC compute platforms, and designing the analytics that turns data with its associated information into knowledge to act. The tasks to exploit the emerging tools require the critical thinking and the interdisciplinary background that is prevalent in CSE training. The CSE practitioner has both the expertise to apply computing tools and the analytical skills to tease out the problems that often are encountered when commercial enterprises seek to design new products, develop new services, and create novel approaches from the wealth of data available. The CSE practitioner knows how to use computational tools and analytics in uncharted areas, often applying previous domain-specific understanding to these new areas. The CSE practitioner, while often a member of a team of others from varying disciplines, is the catalyst driving the change that industry seeks in order not only to remain competitive but also to be first to market, providing the necessary advantage to thrive in a rapidly evolving technological ecosystem.

2.2.2 Future Landscape of CSE and HPC Educational Programs

New CSE and HPC educational programs are needed in order to create young professionals who satisfy this growing demand and who support the growing CSE research field. These include CSE and HPC programs at both the undergraduate and graduate levels, as well as continuing education and professional development programs. These also include programs that are “HPC and CSE focused” and those that follow more of a “CSE and HPC infusion” model. The former includes programs that have CSE and HPC as their primary focus (e.g., a B.S., M.S., or Ph.D. in computational science and engineering), while the latter includes programs that infuse some CSE and HPC training within another degree structure (e.g., a minor, emphasis, or concentration in CSE/HPC complementing a major in mathematics, science or engineering or a degree in a specific computational discipline such as computational finance or computational geosciences).

At the undergraduate level, the breadth and depth of topics covered will depend on the specific degree focus. However, the following high-level topics are important content for an undergraduate program:

1. Foundations in mathematics and statistics, including calculus, linear algebra, differential equations, applied probability, and discrete mathematics.
2. Simulation and modeling, including conceptual models, accuracy, use of modeling tools, assessment of computational models, data-based models, and physics-based models.
3. Computational methods and numerical analysis, including errors, nonlinear equations, solution of systems of linear equations, interpolation, curve fitting, optimization, Monte Carlo, numerical methods for ODEs, and numerical methods for PDEs.
4. Computing skills, including compiled high-level languages, algorithms (numerical and nonnumerical), elementary data structures, analysis, parallel programming, scientific visualization, and awareness of computational complexity and cost.

At the graduate level, again the breadth and depth of topics covered will depend on the specific degree focus. In the next section, we make specific recommendations in terms of a set of learning outcomes desired for a CSE and HPC graduate program. We also note the growing importance of and demand for terminal master's degrees, which can play a large role in fulfilling the industry and national laboratory demand for graduates with advanced HPC and CSE skills.

All CSE graduates should possess the attributes of having a solid foundation in modern mathematics; an understanding of probability and statistics; a grasp of modern computing, computer science, programming languages, principles of software engineering, high-performance computing, and an understanding of foundations of modern science and engineering, including biology. These foundations should be complemented by deep knowledge in one specific area of science, engineering, mathematics and statistics, or computer science. CSE and HPC graduates should also possess skills in teamwork, self-organization, multidisciplinary collaboration, and leadership..

A third area of educational programs is that of continuing and professional education. Opportunities exist for the European HPC community e.g. to engage with industry to create and offer short courses, including those that target general HPC skills for the non-HPC specialist as well as those that target more advanced skills in timely opportunity areas (such as extreme-scale computing, and computing with massive data). Often one assumes that much of the workforce for industry in HPC will come at the postgraduate level; increasingly, however, industry needs people who have an understanding of HPC and CSE even at the undergraduate level in order to realize the full potential growth in a rapidly expanding technological workplace. Future managers and leaders in business and industry must be able to appreciate requirements to practice CSE and HPC and the benefits that accrue from the use of HPC technology in CSE. Continuing education can play a role in fulfilling this need. The demand for training in CSE-related topics exists broadly among graduate students and researchers in academic institutions and research laboratories, as evidenced by the growing number of summer schools worldwide, as well as short courses aimed at the research community. Continuing education programs in HPC can build on the already successfully offered courses at European HPC centers, in particular via PRACE.

2.2.3 Graduate Program Learning Outcomes

A learning outcome is defined as what a student is expected to be able to do as a result of a learning activity. In this section, we describe a set of learning outcomes desired of a student graduating from a CSE Ph.D. program. We focus on outcomes because they describe the set of desirable competencies without attempting to prescribe any specific degree structure. These outcomes can be used as a guide to define a Ph.D or Master level program that meets the needs of the modern CSE graduate; they can also play an important role in defining and distinguishing the CSE identity and in helping employers understand the skills and potential of CSE graduates.

In Table [1], we focus on the "CSE and HPC Core Researchers and Developers" category in Figure [4]. We distinguish between a CSE education with a broadly applicable CSE focus and a CSE degree with a domain-driven focus. An example of the former is a "Ph.D. in computational science and engineering" while an example of the latter is a "Ph.D. in computational geosciences." The listed outcomes relate primarily to those CSE-specific competencies that will be acquired through classes. Of course, the full competencies of the Ph.D. graduate must also include the more general graduate-level skills, such as engaging deeply in a research question, demonstrating awareness of research context and related work, and producing novel research contributions, many of which will be acquired through the dissertation. We also note that it is desirable for graduates of a CSE master's degree program to also achieve most (if not all) of the outcomes in Table [1]. In particular, in educational systems where there is no substantial classwork component for the Ph.D., the learning outcomes of Table 1 would also apply to the master's or honors degree that precedes the Ph.D.

In the next two subsections, we elaborate more on the interaction between CSE education and two areas that have seen considerable change since the design of many existing CSE programs: extreme-scale computing and computing with massive data.

<table>
<tr><th>CSE PhD with broadly applicable CSE focus</th><th>CSE PhD with domain-driven focus in field X</th></tr>
<tr><td>Combine mathematical modeling, physical principles, and data to derive, analyze, and assess a *model across a range of systems (e.g., statistical mechanics, continuum mechanics, quantum mechanics, molecular biology).*</td><td>Combine mathematical modeling, physical principles and data to derive, analyze, and assess a *range of models within field X.*</td></tr>
<tr><td colspan="2">Demonstrate graduate-level depth in devising, analyzing, and evaluating new methods and algorithms for computational solution of mathematical models (including parallel, discrete, numerical, statistical approaches, and mathematical analysis).</td></tr>
<tr><td>Demonstrate *mastery* in code development to exploit parallel and distributed computing and other emerging modes of computation in algorithm implementation.</td><td>Demonstrate *proficiency* in code development to exploit parallel and distributed computing and other emerging modes of computation in algorithm implementation.</td></tr>
<tr><td colspan="2">Select and apply techniques and tools from software engineering to build robust, reliable, and maintainable software.</td></tr>
<tr><td colspan="2">Develop, select, and use tools and methods to represent and visualize computational results.</td></tr>
<tr><td colspan="2">Critically analyze and evaluate results using mathematical and data analysis, physical reasoning, and algorithm analysis, and understand the implications on models, algorithms and implementations.</td></tr>
<tr><td colspan="2">Identify the sources of errors in a CSE simulation (such as modeling errors, code bugs, premature termination of solvers, discretization errors, round-off errors), and understand how to diagnose them and work to reduce or eliminate them.</td></tr>
<tr><td>Appreciate and explain the context of decision-making as the end use of many CSE simulations,
and as appropriate be able to formulate, analyze, and solve CSE problems in control, design, optimization, or inverse problems.</td><td>Appreciate and explain the context of decision-making as the end use of many CSE simulations,
and as appropriate be able to formulate, analyze and solve CSE problems in control, design, optimization or inverse problems *as relevant to field X.*</td></tr>
<tr><td colspan="2">Understand data as a core asset in computational research and demonstrate appropriate proficiencies in processing, managing, mining, and analyzing data throughout the CSE/simulation loop.</td></tr>
<tr><td colspan="2">Demonstrate the ability to develop, use, and analyze sophisticated computational algorithms in data science and engineering, and understand data science and engineering as a novel field of application of CSE.</td></tr>
<tr><td>Demonstrate graduate-level *proficiency in one domain in science or engineering.*</td><td>Demonstrate graduate-level *depth in domain knowledge in field X.*</td></tr>
</table>

Communicate across disciplines and collaborate in a team.

Table 1: ***Learning outcomes desired of a student graduating from a CSE PhD program. Italicized text denotes differences in learning outcomes for programs with a broadly applicable CSE focus (left) and a domain-driven focus in a particular field of science or engineering (right). Learning outcomes that are common to both types of PhD programs span left and right columns.***

2.2.4 Education in Parallel Computing and Extreme-Scale Computing

In addition to the broader educational goals discussed in the previous section, the following analysis will present the necessary components of education specifically in HPC and for parallel computing. Extreme-scale computing poses new challenges to education, but there is also the broader and fundamental need to educate a wide spectrum of engineers and scientists to be better prepared for the age of ubiquitous parallelism, Parallelism has become the basis for all computing technology, which necessitates a shift in teaching even the basic concepts. Simulation algorithms and their properties have been in the core of CSE education, but now we must emphasize parallel algorithms. The focus used to be on abstract notions of accuracy of methods and the complexity of algorithms; today it must be shifted to the complexity of parallel algorithms and the real-life cost to solve a computational problem, which is a completely different notion. Additionally, the asymptotic complexity and thus algorithmic scalability become more important when the machines grow larger. At the same time, the traditional complexity metrics increasingly fail to give guidance about which methods, algorithms, and implementations are truly efficient. Designing and maintaining simulation software has become an extremely complex, multifacetted art. The education of future computational scientists must address these topics that arise from the disruptive technology that is dramatically changing the landscape of computing.

Today’s extreme scale is tomorrow’s desktop. An analogous statement holds for the size of the data that must be processed and that is generated through simulations Education in programming techniques needs to be augmented with parallel programming elements and a distinctive awareness of performance and computational cost. Additionally the current trends are characterized by a growing complexity in the design of the computer architectures. They are hierarchical and heterogeneous, These architectures are reflected by complex and evolving programming models that should be addressed in a modern CSE education. Programming such systems will typically require using different languages and programming paradigms explicitly, having to interface them with the heterogeneous architecture on the one side and the algorithmic requirements on the other side, leading to a complex, interdependent software architecture that is not easy to make flexible, efficient, and highly performant. Overall this goal can be achieved only through a co-design process where all aspects and design decisions are balanced and weighed against each other. Graduates of CSE and HPC programs must be aware of these issues and should be capable of collaborating with specialists in these topics.

In defining the educational needs in parallel and high-performance computing, we must distinguish between different intensities. Any broad education in CSE will require an understanding of parallel computing, simply because sequential computers have ceased to exist. All students must be trained to understand concepts such as concurrency, algorithmic complexity, and their relation to scalability, elementary performance metrics, and systematic benchmarking methodologies. Unfortunately many relevant educational programs across Europe currently fail to address these issues adequately.

In more demanding applications, parallel computing expertise and performance awareness are necessary and must go even further beyond the content of most current curricula. We point out that this requirement is equally true in those applications that may be only of moderate scale but that nevertheless have high-performance requirements, such as those in real-time applications or those that require interactivity. Especially these applications include many of immediate industrial and commercial relevance, such as real-time predictive control, medical imaging and computational medicine, and many more. Here, education must include a fundamental understanding of computer architectures and the programming models that are necessary to exploit these architectures for innovative products and services.

Besides classification according to scientific content and HPC intensity, educational structures in CSE must also address the wide spectrum of the CSE community. This is elaborated below:

CSE Domain Scientists and Engineers – Method Users. Users of CSE technology will typically use dedicated supercomputer systems and specific software on these computers; they will usually not program HPC systems from scratch. Nevertheless, they need to understand the systems and the software they use, in order to achieve leading-edge scientific results. If needed, they must be capable to extend the existing applications, if necessary in collaboration with CSE and HPC specialists.

An appropriate educational program for CSE Users in HPC can be organized in courses and tutorials on specific topics such as are regularly offered by computing centers and other institutions. These courses are often taught in compact format (ranging from a few hours to a week) and are aimed at enabling participants to use specific methods and software or specific systems and tools. They naturally reach only limited depth, but a wide spectrum of such courses is essential in order to widen the scope of CSE and HPC technology and to bring it to bear fruit as widely as possible.

CSE Domain Scientists and Engineers – Method Developers. These are often domain scientists or engineers who have specialized in using computational techniques in their original field. They often have decades of experience in computing and using HPC, and thus, historically, they are mostly self-taught. Regarding the next generation of scientists, students of the classical fields (such as physics, chemistry, or engineering) will increasingly want to put stronger emphasis on computing within their fields.

The more fundamental knowledge that will be needed to competently use the next generation of HPC systems thus can then not be adequately addressed by compact courses as described above. A better integration of these topics into the university curriculum is necessary, by teaching the use of computational methods as part of existing courses or by offering dedicated HPC- and simulation-oriented courses (as electives) in the curriculum. An emphasis on CSE and HPC within a classical discipline may be taught in the form of a selection of courses that are offered as electives by CSE or HPC specialists, or—potentially especially attractive—by co-teaching of a CSE specialist jointly with a domain scientist.

CSE Core Researchers and Developers. Scientists who work at the core of CSE are classified in two groups. *Domain-driven* CSE students as well as those focusing on *broadly applicable methods* should be expected to spend a significant amount of time learning about HPC and parallel computing topics. These elements must be well integrated into the CSE curriculum. Core courses from computer science (such as parallel programming, software engineering, and computer architecture) may present the knowledge that is needed also in CSE, and they can be integrated into a CSE curriculum. Often, however, dedicated courses that are especially designed for students in CSE will be significantly more effective, since they can be adapted to the special prerequisites of the student group and can better focus on the issues that are relevant for CSE. Often again co-teaching such courses, labs, or projects may be fruitful, especially when such courses cover several stages of the CSE pipeline.

These three levels of CSE education are naturally interdependent, but we emphasize that all three levels are relevant and important. In particular, the problem of educating the future generation of scientists in the competent use of computational techniques cannot be addressed solely by offering one-day courses on how to use the latest machine in the computing center.

2.2.5 CSE Education in Uncertainty Quantification and Big Data

The rising importance of massive data sets in application areas of science and engineering and beyond has broadened the skillset that CSE graduates may require. For example, data-driven uncertainty quantification requires statistical approaches that may include tools such as Markov chain Monte Carlo methods and Bayesian inference. Analysis of large networks requires skills in discrete mathematics, graph theory, and combinatorial scientific computing. Similarly, many data-intensive problems require approaches from inverse problems, large-scale optimization, machine learning, and data stream and randomized algorithms.

The broad synergies between computational science and data science offer opportunities for educational programs. Many CSE competencies translate directly to the analysis of massive data sets at scale using high-end computing infrastructure. Computational science and data science are both rooted in solid foundations of mathematics and statistics, computer science, and domain knowledge, and this common core may be exploited in educational programs that can prepare the computational and data scientist of the future.

2.2.6 Changing Educational Infrastructure

As we think about CSE educational programs, we must also consider the changing external context of education, particularly with regard to the advent of digital educational technologies and their associated impact on the delivery of education programs.

One clear impact is an increased presence of online digital materials, including digital textbooks, open educational resources, and massive open online courses (MOOCs). Recent years have already seen the development of online digital CSE resources, as well as widespread availability of material in fields relevant to CSE, such as HPC, machine learning, and mathematical methods. An opportunity exists to make better community use of current materials, as well as to create new materials. There is also an opportunity to leverage other resources, such as CSGF essay contest winners (https://www.krellinst.org/csgf/outreach/cyse-contest) and archived SIAM plenaries and other high-profile lectures. It would be timely to create a SIAM focus group to create and curate a central repository linking to CSE digital materials and to coordinate community development of new CSE online modules. This effort could also be coordinated with an effort to pursue opportunities in continuing education.

Digital educational technologies are also having an impact on the way residential courses are structured and offered. For example, many universities are taking advantage of digital technologies and blended learning models to create “flipped classrooms”, where students watch video lectures or read interactive online lecture notes individually, and then spend their face-to-face class time engaged in active learning activities and problem solving. Digital technologies are also offering opportunities to unbundle a traditional educational model—introducing more flexibility and more modularity to degree structures. Many of these opportunities are well suited for tackling the challenges of building educational programs for the highly interdisciplinary field of CSE.

2.3 Recommendations and Conclusions

Based on the findings in the above report, the following recommendations are made:

1. Strengthening and broadening education in parallel computing and HPC is of vital importance for research and development in Europe, since parallelism has become ubiquitous for all uses of computers in science and technology.
2. A well-balanced system of educational offerings is the basis for shaping the *HPC ecosystem* that is necessary to give computing-based research its foundations and which is necessary to leverage its benefits.
3. HPC training must be seen embedded in the larger problem of education in Computational Science and Engineering (CSE). Considering HPC education without CSE is like considering the engines but forgetting the car that it should power. The engines without the cars would get us nowhere.
4. HPC and CSE education must respect the three levels of expertise that are defined analogous to the structure of the HPC community. Thus HPC education must be organised according to the following classification:

    - HPC **method users** will need a well-balanced system of training courses that enables them to use the most recent HPC systems and software for their research and development purposes competently.

- HPC **method developers** in the domain sciences must in the future be better educated on the university level, e.g. by offering electives in the standard curricula of their fields. Here universities must become active and develop suitable modules. The EU can support this by giving suitable incentives. In particular the existence of suitable educational offerings must be made a prerequisite for institutions and consortia to receiving research funding in HPC from the EU.
- HPC and CSE **core researchers**, both domain driven as well as methods driven, can only be educated efficiently by dedicated course programs. Such programs have been created up to now only in a few universities across Europe. More such undergraduate and graduate level (Master or PHD) programs in CSE and HPC are urgently necessary throughout Europe to train and to prepare the future generation of core HPC- and CSE-scientists. Such a broadening of the competence basis will be essential to move the field forward and to leverage the potential of HPC for European society. Progress in this area will be a long-term key factor to European competitiveness in science and technology. The EU is challenged to develop a system of incentives so that universities invest more aggressively in what will become the foundation of science and technology for the ongoing century.

3. Acknowledgments

This report is based on joint activities of the Society for Industrial and Applied Mathematics and the European Exascale Software Initiative. The two sponsoring organizations have funded a workshop which was held August 4-6, 2014 in Breckenridge, Colorado and which brought together leaders in the field to discuss aspects of education in CSE and HPC. A draft version of the longer report is accessible from the internet at http://wiki.siam.org/siag-cse/index.php/Main Page

Based on this draft version, extensive community feedback was collected during panel discussions and a dedicated miniysmposium held during the SIAM conference on Computational Science and Engineering, March 14-18, 2015 in Salt Lake City, USA. Based on this process, an extended final version of the report will be compiled by the end of June 2015. For this extended version, a publication in SIAM Review is being planned. The authors of the report are

1. Ulrich Rüde, Universität Erlangen-Nürnberg
2. Karen Willcox, Massachusetts Institute of Technology
3. Lois Curfman McInnes, Argonne National Laboratory
4. Hans De Sterck, University of Waterloo

Additional Contributors: Hans Bungartz, Technische Universität München; James Corones, Krell Institute; Evin Cramer, Boeing; James Crowley, SIAM; Omar Ghattas, University of Texas at Austin; Max Gunzburger, Florida State University; Michael Hanke, KTH Stockholm; Robert Harrison, Brookhaven National Laboratory and Stonybrook University; Peter Jimack, University of Leeds; Chris Johnson, University of Utah; Kirk Jordan, IBM; David Keyes, KAUST; Rolf Krause, Universita della Svizzera Italiana, Lugano; Vipin Kumar, University of Minnesota; Stefan Mayer, MSC Software, Munich;Juan Meza, University of California, Merced: Knut Martin Morken, University of Oslo; J. Tinsley Oden, University of Texas at Austin; Linda Petzold, University of California, Santa Barbara; Padma Raghavan, Penn State University; Anne Trefethen, University of Oxford; Peter Turner, Clarkson University; Vladimir Voevodin, Moscow State University; Barbara Wohlmuth, Technische Universität München; Carol Woodward, Lawrence Livermore National Laboratory

We gratefully acknowledge La Joyce Clark and Cheryl Zidel of Argonne National Laboratory for workshop support. We also thank Jenny Radeck of Technische Universität München for assistance with graphics and Gail Pieper of Argonne National Laboratory for editing an earlier version of this document. We further thank Dr. Harald Lanig and Gaby Fleig of FAU for editorial support.