EMMC-CSA

European Materials Modelling Council

Report on

Expert group meeting on

Coupling & Linking

November 8-9, 2018

CECAM-HQ-EPFL, Lausanne, Switzerland

TABLE OF CONTENT

This workshop and report has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 723867

PROPRIETARY RIGHTS STATEMENT

This document contains information, which is proprietary to the EMMC-CSA Consortium. Neither this document nor the information contained herein shall be used, duplicated or communicated by any means to any third party, in whole or in parts, except with prior written consent of the EMMC-CSA consortium.

1. Executive summary

1.1 Description of the deliverable content and objectives

Algorithms and methods to couple and link multiple model types are often needed to tackle some of the most challenging problems in materials' simulations. Developing these represents both a complex and fascinating intellectual challenge, and a pressing need to tackle realistic systems in applied and industrial research.

1.2 Major outcome

We summarize here the conclusions of the meeting, where we have identified the areas where coupling-and-linking efforts would have the greatest impact, and which should be the policy recommendations to make to the Commission to facilitate this impact.

The following areas are of strategic industrial and technological importance, greatly benefit from simulations, and require coupling-and-linking:

1) Predicting the strength of novel metal alloys
2) Predicting storage performance of devices to leverage renewable energies
3) Enabling materials for future information-and-communication technologies/quantum computing
4) Enabling environmental remediation through engineered materials: nanoparticles, metal-organic frameworks, nanoporous materials
5) Predicting manufacturability of novel materials
6) Enabling superhydrophobic / superhydrophilic materials, new colloids, lubricants, inks, machining fluids

There are a number of policy and funding recommendations that should be made, and that we divided according to being relevant for the “Excellence in Science” pillar, or the “Industrial” pillar.

Policy/funding recommendation for the “Excellence in Science” pillar:

1) We need a sustainable model for the development of scientific software, since currently there are no careers in the field – we need to have careers paths for those experts
    a. Embedded in existing academic/research institutions
    b. Embedded in supercomputer centers or national facilities
    c. In a distributed model (a virtual Joint Research Center)
2) Development of scientific software can be fostered by the Commission, asking specific requirements: to be public, verified and validated, and embedded/representing major communities
3) The ERC categories, especially PE3 or PE5, do not recognize theory and simulation of materials as a field – this has a broad and negative impact, since these categories are adopted also by many national bodies

Policy/funding recommendation for the "Industrial" pillar:

1) We need enriched/integrated modelling workflows by using AI: making the modelling faster, better, more accurate and more predictive, predicting more complex properties
2) We need to understand how to deal with heterogeneous low-quality and high-quality data Use outputs of models for improving the quality of experimental data, for reducing the heterogeneity of data, for integrating data coming from different sources
3) We need to integrate modelling with equipment (e.g. robots, high throughput synthesis, measurements and characterization)

2. Progress report (main activities)

2.1 Description of the meeting content and objectives

This expert meeting, convened by the European Materials Modelling Council and hosted by CECAM (https://www.cecam.org/workshop-1667.html), gathered some of the foremost experts across different domain areas, with the goal of identifying a roadmap for coupling and linking simulations that can act as a manifesto for the field, and inform future policy makers of the algorithmic and methodological developments needed and of their relevance and role in accelerating simulation-driven innovation.

2.2 Sessions

We provide here a summary of all the contributions, where for each presenter we summarize the challenges, scientific examples, software used, vision for the field, policy recommendations, and a summary of the ensuing discussion.

- Kurt Kremer Multiscale Simulations of Soft Matter

Challenges: Adaptive simulations; grand canonical simulations, open system simulations. Nonbonded interactions. Correct conformations, usually not in equilibrium, determine electronic/optical etc. properties, which couple to details of local molecular morphology. Structure formation, aggregation, dynamics, accuracy of all atom force fields (also state point dependent!)

Science: Example: photo switchable azobenzene, liquid crystal phase transitions. Dynamics challenging (parking lot mechanism versus direct jump between layers of sematic system; incorrect ratio of different mechanisms for coarse grained models). Structure (and thus function) is history dependent, crucial problem for structure and function development. Other example: Top down/bottom up for P3HT for organic electronics; generic models of polymer melts to be equilibrated to understand non-linear viscoelastic properties. Coarse graining in chemical space. Sequential sampling, and concurrent sampling (for the latter, adaptive resolution). Two ways of coupling: mixed Hamiltonian, mixed forces, with interpolation. Path to open systems simulations, needed for externally driven molecular systems (both synthetic and biological).

Software: http://www.votca.org/ Coarse-graining toolkit (VOTCA-CSG), Charge Transport toolkit (VOTCA-CTP), Excitation Transport Toolkit (VOTCA-XTP), focuses on the analysis of molecular dynamics data, the development of *systematic coarse-graining* techniques as well as methods used for simulating *microscopic charge (and exciton) transport* in disordered semiconductors. Its C++ core is interfaced to bash and perl flow-control scripts.

Supported file formats:

- xtc, trr, tpr (all formats supported by GROMACS)
- DL_POLY FIELD and HISTORY
- LAMMPS dump files
- pdb, xyz (to use with ESPResSo, ESPResSo++ and HOOMD-blue)

http://www.espresso-pp.de/index.html
Software package for the scientific simulation and analysis of atomistic or coarse-grained bead-spring models as they are used in soft matter research. C++ core and Python user interface, open-source software published under the GNU General Public License (GPL).

Vision: Online coupled experiments and simulations (needs huge speedup!), true open systems simulations including QM (both electronic and nuclei for hydrogens)) to study function of strongly fluctuating synthetic or biological molecular systems, overcoming the problem of dynamics and coarse graining.

Policy: Interoperability of software packages, curating data (field is still very diverse), documentation of force field properties (parametrization conditions…)

Discussion: Coarse graining: you lose degrees of freedom, so everything becomes state (thermodynamic) dependent, but you also generate the best new degrees of freedom that capture at the higher scale the smaller scale (e.g. friction)

References:
G. Zhang et al, JCP, 142,221102 (2015)
P. Gemunden et al, Macrmol. Rap. Comm., 36, 1047 (2015)
A.C. Fogarty et al, Prot.-Struct. Funct. and Bioinf., 12, 1902 (2016)
T, Bereau et al, APL Materials, 4, 053101 (2016)
K. Kreis et al, JCP, 147, 244104 (2017)
B. Mukherjee et al, JCP, 147, 114501 (2017)

- Aatto Laaksonen Inverse Monte Carlo (IMC)

Challenges: (for biological structures and processes) Many orders of magnitude (numbers, length, size, volume, weight, time. Biology: self-assembly in time. Electronic degrees of freedom, all-atom models, 1st order coarse graining, higher-order coarse graining

Science: Ab-initio -> trajectory -> RDF -> inverse problem -> effective all-atom potentials -> run AA simulations -> coarse graining -> . IMC potentials are solvent mediated. An ideal system for hierarchical multi-scale modeling: the chromosome - from a short region of DNA double helix to the entire chromosome. From all-atoms (11M atoms) to 29,000 coarse grained to 98 super beads.

Software: Magic

Vision: Genome-wide modelling

Discussion: Anderson theorem on all-atom <-> RDF

- Ignacio Pagonabarraga Lattice Boltzmann for complex matter

Challenges: Long-time/large-scale. Good scalability. Load balancing. Top down: more realistic functionals, molecular specificity, fluctuations/correlations. Bottom up: more realistic functionals, match equilibrium and dynamics properties. Calibrations: e.g. too compressible, realistic density contrast, separation of length scales. We normally assume, when deriving or proposing coarse grained models a separation of time scales. This is generically not the case.

Science: Mixtures of oil, water and ions. Colloidal phoresis - multiscale transport phenomena.

Software: There exist several available, open source implementations. The functionalities vary among available software. ESPResSo, ESPResSo++, LAMMPS contain lattice Boltzmann implementations. walBERla and ludwig are LB implementations

Vision: Generic framework to address material processing; relevant degrees of freedom: capture proper dynamical coupling, new hybrid models that adapt to relevant questions. Consistent thermodynamic formulation of fully dynamic schemes; develop schemes that can identify relevant variables dynamically consistent.

Policy: Address time scale bottleneck in existing and future hardware facilities. Need for proper training on kinetic-based modelling,

- Berend Smit Classical md, kmc, thermodynamical data

Introduction: The focus of this presentation is on the field of nanoporous materials. These materials are used for (gas) separation, gas storage, catalysis, sensing, etc. Of particular interest are Metal Organic Frameworks (MOFs), which a crystalline materials which have a unique chemical tune ability. By changing the metal node or organic linker researchers have synthesized over 70'000 different materials and databases have been generated that contain over millions of predicted materials.

Challenges: Experimental: we can synthesize many more materials that we can test for all possible applications. Which materials is most promising for a given application, and what can be achieved. Computational: what is the right metrics to rank? Can we computer all properties accurately and efficiently. Can predicted materials be made?

Examples of computational challenges:
Challenge 1: MOFs are relatively large crystals (some have more than 1000 atoms in a unit cell) and have been synthesized with almost all metals in the periodic table. Hence, a computational challenge is to develop robust simulation techniques that allows us to sufficiently accurate, efficient and robust to predict a large range of properties for work for all kinds of MOFs and related materials.

Challenge 2: a multiscale/multidisciplinary approach.
To rank materials we need metrics for performance, this requires close collaborations with engineers that make the actual design for a device in which these materials are used. This will insights in the properties one need to be able to compute and the required accuracy.
Computing these properties requires linking different simulation techniques (optimizing the structure, generating force fields, predicting thermodynamic and transport properties, etc.)

Science: MOFs. Genomic approach to generate millions of materials. Application: Xe/Kr separation - using previously generated materials for a new use, with outstanding performance. Application: methane storage at 65 kbar (no reinforced steel tanks). Screening used to assess if an economic target is reachable.

Software: http://aiida.net , http://materialscloud.org

Vision: to develop a computational framework in which groups can upload and experimental or predicted crystal structure of a MOF and the program makes reliable predictions of the properties of this materials and makes a ranking on a range of performance indicators for a set of properties.

Policy:

- Nanoporous materials genome: Simulations have advanced to an extent that predictions on performance can be made before synthesis, but not for all materials with accuracy and efficiency on all properties.
- Infrastructure: Predictions can be made, but tools are scattered and do not communicate with each other's. Materials databases are scattered, and no agreement on format. Works for many structures but not all.
- Enormous number of materials: importance of big-data science
- The effort to making data useful and cutting edge - it's easy to dump a lot of stuff, but useful is key.

References:
H. Furukawa, K. E. Cordova, M. O'Keeffe, and O. M. Yaghi, *The Chemistry and Applications of Metal-Organic Frameworks* Science 341 (6149), 974 (2013) http://dx.doi.org/10.1126/Science.1230444
K. Lee, J. D. Howe, L. C. Lin, B. Smit, and J. B. Neaton, *Small-Molecule Adsorption in Open-Site Metal-Organic Frameworks: A Systematic Density Functional Theory Study for Rational Design* Chem. Mat. 27 (3), 668 (2015) http://dx.doi.org/10.1021/cm502760q
P. G. Boyd, Y. Lee, and B. Smit, *Computational development of the nanoporous materials genome* Nat. Rev. Mater. 2, 17037 (2017)

- Lucio Colombi Ciacchi Computational materials science, ontologies

Challenges: The drawback of the materials genome is that the genome alone does not tell you how the cell works. Nature 533, 22-25 (2016) "Materials databases, especially if obtained via computational screening, tell very little on how potentially interesting materials can be synthesised, manufactured and perform in a complex environment". Challenge: Modern data is incomplete and heterogeneous - queries can be highly incomplete; multiple data sources extremely difficult. Emerging approach: Ontology-based Data Access.

Science: Emerging field of materials informatics. Ontology is a formal representation of knowledge in term of key concepts and their inter-relationships in a domain, ontology introduces the vocabulary and semantics: names of terms, relationships, meaning of terms and relationships. Bremen: an intelligent software agent for materials development. Long-term vision: Ontology orders the data so that machine learning can identify a global solution. Learn-of-the-fly and now machine-learning potentials as pioneering efforts - the learning domain is QM, and the simulation domain is MM.
To develop good ontologies you need deep domain knowledge, and formal logic - a strong team of curators is required to avoid formal incompatibilities. "Materials and process" domain is extremely rich and complex. Databases are already well developed; mapping strategies are required.

Software: The most common language for ontology development, at least in the Materials field, is OWL (https://www.w3.org/OWL/). As an editor and reasoner, typically Protégé is employed (https://protege.stanford.edu). It is open source and flexible enough for many purposes.

Vision: Beyond mere interoperability, the main benefits of ontology development would be: (i) the possibility of merging different simulation scales on the basis of data-driven, logically consistent and transferable schemes; and (ii) the possibility of exploiting large amount of heterogeneous (and incomplete) sets of data by artificially intelligent software agents.

Policy: Why develop ontologies? For interoperability (common language enables seamless interdisciplinary work). For a machine-readable standard vocabulary. To link datasets. For intelligent software agents. Note EMMO development. Ontology development must be connected and is necessary for materials informatics. Merging different simulations scales, and to exploit artificially intelligent software agents. However, it's debatable if one universal ontology can be developed with reasonable effort and, at the same time, have a lean structure that enables its use by intelligent software agents. Consortial endeavors pursuing data-driven approaches and materials and process design could be the "breeding ground" for ontologies.

If we want to have EMMO at the next level, a major effort is needed. I (Lucio) personally think that a task force of at least 20 scientists working full time on it for the next 10 years, acting as logical structure developers, domain experts, and central curators, would be necessary. However, there is definitely a big risk associated with doing a large investment before a number of tangible "success stories" exploiting the EMMO have been published.

Regarding data accessibility and exploitation: There are ongoing efforts at the national level to set up policies and facilities for long-term storage of research data. However, especially small and medium sized industries may be reluctant to publish whole sets of data related to their own products. The added-value of data sharing is not necessarily larger than the intrinsic value of privately kept intellectual property. Even for scientists, it remains open which reward one gets in sharing collected data, especially if this is associated with additional effort (data curation, complying with a rigid data-storing structure)

Discussion: There is a seeming inconsistency about a "universal" ontology able to map the whole physics domain and the risk of the same ontology not being capable of delivering practical solutions in materials engineering or materials modelling. True, but this is due because the "language" of a community is often at a much higher level than the language of fundamental physics. So, mapping the concepts of a community onto a universal, physics-oriented ontological framework can be cumbersome and lead, paradoxically, to inconsistencies.

A "closed" ontological framework integrating as much as possible the domain knowledge of materials science could prevent the discovery of revolutionary new knowledge. Maybe - although the semantic correlation between heterogeneous and incomplete sets of data may, on the other hand, reveal utterly unknown hidden relationships between materials composition, properties, process and performance, thus delivering new knowledge. Also, a well-constructed ontology should help humans to generate new hypotheses and stimulate the collection of novel data where these are most urgently needed.

References:
On Materials Informatics in general: K. Rajan, Annu. Rev. Mater. Res. 45, 153-169 (2015).
On A.I.-supported materials development: N. Nosengo, Nature 533, 22-25 (2016).
MatSEEC Science Position Paper Ontology/API from OptiMaDe: http://www.optimade.org.
Strong group developing ontological frameworks for Product Lifecycle Management at the EPFL: Prof. Dimitris Kiritsis, https://ict4sm.epfl.ch

- Martin Thomas Horsch Ontologies and metadata in the Virtual Materials Marketplace (VIMMP) project

Challenges: "Design your own bespoke material model in five minutes", different ontologies need translations, ontologies often linked to methodologies/levels of description, link market place and appropriate ontology, MODA workflow graph language needs ontology.

Science: Vapour-liquid equilibrium bulk and interfacial properties of fluid mixtures and modelling homogeneous nucleation by coupling+linking molecular methods (e.g., MD simulation based on intermolecular force fields and density gradient theory coupled to the PC-SAFT molecular equation of state).

Software: Database of molecular models: http://molmod.boltzmann-zuse.de . RDA task group "Semantic Assets for Materials Science" as a part of the RDA Vocabulary Services Interest Group (cf. https://www.rd-alliance.org/groups/vocabulary-services-interest-group.html).

Vision: End-user controlled model design. Multicriteria optimization, that can be put in practice for well defined materials classes. EVMPO formulating fundamental paradigmatic categories. Standardized subdomain-specific metadata.

Policy: Formulate a European Virtual Marketplace Ontology (EVMPO) below the EMMO. On this basis, a joint European Virtual Marketplace Framework will be created, containing the VIMMP platform, the MARKETPLACE project platform, and all other interested initiatives that have an interest in becoming interoperable with each other. The following policy statement was made concerning the topic of semantic assets for molecular methods:

To interoperate with thermodynamic property databases, end-user accessible frontends, and numerical solvers that use thermodynamic data (in particular, process simulation), major efforts need to be directed toward workflows for bespoke model design, as an approach that empowers the model end user, and more generally, toward accessible and automated problem-to-solution translation workflows. Thereby, e.g., an application engineer would first train a multicriteria optimization engine, which can then choose model parameters autonomously, e.g., based on the rough-set method. Applications to be built on this basis include an end-user

accessible frontend ("app" style) providing estimates for thermodynamic properties of fluids from molecular modelling. An end user will not need to possess any concrete knowledge of the molecular models.

The suitable molecular and mesoscopic model classes are data-informed physics-based models, or physics-informed data-based models. Models which are only data-based (correlations) or only physics-based (ab initio) cannot extrapolate and predict data reliably, which is typically the main requirement. The distinction between data-based and physics-based models, up to now quite rigid in EMMC/RoMM terminology, will need some nuance.

The European Materials Modelling Ontology needs to be supplemented to capture all aspects of modelling and simulation. The MODA Workflow Graph Language is an excellent point of departure; it permits the description of complex simulation workflows. So far, MODA has not attained the required level of formalization – it needs to be developed into an Ontology for Simulation, Modelling and Optimization. Beyond the ontological level of reasoning, a quantitative epistemological calculus, defined over modelling and simulation workflows, needs to be developed for model validation and uncertainty quantification.

Note (M.H.): Please consider also the statement on task-based parallelization and complex workflows. Concurrency in linking (rather than coupling) is an upcoming trend.

Discussion: So far top down, what about bottom up? Bottom up more complicated, because of different starting points with different scientific traditions Success stories? Medical field. Huge effort needed to put everything together, what can be done already now? There will be no end, i.e. there will be ongoing extensions, improvements. How can this become interesting to model providers, what do they gain? Providing the link between industrial end user and modeller.

References:

CEN Workshop Agreement 17284, Materials Modelling: Terminology, Classification, and Metadata, 2018; this document defines the MODA Workflow Graph Language as a European standard.
Slides on ontologies and metadata in VIMMP, Lausanne, 2018.
References on bespoke model design by multicriteria optimization:

K. Stöbener *et al.*, *Fluid Phase Equilibria* 373, 100-108, 2014.
K. Stöbener *et al.*, *Fluid Phase Equilibria* 411, 33-42, 2016.

- Marino Arroyo Small scale mechanics / continuum

Tension between “small scale” and continuum. Topic of the talk not centered on coupling between scales. Often continuum models work at an expectedly small scale (e.g. nanotubes)
Computing stress-tension from MD. How to do it systematically? Small effects may have large contribution to the system’s mechanics. How to do data-driven coarse-grained of MD.

Specific challenges of the science case. Important to understand how the cell-level property depend on the microstructural detail of the cell-internal architecture. Experiments can only give partial information to this regard. Connecting microscopic models with active gel models is crucial. At which scale should the experiments be performed in order to reconcile the two models? [Chugh et al., Nature Cell Biology 2017].

Science: Biological interfaces as mechano-chemical active materials. Example: actin cortex and epithelial mechanics. Epithelial layers cover every surface of living tissues. Mechanics depend on the cell-cell interfaces (adhesion complexes and actin). Epithelial sheet: a self-healing, ultra-stretchable material. Crucial: actin network on the cytoplasmic side of the plasma membrane. Cross-linked by other molecules. Under tension there will be actin flow and cell-shape changes. Elastic at small time scale, fluid at long time scale. Drives cell motion in confined environment.
Continuum description of the actin cortex as a viscoelastic material with non-uniform density. Many parameters in the model can be directly measured at the single-cell level. Cell locomotion under confinement can be modelled and compared to experiments.
Going to multi-cellular level: active gel model for the cortex at the lower level. Higher level: coarse-graining the model into a vertex model. And eventually homogenize the model into a continuum, out-of-equilibrium physics.

Model used to understand the mechanical behavior of epithelial monolayers. [Latorre et al., Nature 2018]. Cell seeded on micropatterned PDMS substrates with circular non adhesive regions first grow to fill the non-adhesive gaps, then grow into “domes”. These can be artificially engineered. The resulting structures under stress have interesting distribution of forces. Laplace’s law can been applied and the tension in the layer can be estimated. A tension/strain diagram can be built, revealing so-called “tensional buffering”. Cells strain very non-uniformly. Some of them stretch a lot, other not, although the tension in the dome is uniform. This is superelastic behavior reminiscent of nickel-titanium alloys (nucleation of high-strain phases into a largely strained region).
Model constructed to reproduce the experimentally observed “softening” behavior, namely the stretch-induced dilution of actin cortex after a critical strain is reached. This allows cell to extend their surface areas under constant tension.

Software: CG methods based on variational principles.

Vision: Biomimetic materials engineering. Cells and tissues are proper materials. Need to understand their behavior. Step towards controlling shape, rheology, dynamics. Development of functional bottom-up biological systems.
To actually *engineer* those materials, you need predictive, quantitative models and simulations across scales.

Policy: Working at the interface of disciplines may not always be *rewarding*. Problems: it is slow, have a less-structured audience. It is more difficult to recruit young researcher. There are less specific funding panels. These difficulties should be offset by active policies (at EU level).

Discussion: Can the used approach to the different level of granularity be extended to different systems? E.g. foams?

References:
Latorre et al., https://doi.org/10.1038/s41586-018-0671-4
Chugh et al., https://doi.org/10.1038/ncb3525
Fletcher, http://doi.org/10.1038/d41586-018-07291-3
Zhang et al., https://doi.org/10.1038/s41578-018-0034-7

- Matthew Borg Coupling md continuum, University of Edinburgh

Challenges: Many engineering systems are too expensive to be modelled explicitly, such as in micro/nano flow problems. Typical multiscale modelling challenges in this area is that the conventional fluid equations can no longer be applied, but you may have large length (5 orders of magnitude) and time (9 orders of magnitude) scales that need to be resolved, where particle-alone simulations are too intractable. Sequential methods (molecular, then continuum) may be very resource and time consuming for large variables. Concurrent simulation methods allows single-instance simulations: still problems of scalability. Solution: use an adaptive method based on machine learning.
Coupling in space: domain decomposition is very limited. Heterogeneous, equation-free methods [e.g. Borg et al, J. Comput. Phys. 2013] are better. For high-aspect ratio problems the Internal-flow Multiscale Method (IMM) is the best. Asynchronous (multi-model) time-stepping can help scale problems to engineering time scales, provided the physics is well-separated in time and the time steps are well-chosen.

Science: Many examples of micro/nano flow engineering: filtration membranes made of CNT, acoustothermal atomisation processes, droplets jumping in cooling devices, liquid slip over gas nanofilms, Knudsen pumps, etc.

Knudsen pumps move fluids without moving parts. Non-conventional, rarefied gas effect. Using multiscale methods results in huge savings of particles (300x) and timesteps (40000x). Thus: 10 Million times faster than conventional particle simulations.

Filtration membranes based on carbon NT. Measured flows much higher than the expected predictions. Why? Modelling to answer the question. Fluid is very structured inside the confined nanotubes. But also here, explicit MD methods cannot be used, the systems are simply too large. Concurrent MD simulation methods used for the Knudsen Pump did not work efficiently enough. But: use of IMM in sequential mode allows larger parametric space. Results: flow enhancement factors against membrane thickness, can now start to explain experiments and used in the design of this membranes.

Software (open-source): OpenFOAM (CFD solvers, dsmFoam+, mdFoam+), LAMMPS

Vision: Pushing discoveries into real products, enabled by multiscale flow modelling. Need to provide *scalable* code solution to industries. The molecular specificity should be preserved in the method. Protocols for sharing data and using them in ML approaches are needed. Hardware is important. Industries not always have large-scale computer facilities, cheaper multi-scale solutions need to be provided. Exploitation of the same models in other disciplines. Training at undergraduate, postgraduate level.

Policy: See: UK government report on “Computational modelling: Technological Futures” [2018]. Computational modelling is essential to our future productivity and competitiveness. Sharing approaches across communities. Need greater system integration and interaction with experimentation. Advanced sophisticated modelling and linking across scales is recognised as an important area.

References:

Multiscale Methods:

SY Docherty et al. (2016) International Journal of Heat and Mass Transfer, 98:712-727
Lockerby et al. (2015) Journal of Computational Physics, 284:261–272

Borg et al. (2018) Journal of Membrane Science, 567:115-126
Zhang et al. (2017) International Journal of Heat and Mass Transfer, 115:886-896

Software:

Longhsaw et al. (2018) Computer Physics Communications 224:1-21
White et al. (2018). Computer Physics Communications, 224:22-43

Science:

Ramisetti et al. (2017) Physical Review Fluids, 2:084003
Pillai et al. (2018) Physical Review Letters, 121:104502
Zhang et al. (2016) Langmuir, 32:1542–1549

- William Curtin Mechanics of materials atomistic/mesoscopic

Challenges: Metal mechanics: multiscale, multidiscipline field.
Microstructure lengths span over many orders of magnitude. Grains; dislocation lines; dislocation cores (atoms); nanophases distributed over scales of few 10 nm. No scale can be really ignored!
How do the material performance change if: (i) we change a few of the atoms; (ii) we process the material differently; (iii) we change the microstructure; (iv) the material is exposed to (cycling) loading. Models reliable enough to make decisions are needed, with some level of confidence.
Grand challenge in metallurgy: prediction of evolution of microstructure as a function of process.
Grand challenge in computational metallurgy: first-principles prediction of complex precipitate phases in situ in real multicomponent alloys. Quantitative prediction of properties.
Some basic properties can be computed with DFT. But the numbers do not tell how the material behave.
We need: predictive theories with descriptors that can be computed easily, in order to design and discover new metallic materials. Predictive theories involve some physical mechanism at the nanoscale. Such mechanisms may not be seen directly in experiments. How to validate them? How to test the theories?
Canonical solution: to "bridge scales". DFT, MD, Mesoscopic models (dislocation dynamics, phase field, coarsening, grain growth) up to continuum (crystal plasticity). Mesoscopic models: use fitted or unknown parameters; models may be poor.
Bottleneck: is at the level of MD, before going into Mesoscopic. Good interatomic potentials are scarce/missing. Exception: pure Fe, pure Al, pure Mg [Wu and Curtin, Nature 2015, low ductility of Mg due to dislocation "splitting"]. For binary/ternary alloys potentials are very bad, thus "predictive theories" are very bad.

Science: Hard materials. Easier, but maybe *harder* than fluids. Focus of the talk: mechanics of metals. Example: dislocations in Mg. Paradigm shift: towards DFT-driven Machine Learning potentials. Currently looking at Al-Mg-Si alloys [with M. Ceriotti]. Potentials need to reproduce any feature (at a small scale) that can be computed with DFT. E.g. precipitate energetics, stacking-fault energy, misfit strains. Not fitted, but very well reproduced: vacancy migration path barriers. Smallest precipitates: cannot be accessed experimentally!
Other example: coupled atomistic/discrete simulation in 3D. Full atomistic at the dislocation core; continuum dislocation dynamics far from the core.

Vision: In the future there will be only ML potentials for metal alloys, no more EAM etc.
Link to reach industrial relevance: possible with *good* potentials.

Policy: Industries spend huge amount of human time trying to avoid doing computations. Larger companies start doing much large-scale computations.

- Ingo Steinbach Phase-field modelling and simulation

Challenges: Needed are: high computer resources, memory, parallelization. Also needed: constitutive relations and model data, e.g. from atomistics. Mesoscopic full-field models need the maximum number of inputs of all models at other scales. Calphad databases are very important as an input!
Limitations: Regarding nucleation, it has to be incorporated by separate models. Physically speaking it is determined by the interplay between "deterministics" and "statistics". Regarding system size we can treat hundred thousands of phase-field (objects, phases, grains) for simple cases (e.g. grain growth) down to a view phase-fields for complex, considering anisotropic materials, multicomponent diffusion, fluid flow, stress-strain and complex morphologies (dendrites). Topics of future importance: In a concurrent multiscale setting, use AI to run simulations only "where and when" needed. Use PF simulations to train meanfield models. Run simulations to test hypothesis. E.g. in order to discriminate between different mechanisms acting simultaneously.

Science: Phase-field: indicator function in space and time characterizing the distribution and evolution of a phase. Order parameter: corresponds to the fraction of a phase. Moving boundary problem. Joung (started early 1990th), but growing field, more than 250 publication per year nowadays. Build up with contributions of bulk and boundaries. Bulk contribution may be estimated with Calphad methods, continuum micromechanics, fluid flow, diffusive and advective transport. The model is intrinsically thermodynamically consistent if treated right). Needs much input to feed constitutive relations. Atomistic simulations are welcome for all kinds of data.

Example: microstructure prediction in steel (martensite formation). Tempering simulation, nucleation of carbides (in 20MnCr8_4 steel). Crystal plasticity with concentration dependency parameter. Important point: critical strains for plasticity are made dependent on local composition, not the nominal composition of the material. By this the changing composition during the process history can be considered. Prediction of stress-strain curve, especially in the plasticity region, as a function of tempering time and temperature, with different carbon contents.

Software: OpenPhase, www.OpenPhase.de, open source. Professional support: OpenPhase solutions GmbH www.OpenPhase-solutions.com

Vision: Virtual production to virtual testing in "full field". Prediction of time-temperature-transformation curves. Keeping track of the material history under thermo-mechanical load.

Policy: Making PF a standard engineering tool, as FE is in structure mechanics. It is the "best" tool to investigate evolving microstructures. Establish best-practice examples and benchmarks. Integrate PF into graduate courses in materials science. Support industrial application by providing well-supported and maintained software solutions.

Discussion: Application of PF to additive manufacturing simulations? Yes, although taking care of cooling gradients during solidification is a difficult problem to tackle. In principle it is treatable.

References:
L.-Q. Chen, Phase-field models for microstructure evolution. Annu. Rev. Mater. Res. 32:113–40 (2002)

I. Steinbach, Phase-field models in materials science. Model. Simul. Mater. Sci. Eng. 17:073001 (2009)
I. Steinbach, Phase-Field Model for Microstructure Evolution at the Mesoscopic Scale, Annu. Rev. Mater. Res. 43:89-107 (2013)

- Kersti Hermansson Electronic structure

Challenges: Three challenges are emphasized (1) interactions (and their accuracy), (2) the structure, and (3) diversity.

Science: Uses of Electronic Structure (ES) calculations emphasized in the talk:

1. ES to train other models along multiscale ladder
2. ES to understand chemical reactions in active materials (electron transfer)
3. ES for material characterization (of interest in industry)

Regarding (1), approximate DFT methods such as DFTB are about 300 times faster than DFT, whereas reactive FF trained from DFT methods, such as ReaxFF, are about 50 000 times faster than DFT. This cross-talk between methods consists or parametrization of simpler models and feedback to check accuracy against more fundamental models.

Examples (e.g. relative stability between nanoparticle configurations) illustrate very good, even quantitative, agreement between DFT and DFTB. Connection between DFTB to ReaxFF is not so successful. Two approaches to overcome this difficulty are suggested: (a) combination of DFTB and ReaxFF and (b) use AI-determined FFs.

Regarding 2 and 3, a set of examples are presented where modelling meets experiments. One example involves the mechanisms by which oxygen storage depends on the size of nanoparticles. A second example concerns the characterization of materials using STM, where ES calculations interpret dark spots as fluorine impurities rather than oxygen vacancies.

Software: DFTB+

Vision: AI will solve some of the challenges in the field but physical insight will continue to be important. A past and a future conference are mentioned focusing on the interaction between physics-based and data-driven. [please, add details about these conferences]

Policy: Importance of recognition of method/model/software development (research engineers).
Stop overselling computational methods. Promote interest in discrete models, particularly electronic structure and atomistics.

- Joerg Neugebauer Electronic structure and computational metallurgy

Challenges: In the context of electronic structure in the field of computational metallurgy, the overall goal is to use QM + thermodynamics to overcome material challenges using automated tools. A practical challenge is to combine many different computational tools in complex simulation protocols.

Science: As an example of challenges in metallic alloys, subjected to high temperatures, corrosive environments, high loads, and how to overcome the inverse relation between strength and ductility. One way to proceed is by controlling the material structure at the nanoscale since impurities, phase boundaries, precipitates, etc., influence the motion of dislocations.

Three approaches are emphasized to connect QM with larger scales: (1) Integrated computational materials science (multiscale ladder of models), (2) Digital twin, and (3) Ab initio-computed descriptors. The talk focuses on the latter.

One example of valuable ab-initio computed descriptor are stacking fault energies (as a function of chemical composition, temperature, stress), which is a key parameter for material performance. For super-alloys, large scatter of experimental data at low temperature, lack of data at operational temperatures.

To provide ab initio computed descriptors, several challenges are highlighted, including accuracy (xc functional, convergence of thermodynamic averages, including the relevant physics) and structural complexity. The talk focuses on accuracy, and more specifically on xc functionals (anharmonicity through thermodynamic integration, various couplings such as magnon-phonon interactions, etc). Examples are shown about heat capacity calculations, particularly at high temperature.

The case of multicomponent materials with high configurational entropy is also discussed, in which the number of possible interactions explodes and AI approaches based on moment-tensor potentials becomes advantageous. Application to design of ductile magnesium, where stacking fault energies represent a useful ab initio computed descriptor for ductility.

Software: http://www.pyiron.org

Vision: Establish platforms to develop, share, apply the individual tools as well as complex simulation protocols to compute finite temperature materials properties and phase diagrams.
Use high-precision ab initio data for finite temperature materials properties to benchmark/improve ab initio approaches.

Policy: How to make it attractive for young researchers to make not just exciting physics but to develop tools, databases, user interfaces, documentation? courses in universities, extend/adjust scientific evaluation qualifiers. Simulation protocols, databases, big data applications need a certain degree of standardization. Need to strike a balance between administrative restrictions and research freedom. Important to keep healthy diversity of tools, platforms.

Discussion: the discussion focuses on the assessment of quality of different xc functionals.

- Eric Cancès Solvation, multiscale

Challenges: error estimation for molecular simulations / foster multidisciplinary interactions in the field of molecular simulations

New implementations of implicit solvent models based on domain decomposition are extremely inexpensive in terms of computational time. They are the only tractable option in many cases and most often gives quite accurate results, but also have limitations: they dramatically fail in some cases. Progress in this field can be made by 1) improving existing implicit solvent models and 2) design a posteriori criteria allowing to assess the quality of implicit solvent calculations.

Software: http://www.ddpcm.org (LGPL license)

Vision: The presented work has been done in collaboration between mathematicians and chemists. On a more general level, molecular simulation is an inexhaustible source of possible multidisciplinary collaborations between chemists, physicists, mathematicians and computer scientists.

Policy: actions to foster multidisciplinary work: co-supervisions, postdoc exchanges, joint appointments. Multidisciplinary training: modelling courses in applied math, long schools, multidisciplinary projects. Define a roadmap for computational results verification and validation (very relevant for EMMC).

Discussion: It would be important to include non-homogeneous density distribution in the continuum region. It would be good to include solvent molecules within the cavity. Maybe the biggest error is the modelling error, but numerical errors can also significantly affect the results. In many cases, the modelling error is simply unknown: for example, in the coarse-graining, the main point is which variables to throw away. The modelling error can be estimated by comparing different degree of coarse-graining with a reference model. In the special case of quantum chemistry (e.g. CASSCF, coupled clusters…), modelling errors with respect to the reference N-body Schrödinger equation can in principle be estimated since the computed approximate N-body wavefunction can be used to build a residual, which is a key object to design a posteriori error estimators.

- Francisco Chinesta Twins

Challenges: Digital twin: combining offline models and online data collection, enriched by AI

Software: ESI (https://www.esi-group.com/)

Vision: Real-time simulations by model order reduction. Generating surrogate models by snapshots obtained by effective simulations and then using the surrogate models for estimating error propagation and sensitivity analysis. Hybrid twin by ignorance learner and modeller. The divergence model by additive learning.

Policy: Focus on real-time. Data-enriched models are important. Hybridization by combining physics-based models and data-based models. Smart-data paradigm: collect only appropriate data (smart selection).

Discussion: Importance of hybridizing data with models, because too data are dangerous and expensive. Data can enrich models but they cannot substitute them.

References:
https://link.springer.com/article/10.1007/s00161-018-0677-z

- Christoph Ortner, Mathematical multiscale, University of Warwick

Challenges: Modelling and approximation errors in crystal defect simulations.

Science: Mathematical formalisation of errors due to boundary conditions and model approximation in simulating material defects. Examples: point defects dislocation, CLE boundary conditions, flexible boundary conditions, A/C coupling, QM/MM coupling. Examples of a mathematical contributions: (1) For dislocations, the linearisation error is larger than the atomistic-continuum approximation error, therefore flexible bc and Green's function methods give no (asymptotic) improvement over clamped CLE bc. (2) Energy-mixing schemes for QM/MM coupling. Decomposition of QM energy based on the BOP idea, construct MM model in direct approximation to the QM site energy. This strategy allows QM/MM coupling with controllable errors at the interface, because artifacts have been removed. Higher order expansion allows to increase the order of convergence of the method. (3) Machine learning of interatomic potentials: combine traditional modelling approaches (e.g. many-body expansion) with data-driven learning. Main message: IP construction is an approximation problem, not a machine learning problem.

Vision: Mathematical error analysis can identify bottlenecks, inconsistencies, and optimise model and algorithm design to complement materials modelling. The main challenge for mathematics is that atomistic modelling is a tiny niche in applied mathematics, e.g., CFD/DFT > 100, this severely limits the (potentially significant) impact that applied mathematics can (should?) have in this field.

Policy: (1) Major experimental projects typically require participation by modellers. Similarly, major modelling/theory projects should aim to involve mathematicians.
(2) The current hype/trend around data-driven modelling and machine learning can be exploited by revisiting all the traditional model approximation steps, e.g., QM -> DFT -> TB -> IP -> meso -> macro, but treating these as *approximation problems* rather than modelling problems, keeping the physical insight but complementing it with mathematical analysis and data-driven approximation to fill in the gaps.

Discussion: Interdisciplinary training is required, this is very difficult to set up. Any earlier than PhD level is unrealistic and maybe not even healthy. Even at PhD level it is a challenge. A realistic approach may be to train students in their core discipline as well as in the "art" of cross-discipline communication.

References: http://dx.doi.org/10.1007/s00205-016-1019-6,
http://dx.doi.org/10.1137/15M1041250,
http://arxiv.org/abs/1710.07708,

- Matej Praprotnik Coarse graining between MD and DPD

Science: Adaptive resolution scheme (AdResS): partition of the simulated system into coarse-grained / hybrid / atomistic regions. AdResS allows for dynamical switching between atomistic and coarse-grained MD. Larger length and time scales can be achieved because of the speed-up.

Supramolecular mapping: multiple water molecules mapped to 1 CG bead (e.g. 4 to 1 mapping with MARTINI force-field). SWINGER algorithm: concurrently assembles, dissembles and reassembles water clusters
Supramolecular coupling of atomistic water with DPD. Applications: coupling to CFD, DNA molecule in ionic solutions with explicit ions (open boundary MD in grand canonical framework) and hybrid explicit/implicit water.

Software: Espresso++

Vision:
1) efficient and versatile compute code
2) open simulations that exchange matter and energy with environment
3) nonequilibrium situations (fluid flows)
4) supramolecular bottom-up coarse-graining (structure and dynamics)
5) extension to multicomponent liquid mixtures
6) nonequilibrium and date-driven coarse-graining
7) training of next generation of researches, which will need to be familiar with different modeling approaches from different disciplines

Policy: The multiscale modeling will be used to solve relevant materials problems using multiscale simulations for instance to guide design of novel smart nanomaterials. These advanced materials are highly sophisticated structures with defined materials properties that can be tailored based on multiscale modeling and simulations to reach desired functionality more efficiently and faster. In this respect, the major theoretical challenge is to model in an unified multiscale framework such complex molecular systems over a range of time and length scales that spam over several orders of magnitude. Nowadays, the majority of complex materials modeling is performed at separate length scales. Multiscale approaches ranging from atomistic, mesoscopic to continuum models, will allow the exploration of full range of material properties and their hierarchical architectures at all relevant spatial and temporal scales.
Focusing now on concurrent coupling of atomistic and mesoscopic hydrodynamics via linking of molecular dynamics (MD) and dissipative particle dynamics (DPD), we will be able to bridge atomistic and mesoscopic length scales, which is especially useful for simulations of the transport of nanoparticles through fluids. Computer simulations can provide insight into such systems when they can access, both, the atomistic length scales associated with size of the nanoparticles and the micro/macro scales characteristic of the carrier flow field. MD simulations can capture the atomistic details of the nanoparticle-liquid interface but due to their computational cost they cannot be extended, in the foreseeable future, to the macroscale regime of the full flow field. In turn, mesoscopic approaches using the DPD equations may capture the mesoscale behavior of the flow but they fail to represent accurately the flow field at the nanoparticle surface. The hybrid approaches, on the other hand, combine the powerful features of the both descriptions, i.e., the ability to describe the mesoscale behavior of the flow as well accurate boundary conditions around nanoparticles. They allow us to perform efficient multiscale simulations of fluid flows around nanoparticles at the same accuracy as the single-scale microscopic MD simulations but much faster.

References:
J. Zavadlav, M. Praprotnik, J. Chem. Phys. 147, 114110 (2017).
J. Zavadlav, S.J. Marrink, M. Praprotnik, J. Chem. Theory Comput. 12, 4138 (2016).
R. Delgado-Buscalioni, J. Sablić, M. Praprotnik, Eur. Phys. J. Spec. Top. 224, 2331 (2015).

J. Zavadlav, J. Sablić, R. Podgornik, M. Praprotnik, Biophys.J. 114, 2352 (2017).
L. Delle Site, M. Praprotnik, Phys. Rep. 693, 1-56, (2017).
M. Praprotnik, L. Delle Site, K. Kremer, Annu Rev Phys Chem 59, 545 (2008).

- Gianluigi Rozza Reduced basis methods

Challenges: Reduced Order Methods for PDEs: state of the art and perspectives with applications in Industry, Medicine and Environmental Sciences. The state of the art of Reduced Order Methods (ROM) for parametric Partial Differential Equations (PDEs) with a special interest in parametric problems arising in offline-online Computational Fluid Dynamics (CFD) is characterized by moderate and low Reynolds numbers and intrusive methods, as well as too big parameter spaces.

Science: Efficient parametrizations (random inputs, geometry, and physics) are very important to be able to properly address an offline-online decoupling of the computational procedures and to allow competitive computational performances. Current ROM developments in CFD include: a better use of stable high fidelity methods, considering also spectral element method and finite volume discretization, to enhance the quality of the reduced model too; more efficient sampling techniques to reduce the number of the basic functions, retained as snapshots, as well as the dimension of online systems; the improvements of the certification of accuracy based on residual based error bounds and of the stability factors, as well as the guarantee of the stability of the approximation with proper space enrichments. For nonlinear systems, also the investigation on bifurcations of parametric solutions are crucial and they may be obtained thanks to a reduced eigenvalue analysis of the linearised operator. All the previous aspects are very important in CFD problems to focus in real time on complex parametric industrial, environmental and biomedical flow problems, or even in a control flow setting, including uncertainty quantifications.
A further important task is the reduction in parameter space to eliminate parameters with very low sensitivity (active subspaces).
Model flow problems are also being extended to multiphysics, such as fluid-structure interaction problems. Significant examples of applications are related with shape optimisation applied to industrial problems in aero-naval-mechanical engineering, for example.

Software: mathlab.sissa.it/cse-software

Vision: Offline-online computing, database of parametric solutions, from supercomputers to tablets/ laptop/smartphone. The glue is a computational webserver for “CSE web app”.

Policy: better integration between HPC and reduced order methods, Digital Twin, HPC, data science, SMACT (Social-Mobile-Analytics-Cloud-Things), machine learning.

Discussion: Intrusive vs non-intrusive approaches in reduced order modelling, better integration between design and research teams in industry, portability of the methodology in several fields: industry, applied sciences, medicine, dealing with more and more complex systems including uncertainty quantification.

References:
J.Hesthaven, G.Rozza, B.Stamm: “Certified reduced basis methods for parametrized PDEs”, Springer Briefs in Mathematics, 2016.

A. Quarteroni, G. Rozza "Reduced order methods for modeling and computational reduction", Springer MS&A 2014, Vol. 9.
P. Benner, M. Ohlberger, A.T. Patera, G. Rozza, K. Urban." Model Reduction for Parametrized Systems", Springer MS&A, vol. 17, 2017.

3. Conclusions

1 - Software as scientific infrastructure needs to be put at the center of the scientific activity. Interoperability of different software elements is central and need to remain flexible for extensions.
2 - Quality assurance: there needs to be verification and validation, and error estimates/bars
3 - Physics and data driven models are both needed: an example is in atomistic simulation, where data-driven potentials improve physics-driven potentials. More AI-enabled decision making.
4 - High-level goals: we need bespoke/goal-oriented models, test-stressed against multi criteria optimization.
5 - Emerging mathematics. Algorithm development remains essential: reduced-order models for DFT, low-rank tensor methods, domain decomposition, time decomposition, sampling, rare-events handling.

4. References

In text above

Authors	Nicola Marzari (EPFL), Pietro Asinari (POLITO)
Contributing EMMC-CSA partners	EPFL, POLITO, UU and all partners

EC-Grant Agreement	723867
Project acronym	EMMC-CSA
Project title	European Materials Modelling Council - Network to capitalize on strong European position in materials modelling and to allow industry to reap the benefits
Instrument	CSA
Programme	HORIZON 2020
Client	European Commission
Start date of project	01 September 2016
Duration	36 months

Consortium		
TU WIEN	Technische Universität Wien	Austria
FRAUNHOFER	Fraunhofer Gesellschaft	Germany
GCL	Goldbeck Consulting Limited	United Kingdom
POLITO	Politecnico di Torino	Italy
UU	Uppsala Universitet	Sweden
DOW	Dow Benelux B.V.	Netherlands
EPFL	Ecole Polytechnique Federale de Lausanne	Switzerland
DPI	Dutch Polymer Institute	Netherlands
SINTEF	SINTEF AS	Norway
ACCESS e.V.	ACCESS e.V.	Germany
HZG	Helmholtz-Zentrum Geesthacht Zentrum für Material- und Küstenforschung GMBH	Germany
MDS	Materials Design S.A.R.L	France
Synopsys / QW	Synopsys (former QuantumWise A/S)	Denmark
GRANTA	Granta Design LTD	United Kingdom
UOY	University of York	United Kingdom
SINTEF	SINTEF AS	Norway
UNIBO	ALMA MATER STUDIORUM – UNIVERSITA DI BOLOGNA	Italy
SYNOPSYS	Synopsys Denmark ApS	Denmark

Coordinator – Administrative information	
Project coordinator name	Nadja ADAMOVIC
Project coordinator organization name	TU WIEN
Address	TU WIEN | E366 ISAS | Gusshausstr. 27-29 | 1040 Vienna | Austria
Email	nadja.adamovic@tuwien.ac.at emmc-info@tuwien.ac.at
Project web-sites & other access points	https://emmc.info/