SimBest PMMM Workshop Review

This document is a summary of presentations given at the Predictive Multiscale Materials Modelling Workshop at the Isaac Newton Institute, Cambridge (1st-4th December 2015). The 'relevant areas' tags and bullet points following each review reference general themes and specific industry challenges as outlined in the 'KTN Themes' document.

'Introduction to Expectation Propagation' – Tom Minka, Microsoft Research

Tom discusses his expectation propagation algorithm which he designed as part of his PhD[1]. Expectation propagation is a type of message passing algorithm, and is a fast approximation algorithm for Bayesian inference. It is used to predict the output of models with a given certainty in the results produced. Expectation propagation is well suited to high dimensional problems and distributed computation, by the introduction of latent variables.

Factor graphs are also introduced as a technique to visualise the model and parameters. This graphical representation can be applied to highly coupled systems as an intuitive representation of its dependent and independent parameters.

Some of the more recent developments of the expectation propagation algorithm include improving its predictive accuracy by model conditioning:

- using latent Gaussian models[2]
- using learning to initialise message passing[3–5]

and probabilistic back propagation using neural networks[6].

Relevant areas: *machine learning, uncertainty quantification, highly coupled systems, high performance computing*

- 6. -> Factor graphs as a graphical representation of parameters in model
  - Could be used in the visualisation of design parameters in highly coupled systems to see which are dependant/independent.
- 13. -> expectation propagation algorithm is a message passing algorithms which requires multistage inference (has no ability to reject outliers in sample data)
- F. -> for parallel processing of expectation propagation algorithm, introduce latent variables.
- E. -> probabilistic backpropagation of EP algorithm (list papers)
- 7. -> Expectation propagation algorithm is a fast approximation algorithm for Bayesian inference and is especially well suited to high-dimensional problems due to its easy and efficient parallelisation.

'Adaptive Langevin Algorithms for Canonical Sampling with Noisy Forces in Scale-bridging Molecular Dynamics' - Ben Leimkuhler, University of Edinburgh

Thermostats used in molecular dynamics simulations, such as the Langevin dynamics algorithms, are discussed and their statistics are compared for simulations with white noise. It is found that traditional 'highly accurate' schemes[7,8] fail to correctly sample the phase-space and hence produce unrepresentative results.

To accurately model systems with white noise, an adaptive Langevin dynamics algorithm is proposed[9]. First the system is discretised using the Leimkuhler-Matthews method[9] which is shown to approximate paths with white noise more effectively than Euler-Maruyama. Next, splitting methods for Langevin dynamics are introduced and a new 'BADODAB' integrator[10] is shown to most accurately represent the system, i.e. gives the smallest force error. Bayesian logistic regression is used to obtain the error estimates.

These adaptive thermostats have been used in the modelling of metals[9] (Drude models), in multiscale modelling when bridging QM-MM (quantum mechanics/molecular mechanics) simulations[11] and in machine learning[12].

The future applications for these covariance controlled adaptive Langevin dynamics are:

- applications of these methods in non-equilibrium modelling
- investigation of the accuracy vs. exploration trade-off in rare event modelling
- in the treatment of coloured noise.

Relevant areas: *machine learning, thermostating, multiscale modelling and validation, materials modelling, error estimation.*

- 13/e. -> Using Bayesian regression, error estimates for the interatomic forces can be obtained for systems with white noise using the adaptive Langevin dynamics proposed. The error in the result is shown to be smaller than those produced by Google's stochastic gradient thermostat[13].
- 11/c. -> The adaptive Langevin dynamics technique is implemented in QM-MM multiscale software[11].
- 4. -> The adaptive Langevin dynamics scheme proposed retains the superconvergence property present in previous Langevin schemes[10], which is a desirable property in highly integrative modelling.

'Enhanced free energy based structure prediction in materials science' – Mark Tuckerman, Courant Institute, NYU

Recent developments in the materials modelling which use enhanced free energy based methodologies to predict structure and polymorphism in molecular crystals[14] are discussed. These new methodologies can also be applied to oligopeptides[15] to determine the conformational equilibria.

Two main techniques are discussed:

- Heterogeneous multiscale modelling[16,17] (used to identify critical points in the free energy landscape, termed the START technique[18], "Stochastic Activation Relaxation Technique")
- Temperature-accelerated methods[19,20] (used to calculate relative/adiabatic free energies of the critical points).

The effectiveness of these techniques are demonstrated by successfully predicting structures in the CDC 6th blind structure prediction test. By correctly interrogating the structures' energy landscape[21] using adiabatic free energy dynamics simulations. The correct polymorphs of a variety of compounds used in the pharmaceutical industry are also discussed and predicted using this method.

Ongoing challenges include the accurate prediction of:

- Highly flexible molecules (requires 7. -> high dimensional optimisation)
- High temperature/pressure polymorphs (requires d. -> multiphysics research)
- Complex geometries, i.e. stacking faults, screw axes, mixed phases (requires 1. -> modelling of complex structures)
- Isotropic polymorphism (requires 4. -> highly integrative modelling)

Relevant areas: *multiscale modelling, chemical mechanical materials modelling, multiscale validation, optimisation.*

- C. -> heterogeneous multiscale modelling is used to obtain the free energy surface. Sampling of the thermodynamic force instead of the probability density allows for fast generation of the free energy hyperspace[17].
- A. -> Structures in the CDC 6th blind structure prediction test are accurately reproduced using adiabatic free energy (temperature-accelerated molecular) dynamics.
- B. -> The new force sampling approach allows for rapid free energy surface reconstruction[17].

'Rare Events Methods, Reaction Coordinates, and Useful Rate Theories' - Baron Peters, University of California Santa Barbara

Rare events are of interest to multiscale modelling as they occur at small length scale and infrequently (i.e timescales are large), this means they lie off the diagonal on the standard length vs, timescale graph which frequently appears in discussions of multiscale modelling. The most common approach to studying rare events is to study the free energy as a function of some reaction coordinate[22].

The reaction coordinate can contain some of the most useful information about a chemical mechanism. It is viewed as a summary of the mechanism and holds information on kinetic trends and rates, the thermodynamics of the system and information on manifold which can be used in dimensionality reduction. If the reaction coordinate is not obvious, a committor is used to accurately sample the space[23,24].

Identifying the committor is computationally costly, required a large number of simulations to get a large enough sample[24]. Four new approaches to optimise the search for the correct reaction coordinate are to use:

- Spectral theory (i.e. Markov state models, diffusion maps)
- Transition path theory (i.e. Backward Kolmogorov models)
- Variational principles (i.e. Berezhkovskii-Szabo method)
- Statistical inference (i.e. likelihood maximization, neural networks)

Each method works to describe complicated high dimensional dynamics by a scalar measure of progress along the reaction pathway.

To extract the mechanistic insight from these methods is incredibly difficult. The likelihood maximization approach (i.e. a type of statistical inference) is proposed as the most promising direction to investigate in order to be able to retrieve this mechanistic description from the reaction coordinate[25,26].

Relevant areas: *High dimensional optimisation, multi-objective optimisation, uncertainty quantification research, chemical mechanical materials modelling.*

- E. -> The error in the reaction coordinate can be calculated using the costly histogram approach[24], or more favourably using statistical inference to quantify the uncertainty.
- 7. -> Four techniques which implement dimensionality reduction of the reaction parameter space are discussed. Each uses a scalar value to represent progress of each high dimensional set of variables.
- 3. -> The maximum likelihood approach is proposed as a potential solution to bridging the mechanistic-chemical description gap[26].

'Stochastic simulation as a basis for optimizing microstructural characterization protocols' - Lori Graham-Brady, Johns Hopkins University

Work funded by the United States Airforce, Office of Scientific Research (a centre of excellence for integrated materials modelling) is presented – *as a result, only one research paper is cited in the presentation and little detail of the optimisation techniques are given*. The Airforce lab work in the three dimensional characterisation of materials experimentally, bridging the micro and macro scales (in the continuum model). In order to glean information about a material, it is repeatedly polished and scanned to build up a layer-by-layer picture. However, the nature of this experimental work is destructive. The uncertainty introduced by each stage of the workflow is discussed, with emphasis on the uncertainty in the resultant digital model which is created for the material.

Microstructure simulations are used to conduct sensitivity studies on particular properties or responses of interest. Polycrystal structure growth is investigated by partitioning the crystal into smaller regions, showing that the assumption of ellipsoidal growth captures features of the grains much better than using traditional Voronoi tessellations[27].

If the uncertainty in the computational models can be quantified, there is the potential to greatly optimise the data collection parameters. This is a constrained optimisation problem. An optimisation framework has been proposed to minimise the cost of data collection; an initial set of data is collected via the polishing and scanning of the sample, then the data is run through an optimisation algorithm which selects the cheapest modelling technique within the specified error tolerance.

Relevant areas: *uncertainty quantification for robust design, multi-fidelity simulations, multi-scale modelling, data, optimisation.*

- 10. -> A mesh is introduced to bridge the microscale grains to the macroscale representation of the polycrystal – *explicit details of how this is done are not given*.
- 13. -> The 'Digital Material Workflow' used by the Airforce contains multiple sources of uncertainty. The selection of experimental data is a constrained optimisation problem, the mechanical modelling/testing a forward problem and the initial characterisation of a material is an inverse problem.
- B. -> The Airforce are using an optimisation algorithm to select the least costly computational model in their simulations. This is highly advantageous as there are huge volumes of experimental data collected, which need not be scrupulously analysed as a result of this optimisation.

'Building Probabilistic Models One Constraint at a Time' - Roger Ghanem, University of Southern California

The application of uncertainty quantification to stochastic models is discussed. The importance of understanding the relationship between predicted/expected values and currently available information is stressed, due to the random fluctuations and nature of stochastic systems[28]. Areas for error reduction are highlighted: these are in the numerics, statistics, data and models. It is noted that models should be built with an 'error budget' in mind to correctly capture outliers and features in the data. The uncertainties in the numerics, data and statistics place limits on the predictability of a model and so are of highest importance when investigating a system.

The best practice method for constructing of probabilistic models is discussed and common statistical constraints applied to the model are highlighted (i.e central limit theorem, maximum likelihood, Bayes rule.)[29].

Polynomial chaos, a method which was not developed to deal with probabilistic uncertainties, is in fact proposed as a robust and flexible method to represent them[30]. Coupled systems with white noise with the polynomial chaos description applied are discussed[31]. A method to validate the 'up-scaling' or coarse graining of multiscale stochastic models using quadrature is described[31].

Relevant areas: *uncertainty quantification, high-dimensional optimisation, high performance computing, highly coupled systems, multi-scale modelling.*

- 13. -> The best practise method for the construction of a probabilistic model is discussed and key pitfalls are highlighted (i.e. pay attention to the features of the data so as not to select a model which wrongly smooths these out in its representation)[28].
- 6/7. -> Dimensionality reduction of high-dimensional coupled problems is introduced[31].

'Predictive Coarse-Graining' - Phaedon-Stelios Koutsourelakis, Technische Universität München

Coarse grained models have been developed to ease the computational strain of small time-step molecular dynamics simulations with a large number of degrees of freedom. Prosed is a method to capture the uncertainty arising from information loss in the coarse graining process, and to provide probabilistic predictions of the final fine-scale state.

Unlike previously proposed methods[32,33], a probabilistic mapping from coarse to fine is prescribed in the form of an effective (coarse grained) potential. Using a fully Bayesian treatment, i.e. Gaussian process regression, the uncertainties in the mapping and measure the 'information loss' can be calculated.

To reconstruct the fine scale probabilistic prediction[34], machine learning, i.e. Bayesian inference is used[35]. The proposed probabilistic prediction method is then applied to a simple Ising model as an example to demonstrate its predictive power in the reconstruction of fine scale predictions.

The training data used to inform the posterior distribution (i.e. effective CG potential) is discussed, and the effect of larger training sets and sampling techniques[36] on the fit of the prediction and its error shown. The end-to-end process will ultimately comment of the choice of effective CG potential

selected when defining the fine to coarse mapping. This can then allow for a refinement in the effective CG potential used.

Relevant areas: *multi-scale modelling, high-fidelity simulations, uncertainty quantification, multiscale validation.*

- E. -> the fully Bayesian treatment of a bridging between multiscale models gives the uncertainties in the 'information loss' as a result of the mapping (this is almost like the different between sizes/effectiveness of meshes used on each scale).
- 11. -> the resultant uncertainties comment of the original choice of mapping between scales and so this proposed method can be used to validate the model, where a minimal uncertainty implies the best mapping to bridge scales has been found.
- C. -> The simple Ising model example demonstrates how to correctly apply this new multiscale modelling method.

'Canonical Quantum Observables Approximated by Molecular Dynamics for Matrix Valued Potentials' - Anders Szepessy, KTH Royal Institute of Technology

Historically molecular dynamics simulations are believed to give a good approximation of a system in the canonical ensemble (i.e. temperature is fixed) for specific conditions relating to the temperature and electron potential. If the temperature is small compared to the difference of the second and first eigenvalues of the electron potential, the simulation is said to be a good one.

Presented is a proof that the simulations are in fact a good fit for the canonical ensemble at *all* temperatures[37].

Relevant areas: *materials modelling.*

- A. -> A proof is provided that for systems in the canonical ensemble, molecular dynamics simulations can correctly approximate the quantum observables, irrespective of the temperature of the system[37].

'Disordered Hyperuniform Point Patterns in Physics, Mathematics and Biology' - Salvatore Torquato, Princeton University

Disordered hyper-uniform many-particle systems are considered to be a new state of disordered matter[38]. These systems possess both the properties of a crystal (they supress density fluctuations) and properties associated with fluids (they are statistically isotropic). Examples of such disordered states of matter in physics, mathematics and biology are discussed.

The critical phenomena which dictates the transition of a disordered hyper-uniform system to an ordered state is outlined[39] and the initial creation of such disordered states[40] is discussed.

These systems are becoming increasingly relevant, with some examples of their recent applications being (there are many more in the talk, with citations to each paper):

- In the design of photonic metamaterials with large complete (both polarizations and all directions) band gaps[41]
- The same type of irregular structure as the photoreceptors found in the eyes of birds[42]
- In the description of periodically driven emulsions[43].

Relevant areas: *complex fluid and solid material interaction, multiphysics-modelling.*

- D. -> The description of the construction[40] of these disordered hyperuniform systems in important as they are being increasing found in many areas of physics. This mathematical formulation the formation of these systems will allow for computational modelling of them - it becomes possible to write code to emulate the mathematics giving rise to disordered hyper-uniform structures.
- 2. -> The nature of these structures (possessing both fluid-like and solid-like) qualities may help in the understanding of complex fluid and solid material interaction.

'Virtual Materials Testing' - Karel Matouš, University of Notre Dame

Presented is a framework to model heterogeneous materials on a variety of length and time scales whilst preserving information about its chemo-thermo-mechanical behaviour. The framework is implemented as a highly optimised, hierarchically parallel high-performance simulation which runs on thousands of cores.

The mathematical formulation of the multiscale model is presented, with a description of the treatment of interfaces between scales[44]. The Hill-Mandel Lemma which places boundary conditions on a system to ensure continuity in the energies of each scale for a system under stress or strain is highlighted. The bridge from micro to macro scale uses a nonlinear manifold-based reduced order model as a map from the micro to macro scale.

The framework proposed (using the PGFem$_3$D solver[45]) scales almost linearly with the number of cores used, which implies the model is highly optimised. The PGFem$_3$D solver structure is explained and its hierarchically parallel nature detailed in pictorially[45].

Relevant areas: *multi-scale modelling, multiscale validation, high-dimensional optimisation, high performance computing, chemical mechanical materials modelling.*

- 3. -> The multiscale framework for modelling the chemo-thermo-mechanical behaviours of heterogeneous materials across many length and timescales is presented.
- C. -> The construction of a model to simulate multiscale cohesive failure in heterogeneous materials is discussed, with reference to the speaker's published work on the subject[44,46].
- F/7. -> The PGFem$_3$D solver and discussion of its construction is an example of a highly optimised, high-dimensional multi-scale implementation[45].

'Wavelet Scattering Regression of Quantum Chemical Energies' - Stéphane Mallat, École Normale Supériere

The problem of quantum chemistry energy regression is addressed, and comments on the more general themes of learning in physical modelling. The aim of the talk is to build a linear regression model to minimise the difference in energies between the model and observed data[47]. This is a highly desirable technique for use in molecular modelling, with the majority of applications in the pharmaceutical industry.

Questions arise around how this regression should be performed algorithmically and mathematically for these systems which have high degrees of dimensionality – i.e. the frequently appearing 'curse of dimensionality'. To solve this problem, invariants in the physics are leveraged and the particles represented as wavelets instead of as points[48]. The atomisation energy of a molecule is considered to be invariant to:

- Permutations of the atom indices
- Rigid movements of the atomic positions

This approach reduces the number of parameters for each particle (becoming $\phi_i(x,r)$ instead of $\{x_i,r_i\}$ for particle i), therefore reducing the problem to one which can be implemented with machine learning. An upscaling from a 2D to 3D scattering framework is also outlined[49].

The effectiveness of this wavelet scattering approach[47] is demonstrated in three examples involving ergodic texture reconstructions[50], digit classification[51] and complex image classification[52].

Relevant areas: *chemical mechanical material modelling, multiscale modelling, high dimensional optimisation.*

- C. -> The multiscale regroupment of interactions is implemented using the fast multipole method[53].
- 7. -> Dimensionality reduction using the wavelet method allows for molecular modelling problems to be implemented with machine learning techniques[47].
- 3. -> The wavelet representation has application predominantly in the pharmaceutical industry as the implementation of high dimensional molecular learning is highly desirable.

'Coarse-Graining with the Relative Entropy: Recent Theory and Algorithms' - M. Scott Shell, University of California Santa Barbara

Proposed is a multiscale modelling approach based on quantifying the information loss when applying coarse-graining. The measure of information lost is called the relative entropy[32].

The framework is applied to examples of current difficulties in chemical materials modelling, such as peptide folding[54] and water[55] to demonstrate the potential to improve upon current computational and analytical modelling of these complex systems. Coarse graining is used to reduce the dimensionality of the problem by reducing the number of parameters in the configuration space (i.e. in a similar vein to the method used in the wavelet scattering approach[47]).

To obtain a measure of the relative entropy (i.e. information loss), the framework takes a probabilistic view at the likelihood of a given coarse-grained configuration based on the energies of the fine-scale configurations which map[56] to the same coarse-grained result. This formulation then

lends itself well to statistical methods to obtain an error in the coarse grained result[57] (i.e. the relative entropy).

The future prospects for designing coarse grained models are discussed, including different choices of fine to coarse mapping (i.e. by coarse grained hydrogens, amino acid residues or functional group) and the impact of the mapping to the model[58].

Relevant areas: *high-dimensional optimisation, multi-scale modelling.*

- 7/c. -> The relative entropy method of coarse graining[32] reduced the dimensionality of a complex problem by a mapping to a space with a reduced configuration space. This allows for the computational modelling of larger or more complex multiscale systems.

'Joys and Trials in Multiscale Simulations using Particles' - Petros Koumoutsakos, ETH Zurich

Particle methods across multiple spatiotemporal scales are discussed. The 'imperfect paths to knowledge' graph presented[59] as an introduction gives a fantastic summary of the sources of error found in the predictive science process.

The first example discussed is the mesoscale, an example is shown of a model of a μ-fluidic cell, used to track the circulation of a cancer tumour cell[60] (i.e. metastasis). The volumes of fluid (consisting of blood cells) required to model is on the scale of millilitres (i.e. high-throughput), as the circulation of a tumour cell is rare event.

Secondly the modelling of vortex collisions is presented, with a particle remeshing implemented to obtain the correct flow solution to Euler's equation. This method is applied in the speaker's 'Billion vortex particle direct numerical simulations of aircraft wakes' paper[61].

A very large multi-scale simulation of water flow past a fullerene is presented[62]. The modelling of this systems requires near-seamless transitioning between scales, provided by an overlap region, to ensure the correct phenomenon is simulated.

Our ability to carry out multi-scale simulations is becoming so great that models are now considered to be predictive, and in some cases are being used to disprove experimental results. An incorrect experimental measurement of water transport in carbon nanotubes, was published in the highly regarded journal *Nature* in 2005[63]. The mistake was discovered by computational scientists due to their inability to simulate the same result in their models. The speaker, among others, created a large multiscale model which correctly predicted the result[64], and was later confirmed by a second undertaking of the experimental measurement[65].

Finally the Bayesian treatment of these methods is discussed, to obtain the uncertainties in their predictive power[66].

Relevant areas: *high performance computing, multi-scale modelling, uncertainty quantification.*

- 13. -> The application of the Bayesian uncertainty quantification framework is applied to simulation of granular materials using discrete element methods[66].
- F. -> The sheer scales of the simulations presented are huge – they involve huge numbers of atoms requiring a highly parallelised model. (*Details of the parallelisation are not mentioned in the talk*).

- D. -> Bridging between scales using overlapping meshes[62] and particle remeshing[61] are detailed and shown to provide promising results for these multiscale simulations.

'Propagation of Complex Fracture' – Robert Lipton, Louisiana State University

Proposed is a method to predict the initiation and evolution of complex fracture patterns using peridynamics[67]. The method is a bottom-up approach to quantative non-local modelling of complex fracture – the forward-problem approach from a mesoscopic length scale.

In the linear elastic theory of fracture mechanics has previously been successfully used to study the strengths of materials for cracks where their direction of propagation is already known[68]. Shear panels (i.e. used in aircraft) which have already sustained cracks are now modelled, with the goal being to predict the maximum load which they can sustain before failure (i.e. a fatal crack propagates).

Starting from a generalisation of Silling's non-local formulation for the continuum modelling of fracture[69], the effect of the horizon length scale relative to the sample size on the dynamics is studied[70]. This information is then used to calibrate the cohesive energy (bond softening) between the bonds affected by the crack and those outside of this 'process zone'.

Finally a mathematical formulation for the rapid growth of small fissures, from instabilities within the process zone, is outlined[70]. It is found that in this model, a fracture can be thought of as a phase transition from a predominantly linear elastic phase within the horizon, into a predominantly soft phase[71].

Relevant areas: *modelling of composite failures within complex structures.*

- 1. -> The model proposed[67] uses peridynamics to allow for the modelling of cracks where the direction of propagation is not yet known.

'Predictive Multiscale Modelling for Decision Support in Design of Hierarchical Alloy Systems' – David McDowell, Georgia Institute of Technology

Multiscale modelling is applied to 'multi-level' design and development of metallic materials, focusing predominantly on the mesoscale – where discrete dislocations and dislocation patterns are found. The elements of crystal plasticity found (i.e. discreetness and patterning effects) on the mesoscale are difficult to model.

It is highlighted that in multiscale modelling literature, there had not been a great amount of attention given to the uncertainty in scale linking algorithms[72]. There are currently a variety of two scale hierarchical bridging methods (i.e. on-the-fly methods, ab-initio, informed molecular dynamics, domain decomposition, Monte Carlo) which each bring degrees of uncertainty into a simulation. The traditional industry methods of top down design often conflict with the modelling methods which rely on working from the bottom up. This gives rise to the difference in the time required to design the geometric structure of a new aircraft, for example, with the time required to develop a new material.

Ashby maps, used in materials selection are introduced[73] and the properties and responses related to different scales discussed. Mesoscopic examples showing how multiscale modelling and simulation can assist in decision support for materials development. Four examples are discussed:

- Using high fidelity modelling to challenge accepted understanding of how we interpret experiments conducted at the mesoscale[74,75],
- Examine model form and structure uncertainty at the mesoscale, with a focus on the way dislocations move through grain boundaries[76] (i.e. flip-transfer processes),
- Extreme value property estimates in fatigue, mainly how the high-cycle fatigue strength depends on the microstructure[77,78],
- Combined bottom-up[79] and top-down[80] strategies to inform both model form and parameter estimation.

Relevant areas: *multi-fidelity simulations, multiscale modelling, uncertainty quantification for robust design.*

- 13. -> A combined top-down and bottom-up approach is proposed to deal with differing model degrees of freedom and uncertainty encountered in materials engineering[72].
- 10. -> Two examples of high fidelity simulations are given: a mechanistic model of shock-wave-induced viscoplasticity[74] and the plane wave simulation of elastic-viscoplastic single crystals[75].

'Multiscale Simulations of Soft Matter: Applications and New Developments' – Kurt Kremer, Max Planck Institute for Polymer Research

Sequential multiscale descriptions of nanoscopic molecular assemblies and adaptive schemes are introduced as two techniques of great importance used to model the relationship between function and structure in soft matter[81].

The VOTCA (versatile object-oriented toolkit for coarse graining applications) software package, is a coarse grained molecular dynamics analysis toolkit and can perform a variety of basic standard methods, such as Boltzmann inversion for bonded potentials and inverse Monte Carlo[82]. An example of VOTCA's ability in determining liquid-crystal phase transitions is demonstrated[83,84].

Sequential multiscale simulations are discussed, and the AdResS (adaptive resolution molecular dynamics simulation) approach illustrated through an example simulation of liquid water[85].

Relevant areas: *multiscale modelling, surrogate modelling.*

- 9. -> VOTCA[82] is an analysis package containing systematic coarse graining techniques such as iterative Boltzmann inversion and force matching. It has been successfully used for a variety of soft matter applications, for example to determine the liquid-crystal transition of light-driven azo-materials[84].
- C. -> The AdResS technique which moves through scale simulations of a material by on-the-fly calculations has effectively produced an adaptive resolution simulation of liquid water[85].

'The Parallel Replica Algorithm: Mathematical Foundations and Recent Developments' – Tony Lelievre, École des Ponts ParisTech

The parallel replica algorithm was first proposed by Voter[86] as a method to efficiently probe rare events, i.e. transitions between metastable states by launching an array of probabilistic molecular dynamics simulations with the same initial conditions . The mathematical formulation of the algorithm when applied to quasi-stationary distributions is discussed, with an emphasis on escaping from energetic (potential) barriers and entropic barriers[87].

The algorithms has three main steps:

- Decorrelation step: an initial reference simulation is run until a metastable state is reached (i.e. a particle becomes trapped in some potential well)
- Dephasing step: the reference simulation is used to inform the initial conditions for an array of parallel runs of the particle from this point. To ensure the runs are all different, they are dephased from one another.
- Parallel step: The trajectories are run in parallel until the first transition event is detected.

The generalised parallel replica algorithm is proposed, which uses a Fleming-Viot process for the dephasing step and a Gelman-Rubin statistical test to simultaneously identify when the process converges whilst also generating samples[88].

Relevant areas: *high performance computing.*

- f. -> The parallel replica algorithm[86,88] can be used to study rare events. The method relies on the parallel running of multiple simulations informed with the same initial conditions and then 'dephased' from one another.

'Modern Challenges in Coupled Quantum-Continuum Modelling and Control of Closed and Dissapative Systems' – Roderick Melnik, Wilfrid Laurier University

Nanomaterials are of great interest in many branches of science and engineering as they can be assembles into larger physical and biological structures[89], and can be incorporated into devices and systems[89], such as semiconductors.

In the modelling of low dimensional semiconductor nanostructures, *ab initio* methods are the common route chosen. New technological advances in applications of these nanostructures has shown that these simplistic quantum mechanical approximations are inadequate[90,91]. For meaningful results, coupled multiscale simulations up to the nanoscale are required to gain any real insight.

A brief overview of current multiscale coupling scheme challenges are given. The largest source of uncertainty in these systems comes from the classical systems theory treating input, output and signal flow graphs from *ab initio.* A focus is placed on coupled mathematical models used to examine the interaction of particles with surfaces[92]. The challenges associated with modelling open dissipative systems are also discussed[93], and a model of graphene nanowire superlattices presented to illustrate[94].

Relevant areas: *materials modelling, multiscale modelling, highly coupled systems.*

- c. -> In order to effectively model many types of complex systems on the nanoscale, coupled multiscale simulations must be used. The solely *ab initio* approach has been shown to be inaccurate[90,91].
- 6. -> In highly coupled systems the majority of the uncertainty comes from the *ab initio* treatment of input, output and signal flow graphs in the classical systems theory.

'Path-space Information Metrics and Variational Inference for Non-equilibrium Coarse-grained Systems' - Petr Plechac, University of Delaware

[No presentation or video has been uploaded for review]

"Abstract: We discuss information-theoretic tools for obtaining optimized coarse-grained molecular models for both equilibrium and non-equilibrium molecular dynamics. The presented approach compares microscopic behaviour of molecular systems to para- metric or non-parametric coarse-grained systems using the relative entropy between distributions on the path space. It allows us to formulate a corresponding path space variational inference problem. The methods become entirely data-driven when the microscopic dynamics are replaced with corresponding correlated data in the form of time series."

Relevant areas: *uncertainty quantification, high-dimensional optimisation.*

'UQ Information Inequalities, Variational Inference and Accelerated Sensitivity Screening' - Markos A. Katsoulakis, University of Massachusetts

Proposed are information theory-based inequalities for use in uncertainty quantification[95,96]. The relative entropy approach is used as an information metric to measure the 'information loss' when coarse graining a system, to which variational inference is then applied to minimise the relative entropy. This method results in a quantification of the uncertainty in a quantity of interest (i.e. observable).

The relative entropy formulation of the Csiszár-Kullback-Pinsker inequality was used to investigate the information loss of untrained parameters in the variational inference. Unfortunately this method is shown not to work for driven (i.e. dynamical) systems, and the example given to illustrate this is a mean field model.

The correct approach is presented as a sensitivity analysis of quantities of interest. It is implemented in terms of new goal-oriented divergences through a Monte Carlo simulation[97]. As the derived sensitivity bounds in this new formulation rely on the path Fisher Information matrix, they therefore depend only on the local dynamics and are gradient-free. This allows for a computationally efficient implementation of this method in systems with a high number of parameters[97].

Examples to demonstrate this method are shown, namely accelerated sensitivity screening[98] and Lagrangian molecular dynamics[99].

Relevant areas: *uncertainty quantification, high–dimensional optimisation.*

- E. -> The relative entropy uncertainty quantification approach outlined, which uses goal-oriented divergences, has been shown to correctly measure the information metric (i.e. relative entropy) in coarse grained systems[97].
- 7. -> Due to the nature of the path Fisher information matrix (i.e. high-dimensional optimisation) in this new formulation, (i.e. local dynamics-dependent and gradient free sensitivity bounds) the measure of uncertainty in quantities of interest is computationally efficient for this method[97].

'Materials Structure Prediction from First Principles' – Chris Pickard, University of Cambridge

The '*Ab initio* random structure searching' (AIRSS) method[100] is presented and the recent successes in its application are discussed. The method is used in structure prediction and relies upon letting atoms relax to the nearest lowest energy minimum from an initial random configuration.

AIRSS has been used to successfully predict many experimental results in condensed matter physics, such as silane not being superconducting and ammonia being ionic. The different crystal structures formed by ice at a variety of different pressures has also been reported[101]. Hydrogen at extreme pressures has also been investigated[102]. Finally, AIRSS has also been used to predict interface structures[103].

Relevant areas: *materials modelling, high performance computing.*

- A. -> '*Ab initio* random structure searching' (AIRSS)[100] has been used to predict a vast amount of material structures, and is particularly effective at predicting the atomistic structure of materials at high pressure.

'From Atomistic Environments to Mesoscale Structures. Mapping Complexity in Materials.' – Michele Ceriotti, École Polytechnique Fédérale de Lausanne

As a result of the recent advancements in atomistic computer simulations, it is becoming increasingly difficult to extract useful physical insights due to the sheer quantity of data being produced. Discussed are two computer-aided analysis strategies to help overcome this issue:

- Machine learning: used to recognise reoccurring structural patterns in a material,
- Coarse-graining: non-linear high dimensionality reduction used to reduce the parameter space of a structural landscape.

In the presentation of machine learning techniques, the identification and probabilistic analysis of molecular motifs is presented, where clusters are found in the data, which in turn yield the selection of possible motifs[104]. This method is applied to find the structural patterns of a hydrogen bond[104].

Kernel methods are used to compare chemical environments, based on the overlap of atomic densities (formulated as soft overlap potentials SOAP[48]). The measure of dissimilarity between these environments is averaged over the landscape, preserving the dissimilarities[105]. This results in a low-dimensional map to describe accessible configurations of the chemical environment/phase space[106].

This method is applied to an example investigating the structure and stability of oligopeptides, as well as two other large data sets (*all currently unpublished*).

Relevant areas: *high-dimensional optimisation, data, high performance computing, materials modelling.*

- A/12. -> Machine learning techniques are being used to interrogate the large datasets resulting from atomistic simulations. These techniques are able to predict all possible configurations and stoichiometries of a material from a collection of initial atoms and molecules from clustering of the data points.
- 7. -> A non-linear dimensionality reduction instance of coarse graining is described, which relies on the differences in energy landscapes of potential atomic configurations to create a map from two fine scale environments to a coarse-grained difference of the two.

'How Predictive Are Multiscale Materials' Simulations?' - Dionisios Vlachos, University of Delaware

[No presentation or video has been uploaded for review]

"Abstract: The advances in multi-scale computations and experiments enable us to compare model predictions to experimental data, obtain insights into the mechanisms, predict in silico new materials, and interrogate model predictions experimentally. In this talk, grand challenges in closing the gap between experiments and simulations will be presented. We will demonstrate how descriptor-based modelling can enable search of novel materials and assess this framework with experiments. Outstanding questions include how reliable and robust are model predictions and our quest for searching new materials and how critical are the aforementioned gaps. We will present a new computational formalism that addresses this problem. We will demonstrate this methodology with examples of ammonia decomposition and ethanol steam reforming on single metals and bimetallic catalysts"

Relevant areas: Multiscale modelling.

'The Inverse Problem in Materials Theory: Find the System that has a Given Target Property' – Alex Zunger, University of Colorado

[No presentation or video has been uploaded for review]

"Abstract: The history of material research and condensed matter physics has often proceeded via accidental discovery of materials with interesting physical properties – superconductors, Photo catalytic compounds, magneto-resistors to name a few. Yet, for many applications we know well what type of physical properties we want, except that we do not know a material that has those target properties. The question posed in this talk is: does it make sense to first declare the property you really want, then find the structure and material that has this property[107–114]*.The obvious obstacle is that there are innumerably many possible atomic structures that could, in principle, be made even from a few elements and we do not know which structure would have the desired target property. It turns out that modern atomic-resolution quantum mechanics (i.e., electronic structure*

theory) can be combined with biologically- inspired (evolutionary) "Genetic Algorithms" to scan a truly astronomic number of atomic configurations in genomic-like search of the one(s) that have desired, target materials properties[107,114]*. Once the number of configurations with target property is narrowed down to a few, laboratory synthesis becomes viable. Traditionally, laboratory synthesis often focused on the properties of previously reported compounds, but neglect numerous unreported but chemically plausible compounds that could have interesting properties. For example, the Half Heusler ABX family of compounds features examples of topological insulators, thermoelectrics and piezoelectrics, but only 83 out of 483 of these possible compounds have been made*[108–110]*. An important new development in the field of Materials by Design has been the prediction of which of the "missing compounds" are missing for a good reason (e.g are intrinsically unstable) and which are missing because we did not get around to probe them. Prospects for experiment-theory interaction in this arena will be reported."*

<u>*Relevant areas:*</u> *materials modelling.*

'Bayesian Uncertainty Quantification in Alloy Modelling and Design' Nicholas Zabaras, University of Warwick

Proposed is a Bayesian description of the exchange-correlation (XC) functional used in density functional theory (DFT). This approach allows for the fully Bayesian uncertainty quantification treatment of electronic structures predicted using DFT[115].

Using the relevance vector machine framework, the ideal choice of training data for the Bayesian model is discussed[116]. The description is applied to an example predicting (with a given error) the Lattice constants and bulk modulus for a variety of crystalline structures.

The newly proposed formulation of the XC functional is shown to accurately reproduce the experimental results for material band structures, a result which semi-local XC functionals are unable to reproduce.

The choice of surrogates and their effect of the resulting predictions is discussed, with a focus on their use in alloy modelling[117].

<u>Relevant areas:</u> *Uncertainty quantification for robust design, surrogate modelling, multi-physics modelling.*

- e/13. -> Prosed is a new exchange-correlation functional for use in density functional theory calculations, which permits a fully Bayesian treatment of the uncertainties in the predictions of such simulations[115].
- 9. -> The use of surrogates in the modelling of allows is discussed, with results to be published in the near future[117].
- D. -> A variety of quantities of interest can be calculated using the newly proposed exchange correlations functional. They benefit from quantified uncertainty in their prediction using the formulation introduced.

'Mean Field Methods for Stochastic Dynamics' – Manfred Opper, Technische Universität Berlin

A derivation of mean field methods used in statistical physics is outlined, and their application to stochastic dynamical models, such as coupled stochastic differential equations, is discussed. The primary derivation is for Gaussian process regression of mean field models, giving a maximum likelihood estimate for the prediction, using latent variables.

The expectation propagation algorithm introduced by Tom Mika is also presented[1] as an approximation to 'cavity fields'. The cavity approach[118] is introduced as a means to apply Bayesian inference to mean field methods and a formulation of Bayesian inference for stochastic differential equations is derived[119].

Relevant areas: *uncertainty quantification.*

- E. -> Proposed is a framework for the application of Bayesian inference to mean field methods. This variational method allows for the approximation of an otherwise intractable distribution by an ensemble of Gaussians[119].

'Nonlocal Models as Effective Bridges in Multiscale Modelling' – Qiang Du, Columbia University

A vector calculus for nonlocal operators in defined[120], and used in examples to describe some nonlocal models. Silling's nonlocal alternative to local mechanics: peridynamics[69] is discussed and contrasted to the newly proposed mathematical framework. An example on bond-based peridynamics showing consistency on the continuum level with local models only in the new vector calculus reformulation[121].

An asymptotically compatible scheme for non-local models is introduced, providing numerical scheme with convergent approximations to both the solution of the nonlocal model and its local limit[122]. For crack propagation models it is shown that the correct local limit solution is obtained as long as the discretization (mesh) parameter decreases faster than the modelling (horizon) parameter[122].

Relevant areas: *multiscale modelling, modelling of composite failures within complex structures.*

- 1. -> The newly proposed vector calculus method for nonlocal operators has been shown to correctly obtain local limit solution for crack propagation if the finite-element mesh is decreased faster than the modelling horizon[122].

'Probing the Properties of Soft Matter by Optimally Designed Nonequilibrium Experiments' – Carsten Hartmann, Freie Universität Berlin

A numerical method is presented to optimise the sampling of an energy landscape when searching for rare events[123]. The method works by minimising a cross-entropy-like functional representing the optimal nonequilibrium force, unlike previous methods which rely on sampling based in path space. It is shown that optimally designed nonequilibrium perturbations can mimic thermodynamic equilibrium.

The optimal control of diffusions with slow and fast variables is discussed[124], and an example of the method's application to the conformational transition of butane in water is presented. If

Relevant areas: *optimization research.*

b. -> Proposed is an optical sampling method for out-of-equilibrium simulations/rare events[123]. Its application to materials modelling results in much faster computation times and does not rely on sampling in path space as with previous methods.

'Entropy Descriptors in Advanced Databases and Entropy Stabalised Oxides' – Stefano Curtarolo, Duke University

The AFLOW (Automatic –FLOW for materials discovery[125]) is a repository of *ab inito* density functional theory materials simulations, where anyone can search for over 86,000,000 material properties of more than 956,000 material compounds[126]. Each material is given a unique fingerprint[127], and the repository is enhanced with the ability to automatically generate a variety of plots, such as band structures, phase diagrams and densities of states.

Using the ALOW repository, a variety of entropy-stabilised oxides were predicted, and it was proven that by purposely introducing configurational disorder into a system is a strategic tool to predict new phases of crystalline matter[128].

Materials cartography is discussed, using the unique fingerprinting system implemented in AFLOW to perform a big data, graphical analysis of a variety of materials[127]. The cartography combined with simplex optimisation (to find structural motifs) is used to outline a potential framework for predicting superconductors (*"to be continued..."*).

Relevant areas: *data, materials modelling, visualisation.*

- 12. -> To cope with the huge amount of data in the AFLOW repository (>86,000,000 material properties), each material is given a unique fingerprint[127]. This data is then used to perform a database-wide analysis of trends and clustering among materials and properties (i.e. materials cartography)[127].
- 14. -> The materials cartography resulting from big data analysis of the AFLOW repository is a visual, relatively intuitive visualisation of material grouping dependent on a particular material property[127].

'Bridging the GAP: Fitting *ab initio* potential energy surfaces systematically' - Gábor Csányi, University of Cambridge

interatomic potentials used to inform molecular dynamics simulations are discussed, with a focus on SOAP (smooth overlap of atomic positions) potentials types[129]. SOAP potentials are Gaussian approximation potentials and are particularly effective at simultaneously describing bonding with a varying number of neighbours, unlike traditional two-body potentials.

To quantify the uncertainty in the fit of these potentials a fully Bayesian treatment can be used due to their Gaussian nature. A framework for this uncertainty quantification by Gaussian inference is outlined. An example discussing the Peirels barrier for screw dislocation glide and He-vacancy interaction in Tungsten are shown to illustrate the effectiveness of such Gaussian these potentials[130].

Relevant areas: *uncertainty quantification, materials modelling.*

- E. -> A fully Bayesian treatment of Gaussian approximation potentials[129] for use in molecular dynamics simulations of materials is outlined.

'Numerical Design of Pathways for Addressable Self-assembly' – Daan Frenkel, University of Cambridge

The idea of 'addressable complexity' is discussed, with a focus on colloidal structures. The aim is to create a model for the three-dimensional assembly of a material, from hundreds or thousands of distinct building blocks.

Colloidal matter is rarely modelled using atomistic simulations due to the sheer number of atoms they typically contain (hundreds to millions of atoms, potentially even per colloid). Their lack of dependence on atomistic detail is demonstrated in an example describing the depletion forces between coarse grained representations of each colloid.

An example of structures with addressable complexity would be protein complexes and (functionalised) 'DNA-brick' structures. There are two types of complexity which have to be considered in a model for either: -

- 'Structural' complexity through packing
- 'Addressable' complexity through specific interactions.

The creation and modelling of DNA-functionalised nanoparticles is discussed (i.e. colloids with DNA 'hairs' attached)[131]. This example illustrates the common difficulties with complex self-assembly – as with structures formed by the colloids with DNA 'hairs', the materials which tend to form by this pathway are not the equilibrium phase. The most recent successful advancement in the modelling of complex assembly was of DNA brick crystals with different structures depths[132], which is shown as a closing example.

Relevant areas: *materials modelling, complex solid material interaction.*

- 2. -> The prediction of a structure formed from hundreds or thousands of distinct building blocks is a huge computational task. Outlined is the current view of the field and the recent advancement in the modelling of DNA brick crystals[132] is discussed.

'Data, Manifold Learning, and the Modelling of Complex/Multiscale Systems' – Yannis Kevrekidis, Princeton University

Dimensionality reduction using manifold learning is introduced and the concept of projection integration discussed. Projection integration is a computational optimisation technique, in which a

detailed simulation is performed for a short period of time, and its behaviour extrapolated forward over a large time step before a detailed simulation is run again at the new time, realigning the trajectory with the true path[133]. The application of this technique over both spatial and temporal trajectories is known as patch dynamics.

To reduce the dimensionality of a system, the correct coarse grained variable must be found (i.e. find a manifold). This can be investigated though diffusion maps, an output of manifold learning[134,135]. An example using manifold learning and diffusion maps is shown to yield an insight into the collective non-linear behaviour of a shoal of fish[136].

Finally the computation of distance metrics of a stochastic dynamical system from empirical data using the data mining techniques discussed above is presented[137].

Relevant areas: *surrogate modelling, high-dimensional optimisation*

- 7. -> Manifold learning, projection integration and diffusion maps are introduced as methods to reduce the dimensionality of complex systems[135]. They can also be applied to large quantities of empirical data to find trends upon which dimensionality reduction can be applied, and the nonlinear intrinsic variables identified[137].
- 9. -> Before coarse graining of the system can take place, manifold learning and data mining techniques are used to identify the correct coarse graining variables[137].

[Aside: 8. -> 'Multiobjective evolutionary algorithms: A survey of the state of the art'[138]]

Bibliography

(1) Minka, T. P. *PhD Thesis* **2001**.

(2) Cseke, B.; Heskes, T. *J. Mach. Learn. Res.* **2011**, *12*, 417.

(3) Jampani, V.; Eslami, S. M. A.; Tarlow, D.; Kohli, P.; Winn, J. In *AISTATS Conference*; 2015; Vol. 38.

(4) Eslami, S. M. A.; Tarlow, D.; Kohli, P.; Winn, J. *Adv. Neural Inf. Process. Syst.* **2014**, 1.

(5) Jitkrittum, W.; Gretton, A.; Heess, N.; Eslami, S. M. A.; Lakshminarayanan, B.; Sejdinovic, D.; Szabó, Z. In *Uncertainty in Artificial Intelligence*; 2015.

(6) Hernández-Lobato, J. M.; Adams, R. P. In *IMCL Conference - Issues, Infrastructures and Applications*; 2015.

(7) Leimkuhler, B.; Matthews, C. *Appl. Math. Res. Express* **2013**, *2013* (1), 34.

(8) Leimkuhler, B.; Matthews, C.; Stoltz, G. *IMA J. Numer. Anal.* **2015**, 1.

(9) Jones, A.; Leimkuhler, B. *J. Chem. Phys.* **2011**, *135* (8), 084125.

(10) Leimkuhler, B.; Shang, X. *ArXiv: 1505.06889* **2015**.

(11) Mones, L.; Jones, A.; Götz, A. W.; Laino, T.; Walker, R. C.; Leimkuhler, B.; Csányi, G.; Bernstein, N. *J. Comput. Chem.* **2015**, *36* (9), 633.

(12) Shang, X.; Zhu, Z.; Leimkuhler, B.; Storkey, A. J. *NIPS, ArXiv: 1510.08692* **2015**.

(13) Ding, N.; Chen, C.; Skeel, R. D.; Babbush, R. In *Advances in Neural Information Processing Systems 27*; 2014.

(14) Yu, T.-Q.; Chen, P.-Y.; Chen, M.; Samanta, A.; Vanden-Eijnden, E.; Tuckerman, M. *J. Chem. Phys.* **2014**, *140* (21), 214109.

(15) Tzanov, A. T.; Cuendet, M. a.; Tuckerman, M. E. *J. Phys. Chem. B* **2014**, *118* (24), 6539.

(16) Chen, M.; Yu, T.-Q.; Tuckerman, M. E. *Proc. Natl. Acad. Sci.* **2015**, *112* (11), 3235.

(17) Chen, M.; Cuendet, M. A.; Tuckerman, M. E. *J. Chem. Phys.* **2012**, *137* (2), 024102.

(18) Samanta, A.; Chen, M.; Yu, T.-Q.; Tuckerman, M.; E, W. *J. Chem. Phys.* **2014**, *140* (16), 164109.

(19) Yu, T. Q.; Tuckerman, M. E. *Phys. Rev. Lett.* **2011**, *107* (1), 3.

(20) Rosso, L.; Mináry, P.; Zhu, Z.; Tuckerman, M. E. *J. Chem. Phys.* **2002**, *116* (2002), 4389.

(21) Dunitz, J. D.; Scheraga, H. a. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (40), 14309.

(22) Peters, B. *J. Phys. Chem. B* **2015**, *119* (21), 6349.

(23) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53* (1), 291.

(24) Peters, B. *J. Chem. Phys.* **2006**, *125* (24), 241101.

(25) Peters, B. *Chem. Phys. Lett.* **2012**, *554*, 248.

(26) Peters, B.; Trout, B. L. *J. Chem. Phys.* **2006**, *125* (5), 054108.

(27) Teferra, K.; Graham-brady, L. *Comput. Mater. Sci.* **2015**, *102*, 57.

(28) Ghanem, R. G.; Spanos, P. D. *Stochastic Finite Elements: A Spectral Approach*; 1991.

(29) Ravela, S.; Sandu, A. In *First Intenational Conference, DyDESS*; 2014.

(30) Tipireddy, R.; Ghanem, R. *J. Comput. Phys.* **2014**, *259*, 304.

(31) Arnst, M.; Ghanem, R.; Phipps, E.; Red-Horse, J. *Int. J. Numer. Methods Eng.* **2012**, *92*, 1044.

(32) Shell, M. S. *J. Chem. Phys.* **2008**, *129* (14), 144108.

(33) Chopin, N.; Lelièvre, T.; Stoltz, G. *Stat. Comput.* **2012**, *22* (4), 897.

(34) Katsoulakis, M. a.; Plecháč, P.; Rey-Bellet, L.; Tsagkarogiannis, D. K. *J. Nonnewton. Fluid Mech.* **2008**, *152* (1-3), 101.

(35) Gu, M. G.; Kong, F. H. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (13), 7270.

(36) Schmidt, M.; Babanezhad, R.; Ahmed, M. O.; Defazio, A.; Clifton, A.; Sarkar, A. *ArXiv: 1504.04406* **2015**.

(37) Szepessy, A. **2015**, In preparation.

(38) Torquato, S.; Stillinger, F. H. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* **2003**, *68* (4 Pt 1), 041113.

(39) Torquato, S.; Zhang, G.; Stillinger, F. H. *Phys. Rev. X* **2015**, *5* (2), 021020.

(40) Batten, R.; Stillinger, F.; Torquato, S. *Phys. Rev. E* **2009**, *80* (3), 031105.

(41) Florescu, M.; Torquato, S.; Steinhardt, P. J. *Proc. Natl. Acad. Sci.* **2009**, *106* (49), 20658.

(42) Jiao, Y.; Lau, T.; Hatzikirou, H.; Meyer-Hermann, M.; Corbo, J. C.; Torquato, S. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **2014**, *89* (2), 1.

(43) Weijs, J. H.; Jeanneret, R.; Dreyfus, R.; Bartolo, D. *Phys. Rev. Lett.* **2015**, *115* (10), 108301.

(44) Mosby, M.; Matouš, K. *Model. Simul. Mater. Sci. Eng.* **2015**, *23* (8), 085014.

(45) Mosby, M.; Matous, K. *Int. J. Numer. Methods Eng.* **2015**, No. 102, 748.

(46) Matouš, K.; Kulkarni, M. G.; Geubelle, P. H. *J. Mech. Phys. Solids* **2008**, *56*, 1511.

(47) Mallat Stephane G. *Pattern Anal. Mach. Intell. IEEE Trans.* **1989**, *11* (7), 674.

(48) Bartók, A. P.; Kondor, R.; Csányi, G. *Phys. Rev. B - Condens. Matter Mater. Phys.* **2013**, *87* (18), 1.

(49) Benetto, J. J.; Czaja, W.; Ehler, M.; Hirn, M. *SPIE Proc Wavelet Appl. Ind. Process. VII* **2010**, *7535*.

(50) Bruna, J.; Mallat, S. *ArXiv: 1311.0407* **2013**, 5.

(51) Bruna, J.; Mallat, S. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2011**, *1* (2), 1561.

(52) Oyallon, E.; Mallat, S.; Sifre, L. *ArXiv: 1312.5940* **2014**.

(53) Carrier, J.; GreenGardi, L.; Rokhlin, V. *SIAM J. SCl. STAT. Comput.* **1988**, *9* (4), 669.

(54) Hall, B. a.; Chetwynd, A. P.; Sansom, M. S. P. *Biophys. J.* **2011**, *100* (8), 1940.

(55) Chaimovich, A.; Shell, M. S. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **2013**, *88* (August), 1.

(56) Crooks, G. E.; Sivak, D. a. *J. Stat. Mech. Theory Exp.* **2011**, *2011* (06), P06003.

(57) Chaimovich, A.; Shell, M. S. *J. Chem. Phys.* **2011**, *134* (9), 094112.

(58) Foley, T. T.; Shell, M. S.; Noid, W. G. *J. Chem. Phys.* **2015**, *143* (243104).

(59) Oden, T.; Moser, R.; Ghattas, O. *SIAM News* **2010**, *43* (9), 2517.

(60) Rossinelli, D.; Karniadakis, G.; Fatica, M.; Pivkin, I.; Koumoutsakos, P.; Tang, Y.; Lykov, K.; Alexeev, D.; Bernaschi, M.; Hadjidoukas, P.; Bisson, M.; Joubert, W.; Conti, C. *Proc. Int. Conf. High Perform. Comput. Networking, Storage Anal. - SC '15* **2015**, 1.

(61) Chatelain, P.; Curioni, A.; Bergdorf, M.; Rossinelli, D.; Andreoni, W.; Koumoutsakos, P. *Comput. Methods Appl. Mech. Eng.* **2008**, *197* (13-16), 1296.

(62) Walther, J. H.; Praprotnik, M.; Kotsalis, E. M.; Koumoutsakos, P. *J. Comput. Phys.* **2012**, *231* (7), 2677.

(63) Majumder, M.; Chopra, N.; Andrews, R.; Hinds, B. J. *Nature* **2005**, *438* (7064), 43.

(64) Walther, J. H.; Ritos, K.; Cruz-Chu, E. R.; Megaridis, C. M.; Koumoutsakos, P. *Nano Lett.* **2013**, *13* (5), 1910.

(65) Holt, J. K.; Park, H. G.; Wang, Y.; Stadermann, M.; Artyukhin, A. B.; Grigoropoulos, C. P.; Noy, A.; Bakajin, O. **2006**, *312* (May), 1034.

(66) Hadjidoukas, P. E.; Angelikopoulos, P.; Rossinelli, D.; Alexeev, D.; Papadimitriou, C.; Koumoutsakos, P. *Comput. Methods Appl. Mech. Eng.* **2014**, *282*, 218.

(67) Lipton, R. *ARXiv: 1411.4609* **2014**.

(68) Xu, X.-P.; Needleman, A. *J. Mech. Phys. Solids* **1994**, *42* (9), 1397.

(69) Silling, S. A. *J. Mech. Phys. Solids* **2000**, *48* (1), 175.

(70) Lipton, R. *J. Elast.* **2014**, *117* (1), 21.

(71) Du, Q.; Lipton, R. *SIAM news* **2014**, *47* (3).

(72) Panchal, J. H.; Kalidindi, S. R.; McDowell, D. L. *Comput. Des.* **2013**, *45* (1), 4.

(73) Ashby, M. F. *Materials Selection in Mechanical Design*; 1999.

(74) Austin, R. A.; McDowell, D. L. *Int. J. Plast.* **2012**, *32-33*, 134.

(75) Lloyd, J. T.; Clayton, J. D.; Austin, R. A.; McDowell, D. L. *J. Mech. Phys. Solids* **2014**, *69*, 14.

(76) Xu, S.; Che, R.; Xiong, L.; Chen, Y.; McDowell, D. L. *Int. J. Plast.* **2015**, *72*, 91.

(77) Przybyla, C. P.; McDowell, D. L. *Procedia Eng.* **2010**, *2* (1), 1045.

(78) Sweeney, C. A.; Vorster, W.; Leen, S. B.; Sakurada, E.; McHugh, P. E.; Dunne, F. P. E. *J. Mech. Phys. Solids* **2013**, *61* (5), 1224.

(79) Narayanan, S.; McDowell, D. L.; Zhu, T. *J. Mech. Phys. Solids* **2014**, *65*, 54.

(80) Patra, A.; Zhu, T.; McDowell, D. L. *Int. J. Plast.* **2014**, *59*, 1.

(81) Praprotnik, M.; Site, L. D.; Kremer, K. *Annu. Rev. Phys. Chem.* **2008**, *59*, 545.

(82) Rühle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. *J. Chem. Theory Comput.* **2009**, *5*, 3211.

(83) Kreis, K.; Fogarty, A. C.; Kremer, K.; Potestio, R. *Eur. Phys. J. Spec. Top.* **2015**, *224* (12), 2289.

(84) Böckmann, M.; Marx, D.; Peter, C.; Site, L. D.; Kremer, K.; Doltsinis, N. L. *Phys. Chem. Chem. Phys.* **2011**, *13* (17), 7604.

(85) Praprotnik, M.; Matysiak, S.; Site, L. D.; Kremer, K.; Clementi, C. *J. Phys. Condens. Matter* **2007**, *19* (29), 292201.

(86) Voter, A. *Phys. Rev. B* **1998**, *57* (22), R13985.

(87) Lelièvre, T. *ArXiv: 1501.01034* **2015**.

(88) Binder, A.; Lelièvre, T.; Simpson, G. *ArXiv: 1404.6191* **2014**.

(89) Prabhakar, S.; Melnik, R.; Bonilla, L. L.; Badu, S. *Phys. Rev. B Condens. Matter Mater. Phys.* **2014**, *90* (20), 205418/1.

(90) Feng, X.; Xiao, J.; Melnik, R.; Kawazoe, Y.; Wen, B. **2015**, *104503*.

(91) Cheng, Y.; Du, J.; Melnik, R.; Kawazoe, Y.; Wen, B. *Carbon N. Y.* **2016**, *98*, 468.

(92) Dhote, R. P.; Gomez, H.; Melnik, R. N. V.; Zu, J. *Comput. Phys. Commun.* **2015**, *192*, 48.

(93) Prabhakar, S.; Melnik, R.; Bonilla, L. *arXiv Prepr. arXiv1211.2936* **2012**, *2* (2), 7.

(94) Prabhakar, S.; Melnik, R.; Bonilla, L. L. *J. Phys. D. Appl. Phys.* **2013**, *46* (26), 265302.

(95) Pantazis, Y.; Katsoulakis, M. a. *J. Chem. Phys.* **2013**, *138* (5), 054115.

(96) Tsourtis, A.; Pantazis, Y.; Katsoulakis, M. A.; Harmandaris, V.; Mathematics, A.; Mathematics, A.; Katsoulakis, M. A.; Analysis, S.; Matrix, F. I.; Dynamics, M. *ArXiv: 1412.6482* **2014**.

(97) Dupuis, P.; Katsoulakis, M. A.; Pantazis, Y. *ArXiv: 1503.05136* **2015**.

(98) Arampatzis, G.; Katsoulakis, M. A.; Pantazis, Y. *ArXiv: 1412.2153* **2014**, 1.

(99) Kalligiannaki, E.; Harmandaris, V.; Katsoulakis, M. a.; Plechac, P. *ArXiv: 1504.02152* **2015**, 1.

(100) Pickard, C. J.; Needs, R. J. *J. Phys. Condens. Matter* **2011**, *23* (5), 053201.

(101) Chen, J.; Schusteritsch, G.; Pickard, C. J.; Salzmann, C. G.; Michaelides, A. *ArXiv: 1508.03743* **2015**, 1.

(102) Drummond, N. D.; Monserrat, B.; Lloyd-Williams, J. H.; Ríos, P. L.; Pickard, C. J.; Needs, R. J. *Nat. Commun.* **2015**, *6*, 7794.

(103) Schusteritsch, G.; Pickard, C. J. *Phys. Rev. B* **2014**, *90* (3), 035424.

(104) Gasparotto, P.; Ceriotti, M. *J. Chem. Phys.* **2014**, *141* (17), 174110.

(105) Tribello, G. A.; Ceriotti, M.; Parrinello, M. *Proc. Natl. Acad. Sci.* **2010**, *107* (41), 17509.

(106) Ceriotti, M.; Tribello, G. A.; Parrinello, M. *Proc. Natl. Acad. Sci.* **2011**, *108* (32), 13023.

(107) Franceschetti, A.; Zunger, A. *Nature* **1999**, *402* (November), 60.

(108) Gautier, R.; Zhang, X.; Hu, L.; Yu, L.; Lin, Y.; L., S. O.; Chon, D.; Poeppelmeier, K. R.; Zunger, A. *Nat Chem* **2015**, *7* (4), 308.

(109) Yan, F.; Zhang, X.; Yu, Y. G.; Yu, L.; Nagaraja, A.; Mason, T. O.; Zunger, A. *Nat. Commun.* **2015**, *6*, 7308.

(110) Vermeer, M. J. D.; Zhang, X.; Trimarchi, G.; Donakowski, M. D.; Chupas, P. J.; Poeppelmeier, K. R.; Zunger, A. **2015**.

(111) Yu, L.; Zunger, A. *arXiv.org* **2014**, *cond-mat.m*, 1.

(112) Zhang, X.; Liu, Q.; Luo, J.-W.; Freeman, A. J.; Zunger, A. *Nat. Phys.* **2014**, *10* (5), 387.

(113) Trimarchi, G.; Zhang, X.; Freeman, A. J.; Zunger, A. **2014**, *161111*, 10.

(114) Zhang, L.; Luo, J.-W.; Saraiva, A.; Koiller, B.; Zunger, A. *Nat. Commun.* **2013**, *4*, 1.

(115) Aldegunde, M.; Kermode, J.; Zabaras, N. *J. Chem. Phys.* **2015**, *Submitted*.

(116) Bilionis, I.; Zabaras, N. *Inverse Probl.* **2014**, *30* (1), 015004.

(117) Kristensen, J.; Bilionis, I.; Zabaras, N. *J. Chem. Phys.* **2016**, *Submitted*.

(118) Mezard, M.; Parisi, G.; Virasoro, M. A. *Spin glass theory and beyond*; 1987.

(119) Vrettas, M. D.; Opper, M.; Cornford, D. *Phys. Rev. E* **2015**, *91*, 012148.

(120) Du, Q.; Gunzburger, M.; Lehoucq, R. B.; Zhou, K. *Math. Model. Methods Appl. Sci.* **2013**, *23*, 493.

(121) Mengesha, T.; Du, Q. *Nonlinearity* **2015**, *28* (3999).

(122) Tian, X.; Du, Q. *SIAM J. Numer. ANAL* **2014**, *52* (4), 1641.

(123) Hartmann, C.; Schütte, C. *J. Stat. Mech. Theory Exp.* **2012**, *11* (4).

(124) Hartmann, C.; Latorre, J. C.; Zhang, W.; Pavliotis, G. A. *J. Comput. Dyn.* **2014**, *1* (2), 279.

(125) Automatic - FLOW for Materials Discovery: http://aflowlib.org/.

(126) Curtarolo, S.; Hart, G. L. W.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O. *Nat. Mater.* **2013**,

12 (3), 191.

(127) Isayev, O.; Fourches, D.; Muratov, E. N.; Oses, C.; Rasch, K.; Tropsha, A.; Curtarolo, S. *Chem. Mater.* **2015**, *27* (3), 735.

(128) Rost, C. M.; Sachet, E.; Borman, T.; Moballegh, A.; Dickey, E. C.; Hou, D.; Jones, J. L.; Curtarolo, S.; Maria, J. *Nat. Commun.* **2015**, *6*, 1.

(129) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. *Phys. Rev. Lett.* **2010**, *104* (April), 1.

(130) Szlachta, W. J.; Bartók, A. P.; Csányi, G. *Phys. Rev. B* **2014**, *90* (10), 104108.

(131) Leunissen, M. E.; Frenkel, D. *J. Chem. Phys.* **2011**, *134* (8), 1.

(132) Ke, Y.; Ong, L. L.; Sun, W.; Song, J.; Dong, M.; Shih, W. M.; Yin, P. *Nat Chem* **2014**, *6* (11), 994.

(133) Kevrekidis, I. G. **2003**, *24* (4), 1091.

(134) Coifman, R. R.; Lafon, S.; Lee, a B.; Maggioni, M.; Nadler, B.; Warner, F.; Zucker, S. W. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (21), 7426.

(135) Nadler, B.; Lafon, S.; Coifman, R. R.; Kevrekidis, I. G. *Appl. Comput. Harmon. Anal.* **2006**, *21* (1), 113.

(136) Couzin, I. D.; Krause, J.; Franks, N. R.; Levin, S. a. *Nature* **2005**, *433* (7025), 513.

(137) Dsilva, C. J.; Talmon, R.; Gear, C. W.; Coifman, R. R.; Kevrekidis, I. G. *xx*, 1.

(138) Zhou, A.; Qu, B.-Y.; Li, H.; Zhao, S.-Z.; Suganthan, P. N.; Zhang, Q. *Swarm Evol. Comput.* **2011**, *1* (1), 32.