Whitepaper: IPAM Program "Complex Energy Landscapes" (fall 2017)

Conference Paper · January 2018

CITATIONS
0

READS
141

28 authors, including:

Nestor F. Aguirre
Los Alamos National Laboratory
47 PUBLICATIONS **518** CITATIONS
SEE PROFILE

Mauricio J. del Razo
Freie Universität Berlin
34 PUBLICATIONS **290** CITATIONS
SEE PROFILE

Marco Di Gennaro
Toyota Motor
10 PUBLICATIONS **977** CITATIONS
SEE PROFILE

Florent Hédin
Qubit Pharmaceuticals
46 PUBLICATIONS **233** CITATIONS
SEE PROFILE

Whitepaper: IPAM Program "Complex Energy Landscapes" (fall 2017)

Authors list (alphabetically)

Nestor F. Aguirre, Chris Anderson, Lorenzo Boninsegna, Gábor Csányi, Mauricio del Razo Sarmina, Marco Di Gennaro, Florent Hédin, Graeme Henkelman, Richard G. Hennig, Jan Janßen, Tony Lelièvre, Hao Li, Mitchell Luskin, Noa Marom, Jörg Neugebauer, Feliks Nüske, Joshua Paul, Danny Perez, Giovanni Pinamonti, Petr Plechac, Biswas Rijal, Gideon Simpson, Justin C. Smith, Anne Marie Z. Tan, Mira Todorova, Dallas R. Trinkle, Stephen Xie, Ping Yang

Abstract

Recent advances in computational resources and the development of high-throughput frameworks enable the efficient sampling of complicated multivariate functions. This includes *energy* and *electronic property landscapes* of *inorganic, organic, biomolecular, and hybrid materials and functional nanostructures*. Combined with new developments in data science, this leads to a rapidly growing need for *numerical methods* and a *fundamental mathematical understanding* of efficient sampling approaches, optimization techniques, hierarchical surrogate models, coarse-graining techniques, and methods for uncertainty quantification. The complexity of these energy and property landscapes originates from their simultaneous dependence on discrete degrees of freedom – i.e., number of atoms and species types – and continuous ones – i.e., position of atoms. Moreover, dynamical behavior governed by complex landscapes involves a rich hierarchy of timescales and is characterized by rare events that often are key to understanding the function of the materials under investigation.

To make significant advances in the crucial field of complex energy landscapes, we *identify scientific and mathematical challenges* whose solutions impact fields spanning from *machine learning* to *materials science* and *chemistry* to *large-scale computational simulation*. This potential impact includes the discovery of new materials with novel properties, enabling simulations across increasingly larger length and time scales, and finding new physical and chemistry principles that guide how materials work. It also offers the potential for innovative insights into how complex machine learning models interpolate data and identify patterns, and to develop new methodologies describing uncertainty in computational models and efficiently propagating that uncertainty through different models and scales. We identify a range of key issues in the field, along with promising directions to make significant progress.

1. Optimization Methods

Much of the discovery of novel materials structures and processes is obtained through the use of optimization methods to explore complex energy landscapes. In past decades, the development of optimization methods proceeded mostly independently in mathematics, engineering, and physical sciences. However, advances in optimization methods over the last decade, such as the development of interacting particle methods and surrogate models, illustrate the benefit of multidisciplinary research efforts. Also, the rapid increase in computational power has led to an explosion in available data and the capability to create high-dimensional models based upon this data, e.g., reactive force fields and machine-learning models. To reap the benefits of these

developments requires both a better understanding of the mathematical properties of current optimization methods and the development of new and improved optimization methods that specifically incorporate problem specific knowledge. Many theoretical questions remain unanswered.

Current status of optimization methods. Optimizers can be classified into local and global optimization methods. Local optimizer methods are based on gradient descent and Newton methods such as BFGS. Local optimization methods are often used to identify the optimal structure of materials, saddle points for reaction pathways, and phase transformations. Global optimization methods facilitate structure and composition searches for materials and the selection of model parameters. Several global optimization algorithms are presently available and include methods such as genetic algorithms, particle swarm, basin hopping, covariance matrix adaptation evolution strategy, and Gaussian Process Regression (GPR). Different algorithms employ different strategies to achieve a balance between exploration and exploitation. Exploration is performed by making significant changes to the structure, which can move between basins, while exploitation is performed by making small changes in a basin.

Optimizers can also be classified into single and multi-particle methods. Single particle methods evolve with no inter-particle communication as in steepest descent or Monte Carlo methods. On the other hand, multi particle methods employ multiple parallel processes communicating among each other, and include genetic algorithms or parallel replica exchange.

Use of prior information and surrogate models to accelerate convergence. The configurational spaces of interest are often high-dimensional and typically possess multiple maxima, minima, and saddle points scattered around in the configurational space. However, we are generally interested in just one or a few of those extrema, and as a result the typical strategy of using a random search is highly inefficient. The exploration can be greatly improved by using some external information. For example, in the case of geometry optimization we may improve the convergence rate by exploiting chemical intuition, databases, and previous calculations. Surrogate models constructed using machine learning can also be used to accelerate searching by providing good initial estimates of local minima.

Suggestions for future improvements. As we are considering larger and more realistic systems, efficiency is critical for the global optimization methods. Successful attempts will require the integrated efforts from a variety of subjects including applied mathematics, computer science, physics, chemistry, and biology. One way to improve the efficiency is to harvest the vast prior knowledge of chemistry and materials that has been previously reported or learned on-the-fly.

One way to harvest this knowledge could be to mine existing data and determine structural motifs such as bonding, octahedra, clusters, molecular shapes that can be identified with features of the energy landscape, such as local minima or maxima. This information could then be used to guide Monte Carlo searches or genetic operators in structure optimization.

Most materials need to be represented by many observables. In practice, objective functions that include multiple properties, such as energy, forces, geometries are often needed for better optimization. Some of these properties are highly correlated. For that reason, it is necessary to make a preprocessing of the data to remove redundancy, renormalize different units and reduce intrinsic noise. Additionally, a metric is required to measure structural similarities (relation

between structure and property prediction) in combination with clustering methods in order to increase the diversity of the training set as much as possible and also to detect possible over or under coverage regions in the configuration/parameters space.

Significant opportunities remain for the importation and application of tools from the mathematical sciences. For instance, potential fitting methods could be enhanced by applying a fully Bayesian framework. Currently, loss functions are quadratic, all data points are treated equally, and regularization of the coefficients is performed in an *ad hoc* manner. Introducing a prior distribution and an assumed observational noise on the measurements would allow for a statement of the problem in a probabilistic framework. In this setting, the optimization problem then corresponds to obtaining the maximum a posteriori estimate. While the observational noise may be somewhat artificial, we can then consider the behavior of the coefficients as we reduce the noise.

A generic mathematical formulation of the interacting-particle (replica) methods used in the community should also be developed. Common features of mutation and evolution steps are apparent in all these algorithms. Such a mathematical study would establish well-posedness of the algorithms, e.g., existence of a minimizer, convergence to the minimizer, and stability, and clarify the necessary conditions to obtain a solution. The analysis would also relate performance to parameter and algorithm selection. This work would guide practitioners and could inspire new methods.

While many positive results have been obtained with machine learning potentials, questions remain. In particular, though the potentials may pass standard statistical tests such as cross validation analysis, how these potentials extrapolate to unknown data remains unclear – is there overfitting? If there is overfitting is it catastrophic? These questions could be explored in carefully designed numerical experiments where the potentials are "stress" tested by validating at points very different from the training data.

2. Challenges for Machine Learning of Energy Landscapes

The field of machine learning (ML) was conceived in the mid-20th century and in the past 30 years has seen burgeoning use in the fields of mathematics, physics, chemistry, and materials science. The goal of ML is to bypass expensive calculations from various physical theories with minimal loss of accuracy. Landmark successes of ML methods include learning the exchange-correlation functional in density functional theory, obtaining accurate atomization energies, and screening databases of materials and molecules for application in technology and drug design. These successes have been achieved using several different approaches to ML including, among others, regressions models and neural networks (NNs). Despite this success, there is no good way to establish which model performs better on a given problem in advance and the choice depends mainly on the shape of the dataset and the desired output. Kernel based methods can be solved exactly since they can be expressed in a closed mathematical form with the use of a regularizer, but also can be slow to solve, since large matrix diagonalization is involved. NNs are an attractive alternative to represent general nonlinear problems, though they need a large dataset to achieve good predictions and are harder to train. A mathematical foundation for many ML models is missing and the possibilities in this direction are potentially huge and with applications beyond materials science. We identify five primary challenges that are relevant to our community. The first two are general for the field of machine learning, while the latter three are specific to the

problem of learning effective potentials to describe the energy landscapes of materials and molecules.

Learning from the Machine. Machine learning is clearly a powerful computational tool, which at its core is simply a clever procedure for fitting data. Black-box fitting tools can be useful, but we are also interested in questions about why the data has the form that it does. As the complexity of the ML model grows, the ability to interpret the fit and learn simple relationships that explain the data becomes increasingly difficult. The community would benefit tremendously from tools and/or procedures that could provide understandable structure from complex ML models. As well as being useful for providing scientific understanding, it would be helpful-to inform researchers about the quality of the descriptors being used and how to optimize the hyperparameters in the model. Examples such as autoencoders could be useful for extracting concise information from a network.

Hyperparameter Determination. To apply machine learning requires a "machine", which is a mathematical model defined by a set of parameters that can be varied to achieve a specified performance objective, and some optimization method to determine the optimal values of such parameters. In the case of a NN, the machine is the network of layers and nodes, the parameters are the weights of the connections and the optimization method would be a gradient descent algorithm. Beyond the optimized fitting parameters constrained by data, there are a smaller number of hyperparameters controlling structural aspects of the machine: e.g., the number of layers and nodes, choice of activation functions, network connectivity, etc. While ML's success is due to algorithms and hardware that can determine the fitting parameters efficiently, e.g., TensorFlow and Theano, we lack efficient methods for the determination of hyperparameters. There is strong demand for efficient hyperparameter optimization since the accuracy of the model is greatly influenced by hyperparameter choice. Current approaches include grid searches or random searches which are simple to implement and highly parallelizable, but are not particularly efficient. The computation of gradients with respect to hyperparameters can be very expensive – or even impossible – hence, future development of efficient gradient free techniques will be key to continued high impact of machine learning.

Descriptors. In the context of energy landscapes of materials, key to any ML approach is an appropriate choice of input descriptors which encode representative information about the local environments and/or interactions between atoms, but for practical purposes, are computationally efficient and mathematically simple. Current descriptors include the Behler-Parrinello (BP) symmetry functions, smooth overlap of atomic positions (SOAP), and bond lengths between atomic clusters. A bottleneck for future work is the lack of a systematic study of which descriptors work best for different systems. An important consideration is growth in complexity of the descriptors with the number of elements. For example, BP symmetry functions encode the local environment in tens of inputs to the neural network, while SOAPs generate thousands of inputs. Existing descriptors can scale combinatorially with the number of elements in the system. An important open question is whether or not this increasing complexity is necessary for the construction of accurate machines.

Long Range Interactions. Current methods for constructing machine-learned energy functions in materials start with local atomic descriptors; however, a wide variety of systems involve interactions between charged species, including charge transfer. To improve the accuracy and impact of ML potentials, we need to include our physical intuition about charged interactions that

are already captured in variable charge empirical potentials and reactive force-fields. This provides a unique challenge, as charge interactions are long ranged, while ML potentials are short ranged by their very nature. "Hybrid" potentials that combine long-range interactions and machine-learned interactions – along with the algorithms to optimize them – are required to apply ML potentials for a wide variety of important material systems including complex oxides, interacting metal/oxide materials, and any system where charge transfer controls material properties.

Challenge of Multicomponent Systems. Handling multicomponent systems with varying size and chemical composition requires non-standard network architecture as the input dimensions cannot be fixed. Variable-size sentences and images are usually preprocessed by padding with constants, which are later masked during training; however, such an approach runs counter to basic physical and chemical intuition. Implementation of masking for neural networks is an area of active research, and the backpropagation procedure must be altered to train composition-specific models simultaneously without padding.

Uncertainty Quantification of ML Models and Dealing with Sparse Data. The most common goal for a ML potential is to reproduce the energy of atomic structures as compared to a higher level theory, such as density functional theory; moreover, producing a physically reasonable description is crucial for wide applicability. This is difficult without sufficient input data to span the high-dimensional spaces of atomic structures. One sensible procedure is to fit forces as well as energies as there is more force data which provides additional information about the gradient of the landscape. Additionally, it would be helpful to provide physically-motivated limits for energies, such as a positive divergence as atoms approach each other and an asymptotic approach to zero as atoms become far apart. Moreover, by using ML techniques in the context of materials problems, we have the option of generating the data that is used to train the machine. To exploit this advantage, we need to address important mathematical and practical questions for different machines and problem classes: (i) How many, and which data points, should be generated in order to obtain a prescribed accuracy? (ii) What is the maximal accuracy that can be obtained with a data set of a given size? (iii) What are the techniques for identifying when the machine evaluation is likely to be erroneous? The answer to these questions would facilitate the creation of adaptive data generation where uncertainty is quantified so that when a part of configuration space is reached with sparse data and high uncertainty, new data can be obtained in an automated manner to retrain and reduce the error of the model to a specified level.

3. Long-time dynamics

Investigating the long-time dynamics of materials at the atomic scale is one of the fundamental challenges of contemporary computational sciences. While the problem is conceptually straightforward to solve, standard algorithms such as molecular dynamics exhibit poor parallel scaling, typically limited to 10-100ns of trajectory per day. As a result, the development of specialized techniques that are able to fully exploit petascale and future exascale computational resources is an active area of research. In the following, we review the state-of-the-art and highlight upcoming challenges that face long-timescale methods, focusing on three important aspects: the definitions of the effective coordinates in which the problem is cast, the problem of obtaining coarser models, and that of coupling between scales so as to extend the spatio-temporal reach of atomistically informed models. Mathematically, these questions require careful understanding of Langevin-type dynamics related to such models. These questions include understanding of exit

times from neighborhoods of local minima. Spectral analysis of the associated Laplace-like operators entering in the Fokker-Planck equation has been useful in finding lower-dimensional approximations to these models.

Coordinates/simulation space. In complex systems, extracting long-time information often requires the definition of reaction coordinates or metastable states to accelerate sampling of configuration space. Distinct, but related, approaches have appeared in both the soft and condensed matter communities.

In soft-matter systems, such as proteins or nucleic acids, defining good metastable states is non-trivial because of the highly heterogeneous energy landscape. Techniques that learn these good collective descriptors from molecular dynamics trajectories have made great strides over the last few years, driven in particular by Markov State Models (MSM). However, the search for a general and automated workflow to identify descriptors is an outstanding challenge. Approaches that have shown early promise combine current methods with advanced machine learning strategies such as deep neural networks or autoencoders, though much work remains to incorporate these methods into modern simulation techniques.

In contrast to soft materials, the metastable states of hard materials tend to be well defined in terms of the local minima of the energy landscape. This regime has been well addressed by a variety of methods, including adaptive-Kinetic Monte Carlo (AKMC) and accelerated molecular dynamics. However, systems with strong kinetic heterogeneity, i.e., when both very fast and very slow processes coexist, remain challenging to simulate.

Modern simulation techniques can recover performance by coarsening states into larger groups, but general application of this scheme will require robust methods to define states beyond the metastability criterion. In this respect, insights from the simulation of soft materials, including recently developed kinetic state definitions, offer a promising way forward.

Coarse-graining. High-dimensional systems are often governed by low-dimensional dynamics. Coarse graining into a lower dimensional space introduces its own challenges. Defining a low-dimensional model requires the identification of appropriate 'effective' degrees of freedom and of effective interactions that are expressed in this reduced space.

A key requirement is that coarse graining preserves fundamental quantities, such as thermodynamic or kinetics. Traditional approaches guided by physico-chemical intuition have been very successful, but do not allow for systematic improvement. Leveraging machine learning and data-driven approaches offers a promising solution, but full exploitation requires the development of advanced metrics for quantifying the quality of the approximation.

Scale-bridging. Efficient mesoscale methods must be used when system sizes exceed what can be simulated with direct approaches (e.g., when modeling signal transduction at cellular scales or the annealing of irradiated microstructures (both ~100 μm)). The inclusion of atomistic information, without requiring a full atomistic simulation, is highly desirable to systematically improve model accuracy. Ensuring the accuracy of upscaling methods is challenging due to the lack of corresponding atomistic results, and timescale separations between the meso and micro scales. Robust UQ methodologies are therefore required that can flag when a mesoscale model is

stretching its validity range, and eventually trigger additional microscale simulations to preserve the prescribed accuracy.

4. Surrogate models

From excitation spectra to thermodynamic quantities. Thermodynamic modelling of advanced structural and functional materials typically requires free energy calculations of meV accuracy, i.e., in order to resolve phase transition temperatures to within a few Kelvin. Direct computation to these tolerances requires millions of independent samples of configuration space. High-throughput searches of the vast material space are thus restricted to zero temperature properties. This represents a severe limitation of these approaches to address real word engineering challenges.

A promising solution is to employ surrogate models and efficient/enhanced sampling techniques. Presently, effective harmonic models are the method of choice due to the existence of analytic relations between the phonon (excitation) spectra and thermodynamic quantities. However using this relation for relevant operational conditions leads to unacceptable errors. The design of next generation surrogate models must therefore capture vibrational anharmonicity without sacrificing numerical efficiency nor interpretability. This is distinct from the objective of traditional force field fitting, which attempts to capture the full potential energy surface.

Open questions are: (i) Can we formally define criteria for the optimal surrogate model and use this insight in the construction of such models? (ii) Are there mathematical tools to identify the most compact representation of anharmonic degrees of freedom? (iii) Can models of independent oscillators be constructed that capture the full high-dimensional distribution function? Possible research directions could include anharmonic Einstein models or local (e.g., Gaussian) basis sets as used in quantum chemistry.

Knowledge transfer between disciplines. Active research in force field optimization and kinetic modeling offers the potential for synergy that enhances both methods:

- Modern techniques of force field fitting have developed sophisticated similarity measures between configurations of atoms and molecules. Can these be used to accelerate the building of kinetic models in path sampling research?
- Force fields are built to reproduce equilibrium statistics, but then are usually used to study mechanisms. Can we combine the study of paths and force field building by making force fields that are targeted to reproduce the path ensemble?

Optimal inclusion of experimental data in surrogate models. Even for well characterized materials, experimental data is often too sparse for the construction of surrogate models. However, all practical uses of *ab initio* methods applicable to realistic materials have systematic errors. Thus, the experimental and *ab initio* data sets possess an intrinsic incompatibility which prevents a naïve combination when constructing surrogate models. Methods from the field of uncertainty quantification are a potential route to *designing optimal measures to incorporate experimental data*. The availability of such methods would significantly boost the predictive power and impact of high-throughput simulations.

Can we learn something from the dynamical coarse graining community who regularly confront the problem of determining the relevant states in very high-dimensions? What is the relationship

between partitioning configuration space in order to build a fitting database and the basis functions used in the fit (which inherently provide a distance measure between points in configuration space)? Can the latter be used for the former in some consistent way?

5. Efficient algorithms for spatially localized structural perturbations

Existing *ab initio* electronic structure simulations have yielded, through years of development and improvement, results that are both accurate and efficient. These simulations provide the foundation for the construction and exploration of complex high-dimensional energy surfaces as well as constitute benchmark data for high level modeling activity. Many of the simulations depend on algorithms that assume structural uniformity and periodicity in one or more directions. Under these assumptions, results obtained for small regions at greatly reduced computational cost would then be applicable to much larger regions. However, there is a class of problems where structural perturbations, e.g., defects, occur either through intentional manufacturing or through natural processes. These structural perturbations can strongly influence the resulting properties of the material. The challenge is to develop methods and algorithms that allow one to leverage the computational procedures developed for structurally uniform material to create efficient and accurate simulations of structurally non-uniform materials.

The general approach of combining efficient simulations of less complex problems with a correction step has already been used in the development of simulations of fluid and plasma motion. In particular, the general approach applied to tasks involving linear operators has been quite successful, e.g., variance of domain decomposition techniques. However, there is a considerable amount of work to be done in adapting and extending these procedures to the linear operators occurring in the context of materials simulation. In addition, these approaches could be applied not only at the quantum-mechanical level but also for coarse-grained models such as atomistic and continuum approaches. Two main challenges consist of adapting these techniques in such a way that the high accuracy requirements of materials simulation can be met and, just as importantly, determining how to integrate any of these developed procedures into existing simulation packages. For materials simulations based on density functional theory, the non-linearity of underlying equations gives rise to a whole collection of mathematical and computational problems that must be addressed in order to utilize this general approach even more broadly. If such problems are successfully overcome, the resulting increase in computational efficiency would have a dramatic impact by providing high quality simulation data necessary for a wide range of activities associated with the characterization, design, and fabrication of materials with structural non-uniformity.

6. Efficient Hessian estimation for accelerated local optimization

A common feature in atomic-scale simulation is the search for local minima of a system – whether molecules, nanostructures, or bulk material – that has an O(1000) degrees of freedom. Gradient-based methods are used extensively with density-functional theory, where Newton-based methods for optimization rely on estimation of the (inverse) Hessian. The algorithms start with no, or very little, information about the Hessian for the system, and constructing the Hessian matrix from changes in force with displacements or perturbation theory is prohibitively expensive. Acceleration in the convergence of local optimizers by including information about the Hessian has the potential to impact multiple areas: (i) Molecular systems with a range of stiffnesses from covalent bonds along the backbone to dispersion interactions between more distant atoms; (ii)

Weakly constrained portions of the system that arise from different interactions, such as a molecule that is weakly bound to the surface of a material; (iii) Soft long-range elastic interactions combined with stiff short-range interactions as relevant, e.g., for defects in materials.

In all of these cases, we have physical or chemical intuition about the types of interactions that are present, and the structure of the Hessian; however, the optimization algorithms commonly used in solid state simulations do not take advantage of this intuition. The common feature across all of these systems is a wide spectral width in the Hessian along with significant "off-diagonal" terms for the interaction of atoms, while the algorithms instead start with an estimate of the Hessian that is diagonal and isotropic. Bottlenecks are:

1. Constructing a model that better respects the internal structure of the problem by using internal coordinates rather than using the commonly employed Cartesian coordinates;

2. Constructing a model of information about the Hessian for a given initial structure; and

3. Providing standardized interfaces for widely-used software to take advantage of tools/modules that provide such optimization algorithms.

We believe that significant progress is possible as the stiffness of atomic interactions is approximately known in many cases, and acceleration may not require highly accurate estimates of the Hessian to be effective.

7. Benchmark problems

Assessing the quality of novel simulation methods involves two crucial aspects: 1) comparing the algorithmic performance (e.g., execution time, number of iterations) against established methods on reference problems, and 2) assessing the accuracy of the results against a series of gold standards (i.e., well characterized and generally accepted) test problems established through extensive unbiased direct simulations. As of now, a large number of techniques to explore and characterize high-dimensional energy landscapes have been proposed, but it is uncommon to see systematic comparisons between methods, making it difficult for practitioners to clearly assess the tradeoffs in performance and error inherent to different approaches.

Four major classes of computations are typically performed on such landscapes: optimization, dynamics, sampling, and approximation. Optimization algorithms are used to find both energy minimizing configurations and saddle points. Long-timescale techniques are used to probe the dynamical evolution of the system. In order to extract thermodynamic information, algorithms that efficiently sample the Boltzmann distribution induced by the energy landscape are needed. Finally, approximation methods, from statistics and machine learning, are also often used to obtain computationally cheap surrogates for the true energy landscape.

We therefore propose to adopt and publish nontrivial benchmark problems. This will serve two purposes. First, it will allow one to select the algorithm that provides an optimal tradeoff between performance and errors. Second, it will allow method developers to verify that any newly developed algorithm is successful at solving well studied problems, guarding against method and software development errors in increasingly complicated codes.

We propose to formulate a detailed set of criteria by which test problems and sample output can be submitted to a publicly accessible website, such as http://optbench.org. For example, in order to insure the reproducibility of the benchmarks, all metadata, such as algorithm initialization, stopping criteria, and tolerances, along with notes on compilers and architectures used to generate the results, should be included. Benchmarks should include converged (to within predefined error bars) simulations on reference systems, e.g., first-passage time distributions for kinetics, equilibrium distribution functions for thermodynamics, in order to establish well-defined gold-standards. The website should provide the capability to automatically convert all data to different formats (e.g., plain text, xml, html) to allow for easy comparisons.

8. Uncertainty quantification

The importance of uncertainty quantification (UQ) in the broad areas of materials research and drug design is ever increasing. The common toolchain consists of some of the following upscaling steps: DFT → force-field development → molecular dynamics → finite element/kinetic Monte Carlo → higher level applications, design, and optimization.

Unfortunately, the provision of UQ information is not well established at the initial DFT level of this chain. This differs from most other areas of physics and engineering and is rooted in the specific challenges of the choice of the used functional for DFT calculations, as well as related numerical and modeling errors. Besides being unsatisfactory on its own, this also precludes almost all attempts of uncertainty quantification at subsequent stages of the upscaling chain and may affect both the predictive power of models and the ability to carry out design and optimization tasks reliably. Moreover, current practices focus on energy convergence, while recent studies by NIST show that convergence in energy alone may be insufficient to quantify numerical errors in DFT predictions.

Challenges to advance the current state-of-the-art and impact broad areas of computational science include:

1. Quantifying uncontrolled DFT errors: Establishment of community guidelines of "best-practices" for DFT to determine data scatter – a simple estimate of DFT uncertainty – due to the chosen functional, basis set choice, and pseudopotential. At a minimum, one should compare at least LDA and GGA calculations due to their relatively low computational cost.

2. Quantifying controlled DFT errors: Establishment of community guidelines of "best-practices" for DFT with convergence studies for other quantities of interest, rather than just energy. Developers are encouraged to implement high-level routines to automate scientifically indicated routine tasks, e.g., a lattice constant or band gap optimization, to a given accuracy.

3. Publication guidelines: A protocol for best-practice standards for DFT-simulations should be put forward as a recommendation to journals, e.g., guidelines for referees and organizations (NIST, IAEA, APS, ACS, etc.). Articles should provide sufficient information to reproduce the published results (e.g., providing input files as supplementing information or available in public repositories, see Section 9 *Cyberinfrastructure*).

4. Benchmarks: The community needs to develop a small but sufficient number of benchmarks to test and assess DFT software, and best practices, see Section 7 *Benchmark problems*.

5. Algorithms: Development of efficient algorithms to sample computational parameters, e.g., functional choice, k-point mesh, etc., to efficiently propagate uncertainty from DFT through material models. Overcoming this barrier offers the promise of making uncertainty quantification routine for a wide variety of material computations.

6. Applying stochastic frameworks: Coarse graining problems and dynamical path problems are inherently statistical problems, so fits need to be robust to noise. Force field fitting problems appear to be noise free, but this is only in the statistical sense, as the limitations of model representation makes fits inexact. Is a stochastic framework still suitable for quantifying errors?

9. Cyberinfrastructure

Leveraging available computational resources on different platforms with high-throughput workflows requires increasingly complex simulation protocols. This complexity hampers not only the development but also the interdisciplinary exchange between fields. Therefore, a shared cyberinfrastructure ecosystem is essential to foster sharing data and methods, including several needs in software development.

Optimization Algorithm Ecosystem: An extensible software framework for landscape searches and model optimization would advance the development and application of current and newly developed methods. Currently, users – whether amateurs or experts – have difficulty comparing different optimization approaches due to a lack of interoperability. An *optimization algorithm ecosystem* that could provide access to many different types of optimization methods and model representations for versatile applications can be leveraged for many of the challenges previously identified. Benchmarking requires spanning: one-particle methods, such as simulated annealing and basin hopping; many-particle algorithms, such as replica exchange/parallel tempering, genetic algorithms, and swarm algorithms; and frameworks such as Bayesian optimization and bandit approaches. In addition, new hybrid approaches to optimization would become possible, e.g., switching between and coupling of optimization methods in a nested way, or using active learning to make autonomous decisions on-the-fly on which optimization algorithms and surrogate models to use. If a range of model representations are included for (i) structures, ranging from molecules and clusters to fluids and crystals, and (ii) models ranging from empirical potentials to machine-learning models, then the suitability of different optimization approaches for different problem domains can be established for the community. To be sustainable, software development will require modern software engineering principles and integration with various existing optimization and surrogate model packages (see Community Building below). Finally, integrating materials and model databases with the optimization ecosystem will synergistically increase the impact of both.

Machine Learning Software for Energy Landscapes. There is a growing list of machine-learning software packages with overlapping feature sets. Keras, for example, is a high-level API for the construction of NNs using TensorFlow as the computational back end that is particularly attractive due to its continuous development and support by Google as well as its ubiquity across the ML community. However, since Keras was not developed with application to molecular and material

systems in mind, it lacks the flexibility to design variable hierarchies to learn features at different scales, and to extract derivatives for atomic forces. A ML software toolkit that builds upon the existing Keras functionality and TensorFlow efficiency, and extends it for the study of energy landscapes and related models, would provide an important tool for the study of materials.

Databases. High-quality databases are necessary for exploring energy landscapes that benefit several key areas such as building atomistic and coarse grained force fields, validating and benchmarking new search and optimization algorithms, and accumulating prior knowledge for speed-up. Currently, there are some cases that the same distribution is used for testing a model as for training because of the lack of available credited databases. When transferability to different sampling distributions is desired, it necessitates the creation of combined databases to cover properties and observables. Additionally, quantitative measures of transferability are needed for proper evaluation. Several efforts of individual groups and small teams are providing databases for small molecules, nanoscale, and bulk materials, e.g., https://materialsproject.org, https://aflowlib.org, https://oqmd.org, http://www.crystallography.net/cod/, and https://materialsweb.org. It would be helpful for the community to have common API's to access these database and to provide access to all these databases from a central webserver. Databases are also needed for the surrogate models such as for machine-learning and empirical potentials and for benchmarks, where the OptBench site http://optbench.org/ presents an excellent place to start.

Community Building. The development of sustainable cyber-infrastructures requires community participation, continuity, and leadership. Open-source development, establishing software design principles, and coding best practices are essential for community participation in software development. Regular hackathons and coding workshops can provide the needed continuity for building, maintaining, and expanding shared code infrastructure and databases. These efforts can also serve as a vehicle for community building, training of the next generation of scientists and engineers, and broadening the participation of underrepresented groups. Funding of software development efforts, centers, institutes, and training by NSF and other agencies provides the needed leadership for the advancement of these cyber-infrastructures and ensures that we continue to harvest the benefits of these investments.