

Codesign for materials science: An optimal learning approach

Turab Lookman, Francis J. Alexander and Alan R. Bishop
Los Alamos National Laboratory, Los Alamos, NM 87545

Abstract

A key element of materials discovery and design is to learn from available data and prior knowledge to guide the next experiments or calculations in order to focus in on materials with targeted properties. We suggest that the tight coupling and feedback between experiments, theory and informatics demands a codesign approach, very reminiscent of computational codesign involving software and hardware in computer science. This requires dealing with a constrained optimization problem in which uncertainties are used to adaptively explore and exploit the predictions of a surrogate model to search the vast high dimensional space where the desired material may be found.

Introduction

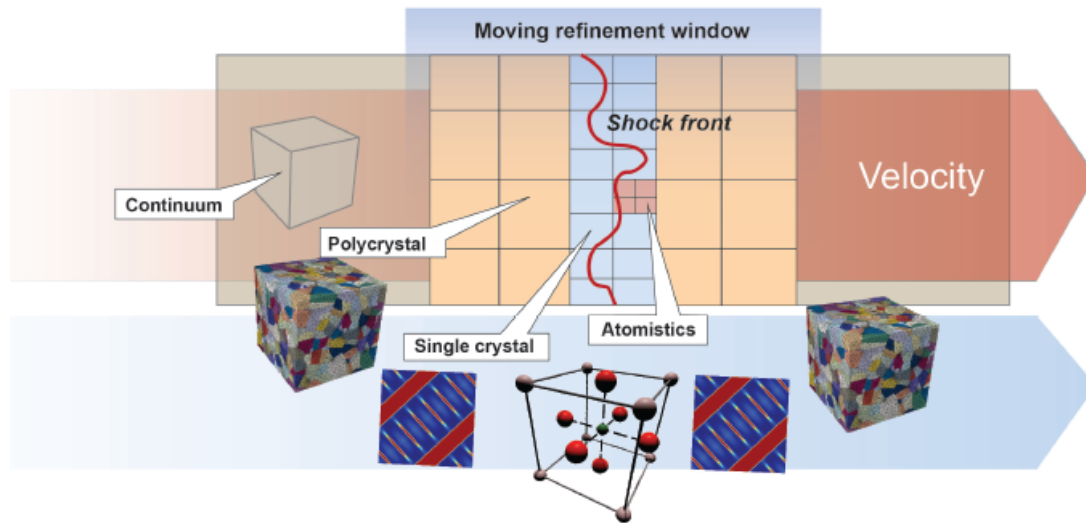


Figure 1 Linking scales via Adaptive Physics Refinement. A macroscale simulation of a shock wave through a bulk sample can spawn higher fidelity smaller length and faster temporal scale simulations dynamically and on demand [9].

Over the last half century there has been enormous progress in our understanding and control of single crystal and homogeneous materials systems. Over the same period there have also been significant advances in both theoretical and computational tools to determine the equilibrium properties of more complex materials. More recently, the roles of extrinsic and intrinsic heterogeneities and disorder and the behavior of materials out of equilibrium have received considerable attention [1]. This progress has led to a strong emphasis on the mesoscale (nanometers to microns) where there is a need for predictive

theories of collective and self-organized behavior. However, it remains an outstanding challenge to model accurately mesoscale physics in terms of coarse-grained fields which are determined from experimental data and/or higher fidelity models (e.g., first principles) at finer scales [2-4]. Currently, there are pressing demands upon U.S. industry both to mitigate supply risks of rare earth and critical materials [5], as well as to accelerate the materials discovery process (i.e., significantly cutting the time and costs from discovery or design to product development and market). These demands have led to the US government initiatives in materials science such as the Materials Genome Initiative (MGI) [6]. Suggested here is a codesign paradigm for materials discovery that optimally integrates theory, experiments and computation as a means to address the demands of these programs

What is Codesign?

Codesign, as originally practiced in computer science, judiciously integrates algorithms, software and hardware to make efficient use of throughput, data access, control and communication in order to solve a specific problem, often under size and weight constraints [7]. Successful computational codesign therefore requires domain scientists (e.g., condensed matter physicists, materials scientists), computer scientists and applied mathematicians working in concert to solve optimally the problem at hand. To demonstrate the feasibility of this concept in a number of different domain sciences, the Department of Energy has funded Codesign centers in the US. ExMatEx, the Codesign consortium led by Los Alamos and Livermore National Laboratories [8] focuses on nonequilibrium shock propagation in solids. To address this strongly nonequilibrium and multiscale problem, physics models at various spatial and temporal scales (polycrystal, crystal, atomistic) generate as needed higher fidelity calculations at successively shorter length and time scales in order to obtain reliable estimates of quantities required for the final macroscale constitutive response under shock conditions (Fig. 1). This “adaptive physics refinement” generalizes the more familiar method of adaptive mesh refinement that is so often successfully applied to solving numerically time dependent multiscale partial differential equations. We believe that many computational grand challenge problems like those in climate and plasma fusion will only be handled effectively using a codesign loop with exascale computers. In this scenario computing resource allocation must be optimized in conjunction with problem formulation.

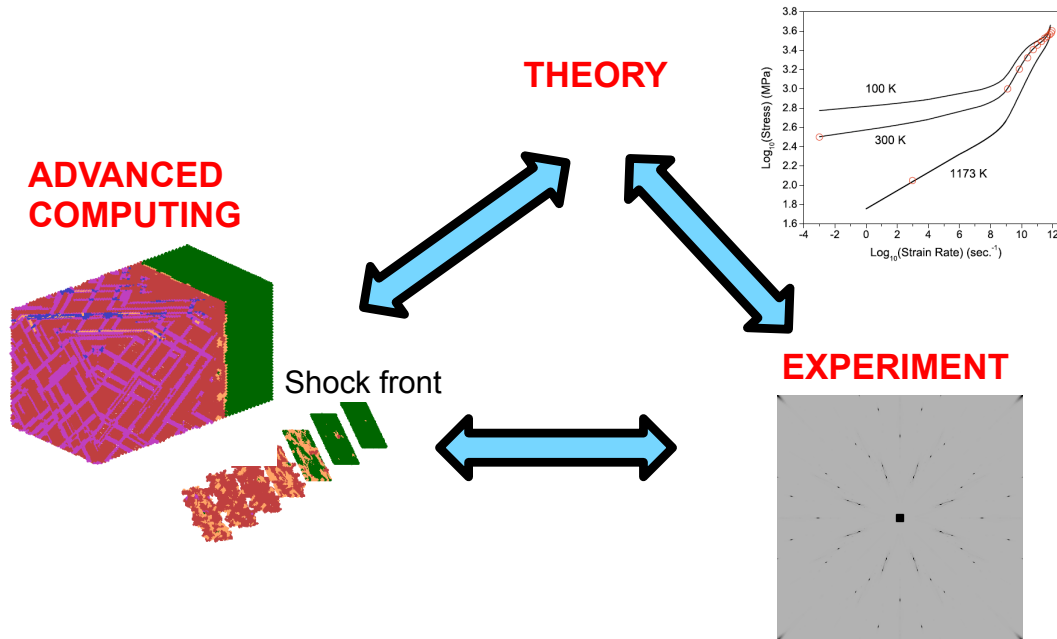


Figure 2 Codesign in a broader sense. Real time feedback between the diffracted images obtained from an experiments, large scale atomistic computation, and expectations from theory.

Recently, the term codesign has begun to be adopted in a broader context in science to refer to the “optimal” integration of theory, modeling, computation *and experiments /observations – an optimal control approach to the scientific method*. In this more expansive interpretation, codesign involves feedback in “real” time between experiments and modeling and simulations. Fig. 2 illustrates a hypothetical but quite feasible x-ray diffraction experiment guided adaptively by a 3D atomistic simulation. The experimentally obtained diffraction patterns in conjunction with additional structural diagnostics, are compared to those calculated from the simulation in “real” time and used in a controller. This controller could modify the energy or direction of the incident x-ray beam in order to probe in more detail certain structural signatures, such as twins. Moreover, the atomistic potential used in the simulations may be modified on the fly by real-time diffraction data in order to improve the predictive capacity of the model/simulations.

This more encompassing use of codesign involves the optimization of, say, a certain quantity or material property by *iteratively* and optimally carrying out experiments, computation and model selection. The model incorporates as much domain knowledge as possible and makes predictions with quantified uncertainties. These predictions in turn adaptively guide, using uncertainties, the next experiments or calculations to be performed (Fig. 3) by balancing the trade-off between exploration of the search space and exploitation of the model results thus far. The optimal solution may satisfy a single or

multiple objectives, such as finding the material with better combination of properties than so far known. The approach here is reminiscent of the area of machine learning known as “active learning” [10]. In active learning the algorithm is allowed (dynamically) to choose the data from which it learns so that it will perform better (statistically) over the long run. The key is learning from and adapting to the environment, and this is what codesign in the broader context aims to do. We discuss in detail the materials domain application and then review a number of science areas where essentially the same framework can apply.

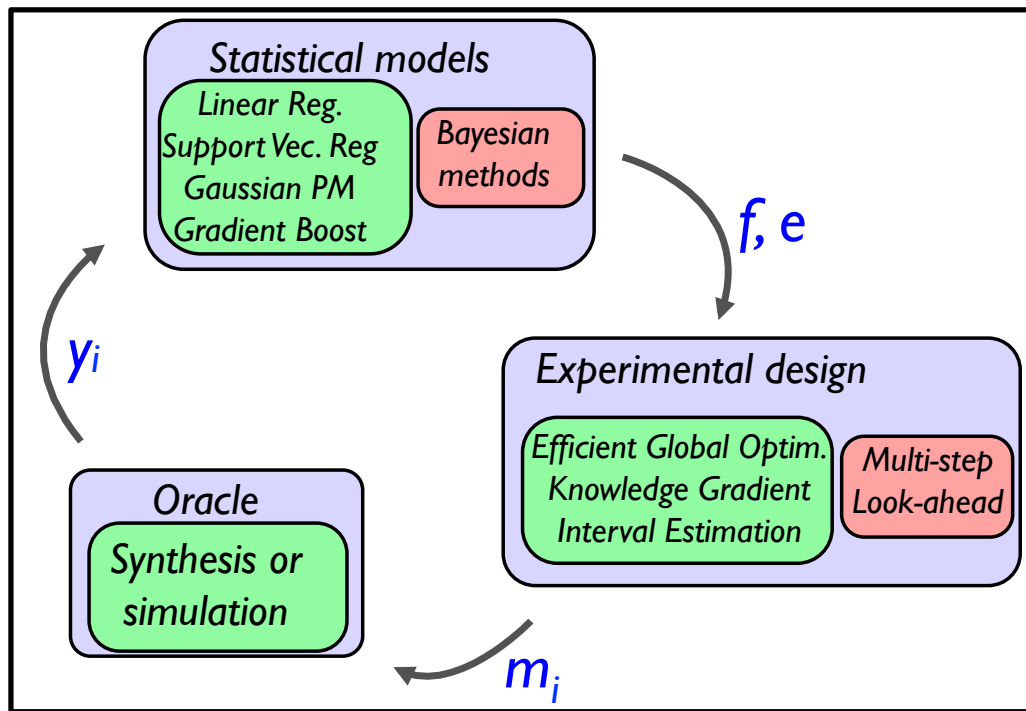


Figure 3 An essential element of codesign. A statistical or surrogate model is iteratively developed from domain knowledge and data and adaptively guides experimental design by motivating the next experiment or calculation to be performed. Statistical inference methods, such as support vector machines (SVMs) or Bayesian approaches for classification and regression, use training data and domain knowledge to “learn” a surrogate model or map, f , for the property, y , to be predicted with uncertainty, e . Experimental design guides the selection of the next material, m , for example, for an experiment or calculation by balancing trade offs between exploiting the current information or exploring the phase space for model improvement. The measured or calculated property, y , typically aids to reduce the uncertainty in the model, thereby improving it in the next iteration [11].

Materials Codesign

Materials design, especially for multifunctional materials such as energy harvesting and storage materials (e.g. ferroelectrics, electrocalorics, magnetocalorics), can be formulated within a codesign framework with its emphasis on optimization and iterative model improvement. Fig. 4 gives an example of an accelerated materials discovery codesign loop [11]. Part of the execution of the codesign process is to guide the "most informative" next experiment or calculation to be performed. The goal may be, for example, to design materials with certain specified equilibrium properties or dynamical features. Consider ferroelectrics as an example. For this class of materials one may wish to find lead-free piezoelectrics with a high transition temperature and/or with a high piezoelectric coefficient. In the case of shape memory alloys, one may be looking for

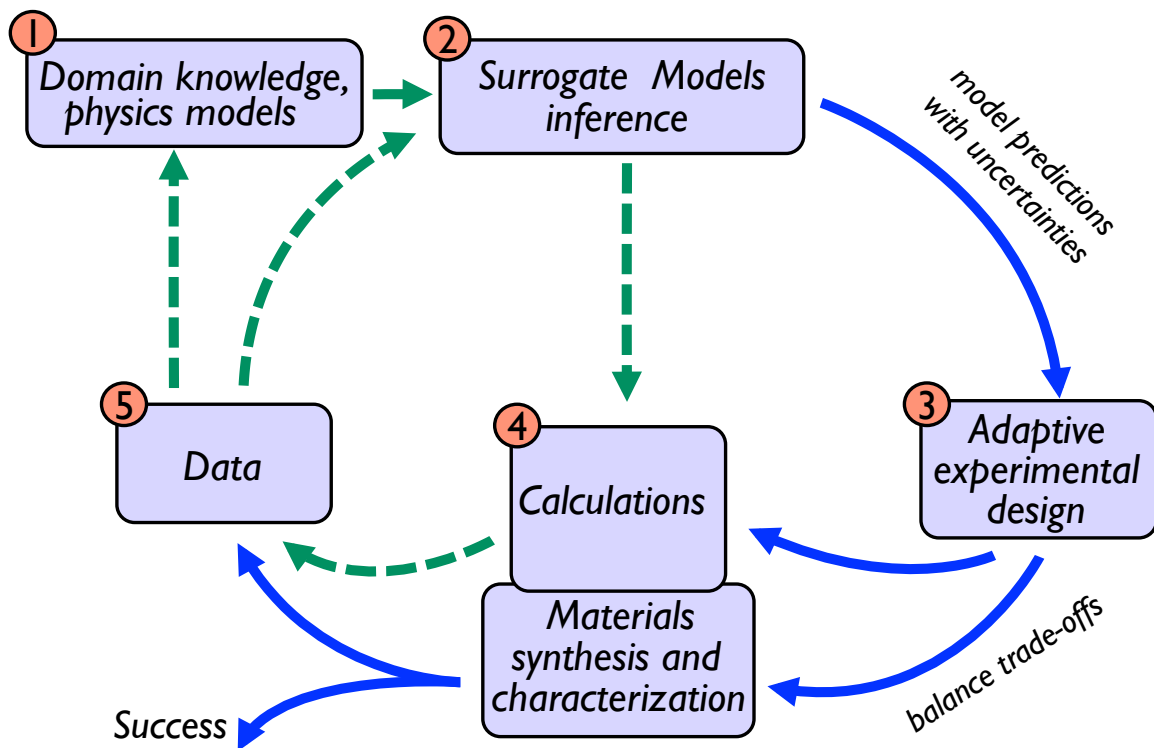


Figure 4 A codesign loop for accelerated materials discovery. The key idea is to iteratively improve model predictions by guiding and using results from experiments and calculations. Green arrows denote the typical individual single steps taken without feedback in most studies today [11].

compounds with greatly reduced thermal dissipation or very low hysteresis in order to minimize fatigue. Traditionally the search for such materials is costly and quite time-consuming, largely driven by trial and error. However, within the last decade both materials/condensed matter theory and high-performance computing have matured sufficiently to be able to predict very accurately some material characteristics from first principles. While current *ab initio* calculations are able to predict elastic constants, inter-atomic distances, crystal structure, polarization, etc, the parameter space is simply too large and there are too many possibilities for a random search for materials [12]. As a result, one needs a principled and efficient method to “learn” from available theory,

simulation and experimental data how to find candidate materials for further experiments and calculations. First one must develop a statistical model for the codesign loop with estimates of materials properties and their associated uncertainties. In executing the loop one should (as in many search algorithms) balance the tradeoffs between exploration and exploitation. The result is an optimal policy given all that is known. This policy suggests the next experiment to be performed on the material predicted to have the largest “expected improvement” [13,14]. This balances the need to exploit the results of the model or to improve the model by performing the experiment or calculation on a material where the predictions have the largest uncertainties.

Essentially the same framework can apply to another problem of much recent interest, namely, additive manufacturing, where the challenge is to predict the optimal combination of material composition, processing, microstructure, properties and performance [15]. In one type of additive manufacturing a powder precursor undergoes rapid melting and solidification as each subsequent layer is deposited during the building process. As the build progresses, these previously deposited layers then experience large temperature gradients and enhanced thermal stresses during the entirety of the part. While the deposit is molten, the material undergoes significant convective flow, potential alloying element vaporization, and steep thermal gradients as it is rapidly heated by a laser. After solidification, the build undergoes complex thermal cycling as new material is placed on top of that previously deposited. Substantial progress has been made by exploring modes of synthesis, characterization and formulating models, some of which are simulated on high performance computers. The current approach, however, is usually based on experience, intuition, performing controlled experiments and incorporating better physics of solidification and phase transformation into empirical or phenomenological models [16]. The state-of-the-art exploits the similarities between physical processes governing laser welding and laser-based advanced manufacturing. Models are used to analyze how changes in laser processing parameters, which maintain a target melt geometry, can impact thermal cycles, fluid flow, and solidification during laser deposition. This allows for the study of how changes in heat input, in order to maintain a target melt geometry, lead to variations in the solidifying microstructure throughout the part. Where possible, knowledge in the form of correlations between composition, processing conditions and microstructure is used to guide exploration. However, this approach is not a recipe for accelerating the design and discovery process or learning how to design a product with given specifications. Acceleration requires posing the inverse question: Given a certain targeted performance (e.g. under shock loading), what constitutive properties (stress-strain) and hence microstructure (grain size), and therefore processing conditions (solidification rate, thermal gradient) and perhaps optimal composition (say Cr to Ni ratio for steels) must the material possess? Why is this a problem? Because the space of possibilities is very high dimensional and this cannot be successfully probed via experimentation alone or by merely using a high fidelity multiscale model with parameters, even if such a model were to exist. Hence, what is required is to use the loop of Fig.4 to successively home in on the set of features (processing conditions, microstructural aspects), within the high dimensional space,

which fulfill the requirements for the targeted material – in as short a time (or with other resource constraints) as possible.

A number of issues need to be addressed and challenges overcome in order to determine the optimal codesign policy. These include specifically which experiments and which calculations are required to train effectively the machine learning algorithms for classification (is a compound thermodynamically stable, is it a piezoelectric, does it have an hcp crystal symmetry?) and regression (what is the material's piezoelectric coefficient? what is its transition temperature?) Although we live in an era of big-data, there is actually quite a limited amount of training data in materials science (and other fields as well). Moreover, the space of composition/processing possibilities is of such high-dimension, that it is necessary to incorporate as much domain knowledge as possible to construct and execute the codesign loop. A Bayesian approach, where domain knowledge is encoded into prior probability distributions offers a natural strategy [13,14,17]. Since the numerical calculations of material properties inform the exploration-exploitation process, accurate error estimates of these calculations are crucial for effective predictions. Unfortunately, when the data is limited, accurate error estimation is a difficult undertaking. Cross-validation, an often used error estimator, can perform extremely poorly on small data sets [17]. We can partially remedy our ability to provide confidence measures by incorporating domain knowledge into the error estimator. This domain knowledge constrains statistical outcomes and yields classifiers that are superior to those designed from data alone. The need to construct better theories in condensed matter and materials science is by no means obviated by increases in data. We will likely be relatively data-poor for quite a while.

Finally, given a choice of possible experiments, we need to consider which experiment is optimal relative to finding a classifier (or regressor) that possesses minimum error. The preceding Bayesian framework is ideal for solving this problem because it provides a mathematical representation of the classifier and its error, as well the potential conclusions from any experiment (in genomics this work has led to the design of optimal experiments for improving drug intervention in genetic regulatory networks). Other strategies balance exploitation (maximizing desired properties) with exploration (improving our understanding of completely different materials). For example, compounds that are similar to the best known piezoelectrics are likely to be well predicted by the models, and are likely to be good piezoelectrics; but learning about materials that are not so similar to those with known properties can potentially provide more information, and possibly uncover a material with extremely large piezoelectricity. Techniques based on solving ranking and selection problems, which allow for a form of global optimization, such as efficient global optimization (EGO) [13,14] and its variant the knowledge gradient [18] algorithm, choose a measurement to maximize the “expected improvement” or gradient with respect to the knowledge gained from the measurement. For each possible compound X in the search space, the statistical models inferred from the machine-learning step estimate its payoff Y within some confidence interval ΔY . New experiments are selected to maximize the estimated value $Y + 2\Delta Y$; this has been shown to be very effective in a number of problems. In this way, one favors materials with a large payoff (“exploitation”), and also ones that could potentially

surprise us because of their large uncertainty (“exploration”).

The codesign approach outlined above can be applied across the sciences. It can be adopted where one has a clear target in mind and the complexity of the experiment, simulation, theory, uncertainty landscape is such that one cannot navigate based simply on intuition alone. For example, such an approach is likely to reap benefits in cancer [19] in particular and medical research more generally. In the biological arena the systems are generally exceptionally complex, and while models exist they tend to contain a great deal of uncertainty. Moreover, despite the deluge of data coming from high-throughput sequencing and experiments, the predictive capacity of these models is poorly characterized because the actually useful data is often still quite limited. A codesign approach to time-domain astronomy is also likely to pay dividends. In this emerging area one searches the night sky for transients such as e.g., gamma ray bursts. Success relies on rapid machine learning /anomaly detection algorithms, tied to high-throughput hypotheses testing and near real-time reallocation of observational resources.

Codesign represents a merging of ideas from operations research, design of experiments, statistics and machine learning. It is an optimal approach to the scientific method and the solution of the optimal design problem may itself be computationally demanding. Of course, as with most interesting stochastic control problems, approximations will have to be made and a suboptimal solution will result. Nevertheless, the expectation is that the suboptimal solution will be significantly superior (in accuracy, use of limited resources etc.) than an ad-hoc empirical approach.

References

1. A.R. Bishop, K.O. Rasmussen, J. Roder, T. Lookman, A. Saxena, A. Vanosi, P. Kevrekidis, in Complexity at the Mesoscale, in Nonlinearity and Disorder: Theory and Applications. eds. by F. Abdullaev et al. (Kluwer, Norwell, 2001), pp. 99–113
2. J.A. Krumhansl, Viewpoints on microscopic and mesoscopic transformation processes in martensite. *J. de Physique IV*, 5, Coll. C2, C2-3:14 (1995)
3. G.W. Crabtree, J.L. Sarrao, Opportunities for mesoscale science. *MRS Bull.* 37, 1079–1088 (2012)
4. T. Lookman, Heterogeneities, The Mesoscale and Multifunctional Materials Codesign, in “Mesoscopic Phenomena in Multifunctional Materials”, Springer Series in Materials Science 198, 2014.
5. US Department of Energy Report “Critical Materials Strategy”, 2011, http://energy.gov/sites/prod/files/DOE_CMS2011_FINAL_Full.pdf.

6. Materials Genome Initiative for Global Competitiveness, White house Office of Science and Technology Policy, June 2011
https://www.whitehouse.gov/sites/default/files/microsites/ostp/materials_genome_initiative-final.pdf

[7] G. De Micheli, Computer-Aided Hardware-Software Codesign, IEE Micro (Stanford University),
<http://icwww.epfl.ch/~demichel/publications/archive/1994/IEEEMICROvol14iss4Aug94pg10.pdf>

[8] ExMatEx, Extreme Materials at Extreme Scale,
<http://www.exmatex.org/>

[9] MaRIE: Matter-radiation interactions in extremes, LANL facility future, Vistas, LALP-10-059; MaRIE Theory, Modeling and Computation Roadmap: the Materials World beyond Bloch and Boltzmann, LA-UR-10-03507 (2010)

[10] M. Prince, Does Active Learning Work? A Review of the Research, J. Engr. Education, 93(3), 223-231 (2004).

[11] T. Lookman *et al.* A perspective on Materials Informatics: State of the Art and Challenges, p 3-12, in Information Science for Materials Discovery and Design, Springer Series in Materials Science 225, 2015.

[12] R. Armiento, B. Kozinsky, M. Fornari and G. Ceder, "Screening for high-performance piezoelectrics using high-throughput density functional theory", Phys. Rev. B 84, 04103-04115 (2011).

[13] D.R. Jones, M. Schonlau, M. and W.J. Welch, Efficient global optimization of expensive black-box functions, Journal of Global Optimization, 13, 455–492, (1998).

[14] D.R. Jones, A taxonomy of Global Optimization Methods based on Response Surfaces, Journal of Global Optimization, 21, 345–383, (2001).

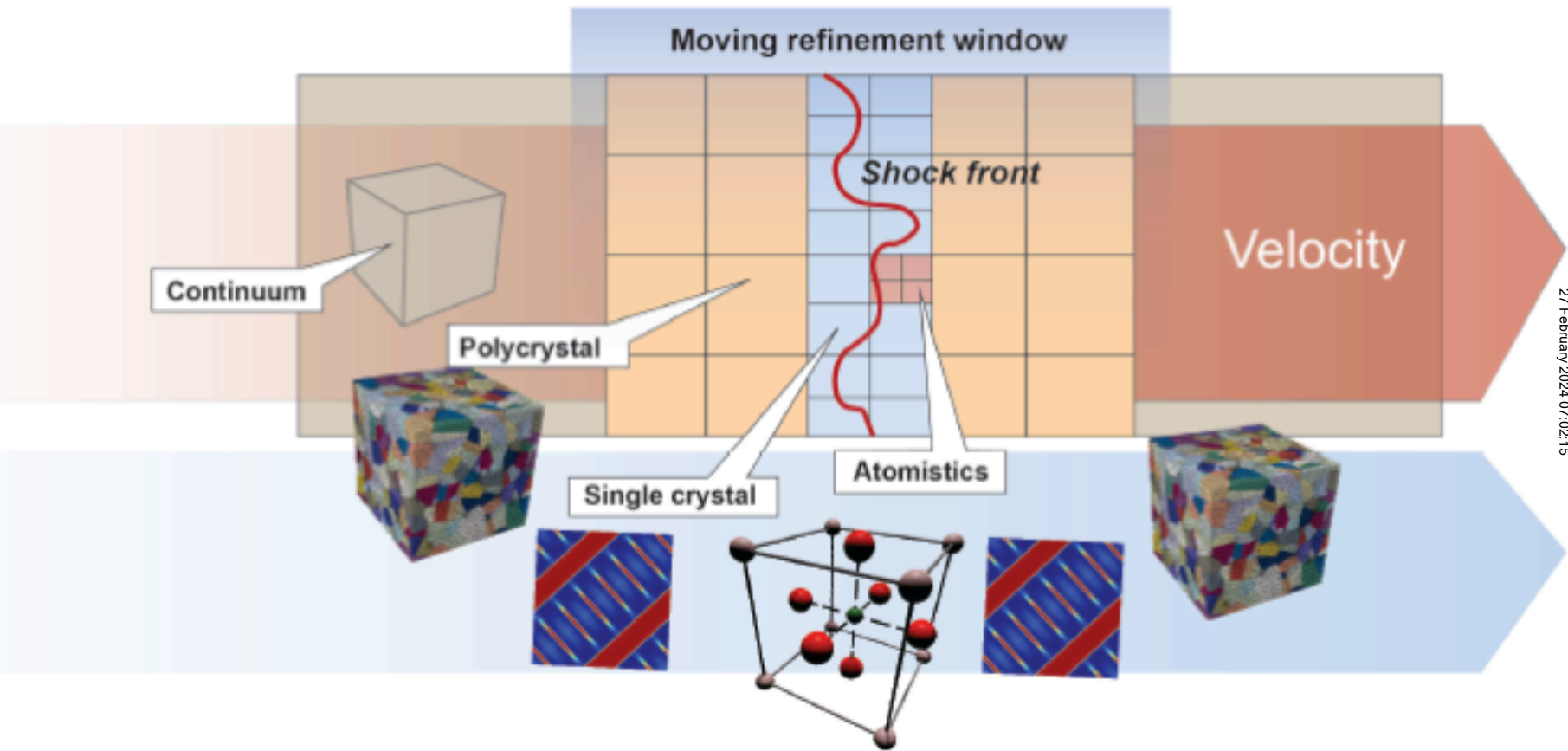
[15] Frontiers of Additive Manufacturing Research and Education, Report of NSF Additive Manufacturing Workshop, Y. Huang and M.C. Leu, (2014).
<http://nsfam.mae.ufl.edu/2013NSFAMWorkshopReport.pdf>

[16] J.J. Blecher, T.A. Palmer, and T. Debroy, Solidification Map of a Nickel-Base Alloy, Metallurgical and Materials Transactions A, 45a, 2142, (2014).

[17] L. A. Dalton and E. R. Dougherty, "Bayesian minimum mean-square error estimation for classification error—Part I: Definition and the Bayesian MMSE error estimator for discrete classification", IEEE Transactions on Signal Processing, 59(1):115–129, January 2011.

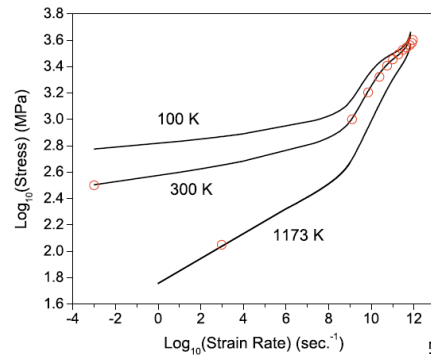
[18] W.B. Powell and P. Frazier, Optimal Learning, Tutorials in Operational Research, INFORMS, 2008; W. B. Powell, “The Knowledge Gradient for Optimal Learning”, Encyclopedia for Operations Research and Management Science, Wiley, 2011.

[19] Dougherty, E.R. and Bittner, M.L. “Epistemology of the cell: a systems perspective on biological knowledge”, 34, (2011).



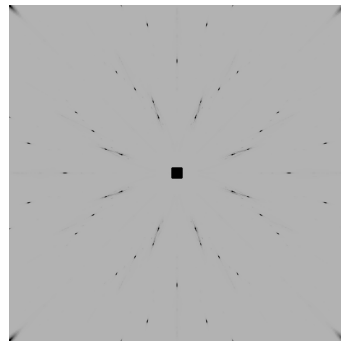
**ADVANCED
COMPUTING**

THEORY

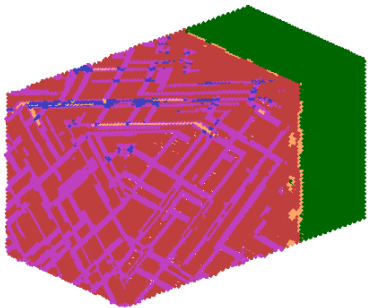
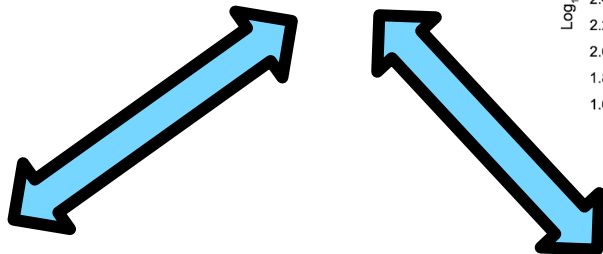


27 February 2024 07:02:15

EXPERIMENT



Shock front



Statistical models

Linear Reg.
Support Vec. Reg
Gaussian PM
Gradient Boost

Bayesian
methods

f, e

Experimental design

Efficient Global Optim.
Knowledge Gradient
Interval Estimation

Multi-step
Look-ahead

m_i

Oracle

Synthesis or
simulation

y_i

