

MedeA: An environment for ICME starting with atoms, and what we can learn from it.

Paul Saxe

Materials Design, Inc., 6 First National Place, Angel Fire, NM, USA psaxe@materialsdesign.com

ABSTRACT

ICME is a grand vision, but there is an enormous amount of work needed to bring together experiment and simulation across the entire spectrum of length- and time-scales to go from atoms to cars, airplanes, computer chips or any other manufactured product. In this talk I will briefly cover what we have accomplished to date, focusing mainly on the common themes that emerge over time. The attention will then turn to the challenges ahead – what they are, and what the technical, societal and financial hurdles that will have to be overcome to sustain this effort, as well as the more general ICME effort, over the long time it will require to reach the goals laid out by e.g. the Materials Genome Initiative.

Keywords: icme, software, modeling, simulation, challenges

1 INTRODUCTION

Integrated Computational Materials Engineering (ICME) as it is laid out in the National Research Council of the National Academies of Sciences [1] report and the subsequent Materials Genome Initiative [2] (MGI) of the White House is a grand vision and a most worthy goal. However, between here and there is an enormous amount of work needed to bring together experiment and simulation and bridge the gaps between the many time- and length-scales from electronic structure to final manufactured products. Eighteen years ago we set ourselves precisely this task, though we did not call it ICME because the term was yet to be coined. But that was definitely our goal, starting from the bottom with the shortest time- and length-scales, those of electrons in *ab initio* quantum calculations. We have progressed up the scales, albeit not very far, creating a link from the electronic structure world to empirical potentials, or forcefields, which give the ability to handle systems ten's of nanometers in size and times up to microseconds.

While this is still small, it allows us to work with defects in solids though not yet microstructure; with fluids; and with amorphous materials such as polymers as well as alloys, investigating whether they are disordered or ordered. And it allows us to look at transport properties such as diffusion, thermal conductivity and viscosity. It also allows us to incorporate temperature into the simulations in a more practical way than we can with quantum mechanical methods, so we can predict the temperature dependence of mechanical properties, etc.

Yet there is vastly much more to do, so now that we have a firm footing at these atomistic levels we are reaching up to larger scales, with TCAD in electronic, CALPHAD methods for alloys and metals, and are starting to investigate connections with computational fluid dynamics (CFD) and finite element analysis (FEA).

Accomplishing the lofty goals of ICME will require a large sustained effort. If our experience is any guide this will be some decades of work, and as far as we can see the heart of the problem is understanding. Understanding each other, gaining a practical understanding of simulations, understanding how to connect simulation with experiment, how to mine all the data for solutions, and in general trying to understand a very complicated physical reality. Bluntly put, this boils down to a question of funding. Unfortunately the current approach by funding agencies, combined with the unthinking promotion of open – open-source, open-data, etc. – clearly and obviously cannot reach the goals laid out. To compound the problem, since the vision of ICME relies substantially on modeling and simulation, the current approach to high-performance computing with large centers is also working against the goal. The security and access restrictions force using the command-line and homebrewed scripts to accomplish otherwise straightforward tasks, and the emphasis on large calculations – capability rather than capacity – both mitigated against the very concept of 'integrated'.

2 CURRENT CAPABILITIES

The place to start is where are we now. I will describe the state-of-the-art in a small corner of the ICME space, quantum and atomistic modeling and simulation, in part because it is the area that I know well, but mainly because it is a good place to start because in many ways it is easier than other areas. It is easier because quantum simulations generate physical property data from general equations without recourse to experiment beyond the values of fundamental constants such as the speed of light and mass of an electron. Surely, can we not describe these simulations precisely, such that anyone can reproduce them? And since quantum simulations work on well defined systems of atoms, can we not save and later search these systems and the results? And if we can do this, isn't it just a matter of persistence and patience to build up to longer length and time scales, and to connect this simulation data to experiment?

As we will see, apparently not.

2.1 Capabilities of MedeA: Databases

MedeA[®] [3] is a comprehensive environment for modeling and simulating materials at the level of atoms. It provides a graphical user interface (GUI) for building models, for setting up and running simulations, and for retrieving and analyzing the results. Within the environment the user can access experimental databases of crystallographic structures: the Inorganic Crystal Structure Database (ICSD) [4], Pearson's Crystal Data [5], NIST Crystal Database (NCD) [6], the Pauling File binary phase diagrams [7] and the Crystallography Open Database (COD) [8]. These are handled as a meta-database, meaning that they appear to the user as a single large database containing essentially all the inorganic crystal structures reported in the literature since, well, the invention of crystallography.

The same database framework can be used in the other direction to seamlessly store models built within MedeA or which result from simulations. These stored structures then appear and are accessed on an equal footing with the experimental structures, but are kept separate so there is no inadvertent contamination of the experimental structures.

2.2 Building Tools

The experimental database provide a rich foundation of models for simulations, however, for simulations one often needs hypothetical structures or structures with defects, interfaces, etc. MedeA provides a range of builders and editors for creating such models. These range from simple tools for editing atoms – deleting them, changing the element, etc. – to typical sketchers for creating organic and organometallic models. Beyond this there are a range of more complex builders for building nanotubes and nanostructures, for building interfaces and grain boundaries between dissimilar materials, to builders for amorphous polymers and thermosets. Building models is and will always remain an area of active development because as our simulation tools become more capable, there will be a need to build more complex and realistic models.

2.3 Simulation Engines

After building one or more models, the next step is simulation in order to predict structures and various physical properties. MedeA integrates a number of leading tools for such simulations. On the quantum side, Gaussian [9] and MOPAC [10] handle primarily molecular systems, though both can also handle periodic, crystalline systems. VASP [11] remains the workhorse for periodic density functional (DFT) calculations. Moving up a little in scales brings us to the forcefield based codes LAMMPS [12] for molecular dynamics (MD) and Gibbs [13] for grand-canonical Monte Carlo methods.

2.4 Extended Capabilities

While these are the main energy codes – fundamentally they provide various ways to calculate the energy and forces of an atomic system – MedeA also provides a number of major codes in their own right that wrap around the energy codes, using the energy and forces to calculate useful properties. PHONON, as its name suggests, computes the phonons of a crystal, and from them the vibrational contributions to the thermodynamic functions. This brings temperature into, in particular, the quantum calculations. This is a major step to meeting experiment since the electronic energy, which is what is typically reported from quantum calculations, is approximately the enthalpy at 0 K, and not even that since zero-point effects are left out.

A second major tool layered on the energy codes is the transition state search (TSS) module. The transition state is a key concept in chemical reactivity. Once the transition state is known, we can calculate the barrier height and with the aid of Phonon, the attempt frequency. This information gives us the reaction rate, starting to give us a good handle on reactivity at the atomic level.

A code such as VASP not only brings us the ability to calculate the structure and energetics of a system, but it also produces a great deal of information about the electronic structure of the system. The Electronics module in MedeA takes this a step further, using the band structure calculated in VASP as the input for both understanding the Fermi surface and using it for calculating electronic transport properties such as the conductivity and the Seebeck coefficient, which is a measure of the goodness of thermoelectric devices. These electronic properties clearly bring a connection to the technology computer aided design (TCAD) widely used by the electronics industry for the design of semiconductor chips.

Cluster expansion methods, such as those embodied in UNCLE [14] move us away from regular crystalline materials to alloys. Cluster expansions provide a way to explore the compositional phase diagram, predicting whether there are ordered or locally ordered structures or truly random alloys for a given composition. And by working across the composition range a picture of the phase diagram starts to appear. Through Monte Carlo methods, UNCLE also brings temperature into the picture. Along with the structure and energetics from VASP and Phonon, UNCLE starts to build the bridge to CALPHAD methods such as those from Thermo-Calc.

Since MedeA is an integrated environment, these powerful tools that are layered on top of the energy codes work with any code that provides the needed input. For example, while traditionally used with DFT codes like VASP, both Phonon and UNCLE will work with any other code that provides energies and forces, for example MOPAC and LAMMPS. This allows a user to use a less expensive, though possibly less accurate, method. This can be useful for prescreening, or to get initial guesses for more refined calculations.

2.5 Analysis Tools

The final part of MedeA is analysis. Apart from a range of geometric tools, it is not easy to describe the range of tools since they are specific to the types of calculations. Suffice it to say that there is a wide range of capabilities. As examples it is straightforward to get diffusion constants, viscosities, thermal conductivities and many similar properties.

In summary, MedeA provides an easy-to-use graphical environment with a wide range of database, builders, simulation engines, and analysis tools. While there is always more to do – there are always new systems to build, new analyses and new simulation methods – it is already a large environment with millions of lines of code in a number of languages including Fortran, C, C++ and Tcl.

MedeA is certainly not unique in the field. There are other commercial codes that are similar in outline and size. There is not, however, an open-source framework with anything like the capabilities of MedeA or the other commercial codes, which is an interesting point that we will return to later.

3 EXAMPLES AND COMMENTS

In this brief paper there is not much room for detailed examples, but let's look at three that cover a range of applications.

3.1 Elastic Constants of Alumina

As a very simple and straightforward example, let's look at the elastic constants of alumina ($\alpha\text{-Al}_2\text{O}_3$). Alumina is a well-known widely used material and as such is a natural benchmark for theoretical methods. Table 1 gives the calculated and experimental values for the elastic constants:

	Calculated [15]	Experiment [16]
C_{11}	495	500
C_{12}	171	162
C_{13}	130	111
C_{14}	20	-23
C_{33}	486	502
C_{44}	148	151

Table 1: Elastic constants in GPa of $\alpha\text{-Al}_2\text{O}_3$

There is certainly reasonable agreement between the calculated and experimental values of the elastic constants, with the notable exception of C_{14} , where the magnitude of the values is reasonable but the sign differs. This discrepancy is surprising because the experimental work was a very careful study by a well known, highly regarded group. And indeed all subsequent experimental determinations agreed on the sign for C_{14} . The discrepancy was resolved by a careful joint computational – experimental study [17] where great care was taken to ensure that the orientation of the alumina crystal was maintained. This showed convincingly that the earlier experiments had the incorrect sign for C_{14} . While not in itself terribly significant, this is an important reminder that experiment can be quite wrong and that one role of simulation is to find inconsistencies in experiment. In this

case, the original experiments were most likely performed on a crystal with its orientation reversed from what it was thought to be, or equivalently on the other enantiomer of this chiral crystal.

3.2 Mechanical Properties of Thermosets

Continuing the theme of mechanical properties, the next example is the mechanical properties of three closely related thermosets. The resins diglycidyl ether of bisphenol A (DGEBA), triglycidyl p-aminophenol (TGAP) and tetraglycidyl diaminodiphenylmethane (TGDDM) were cross-linked with 4,4'-diaminodiphenylsulfone (4,4'-DDS). The three resins are chemically similar but have 2, 3 and 4 cross-linking sites respectively. Experimentally this change in the architecture changes the mechanical properties. Simulations reproduce the trend almost exactly:

Resin	Calculated	Experiment
DGEBA	3.49-3.53	2.4-3.2 [18]
TGAP	4.42-4.45	4.396±0.027 [19]
TGDDM	5.18-5.19	5.103±0.033 [19]

Table 2: Tensile moduli in GPa for crosslinked epoxy resins

The extent of reaction for the DGEBA system experimentally was 0.5-1.0, which probably explains the wide range of the modulus. For TGAP and TGDDM the extents of reaction were 0.93 and 0.88, comparable to the simulation. The simulation almost exactly reproduces the experimental values, given the caution about the quality of the DGEBA samples.

3.3 Schottky Barrier

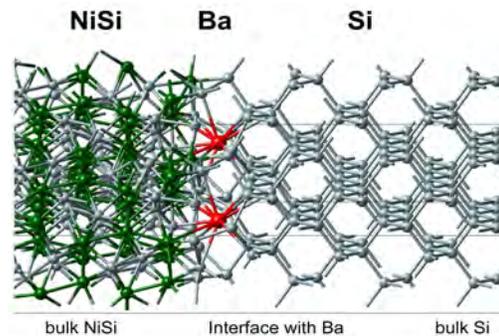


Figure 1: Model of NiSi-Ni interface doped with Ba

Figure 1 shows a model of an interface between nickel silicide and silicon. It is shown doped with barium in the interface. Calculation of the Schottky barrier height for the undoped material gave 0.62 eV, in excellent agreement with the experimentally determined value of 0.65 eV [20]. Doping with barium, as shown in Figure 1 reduces the Schottky barrier height by 0.2 eV to 0.42 eV.

4 DISCUSSION

The first sections of this paper have quickly talked about the capabilities for modeling and simulation of the MedeA environment, followed by three quite disparate examples that show how general and powerful a tool this is. The main message is that, while MedeA covers only a small corner of the ICME landscape, it is nonetheless a large, complicated environment composed of millions of lines of code and hundreds of years of effort. The examples also briefly hinted at the complexity of comparing to and understanding experimental results, but no more than a hint about this complex subject.

What are the challenges going forward? Hopefully I have made the case that developing such a system is difficult, painstaking work that required deep understanding as well as considerable time. While additional resources can certainly decrease the time needed for developing the large code base in such an environment, it is less clear that it can speed up the understanding needed.

As I noted earlier, there is no open-source environment that approaches the capability of any of the commercial solutions. This might be a coincidence, but I am inclined to believe it is a fundamental limitation of open-source in this arena. Open-source is highly successful for widely used codes such as operating systems (Linux), web servers (Apache) and word-processors (Open Office). The common thread is that the application is limited. It does one things, perhaps quite complex, but nonetheless conceptually straightforward. In general we – or at least some group of people – understand almost completely what an operating system or web server needs to do. In that sense, these tasks are understood, at least by some.

ICME is very different. There are huge swaths of the underlying science that are not well understood. While we, or Google, can readily describe a document, we have no detailed description of most materials. What is steel? There are thousands of recognized steels based on composition and processing. The very fact that processing plays a key role is a strong warning of the underlying complexity. Science is not word processing!

Another hallmark of successful open-source projects is that they are self-funding. That there is a large enough community to support the project without external funding. In our ICME world the situation is exactly the opposite. Almost 100% of the effort is funded by governments through funding agencies. A small amount is funded by sales of ICME products to industry. What will happen when government funding diminishes or moves to other areas, as it inevitably will? Who will support the tools then?

In other areas such as CFD and FEA, a distinct community of service companies has developed, companies like MSC, Ansys, Dassault Systemes. This service industry has revenues in the billions of dollars and invests hundreds of millions in R&D. And it needs to be of this size and effort to have the needed impact on such large industry sectors as automotive, which must exceed \$1 trillion worldwide.

Rather than following this model, in ICME the assumption is that a consortium of academic and national laboratory researchers can supply the end users in the manufacturing industry, without a service sector. And that the funding will come from the government in perpetuity. Rather than promoting the development of a service industry, this funding is actively competing with and undercutting the fledgling industry that is trying to form.

The question is will this government funded academic-industrial approach, with no service industry, and no long-term funding based on the commercial use of ICME – will it suffice to radically change massive industries? Will a few hundred million dollars of government funding spent in academia change trillion dollar industries?

REFERENCES

- [1] National Research Council. (2008). *Integrated Computational Materials Engineering*. Washington, DC: The National Academies Press.
- [2] <https://www.whitehouse.gov/mgi>
- [3] MedeA-2.19, Materials Design, Inc., Angel Fire, NM, USA, 2016. MedeA[®] is a registered trademark of Materials Design, Inc.
- [4] <https://www.fiz-karlsruhe.de/en/leistungen/kristallographie/icsd.html>
- [5] <http://www.crystalimpact.com/pcd/>
- [6] <http://www.nist.gov/mml/mmsd/materials-structure/crystallographic-databases.cfm>
- [7] <http://paulingfile.com/>
- [8] <http://www.crystallography.net/>
- [9] <http://gaussian.com/>
- [10] <http://openmopac.net/>
- [11] <https://www.vasp.at/>
- [12] <http://lammps.sandia.gov/>
- [13] <http://www.materialsdesign.com/medea/gibbs>
- [14] D. Lerch, O. Wieckhorst, G.L. Hart, R.W. Forcade and S. Müller, 2009. UNCLE: a code for constructing cluster expansions for arbitrary lattices with minimal user-input. *Modelling and Simulation in Materials Science and Engineering*, 17(5), p.055003.
- [15] Yvon Le Page and Paul Saxe, *Phys. Rev. B* **65**, 104104 (2002)
- [16] J. B. Wachtman, Jr., W. E. Tefft, D. G. Lam, and R. P. Stinchfield, *J. Res. Natl. Bur. Stand., Sect. A* **64**, 213 (1960)
- [17] J.R. Gladden, J.H. So, J.D. Maynard, P.W. Saxe and Y. Le Page, *App. Phys. Lett.*, **85**, 392 (2004).
- [18] S.R. White, P.J. Mather and M.J. Smith, *Polym. Eng. Sci.* **42**, 51 (2002).
- [19] S. Behzadi and F.R. Jones, *J. Macromol. Sci. Part B:Physics*, **44**, 993 (2005).
- [20] Q.T. Zhao, U. Breuer, E. Rije, St. Lenk, and S. Mantl, *Appl. Phys. Lett.* **86**, 062108 (2005).