Materials informatics

by Krishna Rajan

Seeking structure-property relationships is an accepted paradigm in materials science, yet these relationships are often not linear, and the challenge is to seek patterns among multiple lengthscales and timescales. There is rarely a single multiscale theory or experiment that can meaningfully and accurately capture such information. In this article, we outline a process termed 'materials informatics' that allows one to survey complex, multiscale information in a high-throughput, statistically robust, and yet physically meaningful manner. The application of such an approach is shown to have significant impact in materials design and discovery.

Combinatorial Sciences and Materials Informatics Collaboratory (CoSMIC) and Institute for Combinatorial Discovery,
Department of Materials Science and Engineering,
Iowa State University,
Ames, IA 50011, USA
E-mail: krajan@iastate.edu

The search for new or alternative materials, whether through experiment or simulation, has been a slow and arduous task, punctuated by infrequent and often unexpected discoveries[1-6]. Each of these findings prompts a flurry of studies to better understand the underlying science governing the behavior of these materials. While informatics is well established in fields such as biology, drug discovery, astronomy, and quantitative social sciences, materials informatics is still in its infancy[7-13]. The few systematic efforts that have been made to analyze trends in data as a basis for predictions have, in large part, been inconclusive, not least because of the lack of large amounts of organized data and, even more importantly, the challenge of sifting through them in a timely and efficient manner[14].

When combined with a huge combinatorial space of chemistries as defined by even a small portion of the periodic table, it is clearly seen that searching for new materials with tailored properties is a prohibitive task. Hence, the search for new materials for new applications is limited to educated guesses. Data that does exist is often limited to small regions of compositional space. Experimental data is dispersed in the literature, and computationally derived data is limited to a few systems for which reliable data exists for calculation. Even after recent advances in high-speed computing, there are limits to how the structure and properties of many new materials can be calculated. Hence, this poses both a challenge and opportunity. The challenge is to deal with extremely large, disparate databases and large-scale

ISSN:1369 7021 © Elsevier Ltd 2005

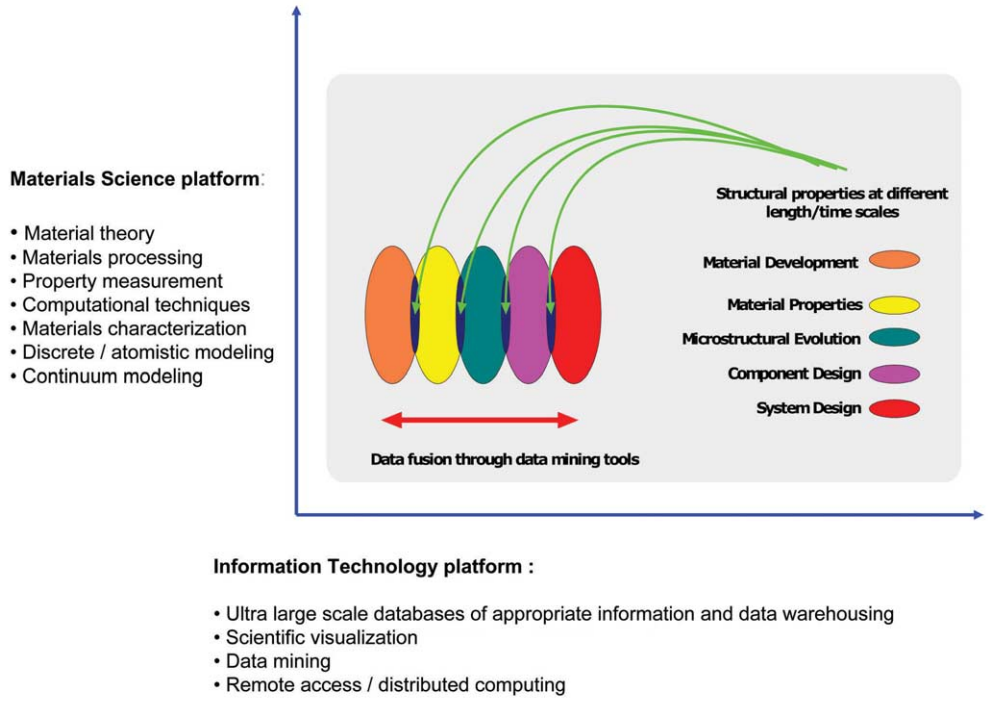


Fig. 1 The role of materials informatics is pervasive across all aspects of materials science and engineering. Mathematical tools based on data mining provide the computational engine for integrating materials science information across lengthscales. Informatics provides an accelerated means of fusing data and recognizing in a rapid yet robust manner structure-property relationships between disparate lengthscales and timescales. A complete materials informatics program will have an information technology (IT)-based component that is linked to classical materials science research strategies. The former includes a number of features that help informatics to be a critical tool in materials research: data warehousing and data management, which involves a science-based selection and organization of data that is linked to a reliable data searching and management system; data mining, providing accelerated analysis of large multivariate correlations; scientific visualization, a key area of scientific research that allows high-dimensional information to be assessed; and cyber infrastructure, an IT infrastructure that can accelerate sharing of information, data, and, most importantly, knowledge discovery.

computation. It is here that knowledge discovery in databases or data mining – an interdisciplinary field merging ideas from statistics, machine learning, databases, and parallel and distributed computing – provides a unique tool to integrate scientific information and theory for materials discovery (Fig. 1). The goal of data mining is the extraction of knowledge and insight from massive databases. It takes the form of discovering new patterns or building models from a given dataset. The opportunity is to take advantage of recent advances in data mining and apply them to state-of-the-art computational and experimental approaches for materials discovery.

Materials science data: feast or famine?

One may naturally assume that large amounts of data are critical for any serious informatics studies. However, what constitutes 'enough' data in materials science applications can vary significantly. In studying structural ceramics, for instance, fracture toughness measurements are difficult to make and, in some of the more complex materials, just a few careful measurements can be of great value. Similarly, reliable measurements of fundamental constants or properties for a given material involve very detailed measurement and/or computational techniques[15-19]. In essence, datasets in materials science fall into two broad categories: datasets on a given materials behavior, related to mechanical or physical properties, and datasets related to intrinsic information based on the chemical characteristic of the material, e.g. thermodynamic datasets.

In the materials science community, crystallographic and thermochemical databases have historically been two of the best-established. The former serves as the foundation for interpreting crystal structure data of metals, alloys, and inorganic materials. The latter involves the compilation of fundamental thermochemical information in terms of heat capacity and calorimetric data. While crystallographic databases are used primarily as a reference source, thermodynamic databases represent one of the earliest examples of informatics, as these databases were integrated into thermochemical computations to map phase stability in

binary and ternary alloys[20-26]. This led to the development of computationally derived phase diagrams – a classic example of integrating information in databases with data models. The evolution of both databases has occurred independently although, in terms of their scientific value, they are extraordinarily intertwined. Phase diagrams map out regimes of crystal structure in temperature-composition space or temperature-pressure space. Yet, crystal structure databases have been developed totally independently. At present, the community must work with each database separately, and information searches are cumbersome. Interpretation of data involving both is very difficult. Researchers only integrate such information on their own for one very specific system at a time, based on their individual interests. Hence there is currently no unified way to explore patterns of behavior across databases that are closely related scientifically.

One of the more systematic efforts to address this challenge has been that of Ashby and coworkers[27-33]. They showed that, by merging phenomenological relationships in materials properties with discrete data on specific materials characteristics, one can begin to develop patterns of classification of materials behavior. The visualization of multivariate data was managed using normalization schemes, which permit the development of 'maps' that provide a way to capture new means of clustering of materials properties. It also provides a methodology for establishing common structure-property relationships across seemingly different classes of materials. While very valuable, this approach is limited in its predictive value and is ultimately based on using prior models to build and seek relationships. In the informatics strategy of studying materials behavior, we approach it from a broader perspective. By exploring all types of data that may have varying degrees of influence on a

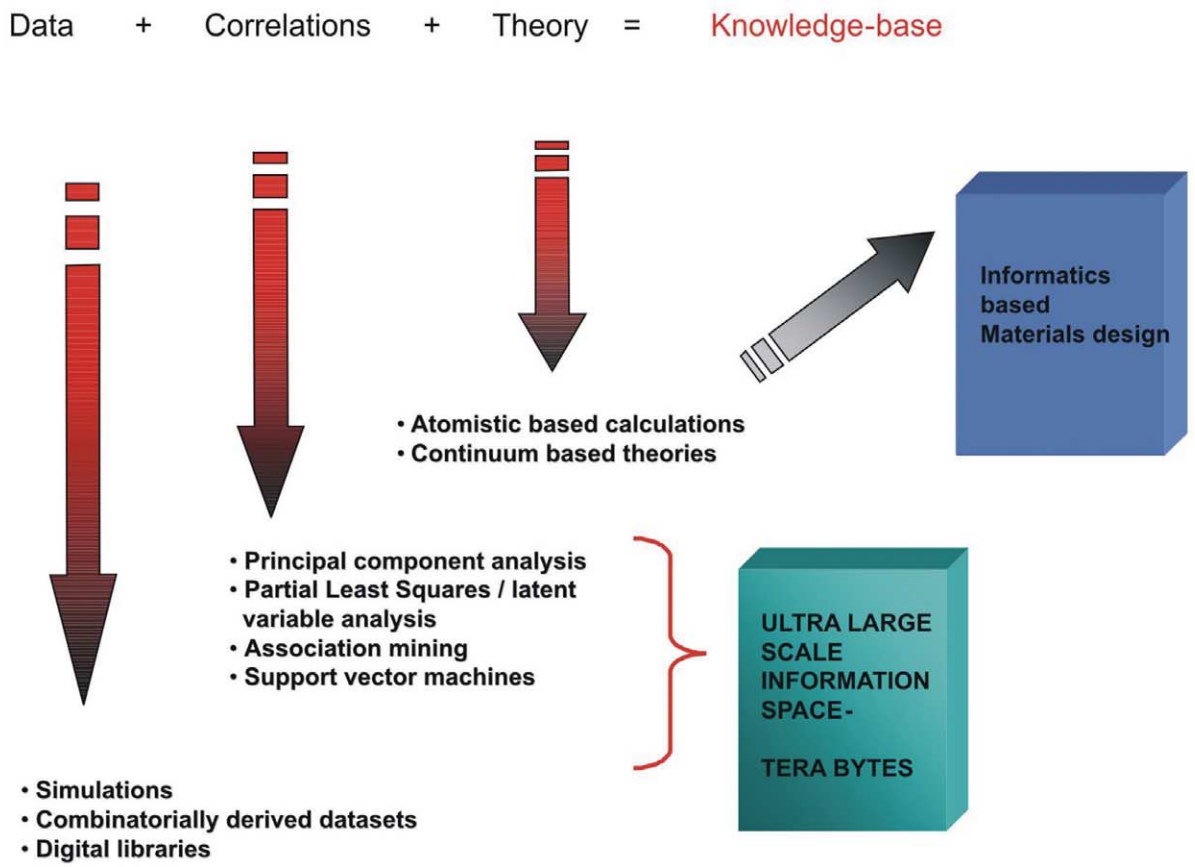


Fig. 2 The ultimate goal of 'knowledge discovery' is achieved through systematic integration of data, correlation analysis developed through data-mining tools and, most importantly, validation by fundamental theories and experiment-based science. The sources of data can be varied and numerous, ranging from computer simulations and high-throughput experimentation through combinatorial experiments and large-scale databases of legacy information. The application of advanced data-mining tools allows processing of very large sets of information in a robust yet rapid manner. The collective integration of statistical learning tools (a few of which are illustrated above) with experimental and computational materials science permits an informatics driven strategy for materials design.

given property or properties with no prior assumptions, one uses data-mining techniques to establish both classification and predictive assessments in materials behavior. However, this is not done from a purely statistical perspective but a perspective where one carefully integrates a physics-driven approach to data collection with data mining, then validates or analyzes it with theory-based computation and/or experiments. Data can come from either experiment or computation. The former, when organized in terms of combinatorial experiments, can screen large amounts of data in a high-throughput fashion[34-38]. However, the materials informatics pathway to knowledge discovery is not a linear process but an iterative one that can, at each 'information cycle', provide new information (Fig. 2).

Data mining: learning from the past and predicting the future

Broadly speaking, data-mining techniques have two primary functions: pattern recognition and prediction, both of which form the foundations for understanding materials behavior. Following the treatment of Tan *et al.*[39], the former, which is more descriptive in scope, serves as a basis for deriving correlations, trends, clusters, trajectories, and anomalies among disparate data. The interpretation of these patterns is tied intrinsically to an understanding of materials physics and chemistry. In many ways, this role of data mining is similar to the phenomenological structure-property paradigms that play a central role in the study of engineering materials, except now we are able to recognize these relationships with far greater speed and not necessarily depend on *a priori* models, provided, of course, that we have the relevant data. The predictive aspect of data mining tasks can serve for both classification and regression operations. Data mining, which is an interdisciplinary blend of statistics, machine learning, artificial intelligence, and pattern recognition, is viewed as having a few core tasks:

- Cluster analysis seeks to find groups of closely related observations and is valuable in targeting groups of data that may have well-behaved correlations and can form the basis of physics-based as well statistically-based models. Cluster analysis, when integrated with high-throughput experimentation, can serve as a powerful tool for rapidly screening combinatorial libraries.
- Predictive modeling helps to build models for targeted objectives (e.g. a specific materials property) as a function of input or exploratory variables. The success of these models also helps refine the usefulness and relevance of the input parameters.
- Association analysis is used to discover patterns that describe strongly associated features in data (for instance, the frequency of association of a specific materials property to materials chemistry). Such an analysis over extremely large datasets has been made possible by the development of very high-speed search algorithms, and can help to develop heuristic rules for materials behavior governed by many factors[40].
- Anomaly detection does the opposite by identifying data or observations that are significantly different from the norm. The ability to identify such anomalies or outliers is critical in materials, since it can identify new classes of materials with unusual properties (e.g. superconducting ceramics as opposed to insulating ceramics) or anticipate potential harmful effects, which are often identified through a retrospective analysis after an engineering failure (e.g. the ductile-brittle transition).

In most materials science studies, we identify *a priori* likely variables or parameters that affect a set of properties. This is usually based on theoretical considerations and/or heuristic analysis based on prior experience. However, it is difficult to integrate information simultaneously from multivariate data, especially when phenomenological relationships cannot always be explained in advance.

A statistical evaluation to search for each descriptor is computationally expensive and most probably ineffective. One basic approach to addressing this problem is to use principal component analysis (PCA). This is a technique for reducing the information dimensionality that is often needed from the vast arrays of data obtained from combinatorial experiments, large databases, or simulations, in a fashion such that there is minimal loss of information (see text box). PCA (also referred to as singular value decomposition) is one of a family of related techniques, including factor analysis and principal coordinate analysis, that provide a projection of complex datasets onto a reduced, easily visualized space. A simple way of imagining this concept is to visualize a three-dimensional cloud of data points that map correlations between datasets based on a multiple set of potential variables or influencing parameters[41-43]. Descriptors can, for example, be physicochemical properties like melting point,

processing variables like sintering temperature, or microstructural descriptors like coordination number. The enormous number of descriptors makes screening through scatter maps a prerequisite. From the screened descriptor space, we can select a region for the solution of our problem. We can also simplify this selection by compressing the dimensionality of the descriptor space by linear combinations of the original descriptors. PCA provides a method for transforming multiple descriptors into a much smaller set of descriptors without losing much information[44-52]. This makes visualization easier, as well as simplifying prediction and classification (Fig. 3).

While PCA is helpful in assessing the relative impact of multiple parameters on properties, it is not a predictive tool. For that, we need to apply other methods. We wish to demonstrate the value of one such approach here using a technique known as partial least squares (PLS). PLS regression is probably the least restrictive of the various multivariate extensions of the multiple linear regression models. This flexibility allows it to be used in situations where the use of traditional multivariate methods is severely limited, such as when there are fewer observations than predictor variables. Furthermore, PLS regression can be used as an exploratory analysis tool to select suitable predictor variables and to identify outliers before classical linear regression[53-55].

Of course, PCA and PLS are just two examples of data-mining methods. There are many others, each suited to different types of datasets in terms of size, skewness,

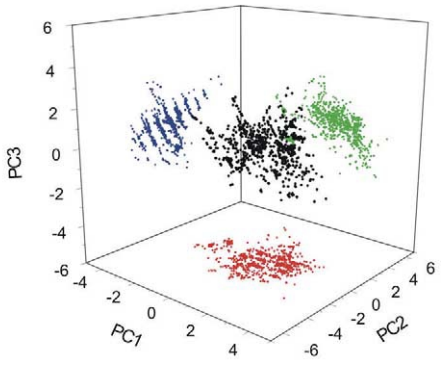
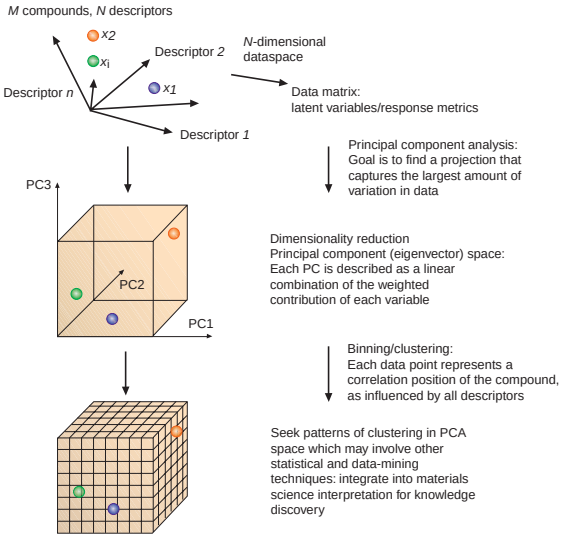


Fig. 3 Principal component plot of correlation data from a superconductor database. This analysis involved hundreds of compounds and the incorporation of dozens of descriptors from each compound. In only one projection (PC3 versus PC2) is a clear clustering pattern seen. Linear clusters were found to be associated with systematic valency changes among the compounds studied. This figure also emphasizes the importance of the visualization of data in aiding interpretation of complex datasets.

Principal component analysis

Principal component analysis (PCA) relies on the fact that most descriptors are interrelated and that these correlations, in some cases, are high. From a set of *N* correlated descriptors, we can derive a set of *N* uncorrelated descriptors (the principal components, or PCs). Each PC is a suitable linear combination of all the original descriptors. The first PC accounts for the maximum variance (eigenvalue) in the original dataset. The second PC is orthogonal (uncorrelated) to the first and accounts for most of the remaining variance. Thus, the m^{th} PC is orthogonal to all the others and has the m^{th} largest variance in the set of PCs. Once the *N* PCs have been calculated using eigenvalue/eigenvector matrix operations, only PCs with variances above a critical level are retained. The *M*-dimensional PC space has retained most of the information from the initial *N*-dimensional descriptor space by projecting it onto orthogonal axes of high variance. The complex tasks of prediction or classification are made easier in this compressed space.



(Adapted from[41]. © 2002 Nature Publishing Group.)

uncertainty within datasets, distribution, and other such features of the data and the type of information that one is ultimately seeking. For example, techniques such as the support vector machine (SVM), which falls under the category of 'supervised learning' methods, is of great value if one has some previous information about the materials under study. These can be combined into a set of training examples used by the SVM to distinguish between members and nonmembers of the class, on the basis of their behavior. For instance, we have used such an approach to develop classification schemes for grouping inorganic high-temperature superconductors[56].

Applications: what can we learn through informatics?

In the following discussion, we provide two brief examples of how data mining can be used to address material science issues.

- Searching through data – what information is really important?

  Chemical tailoring of sintering aids for ceramic processing

  Using the approach described above, we have estimated the effect of a vast array of influences in identifying correlations and key chemistries that control the fracture toughness of silicon nitride ceramics. Using a database that we built that includes over 2000 entries and explores over 20 different sets of latent variables, we have established data-mining procedures to explore correlations between a wide array of parameters across length scales relative to fracture toughness. Based on this, we explored heuristically the impact of a comprehensive history of the role of additives on mechanical properties. We have shown, for instance, that specific rare-earth additives play an important role, and that many others do not. From this, we have been able to assess rapidly the key chemistries that are needed to develop a processing strategy for enhancing the mechanical properties of silicon nitride ceramics (Fig. 4).
- Data mining as a predictive tool

  Establishing 'virtual' materials libraries and structure-property relationships through data mining

  We used predictive models based on first principles calculations to elaborate on unknown entries in a pre-existing library of materials. We then used these descriptors to develop a larger, heuristically-derived database by using a combination of PCA and PLS techniques based on a training set of theoretically derived data. This aided the exploration of a broader range of new

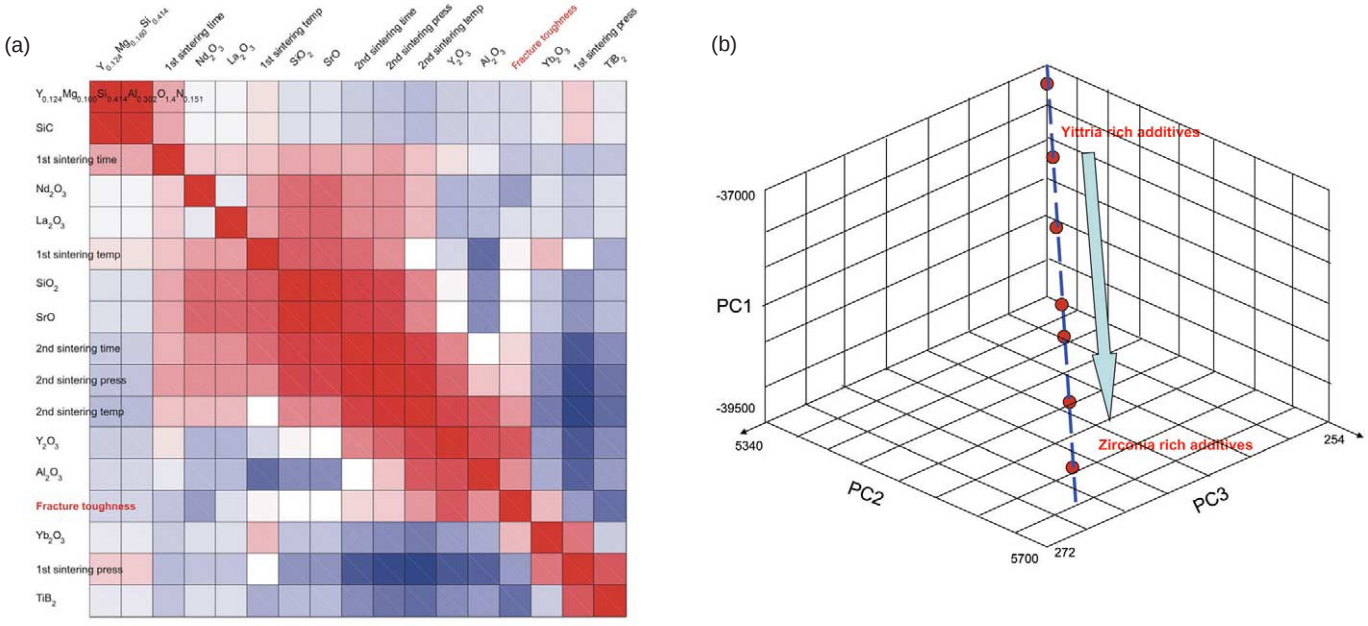


Fig. 4 (a) Correlation map based on thousands of data points plotting correlations from scatter plots of numerous processing and chemistry variables associated with fracture toughness data in SiN. The database was established through a long term survey of literature data. The plots by themselves show little of any trends in structure-chemistry-process-property relationships. (b) Principal component plot of the same data in (a) reveals a striking linear cluster of data for just a few chemistries, indicating that these are the main parameters influencing fracture toughness. The linear plot represents a strong linear clustering of data associated with the specific chemistries that contribute to the targeted property (fracture toughness) of this data-mining analysis. This effectively serves as a data screening tool for identifying important chemistries among a much larger multidimensional dataset.

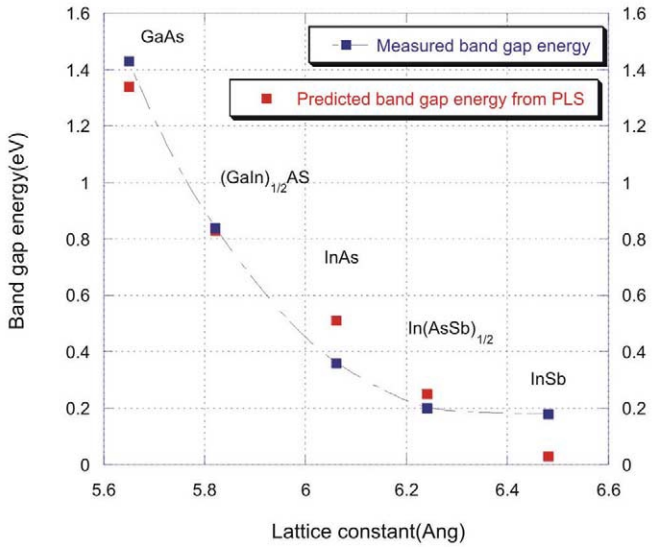


Fig. 5 Structure-property map of 'virtual' compound semiconductors derived from a combination of data-mining techniques. Results compare favorably with theoretical studies from large-scale ab initio calculations. Data-mining results – based on a carefully constructed database of latent variables influencing fundamental properties (in this case bandgap) – provided these results in a very fast and robust manner. Validation of results through independent computational means lends confidence to the use of data mining as a predictive tool.

trends and correlations (in studying chemistry-bandgap/modulus relationships) that have not been established before (Fig. 5), thus creating a 'virtual' library. This virtual library forms a unique database for exploring relationships and trends, from which one can use *ab initio* calculations and, eventually, actual testing to verify these trends and relationships. It is important to recognize that this library can, in principle, be built up by repeating complex atomistic calculations for each chemistry or compound of interest. However, this is an extremely prohibitive approach for screening, since these calculations, even for just one compound, are extremely difficult and time consuming, even after advances in parallel computation. Hence, it is a somewhat recursive process, where the researcher can apply predictive models to generate further data, explore and mine that data, and use this as a guide for where to incorporate *ab initio* descriptors and repeat the information cycle process. This can significantly accelerate the identification of promising materials by cutting down the combinatorial explosion of possible alloys, creating a form of active learning (Fig. 5).

The future: needs and prospects

The role of materials informatics can be pervasive throughout all fields and applications in materials science. Its impact can influence the way we do experiments, analyze data, and even alter the way that we teach materials science. Ultimately, the 'processing-structure-properties' paradigm that forms the core of materials development is based on understanding multivariate correlations and interpreting them in terms of the fundamental physics, chemistry, and engineering of materials. The field of materials informatics can advance that paradigm significantly. It may be helpful to bear in mind a few critical questions in building the informatics infrastructure for materials science[57].

- How can data mining/machine learning best be used to discover what attributes (or combination of attributes) of a material may govern specific properties? Using information from different databases, we can compare and search for associations and patterns that provide ways of relating information among these different datasets.
- What are the most interesting patterns that can be extracted from existing material science data? Such a pattern search process can potentially yield associations between seemingly disparate datasets, as well as establish possible correlations between parameters that are not easily studied experimentally in a coupled manner.
- How can we use mined associations from large volumes of data to guide future experiments and simulations? How does one select from a materials library which compounds are most likely to have the desired properties? Data-mining methods should be incorporated into design and testing methodologies to increase the efficiency of the material application process. For example, a possible test bed for materials discovery can involve the use of massive databases on crystal structure, electronic structure, and thermochemistry. Each of these databases alone can provide information on hundreds of binary, ternary, and multicomponent systems. Coupled to electronic structure and thermochemical calculations, one can enlarge this library to permit a wide array of simulations for thousands of combinations of materials chemistries. Such a massively parallel approach to the generation new 'virtual' data would be daunting, if not impossible, were it not for data-mining tools such as those proposed here.

We conclude this introduction to materials informatics with the question with which we began – what is materials informatics? A most appropriate analogy to the complexity of materials science is the field of astronomy, which offers a vast 'natural' laboratory of data. As noted by Johns Hopkins'

astronomer Alex Szalay, director of the US National Virtual Observatory project, in describing informatics for astronomy[58]: "Science was originally empirical, like Leonardo making wonderful drawings of nature. Next came the theorists who tried to write down equations that explained observed behaviors, like Kepler or Einstein. Then, when we got to complex enough systems like the clustering of a million galaxies, there came the computer simulations – the computational branch of science. Now, we are getting into the data exploration part of science, which is kind of a little bit of them all" – such is materials informatics. MT

Acknowledgments

The work described in this article is based on the doctoral theses of many of my students. I would like to particularly acknowledge the contributions of C. Suh, M. Stukowski, X. Li, and A. Rajagopalan. I gratefully acknowledge support from the National Science Foundation International Materials Institute Program for the Combinatorial Sciences and Materials Informatics Collaboratory (CoSMIC-IMI), grant no. DMR: 0231291.

REFERENCES

1. Saito, T., *et al., Science* (2003) **300**, 464
2. Hugosson, H. W., *et al., Science* (2001) **293**, 2434
3. Zhang, P., *et al., Nature* (2001) **409**, 69
4. Wang, T., *et al., Phys. Rev. Lett.* (1999) **82** (16), 3304
5. Franceschetti, A., and Zunger, A., *Nature* (1999) **402**, 60
6. James, R., *et al., Nature* (2003) **425**, 702
7. Gorse, D., and Lahana, R., *Curr. Opin. Chem. Biol.* (2000) **4** (3), 287
8. Liang, J., and Kachalo, S., *Chemom. Intell. Lab. Syst.* (2002) **62** (2),199
9. Yergey, A. L., and Bierbaum, V. M., *J. Am. Soc. Mass Spectrom.* (2002) **13** (1), 1
10. Vihinen, M., *Biomol. Eng.* (2001) **18** (5), 241
11. Ellis, L. B. M., *Curr. Opin. Biotechnol.* (2000) **11** (3), 232
12. Shlichta, P. J., *J. Cryst. Growth* (1997) **174** (1-4), 480
13. Wesolowski, M. and Konieczynski, P., *J. Pharmaceutics* (2003) **262** (1-2), 29
14. Shiflet, G., *Science* (2003) **300**, 443
15. Song, Y., *et al., J. Comput.-Aided Mater. Des.* (1999) **6** (2-3), 355
16. Grimvall, G., *Thermophysical properties of materials*, Elsevier, Amsterdam, (1999)
17. Ledbetter, H., and Kim, S., Bulk moduli systematics in oxides, including superconductors, in *Handbook of Elastic Properties of Solids, Liquids and Gases* Levy, M., *et al.*, (eds.), Academic Press, (2000)
18. Iwata, S. *et al., J. Nucl. Mater.* (1992) **179-181** (part 2), 1135
19. Davis, J. W., *J. Nucl. Mater.* (1992) **179-181** (part 2), 1139
20. Saxena, S., *et al., Thermodynamic Data on Oxides and Silicates*, Springer, New York, (1993), 428
21. Fabrichnaya, O. B., and Sundman, B., *Geochim. Cosmochim. Acta* (1997) **61** (21), 4539
22. Fabrichnaya, O. B., *et al., Thermodynamic Data, Models, and Phase Diagrams in Multicomponent Oxide Systems*, Springer, New York, (2004)
23. Bale, C. W., *et al., Calphad* (2002) **26** (2), 189
24. Cox, J. D., *et al.*, (eds.) *CODATA Key Values for Thermodynamics*, Hemisphere Publishing, New York, (1989)
25. Soligo, D., *et al., Acta Materialia* (1999) **47** (9), 2741
26. Villars, P., *et al., J. Alloys Compd.* (2001) **317-318**, 26
27. Ashby, M. F., *Materials Selection in Mechanical Design*, Butterworth-Heinemann, Oxford (1999)
28. Shercliff, H. R., and Lovatt, A. M., *Prog. Mater. Sci.* (2001) **46** (3-4), 429
29. Lovatt, A. M., and Shercliff, H. R., *Materials and Design* (1998) **19**, 205 (part 1) and 217 (part 2)
32. Ashby, M. F., *Proc. R. Soc. London, Ser. A* (1998) **454**, 1301
33. Bassetti, D., *et. al., Proc. R. Soc. London, Ser. A* (1998) **454**,1323
34. Landrum, G. A., and Genin, H., *J. Solid State Chem.* (2003) **176** (2), 587
35. Curtarolo, S., *et al., Phys. Rev. Lett.* (2003) **91**, 135503
36. Morgan, D., *et al., J. Phys.: Condens. Matter* (2003) **15** (25), 4361
37. Ceder, G., *Science* (1998) **280**, 1099
38. van de Walle, A., *et al.*, Calphad (2002) **26** (4), 539
39. Tan, P-N., *et al., Introduction to Data Mining*, Addison-Wesley, Boston (2005)
40. Zaki, M., and Rajan, K., *Proc. 17th Int. CODATA Conf.*, Baveno, Italy, (2000)
41. Bajorath, J., *Nat. Rev. Drug Discovery* (2002) **1** (11), 882
42. Quakenbush, J., *Nat. Rev. Genet.* (2001) **2** (6), 418
43. Li, G., *et al., J. Phys. Chem. A* (2001) **104** (33), 7765
44. Rajan, K., An Informatics Approach To Interface Characterization: Establishing a "Materials by Design" Paradigm, In *Science and Technology of Interfaces*, Ankem, S., and Pande, C. S., (eds.), TMS, Warrendale, PA, (2002), 231
45. Rajan, K., *et al., Quantitative Structure-Activity Relationships (QSARs) for Materials Science in Combinatorial and Artificial Intelligence Methods in Materials Science*, Takeuchi, I., *et al.*, (eds.), MRS, Pittsburgh PA, (2004)
46. Suh, C., *et al., Data Sci. J.* (2002) **1** (1), 19
47. Rajan, K., *et al.*, Data Mining and Multivariate Analysis in Materials science, In *Molten Salts: From Fundamentals to Applications, Proc. NATO Advanced Study Institute (NATO Science Series II: Mathematics, Physics and Chemistry)*, Gaune-Escard, M., (ed.), Kluwer Academic Publishers, (2002), 241
48. Suh, C., *et al.*, Chemical Discovery in Molten Salts through Data Mining, In *Int. Symp. Ionic Liquids*, Øye, H. A., and Jagtøyen, A., (eds.), Norwegian University of Science and Technology, Trondheim, Norway, (2003), 587
49. Suh, C., and Rajan, K., *Appl. Surf. Sci.* (2003) **223** (1-3), 148
50. Rajagopalan, A., *et al., Appl. Catal. A* (2003) **254** (1), 147
51. Rajagopalan, A., *et al., Proc. 7th Int. Conf. Systemics, Cybernetics and Informatics*, International Institute of Informatics and Systemics, Orlando, FL, (2003)
52. Suh, C., *et al.*, Combinatorial Materials Design Through Database Science, In *Combinatorial and Artificial Intelligence Methods in Materials Science*, MRS, Warrendale, PA, (2004)
53. Suh, C., and Rajan, K., *QSAR & Combinatorial Sci. J.* (2005) **24** (1), 114
54. Bennett, K. P., and Embrechts, M. J., An Optimization Perspective on Partial Least Squares, In *Advances in Learning Theory: Methods, Models and Applications*, Suykens, J. A., *et al.*, (eds.), NATO Science Series III: Computer & Systems Sciences, IOS Press, Amsterdam, (2003) **190**, 227
55. Rosipal, R., and Trejo, L. J., *J. Machine Learning Res.* (2001) **2** (2), 97
56. O'Connor, A., *et al., Proc. Intelligent Engineering Systems through Artificial Neural Networks*, **12**, ASME Press, New York, (2002)
57. Rajan, K., In *Workshop Report on a Future Information Infrastructure for the Physical Sciences – The Facts of Matter: Finding, understanding and using information about our physical world*, DOE Panel report (2000), www.osti.gov/physicalsciences/index.html
58. Szalay, A., quoted in *New York Times*, May 20, 2003, www.nytimes.com/2003/05/20/science/space/20DWAR.html?ex=1054449062&ei=1&e