Pioneering materials informatics

Clean technologies from molecular design

The Department of Materials Design and Innovation, University at Buffalo, New York, pioneers the development and application of materials informatics.

Professor Krishna Rajan takes an a priori approach to the discovery of new materials.

Tackling global challenges in sustainable energy production calls for the rational design of new generations of clean materials, coupling efficiency with low environmental impact. Materials discovery offers a powerful means to predict structure-property relationships in novel materials, but often lacks the ability to account for the complex interplay of length- and timescales that ultimately determine the potential of a material to exhibit desired functionalities. The Department of Materials Design and Innovation (MDI) at University at Buffalo, New York, pioneers the development and application of materials informatics. Through this unique approach to materials science the MDI is solving problems in environmental sustainability, renewable energy and climate change.

One of the fundamental issues in materials science, and in the design and synthesis of technologically relevant materials, concerns the so-called structure-property relationship. Given the structure of a molecule or crystal, how can we infer which *properties* the system will exhibit? How, moreover, will it behave in realistic conditions and how will it interact with its environment?

In recent years, the development of advanced quantum mechanical methods can provide a huge amount of information concerning the properties of wide classes of materials. These methods can account for a system's stability, chemical bonding patterns, and even electronic or optical properties, along with the use of sophisticated experimental techniques to probe the structure and properties of materials. These approaches can often be used to guide the discovery of new materials with enhanced capabilities, for instance for the clean production or storage of energy, or for the removal of toxic agents from the environment.

However, using large amounts of information for property prediction in a robust and systematic way is highly problematic, even when complex statistical or machine learning based approaches are used to explore the available data. The problem becomes even more complex when the properties of completely new materials, bearing little resemblance with known ones, have to be investigated and predicted.

MATERIALS INFORMATICS
"The challenge", says Professor Rajan, University at Buffalo, New York, "is that

structure-property relationships are often non-linear, and the goal is to seek patterns among multiple length scales and timescales. Rarely can a single multiscale theory meaningfully and accurately capture such information." To address this challenge, we therefore not only need an efficient strategy for searching large data sets, but also techniques to learn from those data, by generating information that can be used outside the models on which the data themselves are based.

Professor Rajan has established a research center, the Department of Materials Design and Innovation (MDI), Collaboratory for a Regenerative Economy, that explores how to develop 'quantum signatures' at a fundamental molecular scale that can guide the rational design of clean materials. With the use of advanced machine learning techniques, they can bridge the gaps in knowledge that theory and heuristics alone cannot capture. His approach marks a change of paradigm in materials discovery by shifting the emphasis from simply searching among large volumes of data to identifying the key materials physics and chemistry (or "inorganic genes") that can be integrated in statistical and machine learning methods. This can provide design rules that govern targeted functionalities in new materials. These inorganic genes encode the fundamental information that collectively characterises the stability and properties of a material. They allow us to map the complexity of a material description of how these fundamental pieces of information self-organise and self-assemble, ultimately giving a material its characteristic properties.

LINKING COMPUTATIONAL CHEMISTRY AND MACHINE LEARNING
The materials informatics approach developed by Professor Rajan aims to design materials with targeted properties. It involves choosing a number of structural, chemical and processing parameters on the basis of experimental and theoretical evidence. During this parameter selection, it is important not to miss critical descriptors that can hold clues for the prediction of the material's properties.

"We have harnessed and enhanced machine-readable representation of the molecular structure," explains Professor Rajan, "allowing us to rapidly explore and successfully predict multi-scale properties of large numbers of materials." The information encoded in this representation allows one to target those properties of a material that are influenced by metrics other than just its chemistry. These macroscopic properties are the result of collective phenomena extending far beyond the microscopic scales that are described by quantum atomistic theories.

SUSTAINABLE MATERIALS DISCOVERY
Despite the tremendous potential of materials informatics in property prediction, there is one important aspect of materials design that needs to be addressed: the impact that new materials or their preparation techniques

can have on the environment, and the hazard they can pose to human health. The focus of Professor Rajan's work is on the chemical design of materials that considers the potential environmental impacts in all phases of a material's life cycle – from synthesis, to manufacturing, and end-of-life replacement. It seeks, in addition, to account for the material's engineering functionality.

This is a daunting task, which requires a careful use of machine learning techniques coupled with computational chemistry. Spearheaded by Professor Rajan, in the approach taken at

MDI, environmental and health considerations are included from the outset in the material's design space. These are treated on the same footing as the technological and engineering aspects. This adds to the complexity of the overall materials discovery

process, but it also offers a pathway to create new materials that exhibit new or improved functionalities *alongside* environmental safety.

PERFORMANCE AND SAFETY
Professor Rajan has developed new accelerated computational methods to establish sets of molecular "fingerprints." These can serve as strategic markers to rapidly identify new materials' chemistries and to provide robust guidelines for assessing the environmental hazard of organic chemicals used during manufacturing.

An important example of the application of this approach has been

in the case of per- and polyfluoroalkyl substances (PFASs), which are man-made chemicals used in a variety of industries around the globe. Once adsorbed, these chemicals act as very persistent and dangerous contaminants, both in the environment and in the

A crucial challenge in accelerating materials discovery is how to rapidly identify patterns among multiple length- and timescales.

Through their unique approach to materials science the MDI is solving problems in environmental sustainability, renewable energy and climate change.

human body. Exposure to PFAS can lead to adverse human health effects, including liver, kidney, and immunological conditions, as well as cancer.

A major challenge in addressing the issue of environmental and health hazards of chemicals is the sparsity of data compared with the vast chemical space that we need to explore. Using informatics, modelling, and machine learning techniques, Rajan's group has developed a framework for estimating the likelihood of toxicity impact of the

of his methods has been the ability to predict structural properties of cubic perovskite crystals. These materials, which are important in solar cell development, are cheap to produce and easy to manufacture.

Professor Rajan has developed a unique machine learning framework that captures the quantum mechanical based descriptions of interactions between chemical bonding and bond geometry using a calculated molecular property. One form of the output of such calculations is visualised in terms of a complex 3-dimensional shape (known as Hirshfeld surfaces) that encodes information concerning properties such as molecular packing, molecular shape, close contact points and inter-molecular interactions in crystals. They are typically used to rationalise the ability of molecular building blocks to assemble and form flexible or rigid structures in multicomponent systems.

Combined with machine learning methods, the Hirshfeld surface provides a fingerprint to reliably predict lattice parameters (and potentially other properties) of perovskite crystals. A machine learning approach based on Hirshfeld surfaces has also been used to study metal-organic frameworks (MOFs) – an extremely rich class of mesoporous materials – which can be synthesised with a variety of pore sizes, geometries, and network structures. This can have application in several fields, including membrane technologies and chemical sequestration.

The method proposed by Professor Rajan permits one to map high-dimensional correlations between the chemical and crystallographic properties of MOFs and to shed light onto how the interplay between chemical bonding and network geometry governs relationships between diverse families of MOFs. This correlative analysis provides solid guidelines on how to engineer new MOF structures, and to optimise their ability to act as materials for applications in renewable energy technologies, like carbon capture.

to quantify the potential toxicity of alternatives to existing PFASs and even to provide guidance on what the most important safety concerns for these systems should be focusing on. Although legacy PFASs have been banned worldwide, the method developed by Professor Rajan provides guidelines for classifying new PFAS compounds and their toxicities. This powerful approach will also help with developing similarity metrics to guide the selection of alternative and safer chemicals; providing an informatics

Linking machine learning and computational chemistry descriptors provides a powerful tool to explore structure-property relationships in new materials.

rapidly increasing amount of PFAS related compounds being introduced into the environment for which hazard data does not exist.

Using a robust machine learning based method for the visualisation of large sets of imbalanced and sparse data, Professor Rajan has been able to develop an accelerated approach for estimating previously unmeasured PFAS properties. This enables him

framework for environmentally conscious selection of safer chemicals in materials design.

MATERIALS FOR RENEWABLE ENERGY TECHNOLOGIES
Professor Rajan has applied his approach to material property prediction to several important compounds which might be of major relevance in sustainable energy production. An example of the success

---

Behind the Research
Professor Krishna Rajan

E: krajan3@buffalo.edu **W:** http://mdi.buffalo.edu

Research Objectives

Harnessing a materials informatics approach, Professor Krishna Rajan takes an *a priori* approach to the discovery of new materials with a heavy focus on minimising environmental impacts without compromising engineering functionality.

References

Su, A Rajan, K (2021). A database framework for rapid screening of structure-function relationships in PFAS chemistry. *Scientific Data*, 8, 14. https://doi.org/10.1038/s41597-021-00798-x

Williams, L Mukherjee, A Rajan, K (2020). Deep Learning Based Prediction of Perovskite Lattice Parameters from Hirshfeld Surface Fingerprints. *J. Phys. Chem. Lett.*, 11, 7462–7468. https://doi.org/10.1021/acs.jpclett.0c02201

Shen, X Zhang, T Broderick, S Rajan, K (2018). Correlative analysis of metal organic framework structures through manifold learning of Hirshfeld surfaces. *Mol. Syst. Des. Eng.*, 3, 826–838. https://pubs.rsc.org/en/content/articlelanding/2018/me/c8me00014j#!divAbstract

Rajan, K (2015). Materials Informatics: The Materials "Gene" and Big Data. *Annu. Rev. Mater. Res.*, 45, 153–169. https://doi.org/10.1146/annurev-matsci-070214-021132

Broderick, S Dongol, R Zhang, T Rajan, K (2021) Classification of apatite structures via topological data analysis: a framework for a 'Materials Barcode' representation of structure maps. *Sci Rep* 11, 11599 (2021). https://doi.org/10.1038/s41598-021-90070-4

Detail

Krishna Rajan
Dept of Materials Design and Innovation
120 Bonner Hall, University at Buffalo
Buffalo NY 14260 USA

University at Buffalo
Department of Materials Design and Innovation
School of Engineering and Applied Sciences
College of Arts and Sciences

Bio
Krishna Rajan is SUNY Distinguished Professor and Erich Bloch Chair of the Department of Materials Design and Innovation, University at Buffalo. Professor Rajan is a leader in the field of Materials Informatics working on many materials science applications including sustainability. He received his ScD from MIT.

Personal Response

Your approach to materials discovery not only enables reliable predictions on technological functionalities of new materials to be made, but also to assess potential health and environmental hazards in their manufacture or utilisation. What are the technological fields in which the methods you have developed can have the largest impact, and what are some of the challenges that you plan to address in the future?

We need a holistic understanding of the molecular mechanisms that cause environmental and human toxicity so that we can develop an expanded definition of performance to include materials functionality as well as the inherent potential hazards involved in the synthesis and fabrication of materials. A key barrier to adopting such a holistic approach is the development and strategic application of informatics techniques that can discover new and unexplored pathways and connections linking chemistries, materials, functionality, and environmental impact. This would help us achieve the goal of an a priori approach to a "benign-by-design' approach to sustainable materials discovery.